



Karlsruhe Institute of Technology

Measurement of the $t\bar{t} + \geq 1$ b-jet cross section using novel multivariate analysis techniques at the CMS experiment

Zur Erlangung des akademischen Grades eines
DOKTORS DER NATURWISSENSCHAFTEN
an der Fakultät für Physik des
Karlsruher Institut für Technologie (KIT)

genehmigte
DISSERTATION

von

Marco Alexander Harrendorf
aus Karlsruhe

Tag der mündlichen Prüfung:	22. Juni 2018
Referent:	Prof. Dr. Ulrich Husemann
Korreferent:	PD Dr. Stefan Gieseke
	Prof. Dr. Dieter Zeppenfeld i.V.



This page intentionally left blank.

License



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

This page intentionally left blank.

Contents

Introduction	1
I Theoretical foundations, experimental basics, and statistical analysis methods	5
1 The Standard Model of particle physics	7
1.1 History of the Standard Model	7
1.2 Elementary particles	8
1.3 Interactions	9
1.3.1 Electromagnetic interaction	10
1.3.2 Weak and electroweak interaction	10
1.3.3 Strong interaction	10
2 Hadron collider physics	13
2.1 Calculation of cross sections	14
2.1.1 Factorization theorem	14
2.1.2 Feynman diagrams and perturbation theory	17
2.1.3 Parton distribution functions	18
2.1.4 Description of a physics process in the four-flavour and five-flavour scheme	19
2.2 Monte Carlo event generation for hadron colliders	20
2.2.1 Event generation	21
2.2.2 Tuning of event generators	25
2.2.3 Matching and merging	25
3 Experimental basics	27
3.1 The Large Hadron Collider (LHC)	27
3.2 The Compact Muon Solenoid detector	29
3.3 Object identification and event reconstruction	35
3.3.1 Tracks and vertices	35
3.3.2 Particle flow event reconstruction and MET	36
3.3.3 Muons	37
3.3.4 Electrons	38
3.3.5 Photon, hadron and jet reconstruction	38
3.3.6 b-Tagging of jets	39
4 Top-quark physics	41
4.1 The beginning of top-quark physics and the discovery of the top quark . . .	41
4.2 Basic properties of the top quark	42
4.3 Top-quark production	42
4.4 Decay modes of the top quark	43

4.5	Top-quark pair production and corresponding $t\bar{t}+X$ processes	45
4.6	Associated production of a Higgs boson with a top-quark pair	46
5	Statistical analysis methods	49
5.1	Maximum-Likelihood method	49
5.1.1	Derivation of the binned likelihood function	50
5.1.2	Maximum-Likelihood fit	51
5.1.3	Properties of nuisance parameters	52
5.1.4	Asimov dataset	52
5.2	Multivariate data analysis	53
5.2.1	Performance of a classifier	53
5.2.2	Linear discriminant analysis	54
5.2.3	Neural networks	55
II	Motivation	63
III	Review of past $t\bar{t}+\geq 1b\text{-jet}$ and $t\bar{t}+b\bar{b}$ measurements	77
IV	Measurement of the $t\bar{t}+\geq 1b\text{-jet}$ signal strength and cross section	89
6	Signal and background processes	93
6.1	General properties of the Monte Carlo samples	94
6.2	Cross sections of the Monte Carlo samples	94
6.3	Choice of the $t\bar{t}$ signal samples	95
6.4	Split of the $t\bar{t}$ samples into $t\bar{t}+X$ sub processes according to the flavour of the additional hadrons and jets	97
6.5	Choice of the $t\bar{t}H$ background samples	98
6.6	Choice of the other background samples	99
7	Measurement data and event selection	101
7.1	Measurement data	101
7.2	Trigger	101
7.3	Jets	102
7.4	B-tagging	102
7.5	MET	102
8	Analysis setup	103
8.1	Analysis framework	103
8.2	Check of Monte Carlo samples and measurement data agreement	104
9	Analysis strategy	107
9.1	B-jet multiplicity based analysis	108
9.2	Input variables for the multivariate analyses	108
9.2.1	Choice of the input variables for the multivariate analyses	112
9.2.2	Discussion of input variables applied for the multivariate analyses	113
9.3	Reweighting of process classes for the training of the multivariate analyses	114
9.4	Linear discriminant based analysis	115
9.5	Neural network based analysis	117
9.6	Application of a binning optimization algorithm	121
9.7	Maximum-Likelihood fit	121

10 Systematic uncertainties	123
10.1 Systematic uncertainties related to the applied Monte Carlo samples	123
10.1.1 Cross section normalization rate uncertainties	124
10.1.2 Extra rate uncertainty due to the cross section normalization of the $t\bar{t}+c\bar{c}$ process	124
10.1.3 PDF variation uncertainties	124
10.1.4 Extra shape uncertainties due to the chosen renormalization and factorization scale for the $t\bar{t}+X$ processes	125
10.1.5 Extra rate uncertainties due to the matrix element generator and parton shower matching for the $t\bar{t}+X$ processes	125
10.1.6 Extra rate uncertainties due to the value of the strong coupling constant α_S chosen in the parton shower for the $t\bar{t}+X$ processes	125
10.1.7 Extra rate uncertainties due to the underlying event tune of the parton shower for the $t\bar{t}+X$ processes	126
10.1.8 Shape uncertainties due to the limited size of the Monte Carlo samples	126
10.2 Systematic uncertainties related to the measurement process	126
10.2.1 Luminosity rate uncertainty	126
10.2.2 Uncertainties due to pileup interactions	126
10.2.3 Lepton ID, tracking, and isolation uncertainties	127
10.2.4 Lepton trigger efficiency uncertainties	127
10.2.5 Jet energy scale uncertainties	127
10.2.6 Jet energy resolution uncertainties	127
10.2.7 B-tagging uncertainties	127
10.3 Summary of systematic uncertainties	128
11 Results	131
11.1 $t\bar{t}+\geq 1b$ -jet signal strength result	131
11.2 Influence of systematic uncertainties on the $t\bar{t}+\geq 1b$ -jet signal strength result	133
11.2.1 Prefit and postfit distributions of the three reference analyses	133
11.2.2 Pull and impact distributions of the three reference analyses	139
11.2.3 Influence of specific systematic uncertainties on the $t\bar{t}+\geq 1b$ -jet signal strength result	142
11.3 Influence of correlations between input variables on the $t\bar{t}+\geq 1b$ -jet signal strength result	144
11.3.1 Strong correlations between pairs of input variables	144
11.3.2 Influence of strongly correlated input variables on the $t\bar{t}+\geq 1b$ -jet signal strength	146
11.4 Influence of binary or multiclassification on the $t\bar{t}+\geq 1b$ -jet signal strength result	149
11.5 Influence of neural network training on the $t\bar{t}+\geq 1b$ -jet signal strength result	151
11.6 Influence of including the $t\bar{t}H$ process as a background process in the training of neural networks	152
11.7 $t\bar{t}+\geq 1b$ -jet cross section	156
12 Discussion of results	159
V Simultaneous measurement of individual signal strengths and cross sections	163
13 Differences between the $t\bar{t}+\geq 1b$-jet, $t\bar{t}+b\bar{b}/t\bar{t}+2b/t\bar{t}+b$, and $t\bar{t}+\geq 1b$-jet/$t\bar{t}H$ signal strength and cross section measurements	167

14 Results	169
15 Discussion of results	175
VI Study of a possible improvement of the $t\bar{t}+\geq 1b$ -jet modeling by Monte Carlo event generators	181
16 Motivation of a combined 5FS $t\bar{t}$ / 4FS $t\bar{t}+b\bar{b}$ sample	185
17 Construction of a combined 5FS $t\bar{t}$ / 4FS $t\bar{t}+b\bar{b}$ sample	189
18 Normalizing 4FS $t\bar{t}+\geq 1b$ -jet events	191
19 Preliminary conclusions	195
VII Summary and outlook	197
Bibliography	207
Appendix	221
A Pull and impact distributions of the three reference analyses	223

*Science as we know it today
is the pursuit of truth and
knowledge for the good of all
mankind.*

Anonymous

Introduction

Preface

Science, and physics in particular, after its establishment as independent field of science in the 17th century, have come a long way since the beginning of mankind.

From the first conscious-thinking early humans trying to grasp their environment over the first astronomers looking up at the sky, followed closely by the early philosophers wondering what holds heaven and earth together via the Scientific revolution and establishment of the experimental method up to the modern science nowadays, human knowledge evolved faster and faster, even though it encountered some obstacles from time to time. Never before in history the technological advancement resting on rapidly growing and swiftly applied scientific insights could provide necessities and comfort of life for so many people. Still some of the oldest questions of mankind remain and still “two mysteries lie at the heart of the human experience. Where do we come from? Where are we going” [1].

It would be presumptuous to claim that particle physics would hold the final answer to these questions. Nevertheless, by providing experimentally-established theoretical models describing the composition of matter up to the smallest and non-directly detectable elementary particles the field of particle physics sheds some light on the matter.

The Standard Model of particle physics, which was developed from the 1960s on and whose current formulation was finalized in the early 1980s, is the best example of this being considered as “the most successful theory ever” [2]. As a successor of the quark model, or rather the Eightfold Way, it could not only explain the existing particle zoo and solve the particle puzzle of the 1960s, but also established the group of elementary particles all known matter is based on, while predicting fundamental particles which were successfully found only later. The latest of these fundamental particles found is the Higgs boson, which was discovered 2012 and which was also the last missing elementary particle predicted by the Standard Model.

Even though all particles predicted by the Standard Model are found, many questions around the Standard Model are still open. By studying physics processes occurring at particle accelerators like the Large Hadron Collider (LHC) particle physicists try to address some of these open questions, while hoping to gain more insight.

Current particle physics experiments would not be possible without the application of state-of-the-art technologies. Therefore, this academic discipline is promoting progress by being an early-adopter of recent advances and, also, in many cases by developing home-

grown innovations due to the need of fulfilling its own challenging demands. Just one of many examples is that particle physicists were among the first who had to handle, prepare and analyze extremely large data sets, which is these days known as Big Data, Data Science, and Data Analytics. All of these three topics are skyrocketing in importance in industry and become part of all areas of life in an unprecedented manner.

A key term which always comes up in this context is artificial neural network as a promising set of tools for a multivariate analysis.

In fact, particle physicists already considered and used neural networks – first mostly for object identification and event reconstruction – since the late 1980s and the early 1990s. Later on, the application of neural networks as a classifier for a particle physics analysis became more popular. However, they were still employed infrequently since the implementation of this multivariate analysis method was often a challenging task and other multivariate analysis methods like boosted decision trees (BDTs) were easier to realize and provided comparable results. Only from roughly 2015 the time was ripe for a widespread application of neural networks, and especially, deep learning techniques, in particle physics and also in industry which is still picking up momentum. The year 2015 can safely be chosen as a starting date since in this time falls the release of the TensorFlow open-source software library, which provided a rather easy framework to adopt neural networks in various software projects, and the NVIDIA company entered the deep learning market, so that artificial neural networks can now be easily trained and evaluated on a graphical processing unit (GPU) which is far more efficient than the computation on a central processing unit (CPU).

Overview

“Standing on the shoulder of giants”¹ this thesis provides a measurement of the signal strength and cross section for the production of a top quark-antiquark pair in association with one or more jets with a bottom hadron ($t\bar{t}+\geq 1b\text{-jet}$), while employing neural networks as multivariate analysis method in such a measurement for the first time. In addition to the neural network analysis the $t\bar{t}+\geq 1b\text{-jet}$ signal strength and cross section is also determined by using a simpler B-jet multiplicity based analysis acting as a baseline analysis and a linear discriminant based analysis, which was used as a cross check of the neural network based analysis.

Furthermore, a simultaneous measurement of the $t\bar{t}+b\bar{b}$ signal strength and cross section, the $t\bar{t}+2b$ signal strength and cross section, and the $t\bar{t}+b$ signal strength and cross section is presented in this thesis. These three processes are subsumed under the term $t\bar{t}+\geq 1b\text{-jet}$ processes.

In similar fashion, a simultaneous measurement of the $t\bar{t}+\geq 1b\text{-jet}$ signal strength and cross section and the signal strength and cross section of the associated production of top quark-antiquark pairs and a Higgs boson ($t\bar{t}H$) was conducted and its results are given.

Moreover, a study of a possible improvement of the $t\bar{t}+\geq 1b\text{-jet}$ modeling by Monte Carlo event generators is presented.

¹Regardless of the fact that many people believe that this quote comes from Sir Isaac Newton, this shorter linguistic expression is a lot older than Newton’s complete sentence of “We are like dwarfs sitting on the shoulders of giants.” and is attributed to Bernard of Chartres by [3] and others.

Outline

In the first part of this thesis (part I) the theoretical foundations, experimental basics, and statistical methods are shortly introduced since they are essential for the further understanding of this thesis:

In chapter 1 a brief overview of the Standard Model of particle physics is given. It is followed by a more in-depth discussion of physics at hadron colliders in chapter 2 with an explanation of the calculation of cross sections and of the Monte Carlo event generation for physics processes at hadron colliders. Subsequently, the experimental basics are introduced in chapter 3 with a discussion of the Large Hadron Collider (LHC), the Compact Muon Solenoid (CMS) detector, and the object identification and event reconstruction employed by the CMS detector. As top-quark physics processes are in the focus of this thesis chapter 4 is devoted to top-quark physics. Finally, chapter 5 completes part I with a discussion of statistical analysis methods, such as the Maximum-Likelihood method or linear discriminants and neural networks as multivariate analysis methods.

An in-depth motivation of the measurements conducted and presented in this thesis is given in part II, while in part III past measurements studying the same physics processes are reviewed.

As all three measurements contained in this thesis rely on the same analysis chain, the analysis chain and setup is introduced in part IV for all measurements.

In chapter 6 the signal and background processes employed in this thesis are presented, before the measurement data and event selection is introduced in chapter 7. A short overview of the analysis setup is given in the subsequent chapter (chapter 8), before in chapter 9 the analysis strategy is presented. In this chapter the three analyses (B-jet multiplicity based, linear discriminant based, neural network based), which were employed for all measurements, are discussed in detail. In chapter 10 follows an overview of the systematic uncertainties which were considered in these measurements. The results of the first measurement are presented in chapter 11, while the results are further discussed in the next chapter (chapter 12).

The simultaneous measurement of the $t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$ signal strength and cross section as well as the simultaneous measurement of the $t\bar{t}+\geq 1b\text{-jet}$ and $t\bar{t}H$ signal strength and cross section are presented in part V. They are largely identical to one another and also to the reference analyses of the first measurement. Therefore, only the differences between the measurements are presented in chapter 13. In chapter 14 the results of the simultaneous measurements are given, before these results are discussed in chapter 15.

In part VI a study of the possible improvement of the $t\bar{t}+\geq 1b\text{-jet}$ modeling by Monte Carlo event generators is presented.

A summary and outlook given in part VII concludes this thesis.

Part I

Theoretical foundations, experimental basics, and statistical analysis methods

1. The Standard Model of particle physics

The Standard Model of particle physics (SM) is a renormalizable quantum field theory unifying the electromagnetic, weak, and strong nuclear interactions in one common framework, but leaving out the fourth important interaction observed in nature, gravity. It can be said that the to-date knowledge of particle physics is aggregated in this framework.

In the next section a historical summary of the SM will be given, before in section 1.2 the elementary particles described by the SM are introduced. In section 1.3 an overview of the three interactions which are unified in the SM will be given.

1.1 History of the Standard Model

The current formulation of the SM was finalized in the 1970s. In 1974 the discovery of the J/ψ meson [4,5] proved the existence of the charm quark as a fourth quark and confirmed the quark model described by the SM. Shortly before in 1973, the correctness of the electroweak theory contained in the SM was already shown by the discovery of neutral currents in neutrino scattering [6–8]. But in any case no later than 1983, by the discovery of the $W^\pm$ gauge bosons [9,10] and the Z gauge boson [11,12], which completed the electroweak theory, the SM became widely accepted as the default theory in particle physics.

The SM is regarded as one of the most successful theories of nature in history: Few generally-accepted theories provided more experimentally testable predictions like the existence of new elementary and composite particles which were found later in experiments and led to further trust in the SM.

Among the many discoveries that were successfully predicted by the SM of particle physics two findings are of particular significance for this thesis:

- **The discovery of the top quark**, which is discussed in more detail in section 4.1.
- **The discovery of the Higgs boson** which was the last missing piece of the elementary particle puzzle anticipated in the SM and which was already theoretically described as consequence of what was later called the Higgs mechanism by Robert

Brout, Francois Englert [13], Peter Higgs [14, 15], Gerald Guralnik, Carl Richard Hagen, and Tom Kibble [16] in 1964. The Higgs boson was finally discovered in 2012 at the Large Hadron Collider (LHC) by two independent measurements [17, 18] of the ATLAS and CMS collaboration.

However, even the successes of the SM in predicting new particles cannot mask the fact that the SM falls short of being a complete theory: The SM lacks the description and explanation of gravity as the fourth important interaction observed in nature. Up to now, there exists no verified and widely accepted way to implement the theory of general relativity into the SM. Furthermore, the SM seems to be a limiting case of a more general theory since it cannot explain the following established findings, among others:

- Observation of dark matter and dark energy in the universe, which are not part of the Standard Model of particle physics.
- Neutrinos are massless in the Standard Model, but the observed neutrino oscillations require that neutrinos have small masses.
- The hierarchy problem, a large discrepancy between the strengths of the weak interaction and gravity by a factor of 10^{32} .
- 19 to 28 free parameters (depending on the counting method) occur in the Standard Model, which is considered unacceptable for a theory of everything.

1.2 Elementary particles

According to the SM of particle physics all known ordinary matter is composed of few elementary particles which can be classified according to the interactions they take part in and also in accordance with their spin.

Fermions are particles with half-integer spins, such as $\frac{\hbar}{2}, \frac{3\hbar}{2}, \dots$, while particles with integer spins, such as $\hbar, 2\hbar, \dots$ are known as bosons.¹ Fermions – in contrast to the bosons – are subject to Fermi-Dirac statistics and obey the Pauli exclusion principle. Therefore, two or more identical fermions cannot occupy the same quantum state at the same time. Out of the elementary particles predicted by the Standard Model leptons and quarks all having spin $\frac{1}{2}$ are fermions and can be paired in three generations (electron, electron neutrino; muon, muon neutrino; tau, tau neutrino and up quark, down quark; charm quark, strange quark; top quark, bottom quark, respectively). Furthermore, each lepton and quark has an associated antiparticle with the same mass and opposite charge quantum numbers. Both the charged leptons and the quarks acquire their mass via the Higgs-Yukawa coupling, while, with the exception of the massless photon, the gauge bosons of the SM gain their mass from the Higgs mechanism.

In contrast to fermions, bosons following Bose-Einstein statistics can occupy the same quantum state in infinite numbers. A special kind of bosons occurring in the SM are the gauge bosons. These bosons carry a spin of one and represent the mediators of the three interactions described by the SM which will be explained in the following.

In addition, the SM includes the Higgs boson H^0 , which is an uncharged spin zero particle and represents the smallest possible quantum excitation of the Higgs field.

¹In the following natural units will be used in which the reduced Planck constant $\hbar$ and the speed of light c are set to $\hbar = c = 1$.

Figure 1.1 gives an overview about all the fundamental fermions and bosons described by the SM of particle physics. Additionally, it is shown which kind of interaction a particle is subject to and which gauge boson is exchanged due to this interaction: The photon as the mediator of the electromagnetic and electroweak interaction is exchanged between all charged particles. The gluons as carriers of the colour charge and gauge bosons of the strong interaction couple only to the quarks and antiquarks, while every fermion interacts via the (electro)weak interaction with the W and Z bosons.

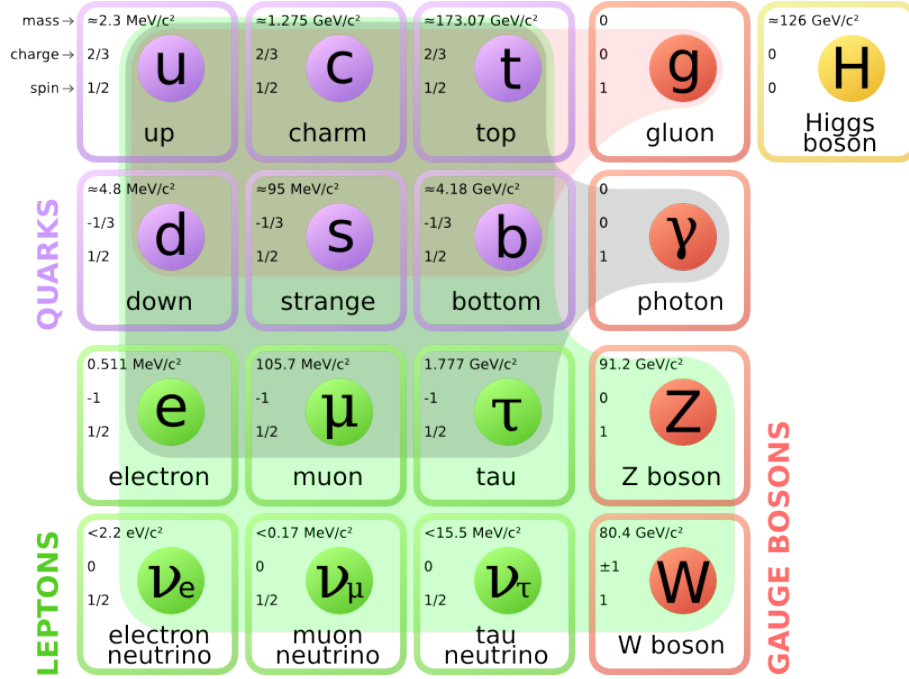


Figure 1.1: Elementary particles described by the SM of particle physics: The leptons, quarks, and gauge bosons described by the SM of particle physics are presented. The corresponding antiparticles are neglected for presentation purposes. The mass, the electric charge, and the spin is specified in the top left corner of each particle. By coloured areas the interaction a particles takes part in is shown (Dark blue colour: Electromagnetic interaction, Red colour: Strong interaction, Green colour: Weak interaction). Adapted work, based on [19].

1.3 Interactions

The SM is a gauge quantum field theory. Therefore, the elementary particles and interactions of the SM are described by quantum fields, while the dynamics of the underlying fields and the transitions between the quantum states are described by Lagrangian densities $\mathcal{L}$ from which the equations of motion for a given field can be obtained by applying the Euler-Lagrange equation.

Moreover, the SM is locally gauge invariant under transformations of the $SU(3) \times SU(2) \times U(1)$ gauge group. The three factors of the gauge group can be used to distinguish three different interactions in the SM, or rather two since the electromagnetic and weak interaction is unified in the electroweak interaction. The corresponding gauge bosons as quanta of the associated gauge fields are:

- The photon for the electromagnetic interaction and also the electroweak interaction,
- the W^+ , W^- , and Z^0 boson for the weak and electroweak interaction,

- and gluons for the strong interaction,
- all having spin one.

1.3.1 Electromagnetic interaction

The electromagnetic interaction is described by the Abelian gauge theory of Quantum ElectroDynamics (QED) [20, 21] with the $U(1)$ symmetry group. The range of its corresponding quantum field is infinite since the photon as the gauge boson is massless and does not carry electric charge itself. Nonetheless, with increasing distance the force of the field decreases according to the inverse square law which is in contrast to the behaviour of the strong interaction, where the field strength increases with distance. Furthermore, the coupling constant of the electromagnetic interaction corresponds to the fine-structure constant $\alpha \approx \frac{1}{137}$ at low momentum transfer. This constant is roughly 100 times smaller than the coupling constant of the strong interaction α_S if one neglects the running of α_S explained later. Therefore, the physics at the Large Hadron Collider (section 3.1) is in most cases in first- and often also in second-order dominated by the strong interaction processes, while electromagnetic interactions and their corresponding physics processes in many cases act as minor corrections.

1.3.2 Weak and electroweak interaction

The weak interaction was first described by the theory of Quantum Flavour dynamics (QFD) [22]. However, the weak interaction can be better understood in terms of the Electroweak Theory (EWT) [23] in which the electromagnetic interaction and weak interaction were unified in a more general, common gauge theory a few years later and which has a $SU(2) \times U(1)$ symmetry group.

The electroweak interaction couples to SM fermions. Hereby, two types of the electroweak interaction can be distinguished: The flavour- and charge-changing electroweak interaction caused by the exchange of the electrically charged W^+ and W^- bosons and the neutral-current interaction mediated by the neutral photon and Z^0 boson, which conserves the flavour and charge of the participating fermions.

By coupling to the Higgs field via the Higgs mechanism, the $W^\pm$ and Z^0 bosons gain a large mass of 80.4 GeV and 91.2 GeV [24], respectively. Furthermore, these mediators of the (electro)weak interaction can couple among themselves. Therefore and due to having a mass, the range of the (electro)weak interaction and its gauge bosons is – in comparison to the other contemplated interactions – rather short. This is also the reason why the strength of the quantum field of the (electro)weak interaction is in general rather small. Hence, the coupling constant of the (electro)weak interaction is some orders of magnitude smaller than the coupling constants of the two other interactions explaining also the name of this interaction.

In the context of the (electro)weak theory the Cabibbo-Kobayashi-Maskawa (CKM) matrix [25] should also be mentioned. It is used to describe quark-quark transitions by flavour-changing charged currents under emission of a $W^\pm$ boson.

1.3.3 Strong interaction

The strong interaction is described by the non-Abelian gauge theory of Quantum Chromodynamics (QCD) and a $SU(3)$ symmetry group. Its gauge bosons are eight gluons with

different colour-charge-states (combination of colour and anticolour). Since in contrast to a photon each gluon is a charge carrier, it cannot only annihilate into a quark-antiquark pair carrying colour charge or emerge from such one, but also interact with itself. This kind of interaction is called self-coupling and gives rise to a number of unique phenomena of QCD:

Fragmentation, hadronization and the phenomenon of confinement The strength of the strong interaction quantum field increases linearly with distance. If the distance between two originally interacting partons (quarks, antiquarks, or gluons) connected by a gluon exchange increases above the so-called fragmentation threshold, the field strength has become so large that instead of a further increase of the strength of the field fragmentation occurs. By this fragmentation a new quark-antiquark pair is created by the exchanged gluon. The freshly created quark and antiquark and the initial particles form bounding states (hadrons). Also on that account single quarks can never be observed in nature and by the phenomenon of confinement the formation of colour-neutral hadrons, termed hadronization, is always preferred.

Due to this behaviour of the strong interaction high-energetic partons (quarks and gluons) emerging in the hard interaction process or subsequent decay processes involving large momentum transfers can lead to a stream of collimated subsidiary hadrons. Due to momentum conservation these hadrons fly in a similar direction as the original partons and result in a so-called jet.

The phenomenon of asymptotic freedom The counterpart to confinement and originating from the gluon self-coupling, too, is the phenomenon of asymptotic freedom. With larger energy of the interacting partons and shorter distance between them the strength of the QCD field and the value of the corresponding strong coupling constant diminishes, thus, the strong interaction between the considered partons becomes negligible and the partons bound in a hadron behave like free particles. For this reason, perturbation theory can be applied to describe hadron-hadron scattering processes involving large momentum transfers at high energies, while the hadronization taking place at low energies and with, in comparison, rather small momentum transfers cannot be described by perturbative approaches. Therefore, it must be treated with non-perturbative approaches and semi-empirical models.

2. Hadron collider physics

Theoretical models like the SM of particle physics are tested by calculating physics observables using the corresponding theory and by comparison of the determined values of these physics observables with data obtained from measurements. One of the most important observables in collider physics is the (total) cross section of a scattering process, which corresponds to the probability that a scattering between two high-energetic particles occurs and which is briefly described in the next section. Additionally, for the more detailed study of such a physics process the simulation of entire collision events is often inevitable. That is the reason why in section 2.2 an overview about the generation of simulation data or, more exactly, the computation of events by employing Monte Carlo simulations is given.

Most of following explanations apply to all types of particle colliders in which two particle beams are collided to produce collision events and observe elementary particles. Nonetheless, there are some specific peculiarities in the case of hadron-hadron colliders like the Large Hadron Collider (LHC) (section 3.1). As in the case of other types of particle colliders like lepton-lepton and lepton-hadron colliders, physics and scattering processes can involve all three interactions described by the SM of particle physics (section 1.3). However, in contrast to the physics at lepton-lepton colliders, the presence of hadrons like protons or lead-ions in the initial state of a hard scattering process means that the Quantum Chromodynamics (QCD) interactions are in general the dominant ones and are of particular importance. Furthermore, since the two partons participating in the hard scattering were originally part of a hadron and, hence, carry only a portion of the total momentum of this hadron, the total energy and momentum of the initial-state process is smaller than the original center-of-mass energy of the collision. Therefore, in general a full event reconstruction is only possible under certain assumptions, if at all. Moreover, the proper description of hadron collider physics processes can be a challenge since the short- and long-distance behaviour of QCD has to be considered, which will be discussed in more detail in the following sections.

More details about the calculation of cross sections and Monte Carlo event generation for hadron colliders can be found in [26–28].

2.1 Calculation of cross sections

In hadron collider physics cross sections σ are used to describe scattering processes between two high-energy partons emerging from colliding hadrons. The scattering cross section $\sigma_{h_1 h_2 \rightarrow cd}$ is a measure of the probability that a scattering occurs, leading to the appearance of the particles c and d in the final state, while in the beginning two partons of the involved hadrons (h_1, h_2) interacted with each other. In a classical view, the cross section is the effective area the hadrons h_1 and h_2 have to hit to create the particles c and d with a given frequency and is stated in units of area. Typically, the unit barn ($1 \text{ barn} = 10^{-28} \text{ m}^2$) is used. The scattering cross section σ is an universal quantity and together with the time-dependent luminosity $\mathcal{L}$ the event rate $\frac{dN}{dt}$ of a specific scattering process occurring at this particle collider is given by:

$$\frac{dN}{dt} = \sigma \cdot \mathcal{L}.$$

The luminosity is a characteristic measure of the total particle collisions taking place per time and area at a given particle accelerator and can be expressed by operating parameters as follows:

$$\mathcal{L} = \frac{n \cdot N_1 \cdot N_2 \cdot f}{A}$$

Here, n is the number of colliding bunches per particle beam, N_1 and N_2 are the number of particles per bunch, f is the revolution frequency, and A is the cross section of the beam at the interaction point.

Usually the time-independent integrated luminosity $\int \mathcal{L} dt$ is used to obtain a prediction of the number of events N which should be observed for a specific physics process in a given time:

$$N = \sigma \cdot \int \mathcal{L} dt.$$

By using this relation, a cross section can be experimentally determined by counting the events of a specified physics process with regard to all the other occurring events. Hence, the cross section which can be derived from a theoretical model represents a link to a simple counting measurement while by consideration of the luminosity the specific properties of the given particle collider are taken into account.

The aforementioned cross section is a total cross section, while one can define also differential cross sections if one considers the angles or energies of outgoing particles of an event as further parameters, for example.

In the following important ingredients to calculate the scattering cross section of physics processes at hadron colliders are explained.

2.1.1 Factorization theorem

The calculation of scattering cross sections of physics processes at lepton-hadron and hadron-hadron colliders usually makes it necessary to include the short- and long-range behaviour of Quantum Chromodynamics (QCD) at different levels of the event evolution. The reason for this is that elementary particles obeying to the strong interaction can appear at any level like the initial state or final state as well at the hadronization level.

The short-distance interactions like the initial hard scattering sub process can be described

by a perturbative QCD approach due to the strong coupling constant α_S being sufficient small. By such a perturbative approach the prediction of an observable like a total scattering cross section σ can be expanded in powers of the strong coupling α_S as:

$$\sigma = \sigma_1 + \alpha_S \cdot \sigma_2 + \alpha_S^2 \cdot \sigma_3 + \dots$$

In general, including higher-orders in the perturbation calculation gets soon very complicated for many particle physics observables. Therefore, usually only leading order (LO), next-to-leading order (NLO), and next-to-next-to-leading order (NNLO) terms are applied in the calculation of particle physics observables.

The strong coupling constant α_S being small holds no longer true for long-distance interactions. Here, due to confinement (section 1.3.3), the strong coupling constant α_S is no longer a small parameter. Thus, a perturbation ansatz is not valid requiring non-perturbative and empirical models to address these long-distance interactions.

For this reason, a factorization theorem is employed for the calculation of the cross section of the scattering between the hadrons h_1 and h_2 :

$$\sigma_{h_1 h_2} = \sum_{a,b} \int_0^1 dx_a \int_0^1 dx_b f_a(x_a, \mu_F^2) \cdot f_b(x_b, \mu_F^2) \cdot \sigma_{ab}(x_a \vec{p}_i, x_b \vec{p}_j, \alpha_S(\mu_R), \mu_R, \mu_F)$$

Part of this factorization ansatz are the process-dependent, short-distance partonic cross section $\sigma_{ab}(x_a, x_b, \alpha_S(\mu_R), \mu_R, \mu_F)$ and the process-independent, long-distance interactions $f_{a/b}(x_{a/b}, \mu_F^2)$:

The calculation of the process-dependent, short-distance, partonic cross section The partonic cross section $\sigma_{ab}(x_a \vec{p}_a, x_b \vec{p}_b, \alpha_S(\mu_R), \mu_R, \mu_F)$ gives the probability of the scattering between two incoming partons which carry the momentum fraction $x_{a/b}$ of the corresponding hadron momentum $\vec{p}_{a/b}$. Nowadays, this calculation of the process-dependent, short-distance, partonic cross section is done by means of dedicated matrix element (ME) event generators like MG5aMC [29] or POWHEG Box Version 2 [30–34], using perturbative methods.

The partonic cross section depends on the center-of-mass energy of the interacting particles. Furthermore, the integrals of loops possibly occurring in the underlying Feynman diagrams (section 2.1.2) can diverge. That is the reason why a renormalization scale μ_R being part of a regularization procedure is introduced which can be seen as a momentum cut-off. By this renormalization ultraviolet divergences occurring in the integrals of the propagator terms are absorbed via a redefinition of particle masses, quantum fields, and coupling constants. Therefore, the running of the strong coupling constant α_S depends on the chosen value of the renormalization scale. Hence, the strong coupling constant is expressed as $\alpha_S(\mu_R)$.

Furthermore, similar to the renormalization scale μ_R , a factorization scale μ_F is introduced. Small-angle and low-energetic parton splittings which would lead to collinear or infrared divergences in the cross section calculation up to this factorization scale μ_F are absorbed by the parton distribution functions, making them part of the process-independent, long-distance interactions.

In the unlikely case that all orders of perturbative expansion are considered, the cross section would not depend on these two scales.

The determination of the process-independent, long-distance interactions by non-perturbative and empirical models In contrast to the ME event generators, so-called general-purpose Monte Carlo (GPMC) event generators, like Herwig7 [35, 36] or Pythia8 [37, 38], are applied to consider the long-distance interactions which are process-independent and which can be described by universal non-perturbative and empirical models.

Furthermore, universal non-perturbative parton distribution functions (see below) are employed to describe the parton content of the hadrons $x_{i/j}$. They are functions of the factorization scale μ_F .

For the illustration of the factorization ansatz in figure 2.1 the γ +jets production at a proton-proton collider is shown together with a schematic representation of the factorization model used for the calculation of the total cross section $\sigma_{h_1 h_2 \rightarrow cd}$. In the left picture, from each of the two protons, one parton participates in the hard scattering interaction while the remaining partons do not interact. The two interacting partons of the left figure are represented in the right figure by the partons a and b which interact at the vertex $\hat{s}$ creating the final-state particles c and d , while the partonic cross section of this scattering process $\sigma_{ab \rightarrow cd}$ can be computed with a ME generator using perturbative methods. The parton distribution functions f_{a/h_1} and f_{b/h_2} , which are part of the schematical view in the right figure, describe the probability of finding a parton of type a or b with a certain fraction of the momentum of the original hadron and factorize the interactions inside of the corresponding hadron by introducing a factorization scale. A further explanation of both will be given in section 2.1.3.

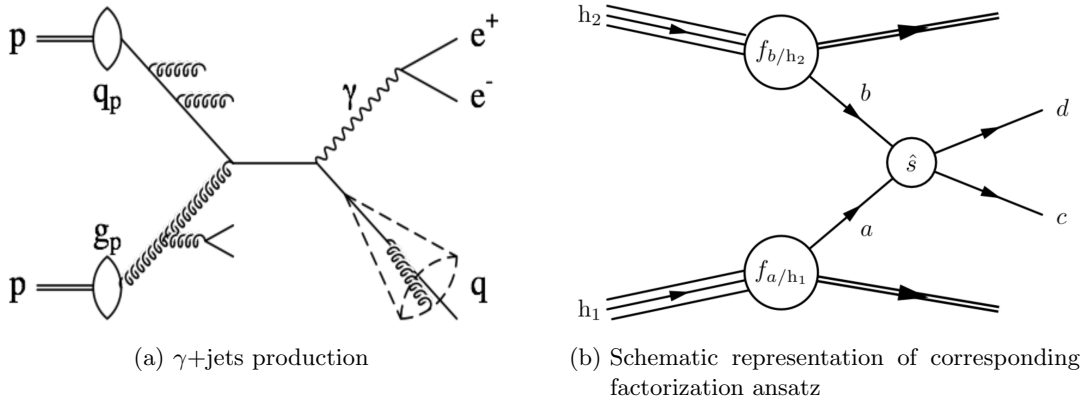


Figure 2.1: Illustrative sketch of the factorization theorem: In figure (a) the γ +jets production at a proton-proton collider is shown as an example, while in the figure (b) the schematic representation of the factorization ansatz used for the calculation of the total cross section $\sigma_{h_1 h_2 \rightarrow cd}$ is displayed. The computation of the total cross section $\sigma_{h_1 h_2 \rightarrow cd}$ computation is split in two parts: A calculation of the partonic cross section $\sigma_{ab \rightarrow cd}$ at the interaction vertex $\hat{s}$ using a perturbative method and the determination of the long-distance interactions making use of two parton distribution functions f_{a/h_1} and f_{b/h_2} . Slightly modified from [39].

2.1.2 Feynman diagrams and perturbation theory

Feynman diagrams, introduced by Richard Feynman in 1949 [40], are an important tool in particle physics to illustrate hard scattering processes since they represent the perturbative transition from an initial to a final quantum state. Together with Feynman rules they also allow the computation of hard scattering processes. Hence, they translate the knowledge of an quantum field interaction described by a theoretical model like the SM of particle physics (chapter 1) into a simpler to understand graphical representation of an underlying mathematical term. The in- and outgoing lines of Feynman diagrams applied in particle physics (figure 2.2) represent the wave functions of on-shell elementary particles in momentum space. Each external line depicts the four momentum and the spin of an individual (in principle measurable) particle, while inner lines correspond to the propagators of the underlying quantum field interaction. They can be seen as virtual particles. Different inner and outer lines are connected to each other via interaction points called vertices. These vertices are the space-time point at which one of the three fundamental interactions described by the Standard Model of particle physics takes place. The associated mathematical interaction term can be derived from the Standard Model Lagrangian as a vertex factor.

Feynman diagrams are a description of a physics process in space-time. They have two axes: One axis is containing the time coordinate, while the other axis represents the space component. That is why the exemplary Feynman diagram shown in figure 2.2 can be understood in multiple ways depending on the choice of the time and space axis. One can make out an electron-positron annihilation followed by a muon-pair production (time scale in horizontal direction) or an electron/positron-muon scattering process (time scale in vertical direction). Both possible processes are mediated by a photon.

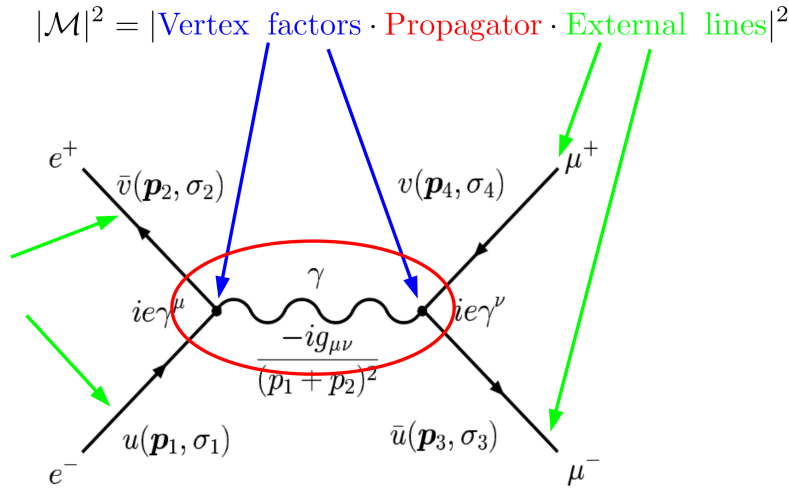


Figure 2.2: Exemplary Feynman diagram containing leptons: The external lines represent the wave functions of possibly measurable particles, while the two outer lines are connected by two vertices via an inner line or rather an electromagnetic propagator term corresponding to a virtual photon. Note that the time and space axis can be chosen either in horizontal or vertical direction. Taken from [41].

Feynman diagrams and the corresponding determination of the cross sections can be classified according to the order of perturbation theory they depict. For example, if in a cross section calculation only leading-order (LO) Feynman diagrams are taken into account for the hard scattering process, the cross section result will also only reach leading-order accuracy. Obviously, it is worth pursuing to include higher orders in the perturbation series

to achieve a more complete and more precise description of a physics process. However, higher orders of perturbation come with a cost. They will involve additional radiation and loops which often present a challenge to include and evaluate in a calculation. That is the reason why the production of Monte Carlo events (section 2.2), as a general rule, is nowadays done mostly in next-to-leading order (NLO) and seldom makes use of orders higher than next-to-next-to-leading order (NNLO) of perturbation theory. In fact, NLO event generation for various physics processes in an automated way was not feasible until in 2007 a program making use of a reduction method to compute one-loop amplitudes [42] was released.

An example of Feynman diagrams containing possible different orders of perturbation theory for the $pp \rightarrow t\bar{t}H$ process is given in figure 2.3. In this figure, besides a LO Feynman diagram which is also called a Born level diagram, three additional NLO Feynman diagrams are presented: One diagram containing a loop in the propagator (virtual) as well as diagrams representing either initial-state radiation (ISR) or final-state radiation (FSR).

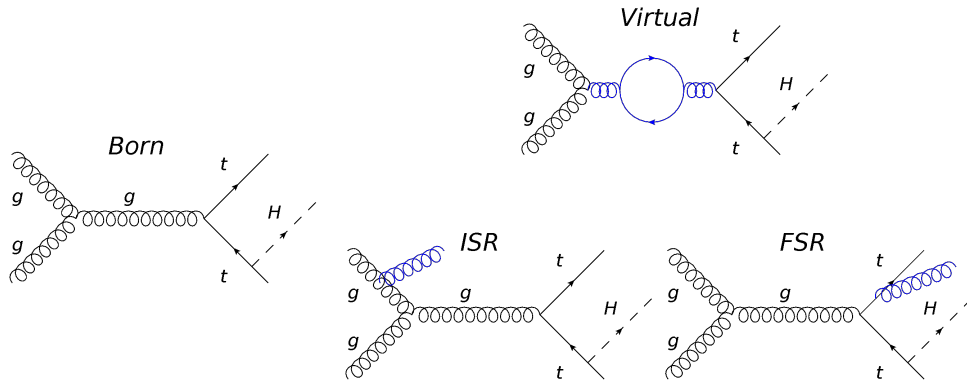


Figure 2.3: Example of Feynman diagrams containing different orders of perturbation theory: A selection of Feynman diagrams describing the $pp \rightarrow t\bar{t}H$ is shown, a LO only, Born level diagram and three additional NLO diagrams. One diagram depicts a virtual loop process and the other diagrams represent either initial-state radiation (ISR) or final-state radiation (FSR).

2.1.3 Parton distribution functions

As shown above, parton distribution functions (PDF) are used to describe the parton content of hadrons like protons in a non-perturbative way. In this way, they conceal interactions taking place inside of the corresponding hadron below a certain scale, the factorization scale μ_F . Since the colliding protons are composite particles their contained valence quarks interact with each other by gluon exchange. At the same time the gluons can temporarily create virtual quark-antiquark pairs in accord with the time-energy uncertainty principle. On this account, not only the valence quarks of the proton but also the exchanged gluons and temporarily created sea quarks can be a collision partner in a hard scattering process with an energy scale Q . Since the content of a hadron should be universal due to the mentioned uncertainty principle the parton distribution functions can be seen as universal. For this reason, parton distribution functions are determined in various experiments studying deep-inelastic scattering (DIS), e.g. at lepton-hadron colliders, and also at different energy scales Q . However, the inner structure of the hadron which is described by a parton distribution function depends on the energy scale Q at which it is probed: At lower energy scales Q the energy of the probe particle like an electron can only be enough to scatter hadrons as a whole. In such a case the hadron appears more like a fundamental particle, while by using a highly-energetic probe particle the inner structure

of the hadron can get resolved. For the evolution of parton distribution functions measured at lower energy scales Q_{low} to higher energy scales Q_{high} the DGLAP evolution equations ([43–45]) can be employed. These equations allow the determination of the change of the parton density in relation to the energy scale variation Q . As a result, parton distribution functions derived from measurements at fixed-target DIS and proton-nucleon experiments as well as DIS data from the HERA $e^\pm p$ -collider and jet data from the Tevatron $p\bar{p}$ collider may be used as the foundation of a set of parton distribution functions applied for the physics analyses at the Large Hadron Collider.

2.1.4 Description of a physics process in the four-flavour and five-flavour scheme

The physics analyses covered in this thesis study physics processes involving bottom quarks occurring in the initial state or final state of the matrix element computation. Due to the non-negligible mass of the bottom quarks, the cross section computation and event generation for such processes must be treated differently than for processes containing only lighter quarks. Two schemes exist to describe such processes, the four-flavour and the five-flavour scheme. Both schemes, having advantages and disadvantages, cover different needs of a physics analyses, but if one would consider all orders in perturbation theory both schemes could be defined in a way to be identical. This section, following the far more detailed reference [46], will give a short introduction to both flavour schemes.

Four-flavour scheme

In the four-flavour scheme (4FS) or also the so-called “massive” scheme it is assumed that due to their large mass bottom quarks cannot be part of the composite proton below the factorization scale μ_F and that they only appear as quark-antiquark pairs in scattering processes at a high energy scale Q_{high} . Furthermore, the assumption is made that the mass of the bottom quark is of the same order as the energy scale Q of the hard scattering process. As a consequence bottom quarks are not considered in the parton distribution function of the colliding protons. Therefore, this scheme can be seen as an effective theory in which the bottom quarks do not enter the evolution of the parton distribution functions via the DGLAP equations or via the computation of the running coupling constant of the strong interaction α_S .

The advantage of this scheme is that “the full kinematics of the heavy quarks are taken into account already at leading order” [46] in the calculation of perturbation theory. Hence, observables which have a strong dependence on kinematics are in general better described by this scheme. Moreover, this scheme can easily be adopted by the parton shower codes of general-purpose Monte Carlo event generators. However, the scheme has also some disadvantages: Due to the presence of final states with massive bottom quarks and multiple legs the computation is usually more complicated than in a scheme with massless bottom quarks. Furthermore, the occurrence of possibly large logarithms in the initial and final state, which are not resummed, can lead to a poor behaviour of the perturbative expansion and can lead to an inaccurate prediction of the total cross section.

Five-flavour scheme

The five-flavour (5FS) or also the “massless” scheme is a natural choice if the typical energy scale Q of the hard scattering is rather large in comparison to the bottom-quark mass. In this scheme bottom quarks can appear both in the initial state and final state since the applied parton distribution function considers also bottom quarks in addition to the light quarks. By including the bottom quarks in the parton distribution functions large

logarithms appearing in the initial state can be resummed into the bottom-quark distribution function of the corresponding parton distribution function. Logarithms occurring in the final state can either be resummed into perturbative fragmentation functions or be possibly avoided by introducing more inclusive observables like using bottom jets and their properties instead of using bottom hadrons.

An advantage of this scheme is usually that the corresponding calculations are highly simplified, since at leading-order of perturbation theory the number of occurring external legs as well as the number of energy scales to consider in the process compared to a four-flavour scheme is reduced. “In addition, as mentioned above, potentially large logarithms [...] arising from collinear splitting of the initial heavy quarks and gluons, are consistently resummed in the heavy quark PDF” [46]. Therefore, this scheme is generally a very suitable choice to predict rather inclusive observables like the total cross section of a given process. However, the downside of this scheme is that the results can differ strongly between the orders of perturbation theory which were chosen for the computation. This is especially the case for observables like kinematic quantities which would benefit from higher orders of perturbation theory.

Due to the occurrence of two top quarks in the final states of the signal processes studied in this thesis, the assumption that was made for the four-flavour scheme, that the energy scale of the hard scattering process Q is of the same order as the bottom quark mass, is questionable. Since the applicability of four-flavour schemes and five-flavour schemes for physics processes involving top quarks and the Higgs boson at the LHC cannot easily be decided on and is an ongoing discussion, a good starting point for an interested reader to learn more about the ongoing studies would be the $t\bar{t}+b\bar{b}$ production studies presented in [47].

For the generation of the $t\bar{t}+X$ samples employed by default in this thesis the five-flavour scheme was applied. This seems a suitable choice since in the physics analyses presented later the major focus will be on the determination of the $t\bar{t}+\geq 1b$ -jet signal strength r comprising of the $t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, and $t\bar{t}+b$ physics processes. Kinematic observables are only used as inputs for the multivariate classifiers to derive this signal strength r .

2.2 Monte Carlo event generation for hadron colliders

As discussed before, single physical observables like the total cross section can be calculated using a factorization ansatz, parton distribution functions, Feynman diagrams and further ingredients. In principle, an analytic solution of the cross section calculation or the determination of another single basic observable like the transverse momentum p_T of an initial-state top quark is possible for many physics processes of importance conceivable up to a few orders in perturbation theory. For example, since 2013 there exists a fully analytic prediction of the total $t\bar{t}$ production at next-to-next-leading order (NNLO) QCD accuracy [48] and since 2017 a prediction of the top p_T observable of the $t\bar{t}$ process among others is available at NNLO QCD accuracy including next-to-leading order (NLO) electroweak corrections [49].

Since the analytical and numerical calculations provide the most accurate predictions of cross section results and some other basic observables they are preferable whenever they exist. However, in many cases the analytical or numerical computation of more complex observables measurable by a particle detector or even differential cross sections is not feasible. Therefore nowadays, a two-prong approach is applied: Entire collision events are simulated using event generators based on Monte Carlo methods allowing to make use of the full event kinematics in physics analyses, while the cross section of the simulated event

data set is obtained by scaling the event weights to the corresponding theoretical cross section obtained from a precise analytical or numerical calculation.

The generation of single collision events by event generators is performed similar to the calculation of a cross section explained before. However, to obtain an event for a given physics process and a corresponding Feynman diagram a Monte Carlo random sampling method is chosen which selects the particles occurring in the initial state and final state of the hard scattering process as well as their energy and momentum distribution. Other decisions on the event topology like the occurrence of gluon splitting are also being made by using Monte Carlo sampling methods. Furthermore, integrals which emerge during the computation of an event are usually solved by Monte Carlo integration since the multi-dimensional integrals can be arbitrarily complex and often cannot be solved by other numerical methods. In comparison to the calculation of a (total) cross section no integration over the full phase-space is conducted.

Subsequently, an outline of the different stages of the event generation procedure will be given. It is followed by a description of the tuning of event generators (section 2.2.2), before this chapter will be concluded with a short discussion of matching and merging methods. These are important techniques to increase the accuracy of the produced event samples.

For more details about Monte Carlo event generation see [26–28].

2.2.1 Event generation

As mentioned above, event generation based on Monte Carlo methods is applied “to simulate the final states of high-energy collisions in full detail down to the level of individual stable particles. The aim is to generate a large number of simulated collision events, each consisting of a list of final-state particles and their momenta, such that the probability to produce an event with a given list is proportional (approximately) to the probability that the corresponding actual event is produced in the real world.” [50]. By taking advantage of the aforementioned factorization ansatz (section 2.1.1) this event generation procedure to simulate events at hadron colliders is usually split into different steps. This allows the usage of the best set of tools in every step and to simulate an event in full detail. A short overview of the different steps, which follows closely the illustration of the event generation process presented in figure 2.4, will be given in the subsequent paragraphs. It starts with the computation of the hard scattering process by a matrix element event generator and ends with the simulation of a detector.

The computation of the hard scattering process

For a given physics process the event generation starts with the computation of the partonic cross section and the simulation of the hard scattering process. As we have seen before, the hard scattering process is defined by the interaction of two highly-energetic partons originating from hadrons (here: protons) which themselves are part of two colliding particle beams. The interaction of these two partons leads to a partonic final state containing a few partons, leptons, and possibly other elementary particles. Hence, events of this stage of event generation procedure are also called parton-level events in experimental particle physics. Since from each of the two protons only one parton takes part in the hard scattering process, the remaining partons of the proton (depicted as black outgoing lines in figure 2.4) do not contribute to the hard interactions, but they will later form the underlying event.

The hard scattering interaction takes place with a high momentum transfer. Therefore,

the strong coupling constant α_S can be assumed to be rather small and the hard scattering process can be described by perturbation theory. The order of perturbation theory used in this step defines the achievable accuracy of the whole event simulation. For this reason, nowadays, dedicated matrix element generators like MG5aMC [29] or POWHEG Box Version 2 [30–34] are used which allow to describe the hard scattering process in leading-order or next-to-leading order QCD accuracy and which in many cases allow to consider quantum interference effects in the initial state and final state. Furthermore, especially processes which have rather highly-energetic partons in the final state like many top physics processes can benefit from higher-order matrix element computations: They get described better by additional matrix elements with larger multiplicities (number of additional partons in a final state besides the nominal partons). For example, their description is improved if one or more additional partons besides the partons of the nominal Born-level process are considered in an additional next-to-leading order matrix element. However, matrix element computations are generally limited to some number of partons in the final state since the computation of the underlying matrix elements can be complex, involves much computing, and increases drastically with the multiplicity of a given physics process. That is one of the reasons why matrix element generators are not applied to describe the evolving parton shower after a hard scattering which involves a multitude of partons. The other reason is that the evolving parton shower cannot be described fully by a perturbative approach.

Parton shower, hadronization, and hadron decay

After the simulation of the hard scattering process follows first a so-called parton shower step which is usually simulated with a general-purpose Monte Carlo (GPMC) event generator like Herwig7 [35,36] or Pythia8 [37,38]. Such a step is necessary to obtain an inclusive event sample. The partons emerging during the hard scattering process and carrying a colour charge can emit QCD radiation in the form of gluons which leads to a cascade of parton creation and parton splitting processes, the so-called parton shower. This parton shower goes on until the energy of the individual colour-charged partons is decreased so much that the parton reaches the hadronization scale which is roughly in the order of one GeV. At this scale the formation of colourless hadrons out of single colour-charged partons is possible due to the increased strength of the strong interaction (cf. pink bubbles in figure 2.4). This process is called hadronization. The arising hadrons can either be stable or decay further. However, these hadrons are colourless and, hence, observable particles, while the single colour-charged partons are not due to colour confinement. Events obtained from this stage of event generation are called particle-level events (also: generator-level events) since in strict terms the event generation is complete at this point. However, in this way, multiple parton interactions and the underlying event as well as the possible need of considering pile-up events and of a subsequent detector simulation would be neglected.

Multiple parton interactions and the underlying event

So far, the remaining partons of the original protons, which did not participate in the hard scattering process, and their interactions were neglected. Since the most common hard scattering process at the Large Hadron Collider is elastic scattering of gluons $gg \rightarrow gg$ and the corresponding total cross section can even be larger than the total proton-proton scattering cross section, which is not a contradiction, multiple parton interactions are highly probable [50]. From these multiple parton interactions arise further partons and subsequent hadrons in an event which have to be considered by a Monte Carlo simulation as so-called underlying event, even though these partons and hadrons do not originate from an identifiable hard sub process.

Pile-Up events

Each proton beam of the LHC consists up to 2808 bunches at full beam intensity, while each bunch itself will roughly contain 10^{11} protons at the start of a nominal fill. As a consequence, it is expected that at each bunch crossing not only one proton-proton collision occurs but in general somewhere between 20 and 50 simultaneous collisions. In most of these collisions only elastic scattering processes or low-energetic inelastic scattering processes will happen. Nonetheless, these unwanted and rather uninteresting collisions and their corresponding partons and hadrons, which are called pile-up events, can emerge at a similar time as or near to an collision containing a hard scattering process of a physics process of interest. Thus, it is quite likely that a collision event measured and recorded at the CMS detector (section 3.2) also contains particles from pile-up events. For this reason, besides the simulation of the physics processes of interest, pile-up events are computed by Monte Carlo event generators, too. For each simulated event of a physics process of interest a number of additional pile-up events is randomly sampled and randomly drawn pile-up events are merged with the initial event.

Detector simulation

In experimental particle physics the events obtained by means of a Monte Carlo simulation have to be compared to the measurement data derived from a particle detector like CMS, which is explained in section 3.2. Such a real-world particle detector cannot directly determine the particles occurring at the aforementioned particle-level. As a result it has to rely on an indirect detection and measurement of these particles by making use of interactions between the produced particles and the purposefully-chosen detector material and detection systems (section 3.3). Furthermore, the single measurement components used in the detector only have a limited measurement precision and object reconstruction probability for physical and technical reasons. Also the hermetic coverage of the collision point by the detector itself is not totally perfect. Thus, particles and decay products emitted during the collision process can escape the detector without being accounted for or only being partly considered in a subsequent reconstruction step.

This has to be taken into account by a simulation of the full detector, its subcomponents, and of the successive object identification and event reconstruction step to obtain simulation events comparable to events measured by the real detector. Since the detector simulation is tailored to the properties of the specific given detector the derived simulation events are no longer generally applicable and are called reco-level events after the final reconstruction step they have passed.

For the simulation of their detector the CMS Collaboration applies the Geant4 framework [51].

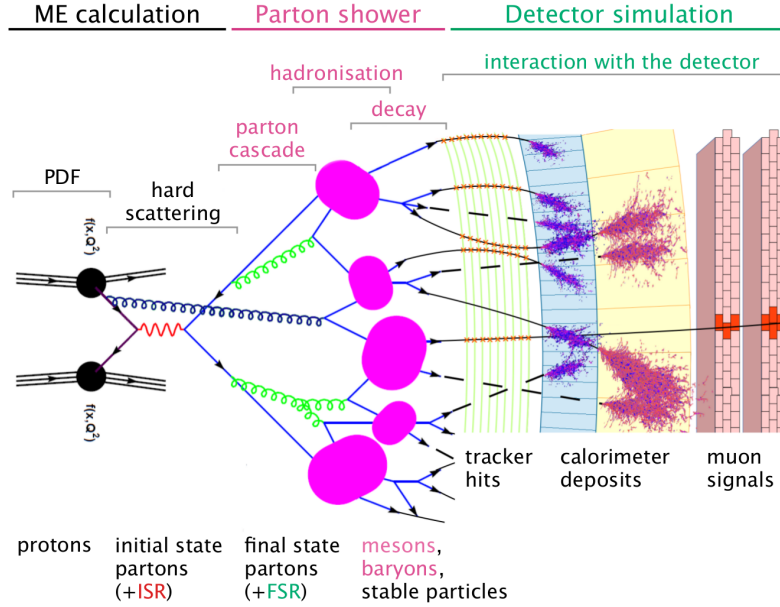


Figure 2.4: Illustration of the event generation procedure: The event generation starts with the computation of the hard scattering process making use of parton distribution functions (PDFs) and of matrix element (ME) calculations and results in a few partons in the final state (two blue lines representing a quark pair and possibly a gluon originating from an interaction in the initial state). It follows a parton cascade since the high-energetic, colour-charged partons appearing either in the initial state or final state will undergo parton splitting (further green and blue lines) until they have lost enough energy to reach the point of hadronization at which they form colourless hadrons (pink bubbles). The hadrons can decay further or can be stable. A simulation of the detection, measurement, and reconstruction of the obtained hadrons by a real-world detector concludes the event generation procedure. Taken from [52].

2.2.2 Tuning of event generators

The approximation of physics processes by Monte Carlo event generators involves a number of free or at least not fully constrained parameters, for example, the previously mentioned energy scales. Many of these parameters like parameters of hadronization described by phenomenological models cannot be deduced or even fixed from first principles. At the same time the choice of these parameters affects the prediction of Monte Carlo generators and, in particular, how well data is described by them. Therefore, these parameters must be matched to large sets of experimental data, so that a wide range of physics can be properly described by a Monte Carlo generator. This adjustment of event generator parameters is referred to as tuning [53].

2.2.3 Matching and merging

Matching and merging methods are an important tool set to increase the accuracy of event samples obtained from a Monte Carlo event generator since the matrix element (ME) generators and the general-purpose Monte Carlo (GPMC) generators used at different stages of the event generation procedure have both advantages and disadvantages: The ME generators which are applied to simulate the hard scattering process deliver sound results if the interacting partons can be resolved in a fixed order calculation by having high energy (also referred to as being hard) and being well separated. In contrast to that the GPMC generators based on non-perturbative methods and empirical models are better suited to describe the low-energetic interactions and collinear splittings happening at a later stage of the event generation.

For this reason, for the production of simulation data both complementary types of event generators are combined by the means of so-called matching methods since a while. By using such matching methods it is ensured that the parts of the phase-space which could in principle be described by both event generator types simultaneously are only filled once by either one of the two event generators or by a clever combination of them. So, the major motivation of a matching method is to avoid the so-called double-counting. Double-counting describes that additional partons occurring in a given physics process can either emerge from the consideration of matrix elements with multiplicity greater than zero in the ME generator or by the subsequent application of a parton shower. For the production of the simulation samples used later in this thesis, either the MC@NLO matching prescription [54] in the case of the MG5aMC samples or the POWHEG matching prescription [31] in the case of the POWHEG Box Version 2 samples was applied.

As mentioned earlier, it is worth striving to include further matrix elements in next-to-leading order or even higher order of perturbation theory in the computations of the matrix element generators because this results in an improved accuracy of the simulation. Besides of adding higher-order matrix elements the description of a physics process can also be improved by considering additional matrix elements of a process which just has a higher multiplicity. Then, initial and final states of the overall simulated physics process containing a few additional partons besides the nominal ones are not only obtained by a combination of matrix element computation and parton shower. They can also be directly derived from the additional matrix elements of the added sub process.

However, all of these improvements in the matrix element description of a given physics process make it necessary to apply further prescriptions in the event generation procedure to avoid an overlap in phase-space, so-called merging prescriptions. For example, these merging prescriptions can take into account that a leading-order matrix element for a sub process with a given multiplicity in principle can correspond to a next-to-leading order

matrix element with the same multiplicity minus one. This would result again in a double-counting in certain areas of the phase-space without the application of a merging technique. Among the best-known merging prescriptions are the MLM merging prescription [55] which is used by MG5aMC in the case of the combination of leading-order sub processes, the FxFx merging prescription [56] used by MG5aMC for the combination of next-to-leading order sub processes, and the various CKKW(-L) merging prescriptions [57–60].

3. Experimental basics

In physics like in any other scientific field the best theory is worthless if the theory is not tested by conducting experiments. Trust in a theory is created if its predictions are established experimentally and proven by results. As science progresses and the generally accepted theoretical models evolve, it seems inevitable that the experimental setups, experimental tools, and experimental methods have to advance, too.

In the case of particle physics and, especially, in the context of the in-depth exploration of the Standard Model of particle physics (section 1) this unavoidably leads to the study of particle processes or more precisely particle collisions involving larger and larger momentum transfer and, closely related, higher energies of the involved particles as well as particle accelerators providing higher luminosities. Meeting the first condition means that the theory can be examined in novel and until now inaccessible phase-space regions. As a result it can be checked if the established theory like the SM holds true and can be extended to these regions without modifications. In the case of the higher luminosity the more frequent occurrence of a particle process can be used to reduce the statistical and in many cases also the systematic uncertainties related to a measurement of such a particle process. This drives the involved experimental and theoretical boundaries further down. The Large Hadron Collider (discussed in the next section) as the till now most powerful and sophisticated particle accelerator fully meets the first condition and also provides the largest-ever luminosity of a proton-proton collider¹. Therefore, at the LHC particle physics analyses can be conducted at hitherto unknown energies and precision.

Studying particle collisions with a high degree of accuracy would not be possible without well-developed tools like the Compact Muon Solenoid detector (discussed in section 3.2) and matching object identification and event reconstruction methods (presented in section 3.3).

3.1 The Large Hadron Collider (LHC)

Particle accelerators relying on the same principles and technological foundations as modern accelerators were built since the 1920s. Electrostatic particle accelerators using static electric fields are the first type of particle accelerators still in use today. However, the

¹A up to ten times larger luminosity of roughly $1 \times 10^{35} \frac{1}{\text{cm}^2 \text{s}}$ will be reached in the high-luminosity upgrade of the LHC [61, 62], while the SuperKEKB e^+e^- particle accelerator [63] being a so-called (bottom-meson-) b-factory provides a larger peak luminosity of $8 \times 10^{35} \frac{1}{\text{cm}^2 \text{s}}$.

achievable final energy of the charged particles accelerated by them is limited due to electrical breakdown. Thus nowadays, these particle accelerators have a mere existence as pre-accelerators or as accelerators in experimental setups requiring rather small particle energies below a few MeV. Meanwhile, electrodynamic particle accelerators make use of electromagnetic fields to accelerate charged particles. They have the advantage that if the accelerator is circular-shaped the charged particles can pass through an acceleration field many times. In this way, the obtainable final energy of these particles is in principle only limited by the magnetic field forcing the particles on circular orbits. For this reason, synchrotron accelerators like the Large Hadron Collider (LHC) [64] represent the technological work horse of modern particle physics.

The LHC is part of the accelerator complex at CERN, the European Organization for Nuclear Research, and the last element of CERN's main accelerator chain (figure 3.1). In this chain a succession of machines is used to accelerate either protons or lead ions to increasingly higher energies until in the LHC a beam energy of up to $\sqrt{s} = 6.5$ TeV and a luminosity in the order of $\mathcal{L} = 2 \times 10^{34} \frac{1}{\text{cm}^2\text{s}}$ is reached in proton-proton collisions.

The LHC was built in the same 27 km long tunnel as the previous Large Electron-Positron Collider (LEP). This former collider was used to accelerate electrons and positrons. In comparison to LEP the LHC was purposely designed as a proton-proton collider. The reason for this is that the energy loss due to synchrotron radiation which goes roughly by $\frac{1}{m^4}$ is a lot smaller for protons than for lighter electrons and positrons. Therefore, far higher collision energies can be obtained by the LHC. For the same reason of reaching higher collision energies while keeping feasible collider dimensions, instead of using a linear collider design, a circular collider design is used. Hereby, the charged protons or lead-ions are forced on a circular trajectory in the LHC by a magnetic field of up to 8.33 T originating from 1232 superconducting dipole magnets. This allows the acceleration of the beam particles in each revolution until they reach their final energy. The circular collider design of the LHC has also the advantage that the beams can be used over and over again. Thus, a far higher luminosity than in a linear collider can be reached. This goal of reaching a large luminosity lead also to the decision to build two separate beam tubes and to collide two opposite-directed proton beams. In contrast to that, a proton-antiproton collider needing only one beam tube would have been a lot easier to construct, but would have provided far less luminosity due to the impossibility of producing the same large number of antiprotons as protons.

Protons gain additional energy if they pass through an accelerating cavity at the same time when the radio frequency (RF) field has the correct orientation. That is the reason why in the LHC not a constant proton stream is used. Instead each proton beam in the LHC consists of up to 2808 bunches containing approximately 1×10^{11} protons. The time spacing between these bunches is at least 25 ns. Furthermore, this bunching of protons is already introduced in the beginning of the accelerator chain: The reason for this is that only a given number of protons can be obtained from neutral hydrogen neutrons simultaneously. Furthermore, only a limited number of protons can be accelerated in the RF cavities of the linear accelerator (LINAC) at the same time. The acceleration in the RF cavities is also direct cause of the bunching. Only protons which are in sync with the RF field get accelerated sufficiently, so that they will reach the BOOSTER as the next accelerator in the accelerator chain. Protons which fall out of sync will be removed by dedicated techniques. Otherwise, bunches would not have a homogenous energy distribution of the protons and would not be properly accelerated as well as inserted into the following accelerators.

At four so-called collision points out of the eight LHC beam crossing points the two counterrotating proton beams cross in such a manner that protons from each of the two bunches pass the collision point at the same time and can interact with each other. The four major particle detectors of the LHC (ALICE, ATLAS, CMS, and LHCb) are placed at these collision points to measure the collision products.

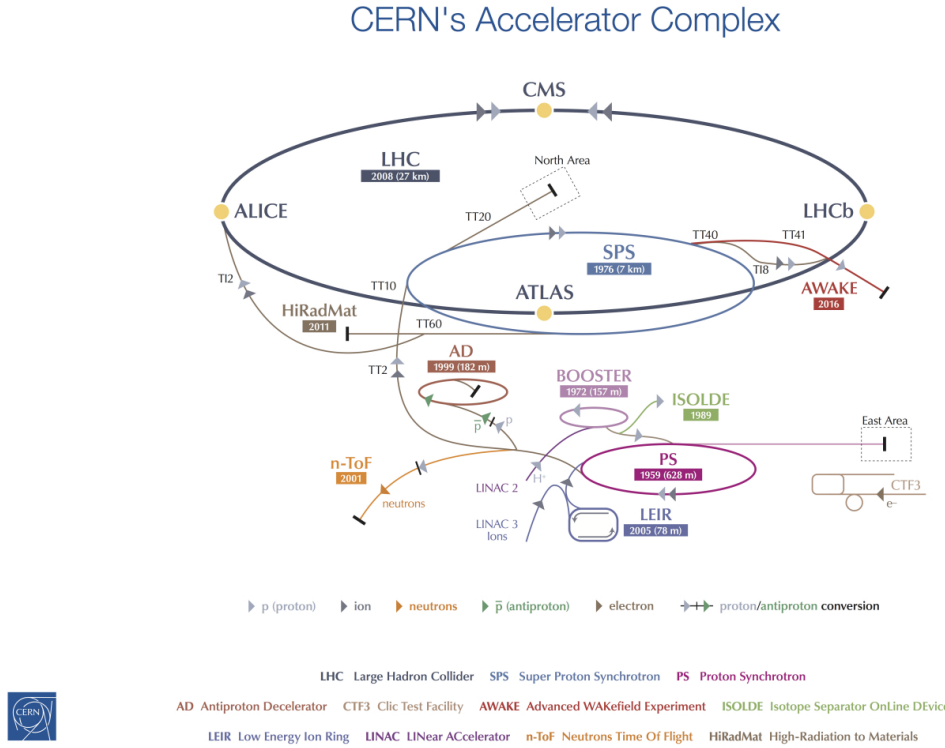


Figure 3.1: CERN's accelerator complex: The Large Hadron Collider (LHC) is the last and most powerful accelerator in the main accelerator chain. The two counterrotating proton or lead-ion beams are accelerated in the LHC up to their final center-of-mass energy after being transferred from the Super Proton Synchrotron (SPS), which itself gets its beams from a chain of smaller accelerators. Taken from [65]

3.2 The Compact Muon Solenoid detector

The occurrence of a multitude of elementary particles and composite particles in the beam-beam collisions and the subsequent particle and antiparticle cascade at the LHC together with the wide variety of possible physics analyses places rigorous demands on the detection and measurement capacities of a particle detector at the LHC. In particular many different particles types have to be measured over a large energy and momentum range, while these particles also leave the collision point at various angles. This need is met by particle detectors arranged in an onion-like structure and comprised of several sub detectors. Each of these sub detectors is an advanced system itself and fulfills varying individual requirements, while the entire detector should accurately detect and measure as many particles involved in the collision as possible to achieve an almost hermetic geometric coverage of the collision point.

The Compact Muon Solenoid (CMS) detector [66] [67], which is located at beam crossing point 5 of CERN’s Large Hadron Collider and which is operated by the independent CMS collaboration, was designed in due consideration of these requirements on a general-purpose particle detector. It consists of a central barrel region complemented by endcaps on both sides of the outward beam line resulting in an overall length of roughly 21.6 m, a diameter of approximately 14.6 m and a total weight of more than 14 000 t. The whole CMS detector is dominated by a superconducting solenoid with an overall length of 13 m and a diameter of 7 m. The solenoid encloses all sub detectors except the muon detection system. The placement of most sub detectors inside the solenoid is also the cause of the rather compact detector outline. The following is a more detailed description of the onion-like structure of the CMS detector from the inner core around the beam pipe to the external muon detection system (figure 3.2). It ends with a subsequent short discussion of the trigger system.

Coordinate system Collisions taking place in the CMS detector are described by an orthogonal right-handed coordinate system: The x axis of this coordinate system points toward the center of the LHC accelerator ring. Since the z axis points counterclockwise in the direction of the beam pipes, the x and z axis form the plane of the accelerator ring. The y axis points upwards in an orthogonal direction to this plane of the accelerator ring. Furthermore, two angles are defined: The angle ϕ defines the azimuthal angle to the x axis in the xy plane, while θ is the polar angle measured from the z axis. However, more commonly the pseudorapidity η is used instead of the polar angle θ . It is defined as:

$$\eta = -\ln \left[\tan \left(\frac{\theta}{2} \right) \right]$$

An advantage of the pseudorapidity η is that it covariant under boosts along the z axis in the limit of massless particles.

Additionally, the distance ΔR

$$\Delta R = \sqrt{(\Delta\phi_{ab})^2 + (\Delta\eta_{ab})^2}$$

as a measure of the distance between two objects a und b is defined in this coordinate system.

Silicon-based tracking system The CMS tracking system [68] is constructed as the first layer of the CMS detector directly around the collision point and beam pipe. It allows to identify the trajectory of centrally ($|\eta| < 2.5$) produced charged particles.

Charged particles having a transverse momentum of at least 1 GeV can be identified with an excellent resolution of up to 10 μm in the x-y direction and 20 μm in the z-direction. With decreasing efficiency the spatial resolution of charged particles with a lower transverse momentum can still be identified up to a transverse momentum in the order of 10 MeV. Similarly, the momentum of charged particles can be precisely determined (e.g. less than 2% uncertainty on the p_T value for muons with a p_T of at least 100 GeV) since the charged particles traverse the detector on bent trajectories due to the magnetic field of the solenoid [68]. By the combination of both pieces of information it is possible to reconstruct the trajectories of single charged particles. This enables a distinction between primary vertices, geometric points at which the primary protons interacted with each other, secondary vertices, which are originating from the delayed decay of particles produced in the primary collision into secondary charged particles, and pile-up vertices, geometric points at which partons other than those of the primary collision interacted with each other (section 3.3.1). Especially, for the identification of jets stemming from b quarks (the b-tagging of jets is discussed in section 3.3.6) this vertex information is very important. Such b-tagged jets quite often occur in physics processes involving Higgs-boson and top-quark decays. Usually

they lead to the occurrence of secondary vertices.

However, since more than 1000 charged particles are produced at each bunch crossing a good spatial resolution, a fine granularity, and a fast response time for the tracking system is required to properly identify vertices and tracks. In addition, the tracking system has to withstand the large particle-flux around the collision point, while it should have a low material budget preventing that passing particles lose energy in inactive detector material. A tracker design based on two types of silicon-based semiconductor detectors was chosen to fulfill this needs and to keep the total costs in line with the budget: The innermost layers are composed of the so-called pixel tracker consisting of approximately 66 million pixels. More precisely, 66 million pn-junctions and corresponding read-out electronic circuits are arranged in three central layers in the barrel region and two endcap disks on either side. This sub detector is located just four to ten centimeters away from the beam line. This is an apt choice to endure the high particle flux and to offer fine granularity at the same time. This setup is completed by layers of the so-called strip tracker consisting of over 9 million silicon strips in roughly 15 000 modules. The strip tracker, even though being coarser-grained, is still a suitable and simultaneously cost-effective solution due to the lower particle flux further away from the collision point .

The CMS pixel detector was replaced in the 2016/2017 year-end shutdown as part of the so-called Phase-1 upgrade [69]: Among many other changes the number of pixel layers in the barrel region was increased from three to four layers and the total number of the pixels was increased to 124 million pixel. Given that this thesis makes use of the 2016 CMS measurement data the basic layout of the old pixel detector was mostly referred to, but the overall requirements and concepts still hold true.

Electromagnetic calorimeter The electromagnetic calorimeter (ECAL) [70] [71] starts only 1.3m away from the collision point and hermetically encloses the tracking system in a pseudorapidity range of $|\eta| < 3.0$. It is an important ingredient to identify and measure the energy of electromagnetically-interacting particles, specifically for the detection and energy measurement of electrons and photons. In the ECAL more than 75 000 PbWO_4 crystals are applied as scintillators in which the energy of electrons and photons is measured by detecting the scintillation light produced by electromagnetic showers. Similar requirements as for the tracking system are fulfilled: The ECAL has a sufficient depth of up to 25.8 radiation lengths. Due to the short radiation length of 0.89 cm of the lead-tungstate (PbWO_4) crystals the ECAL is also compact enough to fit inside of the solenoid together with the tracker and hadronic calorimeter. At the same time fine granularity and the possibility to operate in a strong magnetic field is vital and achieved, too. Again, the detector system is facing rather high particle flux. Thus, the ECAL is also designed to endure such an amount of radiation.

As in the case of the tracking system, the ECAL is split into a cylindric barrel region covering a pseudorapidity range $|\eta| < 1.479$ and two endcaps covering $1.553 < |\eta| < 3.0$. In addition, a preshower detector outwards of the endcaps is installed allowing the distinction of isolated photons from pion decays. Furthermore, it improves the detection as well as the position and energy measurement of electrons and photons. As a result of this sub detector setup “the resolution for $\vec{E}_T \approx 45$ GeV electrons from Z-boson decays is better than 2 % in the central region of the ECAL barrel ($|\eta| < 0.8$), and is between 2 % and 5 % elsewhere” [71].

Hadronic calorimeter Following the electromagnetic calorimeter a hadronic calorimeter (HCAL) [72] is positioned. It is employed to detect and measure the energy, arrival time and position of hadrons and their corresponding jets, which are produced at large rate at the LHC. Moreover, the HCAL allows the indirect measurement of missing transverse energy and missing transverse momentum. As before for the other sub detectors this calls again for a rather hermetic enclosing of the beam line by the HCAL. Hence, besides

dividing the calorimeter into a barrel part reaching up to a pseudorapidity of $|\eta| < 1.4$ and a partly overlapping endcap starting from $|\eta| < 1.3$ an additional forward calorimeter is used, too. Since the forward calorimeter is very close to the beam line covering small opening angles up to $|\eta| < 5.2$ it has to endure an extreme particle flux. That is the reason why it must be very radiation hard which is achieved by the application of hard quartz fibers as active material and embedded steel absorbers.

On the other hand, a sampling calorimeter consisting of plastic scintillator tiles and dense brass plates is used for the barrel and endcap regions since brass is non-magnetic. Thus, the material is not interfering with the outer magnetic field. Meanwhile it has a suitable nuclear interaction length of roughly 16 cm, representing an appropriate depth of 5.15 nuclear radiation lengths by a total thickness of about 80 cm. This leads to a relatively compact design, even though that hadronic showers typically develop slower than electromagnetic ones. To increase the overall thickness and the associated radiation length further from 5.15 to 11.8 an additional barrel part is placed behind the solenoid.

Solenoid The superconducting solenoid being made out of niobium titanium alloy has a magnetic field ranging from 2 T in the outer detector parts to 4 T in the detector core. This configuration was chosen to allow for a precise measurement of the momentum of charged particles by determining the curvature of the charged particle's track inside the magnetic field. The large field strength of up to 4 T is required so that even particles having rather large energies and momentum get slightly deflected. In this way, they can still be measured by dedicated sub detectors. The magnetic field is returned via an iron yoke which is interleaved with the modules of the muon detection system. Hence, the tracks of the muons get bent in two directions: Inside of the solenoid the muons are deflected in either clockwise or anticlockwise direction according to their polarity, while outside of the solenoid they are getting deflected in the opposite direction.

Muon tracking and detection system The CMS detector has a dedicated muon tracking and detection system, which gives the entire detector its name. It is used to identify and trigger on muons as well as to precisely measure the tracks and momenta of highly-energetic muons in combination with the information obtained by the inner tracking system and other inner sub detectors.

The muon detection system, like the inner tracking system, does not rely on the stopping of particles. Instead it determines the trajectory of charged particles by making use of the magnetic field of the interleaved iron yoke. In principle, except for the undetectable neutrinos, only muons, which have a longer mean free path than other particles, can pass the inner detectors and solenoid without losing too much energy and reach the muon detectors. Therefore, the particle flux in the muon system is quite low. However, because the detector system is placed outside of the solenoid and far away from the collision point a robust muon detection and measurement of the muon trajectories requires the instrumentation of a large volume. The aspiration to measure muons with a transverse momentum of up to 1 TeV still precisely lead to the construction of three types of cost-efficient large gaseous detectors: drift tubes, cathode strip chambers, and resistive plate chambers. All of them make use of the principle that charged particles like muons can ionize a gas and create an avalanche of further charged particles which can be measured as an electric signal if a strong electric field is applied.

As in the other sub detectors, the muon detector is divided into a barrel and an endcap region. The drift tubes placed in the barrel region cover a pseudorapidity range of up to $|\eta| < 1.3$, while for the coverage of a pseudorapidity range of $0.9 < |\eta| < 2.4$ cathode strip chambers are positioned in the endcaps. Cathode strip chambers are chosen for the endcaps since they can better endure the radiation coming with higher particle flux than drift tubes. Both of these sub detectors have a spatial resolution between roughly $40 \mu\text{m}$ to $150 \mu\text{m}$ [73]. They are used to provide an accurate measurement of the momentum

of the muons. The uncertainty on the muon momentum is 8 to 15 % for muons having a momentum of about 10 GeV and 20 to 40 % for a muon momentum in the order of 1 TeV if only the information from the muon system is used [73]. In contrast to that, the resistive plate chambers have a poorer spatial resolution in the order of 1 cm, but a fast response time of around 3 ns [73]. They are placed in the barrel and endcap region up to a pseudorapidity range of up to $|\eta| < 2.1$ and are employed as part of the trigger system for events containing high-energy muons.

Trigger and data acquisition system An important part of the CMS detector besides the sub detectors is the specialized trigger system: Up to roughly 60 and on average 35 proton-proton collisions can take place simultaneously every 25 ns at the LHC, resulting in a huge amount of measurement data. Not all data recorded by the CMS detector systems can be written to dedicated computer centers. Besides of that only few events are of scientific interest as most of the times rather low-energetic collinear or elastic scattering occurs during the bunch crossings. That is the reason why the CMS collaboration is making use of a two-staged triggering approach. The level 1 trigger stage is integrated as custom-made hardware into the read-out systems of the sub detectors. It is used to reduce the event rate from 40 MHz to roughly 100 kHz. Since this event rate would still overload the write-out capabilities of the computer centers only events which passed the level 1 trigger get further processed by a subsequent high level trigger. This software-based trigger reduces the number of considered events further, so that an output rate of up to 1 kHz is achieved resulting in a passable amount of measurement data which can be transferred to the computer centers. There, reconstruction algorithms are applied to analyze the events, before the events as measurement data become available for physics analyses.

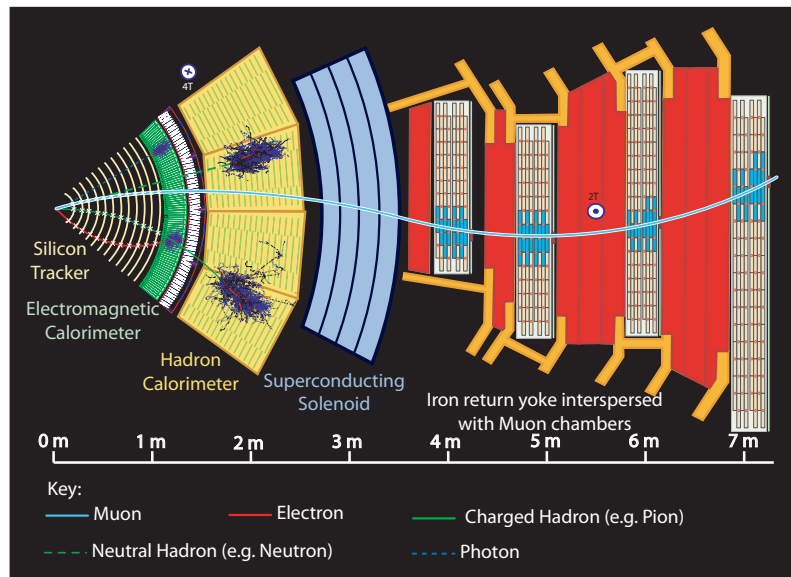


Figure 3.2: Schematic cross section of the barrel region of the CMS detector: The different parts of the CMS detector are shown radially from the inner to the outer parts and going from left to right. The silicon-based tracking system represents the innermost part of the CMS detector. Therefore, it is shown on the left. Note that all sub detectors – except the muon detection system shown on the right – are located inside of the solenoid. Taken from [74]

3.3 Object identification and event reconstruction

Without further processing the output derived from the CMS detector represents only a stream of raw and inscrutable digital signals. Therefore, it must be matched with subsequent object identification and event reconstruction steps to become applicable in physics analyses. During these steps the information gained from the single detector components is combined to define single objects at first. Afterwards the whole event which unfolded during a specific collision in the detector is reconstructed, so that exploitable measurement data is obtained.

The complete event reconstruction is done on dedicated computer clusters and it must be distinguished between the online event reconstruction and the offline event reconstruction. The online event reconstruction takes place in the first seconds after a collision event happens and is recorded due to fulfilling some trigger requirements. Usually measurement data from online event reconstruction is employed only to check the operating status of the CMS detector. It is also used for a few high-priority analyses which require fast feedback but cannot wait for the data of the offline event reconstruction to become available.

In the offline event reconstruction measurement data stored on a dedicated storage site is reprocessed at a later point in time. In contrast to the online reconstruction, time-consuming computation algorithms may be applied, which are not feasible in real time. Furthermore, the offline event reconstruction makes use of additional information like the state and configuration of single detector components during a particular collision, which was not available at event time. In this way, an improved version of the measurement data is provided by this reconstruction. Therefore, the most high-precision analyses – including the analyses in this thesis – employ the measurement data obtained with the offline event reconstruction.

In the following a brief introduction to basic (physics) objects which are used later to reconstruct and classify an event is given. It starts with objects simpler to define and nearer to the collision point and ends with finding, reconstructing, and b-tagging of jet objects. Jet objects are complex objects, consisting of many particles.

3.3.1 Tracks and vertices

If a set of simultaneous collisions occurs inside of the CMS detector, one of the first and most important pieces of information determined by the detector is the reconstruction of the trajectories of charged particles called tracks and of the vertices from which the tracks originate. More precisely, the silicon-based inner tracking system and the outer muon tracking system are mainly used for this kind of reconstruction.

Charged particles traversing the inner tracking system, and in the case of muons also reaching the muon tracking system, produce hits in each active layer by energy transfer. The track of such a charged particle can be reconstructed by a combinatorial track finder algorithm which in the case of the CMS Collaboration is based on the Kalman filtering method [75] [76]. Hereby, the reconstruction algorithms consider the deflection of a charged particle by the Lorentz force due to the strong magnetic field among other effects.

The reconstructed tracks are then used to determine the corresponding vertex candidates. These vertices can be seen as the origins of reconstructible particles decaying into charged particles. Therefore, at first possible vertex candidates are determined via track clustering. Subsequently, the vertices having the largest likelihood are derived via the adaptive vertex-fitter method [77].

At the LHC usually several collisions happen at the same time (in 2016 on average around 23 collisions per bunch crossing). Therefore, more than one vertex is reconstructed in general. Therefore, different kinds of vertices are distinguished:

Primary vertex of the hard interaction The primary vertex of an event defines the center of the coordinate system for a single hard-scattering event. It is found by determining the vertex with the largest sum of squared transverse momentum of its associated tracks, $\sum_{\text{Tracks}} p_T^2$. Therefore, a primary vertex generally corresponds to a hard, inelastic collision. This primary vertex of a hard interaction is important for the reconstruction of the full event since the reconstruction algorithms try to map all outgoing particles to this initial vertex.

Underlying event and pile-up vertices Besides the partons of the two colliding protons participating in the hard, inelastic interaction, further bystander partons either from the colliding protons or other protons in the bunch can interact with each other. This leads to the occurrence of additional vertices and particles not connected to the primary vertex of the hard interaction. Since the resulting particles were not involved in the hard interaction of interest they could distort the overall event reconstruction of the hard collision process if they are not removed. Therefore, the identification of such vertices and associated particles is of importance. The events which arise from such vertices are called underlying events and pile-up.

Secondary vertices of the hard interaction Hadrons emerging from the primary collision, but having a long decay time can move far enough away from the corresponding vertex where they were created before they decay into charged particles. That is the reason why a displaced vertex can be observed. Such a secondary vertex is often related to the occurrence of bottom or charm hadrons usually having long decay times. Therefore, this feature of bottom hadrons is exploited for their identification and for the so-called b-tagging (section 3.3.6).

More information about the track and primary vertex reconstruction can be found in [78].

3.3.2 Particle flow event reconstruction and MET

The CMS collaboration employs a particle flow algorithm [79–81] to allow for a better object identification and reconstruction as well as for the reconstruction of the full event: Instead of using measurement data only from single detector parts, the information of all sub detectors is exploited simultaneously to construct particle objects at first. Then based on the derived particle objects the information from the full detector is combined to reconstruct the full event.

By the particle flow algorithm (figure 3.3) five types of object candidates are identified: Muons, electrons, photons, neutral hadrons, and charged hadrons. They represent all the particles detectable by the detector.

Muons are the first objects which are reconstructed by the particle flow algorithm. Here, hits and tracks from the inner tracking system or hits and tracks in the outer muon tracking system are combined.

Tracks and hits corresponding to a muon candidate are then removed from the collection of particle objects which are available for the reconstruction of further objects. In the next

step electron candidates are determined by combining tracks from the inner tracking system with energy deposition in the ECAL. Again, the corresponding tracks and ECAL energy-deposition towers are removed. Subsequently, photon candidates, neutral hadron, and charged hadron candidates are reconstructed by making use of the remaining information from the inner tracking system, the ECAL, and the HCAL. Finally, by means of the particle-flow algorithm object candidates for all detectable particles are obtained and based on the particle candidates the full event is reconstructed.

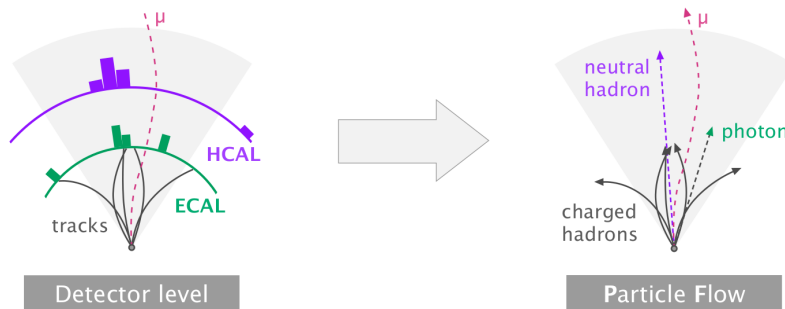


Figure 3.3: Exemplary sketch of the particle flow algorithm: On the left side, information derived by the single sub detectors like tracks in the inner tracking system, energy-deposition towers in the ECAL and HCAL as well as the track of a muon candidate is shown. On the right side, particle objects reconstructed by the particle flow algorithm using the information of the full detector are presented. Taken from [82].

Furthermore, after the reconstruction of a full event the missing energy in the transverse plane $\cancel{E}_T$ can be assessed. It is defined as the negative sum of the momenta of all n reconstructed particles:

$$\cancel{E}_T = - \sum_{i=1}^n \vec{p}_{T,i}$$

Since the partons taking part in the hard collision process only have a small intrinsic transverse momenta, the sum of the transverse momentum p_T of all detected particle-flow particle candidates in the final-state should also be rather small. However, a large $\cancel{E}_T$ is expected, if (highly-energetic) particles escape the detector acceptance or if undetectable particles like neutrinos from weak interactions occur in the final-state. For example, after the decay of bottom or top quarks neutrinos can emerge by leptonic decay of $W^\pm$ bosons. On that score, $\cancel{E}_T$ can be an important observable in top-physics analyses.

3.3.3 Muons

For the CMS detector two different types of muon reconstruction are applied:

Reconstruction of “tracker” muons Muons having a momentum lower than roughly 300 GeV and losing most of their energy in the inner parts of the detector can be better reconstructed by starting from the inner tracking system. These are called “tracker” muons. In the beginning all of the tracks of the inner tracking system are considered for possible muon candidates. The tracks inside of the inner tracking system are then extrapolated to the outer muon tracking system in consideration of the deflection by the magnetic field, energy loss due to bremsstrahlung, or the occurrence of multiple scattering. If an extrapolated track can be matched with at least one hit in the outer muon tracking system, a valid muon candidate is assumed.

Reconstruction of “global” muons In contrast to that, muons having larger momentum can be better reconstructed by using the dedicated outer muon system in the first place. These muons are referred to as “global” muons. They are reconstructed by combination of hits in the outer muon tracking system, so that a muon track in the outer tracking system is obtained. This track is then extrapolated to the inner tracking system and, if applicable, matched with a corresponding track.

In the end muons derived from both approaches and sharing the same track are combined by means of the CMS particle flow algorithm.

Detailed descriptions of the muon reconstruction and its performance can be found in [73, 83].

3.3.4 Electrons

The reconstruction of electrons in the CMS detector is based on their energy deposition in the ECAL and a matching track in the inner tracking system. The electrons get deflected in the tracking system due to the large magnetic field. Furthermore, they lose energy by radiating off photons as bremsstrahlung. Therefore, the finding of their tracks and their subsequent reconstruction is challenging. That is the reason why a supercluster approach combined with a Kalman [75] [76] and Gaussian Sum filtering method [84] is used: In the beginning a seed crystal corresponding to the center of a cluster with high-energy deposition in the ECAL is chosen. Subsequently, neighbouring clusters exceeding a threshold get added, so that a super cluster is formed inside of the ECAL. An electron track is then reconstructed either by an extrapolation of the aforementioned ECAL supercluster in direction of the tracking system or from at least two hits in the inner layer of the tracking system. For the combination of electron traces in the tracking system to a track first a combinatorial Kalman filter method with loose constraints is applied. It is then followed by a Gaussian Sum filtering (GSF) method. This method accounts for the large radiative energy losses of the original electrons due to the creation of bremsstrahlung photons and subsequent electron-positron pair production as well as the curved trajectories of the electrons and their associated charged particles. By geometrical matching electron candidates are then derived by an association of the super clusters with a corresponding track obtained by the GSF method. A detailed description of the electron reconstruction can be found in [85] [86].

3.3.5 Photon, hadron and jet reconstruction

After electrons and muons are identified and removed, together with their tracks, from the particle object collection, photons and neutral hadrons are determined by the particle flow algorithm. Therefore, as a first step topological clusters representing energy deposition in neighbouring detectors cells in the ECAL and the HCAL are determined. Clusters in different layers of the calorimeters are then matched to straight tracks. Afterwards, the tracks from the ECAL and the HCAL are combined, if applicable.

If clusters occur in the ECAL (within the tracker acceptance of $|\eta| < 2.5$) which are not linked to any track, these clusters are converted to isolated photons. Similarly, clusters in the HCAL not linked to any tracks give rise to neutral hadrons. For ECAL clusters outside of the tracker acceptance still an isolated photon is assumed if the ECAL clusters are not linked to any HCAL clusters.

Otherwise, as in the case of a track simultaneously linked to ECAL and HCAL clusters a shower of charged and neutral hadrons is assumed in general. The distinction between a photon, neutral hadron, or charged hadron is then made by comparing the sum of the

associated track momenta to the energy deposited in the calorimeters as clusters: If the observed energy deposit of the associated clusters is larger than the sum of the track momenta by a certain amount, this excess is interpreted as the presence of photons or neutral hadrons. In the case that the energy deposit of the clusters is compatible with the sum of the track momenta a charged hadron is assumed and its energy is determined from a combination of the calorimeter and track measurement. If the energy deposited in the calorimeter is smaller than the sum of the track momenta an extended search for muons or fake tracks is performed. Tracks surviving this extended checks are associated with charged hadrons.

From the particle flow candidates reconstructed before jets are constructed using the anti- k_T clustering algorithm [87, 88] with a radius parameter of 0.4 in the case of this thesis. Hereby, the jet reconstruction algorithm extrapolates the calorimeter tracks to tracks in the inner tracking system up to the primary vertex of the hard interaction. In general, particle flow candidates which cannot be mapped to this primary vertex and originate from other vertices are neglected to remove pile-up and mitigate distortion effects due to the underlying event.

3.3.6 b-Tagging of jets

It is important to identify jets originating from bottom quarks and bottom hadrons. This can be achieved using b-tagging algorithms which determine if at least one bottom-flavoured hadron was part of the parton shower causing the corresponding jet and if the parton shower emerged from a bottom quark in the beginning. These kind of b-tagged jets have some distinct features:

Cone sizes of b-tagged jets Since bottom quarks are a lot heavier than the lighter quarks (up, down; charm, strange), hadrons which are formed between a bottom quark and one or two lighter quarks also have a rather large mass in comparison to the usual hadrons occurring during a parton shower. Therefore, the decay products of such heavier b-hadrons are generally more boosted in one direction compared to jets from hadrons formed only by light quarks. Due to momentum conservation and the relativistic movement of the bottom quark this results in jets with smaller cone sizes.

Lifetime of b-hadrons, displaced and secondary vertices Furthermore, the bottom quark and its corresponding b-hadron usually have a large lifetime in the order of 1×10^{-12} s. The reason for this is that the original bottom quark cannot decay to a top quark via the weak interaction due to their difference in mass. Thus, it can only decay via the weak interaction into an up-type quark of another generation which is suppressed by the off-diagonal matrix element of the CKM matrix. Therefore, if a bottom quark is created inside the inner tracking system its associated b-hadron can travel distances around a few millimeters before decaying. This leads to an observable displaced vertex in the inner tracking system. In fact, these secondary vertices (section 3.3.1) are employed in b-tagging algorithms as an important measure to identify jets from bottom quarks.

In figure 3.4 an event with two light jets and a jet emerging from a bottom quark decay is shown in an exemplary way: The developing jet cones of the decay products of the light jets can be matched to the primary vertex as their origin. However, due to the decay of the bottom quark at a later point in time its corresponding jet partly starts at the secondary vertex. Moreover, it can be observed that the overall direction of the decay products of the b-tagged jet is slightly offset with regard to the primary vertex resulting in a measurable impact parameter d_0 for the b-tagged jet.

Charged leptons inside of a jet A further hint of the occurrence of a bottom hadron in the beginning of a parton shower used by b-tagging algorithms can be the appearance of few charged leptons in the developing jet. This is caused by semileptonic weak decays of the bottom quark or its decay products.

For the analyses presented in this thesis the combined secondary vertex b-tagging algorithm version 2 (CSVv2) [89] is employed. It makes use of a combination of secondary vertex and track-based lifetime information to determine b-tagged jets. Among others, the algorithm uses features like the mass of the considered vertices and the significance of the measurable impact parameter for the tested tracks. One novelty as compared to the LHC Run 1 version is that this b-tagging algorithm now combines the two sets of information with shallow neural networks instead of a likelihood ratio. In future CMS analyses this algorithm will most likely be replaced by b-tagging algorithms which completely depend on neural networks with at least a few hidden layers like the DeepCSV b-tagging algorithm [90]. More information about the b-tagging algorithms available for the CMS physics analyses in the LHC Run 2 and their performance can be found in [89, 90].

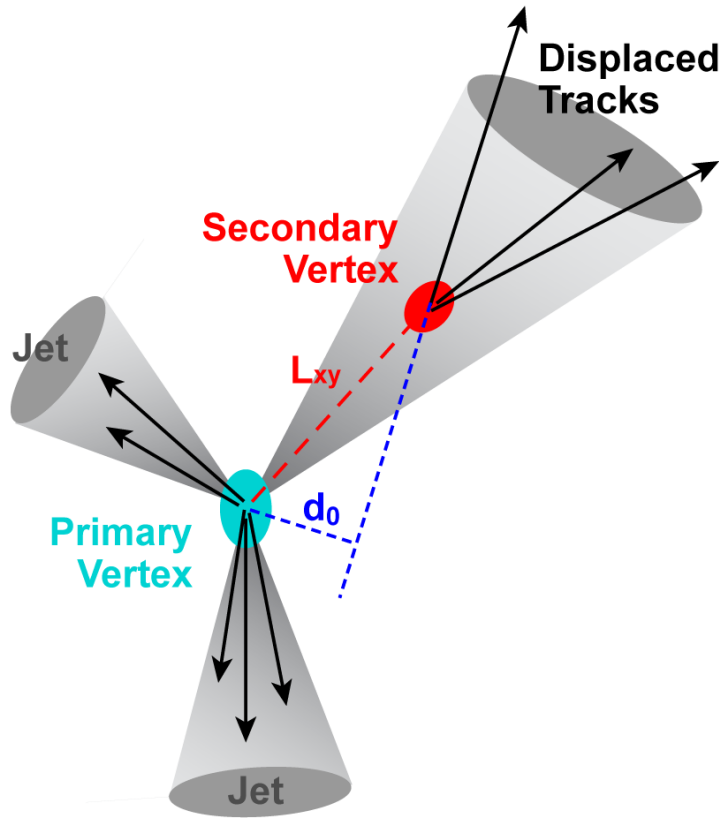


Figure 3.4: Example of an event with two light jets and a b-tagged jet: The jet cones and tracks of the light jets can be matched to the primary vertex as their origin. In contrast to that the cone of the b-tagged jet still starts at the primary vertex, but the displaced tracks originate from the secondary vertex. Furthermore, the cone direction of b-tagged jet is slightly offset resulting in an measurable impact parameter d_0 . Taken from [91].

4. Top-quark physics

Physics involving top quarks, which were discovered only in 1995, is still an important topic of research within the studies of the Standard Model of particle physics (chapter 1). Due to the large mass of the top quark this kind of physics processes feature some specific characteristics as discussed in the following. Moreover, the proper description and the precise measurement of such physics process can still be a challenge (sections 4.5 and II). After a short review of the discovery of the top quark which also represents a starting point for top-quark physics (next section), some information about the basic properties of the top quark will be given in section 4.2. This is followed by a discussion of the top quark production and decay modes (section 4.4). Subsequently, the focus will be on top-quark pair production and possible $t\bar{t}+X$ processes. A short overview of the associated production of a Higgs boson with a top quark-antiquark pair completes this chapter (section 4.6). More detailed descriptions about top physics can be found e.g. in the following review articles [92–94].

4.1 The beginning of top-quark physics and the discovery of the top quark

The prediction of the necessity of a third generation of the quark family by Makoto Kobayashi and Toshihide Maskawa can be marked as the beginning of top-quark physics as a research topic. By this prediction of the top quark as the up-type quark of the third quark generation an observed violation of the CP symmetry in kaon decay [25] could be explained. Later, in 1975, the names top and bottom quark were introduced by Haim Harari [95] for the quarks of this third generation in analogy to the up and down quark of the first generation. Similarly, these two quarks also form a weak isospin doublet. The discovery of the bottom quark followed quickly in 1977 by the E288 experiment [96] [97] at Fermi National Accelerator Laboratory (FNAL). At this time the existence of the top quark was generally assumed and an imminent top-quark discovery was widely anticipated.

Nonetheless, it was not until 1995 that the discovery of the top quark could finally be claimed by the CDF and DØ collaborations [98] [99] making use of the Tevatron accelerator located at FNAL, while first evidence of events containing top quarks already showed up in 1992 [92].

Today, at the LHC roughly 17 top-quark pairs are produced per second due to the large

$t\bar{t}$ production cross section of 832 pb at a center-of-mass energy of $\sqrt{s} = 13$ TeV and an instantaneous luminosity of $\mathcal{L} = 2 \times 10^{34} \frac{1}{\text{cm}^2\text{s}}$.

4.2 Basic properties of the top quark

The top quark is the heaviest of the six quarks predicted by the Standard Model of particle physics (chapter 1). Due to being the up-type quark in the third generation of the Standard Model quarks it has an electric charge of $+\frac{2}{3}$ in units of the elementary charge e . Furthermore, the left-handed top quark has a value of the third component of the weak isospin of $T_3 = +\frac{1}{2}$, while its right-handed counterpart has $T_3 = 0$.

The average mass of the top quark considering direct measurements done at the LHC [100, 101] and the Tevatron [102] is (173.1 ± 0.6) GeV [24]. The average value of the top quark decay width is $\Gamma = (1.41_{-0.15}^{+0.19})$ GeV [24]. Here, measurements of the CMS collaboration done at the LHC and of the DØ Collaboration conducted at the Tevatron were taken into account.

These basic properties of the top quark are also listed in table 4.1 together with the top-quark pair production cross section at a center-of-mass energy of $\sqrt{s} = 13$ TeV and the single-top production cross sections which both will be discussed in the next section. More details about the properties of the top quark can be found in [24, 94].

4.3 Top-quark production

At hadron colliders like the LHC (section 3.1) top quarks are produced dominantly in pairs by QCD process at leading order, this top-quark pair production ($pp \rightarrow t\bar{t}$), which is shown in figure 4.1a, can either emerge from gluon-gluon (gg) fusion or quark-antiquark ($q\bar{q}$) annihilation. At the LHC with a center-of-mass energy of $\sqrt{s} = 13$ TeV of the colliding protons, roughly 88 % of the top-quark pair production arises from gg fusion with the remainder from $q\bar{q}$ annihilation.

The total cross section of the top-quark pair production is known fully analytically at next-to-next-leading order (NNLO) QCD accuracy with next-to-next-leading log (NNLL) soft gluon resummation since 2013 [48]. In this calculation, a top quark mass of $m(\text{top}) = 172.5$ GeV and a center-of-mass energy of $\sqrt{s} = 13$ TeV has been assumed to derive the cross section: $\sigma(t\bar{t}, \text{NNLO}+\text{NNLL}) = (832_{-46}^{+40})$ pb.

There exists a further production mode of top quarks called single top-quark production. Since this production mode involves electroweak interactions the corresponding cross sections are significantly smaller than for pair production. Furthermore, the single top-quark production can be divided into three separate production modes, ordered by virtuality of the $W^\pm$ boson. The production modes are shown in figure 4.1b and the process cross sections are stated in table 4.1.

The t-channel single top-quark production The t-channel single top-quark production involving the process $qb \rightarrow q't$ has a sizable cross section (see below) at the LHC. However, this process is suppressed due to the smaller coupling strength of the weak interaction compared to the $t\bar{t}$ production by the strong interaction. The cross sections of this process for the top quark and its corresponding antiquark are different at the LHC due to the charge-asymmetric initial state. This difference in the cross section for top quark and antiquark at the LHC can also be seen for the s-channel single top-quark production. The total cross sections of the t-channels single top-quark production are known to NNLO

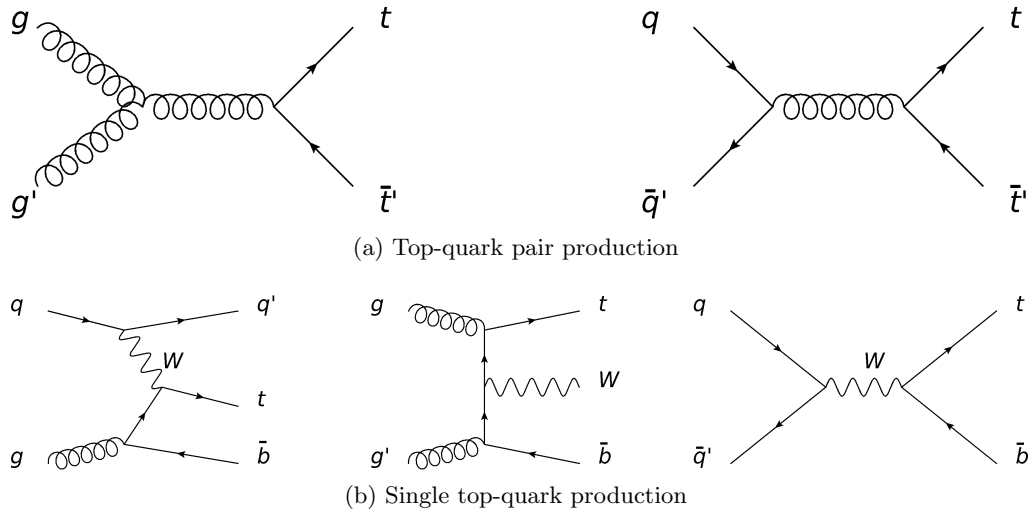


Figure 4.1: Exemplary Feynman diagrams of top-quark pair production and single top-quark production: In the left plot of figure (a) the gluon fusion initiated top-quark pair production is shown, while in the right plot the production via quark-quark annihilation. In the sub figure (b) the t-channel, the Wt-associated, and the s-channel single top-quark production is presented from left to right.

QCD accuracy since 2014 [103] and the value of the cross section combining the top quark and top antiquark contributions to the t-channel production is $\sigma(t+\bar{t}, \text{NNLO}) = (214.8^{+2.1}_{-1.3}) \text{ pb}$ [104] at a center-of-mass energy of $\sqrt{s} = 13 \text{ TeV}$.

The Wt-associated single top-quark production The Wt-associated production, corresponding to the $b\bar{g} \rightarrow W^- t$ process, has, due to the additional occurring real $W^\pm$ boson in the final state, a smaller combined total cross section of $\sigma(t+\bar{t}, \text{NNLO approx.}) = (71.7 \pm 1.8 \pm 3.4) \text{ pb}$ [105]. This value was also derived from an NNLO approximate calculation and an assumed top mass of $m(\text{top}) = 172.5 \text{ GeV}$.

At NLO, this production mode interferes with top-quark pair production. It can be a very interesting production mode because of this interference, but also due to its sensitivity to new physics [106–108]. Furthermore, in many Higgs and SUSY searches it has a not insignificant role as a background process [109].

The s-channel single top-quark production With a value of $\sigma(t+\bar{t}, \text{NNLO approx.}) = (11.4 \pm 0.2 \pm 0.4) \text{ pb}$ [105] the s-channel has the smallest combined total cross section of the three single top-quark production modes at the LHC. The main reason is that this process is quark-antiquark initiated, while the probability of gluons occurring in the initial state is quite larger at the LHC.

4.4 Decay modes of the top quark

The large mass of the top quark results in a large decay width of the top quark of $\Gamma = (1.41^{+0.19}_{-0.15}) \text{ GeV}$ [24]. Accordingly, it has a corresponding very short life-time of roughly $5 \times 10^{-25} \text{ s}$. Since this is below the time required to form hadronic bounding states via hadronization, the physics of the top quark is fundamentally different compared to the other quarks. Hadron spectroscopy like for the other quarks is not possible, but instead the top quark offers “the unique opportunity to study the production and decay of a ‘bare’ quark, at energy scales much larger than those which are typically involved for the other quarks” [93].

Table 4.1: Basic properties of the top quark: First a description of the property is given, followed by the corresponding symbol of the property, finally the value of the property is stated. The values are taken from: ¹ [24], ² [48], ³ [104], and ⁴ [105].

Description of property	Corr. symbol	values
Average mass of the top quark ¹	$m(\text{top})$	$(173.1 \pm 0.6) \text{ GeV}$
Average decay width of the top quark ¹	Γ	$(1.41^{+0.19}_{-0.15}) \text{ GeV}$
Top-quark pair production cross section prediction at $\sqrt{s} = 13 \text{ TeV}$		
Total cross section prediction ²	$\sigma(t\bar{t}, \text{NNLO+NNLL})$	$(832^{+40}_{-46}) \text{ pb}$
Single-top production cross section prediction at $\sqrt{s} = 13 \text{ TeV}$		
t-channel prediction ³	$\sigma(t+\bar{t}, \text{NNLO})$	$(214.8^{+2.1}_{-1.3}) \text{ pb}$
Wt-associated production prediction ⁴	$\sigma(t+\bar{t}, \text{NNLO approx.})$	$(71.7 \pm 1.8 \pm 3.4) \text{ pb}$
s-channel prediction ⁴	$\sigma(t+\bar{t}, \text{NNLO approx.})$	$(11.4 \pm 0.2 \pm 0.4) \text{ pb}$

Off-diagonal decays of the top quark are suppressed by the CKM matrix. Therefore, the top quark decays into a W^+ boson and a bottom quark in almost of all cases (branching ratio: $\mathcal{BR} = 1.014 \pm 0.003(\text{stat}) \pm 0.032(\text{sys})$ [110]). According to the further decay of the involved $W^\pm$ boson one distinguishes two top-quark decay modes: If the $W^\pm$ boson decays into a charged lepton and a neutrino, this is called a leptonic decay of the top quark. Meanwhile, a hadronic decay of the top quark is assumed if the $W^\pm$ boson decays into a quark-antiquark pair. The hadronic decay of the $W^\pm$ boson and the associated hadronic decay of the top quark is twice as likely as the leptonic decay. The reason for this lies in the three different colour charges of the quarks in comparison to the single electric charge of the leptons. Therefore, the $W^\pm$ boson can decay into three times more quark-antiquark pairs than fermion pairs. The two types of $W^\pm$ boson decay are shown in figure 4.2a. Following the above categorization of the top quark decay, one can also define three different decay modes of the top-quark pair production ($t\bar{t}$), ordered by the size of their branching ratio $\mathcal{BR}$:

Fullhadronic $t\bar{t}$ decay mode In the fullhadronic $t\bar{t}$ decay mode both top quarks decay hadronically. This decay mode features the largest branching ratio of 45.7 % [24].

Semileptonic $t\bar{t}$ decay mode In the semileptonic $t\bar{t}$ mode one top quark or antiquark decays hadronically, while the other top parton decays leptonically (see also figure 4.2b for an example of such a kind of $t\bar{t}$ decay). The branching ratio $\mathcal{BR}$ of this decay mode is 43.8 % considering all three charged leptons (electron, muon, tauon) [24].

In this thesis only $t\bar{t}$ processes involving the semileptonic decay mode limited to electrons and muons will be considered. The reason for this is that these kind of $t\bar{t}$ processes offer a good compromise between a rather high probability of occurrence compared to $t\bar{t}$ processes of the dileptonic decay mode and a still rather good possibility to reconstruct the full event in contrast to fullhadronic $t\bar{t}$ processes. Furthermore, in this way a difficult tauon reconstruction is avoided.

Dileptonic $t\bar{t}$ decay mode In the dileptonic $t\bar{t}$ decay mode which has the lowest branching ratio $\mathcal{BR}$ of 10.5 % [24] both top quarks decay leptonically.

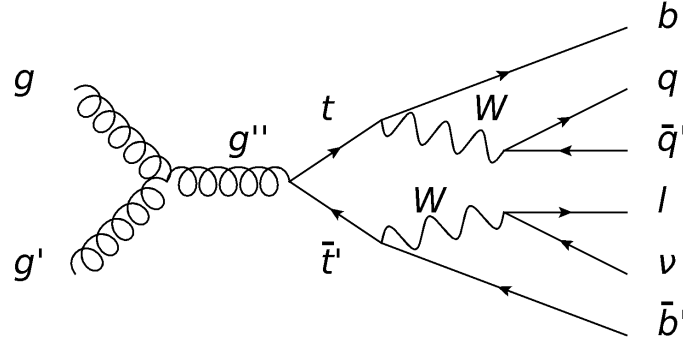
(a) The two types of $W^\pm$ boson decay(b) Semileptonic $t\bar{t}$ decay mode

Figure 4.2: Exemplary Feynman diagrams of the $W^\pm$ boson decay and the semileptonic decay mode of the $t\bar{t}$ process : In the left part of the upper figure the leptonic decay of the $W^\pm$ boson is shown, while in the right plot the hadronic decay mode. In the lower figure the semileptonic decay mode of the $t\bar{t}$ process is presented.

4.5 Top-quark pair production and corresponding $t\bar{t}+X$ processes

Of greatest interest are processes where the top-quark pair is created in association with additional quarks. These $t\bar{t}+X$ processes can be distinguished by the number of heavy-flavour quarks in the final state at matrix-element level (section 2.2.1), which is done in the following to provide a short description of their differences.

However, in this thesis a classification of $t\bar{t}+X$ events based on the mapping of heavy-flavour particle-level jets to heavy-flavour hadrons is employed as described in section 6.4. The group of all $t\bar{t}+X$ processes (including those with no additional partons) is referred to as the inclusive $t\bar{t}+X$ processes, while the single $t\bar{t}+X$ processes are also often called exclusive processes.

Commonly, processes in which a boson emerges besides the top-quark pair in the final state instead of quarks, like $t\bar{t}H$ or $t\bar{t}Z$, are not counted to the $t\bar{t}+X$ processes. They are considered as separate processes.

The $t\bar{t}+b\bar{b}$ processes

A $t\bar{t}$ process is spoken of as $t\bar{t}+b\bar{b}$ process if besides the two bottom quarks originating from the decay of the two top quarks two additional bottom quark occur in the final state of a matrix element.

This kind of processes are of particular interest since they are among the SM processes occurring at the LHC which have the heaviest particles in the final-state. Furthermore, by comparing the Feynman diagrams of these processes with the diagrams of the $t\bar{t}H(b\bar{b})$ process (discussed in the next section) it is obvious that in most cases the only difference between both processes is the origin of the additional bottom-quark pair. The bottom-quark pair in general either originates from a massless gluon or light quark in the case of

the $t\bar{t}+b\bar{b}$ process or from the massive Higgs boson. Hence, the kinematics of the bottom-quark pair are the major difference. That is also the reason why the differentiation between the two processes is difficult. Therefore, both processes are an important object of study as will be motivated in more detail in part II.

The $t\bar{t}+2b$ process

The $t\bar{t}+2b$ process cannot be defined by the number of additional quarks at matrix-element level. It can only be understood at particle level since it is defined by mapping the same particle-level jet to at least two bottom hadrons (section 6.4). Nonetheless, the introduction of this process later in this thesis is useful: Events which originate from $t\bar{t}+b\bar{b}$ matrix elements and which would be misidentified as $t\bar{t}+b$ events (see below) at particle level will be considered by this process class. Therefore, in principle this process defines a sub class of the $t\bar{t}+b\bar{b}$ processes in which the two additional bottom quarks are radiated in a similar direction and, hence, their bottom hadrons will lead to a single particle-level jet.

The $t\bar{t}+b$ processes

An event is classified as a $t\bar{t}+b$ event if exactly one additional bottom quark occurs in the $t\bar{t}$ final state at matrix-element level. At the matrix-element level it is quite different to the $t\bar{t}+b\bar{b}$ process since it is always quark-gluon initiated in leading-order QCD, while the $t\bar{t}+b\bar{b}$ process is either originating from gluon fusion or quark-antiquark annihilation.

The $t\bar{t}+c\bar{c}$ processes

$t\bar{t}$ events which do not contain any additional bottom quarks in the matrix element, but still have at least one charm quark at matrix-element level are counted as $t\bar{t}+c\bar{c}$ events. The softened requirement of only one charm quark is motivated by the difficult identification of charm hadrons and charm-tagged jets at reconstruction level. It is quite likely that only one charm quark and its successors are identified later. Therefore, it seems not reasonable to split this class of processes further like in the case of the aforementioned processes involving bottom quarks.

Compared to the $t\bar{t}+b\bar{b}$ processes, a possibly emerging charm-quark pair usually carries a smaller transverse momentum than the heavier bottom quarks and, hence, it is far more likely that the two developing jets of the charm quarks would overlap.

The $t\bar{t}+lf$ processes

The $t\bar{t}+X$ processes which remain and do not fall in one of the other process classes are referred to as $t\bar{t}+lf$ processes. These processes represent the lion's share of all $t\bar{t}+X$ processes and are important background processes for many physics analyses conducted at the LHC.

4.6 Associated production of a Higgs boson with a top-quark pair

The associated production of a Higgs boson with a top-quark pair, $pp \rightarrow t\bar{t}H$, is a physics processes of particular interest. It allows a direct measurement of the top-Higgs-Yukawa coupling which also corresponds to the largest Yukawa coupling of all SM fermions. The top-Higgs-Yukawa coupling can also be measured by means of the Higgs-boson production via gluon-gluon fusion or the Higgs-boson decay into two photons. However, these are only

indirect measurements since besides the top quarks other non-SM particles could be part of the occurring virtual loops. Therefore, all three measurements are an important test of the Standard Model of particle physics (SM). Furthermore, findings in such measurements can be a possible gateway to new physics.

The $t\bar{t}H$ production mode of a Higgs boson has the smallest cross section of all four major Higgs-boson production modes (gluon fusion, vector-boson fusion, associated production with a vector boson, and associated production with a top-quark pair) described by the SM. The reason for this is that the final state contains three heavy massive particles (figure 4.3) which together have an invariant mass of roughly 470 GeV. Hence, the predicted total cross section of the $t\bar{t}H$ process at a center-of-mass energy of $\sqrt{s} = 13$ TeV and calculated in NLO QCD including electroweak corrections according to [47] is: $\sigma(t\bar{t}H, \text{NLO}) = (507.1^{+81.6}_{-98.9})$ pb.

The observation of the $t\bar{t}H$ process was recently made by the CMS collaboration with a significance of 5.2σ , while the expected significance was 4.2σ [111]. Shortly before the ATLAS collaboration claimed evidence of this process with a significance of 4.2σ observed standard deviations (3.8σ expected) [112, 113].

In this thesis only the $t\bar{t}H$ sub processes in which the Higgs boson decays into a bottom-quark pair will be considered [114, 115]. Even though this class of $t\bar{t}H$ sub processes has the largest branching ratio $\mathcal{BR}$ it has a roughly 1600 times smaller cross section than the $t\bar{t}$ process. Nonetheless, $t\bar{t}$ production can still be seen as an important background, especially to the $t\bar{t}+b\bar{b}$ sub processes since a $t\bar{t}+b\bar{b}$ process and a $t\bar{t}H(b\bar{b})$ processes can share the same final state but with different kinematics.

In principle, with the same argumentation of possibly having similar final-states, the $t\bar{t}Z$ and $VV+\text{jets}$ process can also be considered important backgrounds. However, the major difference is that these processes can also be measured and constrained well by making use of other final states. In contrast to that, the $t\bar{t}H$ process still has to understood better and is associated with uncertainty.

As can be seen from the Feynman diagram in figure 4.3 the $t\bar{t}H(b\bar{b})$ process under the assumption of a semileptonic $t\bar{t}$ decay usually features a very “busy” event. Such an event involves several quarks, leptons, and missing transverse momentum due to at least one emerging neutrino because of one top quark decaying leptonically. It is for this reason that a proper object identification and event reconstruction represents a challenging combinatorial problem. Especially the assignment of observed and possibly b-tagged jets to the originally occurring top quarks and the Higgs boson can be highly ambiguous.

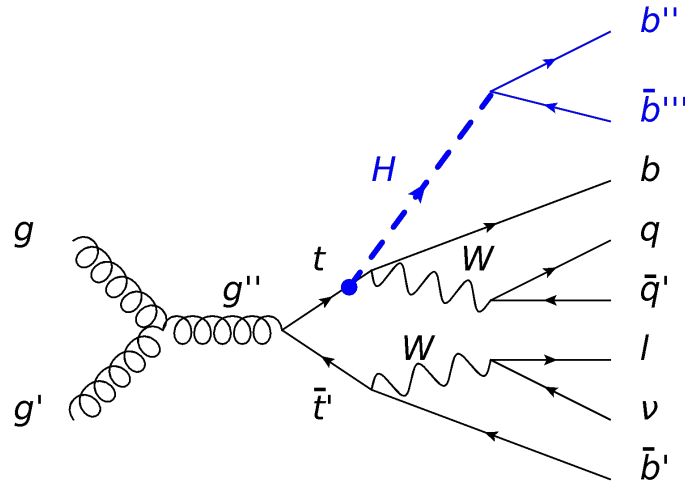


Figure 4.3: Exemplary Feynman diagram of the associated Higgs boson production with a top-quark pair, in which a semileptonic decay of the $t\bar{t}$ system occurs.

5. Statistical analysis methods

Statistical analysis methods are an important set of tools in particle physics:

On the one hand, they provide methods to interpret the data obtained in a physics analysis for example with the maximum-likelihood (ML) method (next section). In this way, by comparing the results derived from such a physics analysis with a given prediction these methods lead to a quantitative result.

In this thesis the ML method is applied to obtain the value of the signal strength modifier r for top-quark pair production in association with b-tagged jets. This modifier states the deviation of the analysis result from a theoretical prediction.

Among others, the $t\bar{t} + \geq 1\text{b-jet}$ signal derived from physics analyses is compared to the prediction of the $t\bar{t} + \geq 1\text{b-jet}$ signal by theoretical calculations in the framework of the Standard Model of particle physics in this thesis.

On the other hand, multivariate data analysis (section 5.2) is also based on statistical methods and is also often applied in particle physics. Multivariate analysis methods allow the analysis of data using more than one variable simultaneously. The application of such methods takes into account that usually the separation between signal and background processes is a multidimensional problem in particle physics. Furthermore, often a better separation between signal and background processes can be obtained if multiple variables and correlations between them are used.

In this thesis linear discriminants and neural networks are applied as multivariate classifiers to provide a separation between signal and background events.

5.1 Maximum-Likelihood method

This section begins with a derivation of the binned likelihood function as part of the explanation of the maximum-likelihood (ML) method in the subsequent section. Properties of nuisance parameters are shortly discussed in section 5.1.3. Furthermore, the Asimov dataset is introduced in section 5.1.4.

5.1.1 Derivation of the binned likelihood function

The number of events measured in data n_{data} in a physics analysis can be seen as the outcome of a counting experiment. In general, Poisson distributions are used to describe the probability of a given number of events or, in our case, data events n_{data} in a such counting experiment:

$$\text{Poisson}(n_{\text{data}}|\nu) = \frac{\nu^{n_{\text{data}}} \cdot e^{-\nu}}{n_{\text{data}}!},$$

in which the probability depends on the number of predicted events ν .

In particle physics, the number of predicted events ν is usually obtained from Monte Carlo samples. Furthermore, this prediction is split into signal events n_{sig} originating from signal processes of interest and background events n_{bkg} which emerge from all other processes, so that $\nu = n_{\text{sig}} + n_{\text{bkg}}$. Therefore, the Poisson distribution for a single counting experiment in particle physics would be given by:

$$\text{Poisson}(n_{\text{data}}|n_{\text{sig}} + n_{\text{bkg}}) = \frac{(n_{\text{sig}} + n_{\text{bkg}})^{n_{\text{data}}} \cdot e^{-(n_{\text{sig}} + n_{\text{bkg}})}}{n_{\text{data}}!}.$$

The predicted number of signal events n_{sig} and background events n_{bkg} is generally only known within some uncertainty. These uncertainties on the predictions are considered by introducing so-called nuisance parameters θ_i . For each uncertainty such a nuisance parameter θ_i together with its associated probability density function (PDF) $P(\theta_i)$ is considered. The number of signal events $n_{\text{sig}}(\vec{\theta})$ and of background events $n_{\text{bkg}}(\vec{\theta})$ are then given as functions of these nuisance parameters. Furthermore, from this the likelihood function for a single counting experiment can be derived:

$$\mathcal{L}(n_{\text{data}}|n_{\text{sig}}, n_{\text{bkg}}, \vec{\theta}) = \text{Poisson}(n_{\text{data}}|n_{\text{sig}}(\vec{\theta}) + n_{\text{bkg}}(\vec{\theta})) \cdot \prod_i^N P(\theta_i).$$

So far, it was neglected that in the analyses of this thesis as in the most other particle physics analyses binned data is analyzed. However, the derived likelihood function $\mathcal{L}(n_{\text{data}}|n_{\text{sig}}, n_{\text{bkg}}, \vec{\theta})$ can easily be extended to binned data. This is done by describing the observed data $\vec{n}_{\text{data}}$ as well as the predicted signal events $\vec{n}_{\text{sig}}(\vec{\theta})$ and predicted background events $\vec{n}_{\text{bkg}}(\vec{\theta})$ via vectors of individual counting experiments:

$$\mathcal{L}(\vec{n}_{\text{data}}|\vec{n}_{\text{sig}}, \vec{n}_{\text{bkg}}, \vec{\theta}) = \prod_j^M \text{Poisson}(n_{\text{data},j}|n_{\text{sig},j}(\vec{\theta}) + n_{\text{bkg},j}(\vec{\theta})) \cdot \prod_i^N P(\theta_i).$$

In this way, each value of a vector corresponds to a bin in the binned output distribution used for the construction of the likelihood function.

Often in particle physics a signal strength modifier r is introduced. The number of predicted signal events $\vec{n}_{\text{sig}}(\vec{\theta})$ is scaled by r during the ML fit, so that it corresponds to the number of observed signal events in data:

$$\vec{n}_{\text{data},\text{sig}} := r \cdot \vec{n}_{\text{sig}}(\vec{\theta}).$$

Therefore, given the number of signal events in data is properly predicted by the Monte Carlo samples for the signal processes this signal strength modifier r should be one. This

value is also chosen as initial value in the ML fit.
The corresponding likelihood function

$$\mathcal{L} \left(\vec{n}_{\text{data}} | r, \vec{n}_{\text{sig}}, \vec{n}_{\text{bkg}}, \vec{\theta} \right) = \prod_j^M \text{Poisson} \left(n_{\text{data},j} | r \cdot n_{\text{sig},j}(\vec{\theta}) + n_{\text{bkg},j}(\vec{\theta}) \right) \cdot \prod_i^N P(\theta_i)$$

is then used as input for the binned ML fit.

5.1.2 Maximum-Likelihood fit

The ML method is used to estimate the parameters of the likelihood function, the signal strength parameter r and the nuisance parameters $\vec{\theta}$. This parameter estimation is done by finding the extremum of the likelihood function. Often instead of maximizing the likelihood function the negative logarithmic-likelihood function is minimized. The advantage of this approach is that the logarithmic-likelihood function has the extreme value at the same spot as the likelihood function, but it is easier to compute due to occurrence of sum terms instead of product terms, the vanishing of exponential functions and constant terms like the factorial $n!$ being part of the definition of the Poisson distribution:

$$-\log \mathcal{L} \left(\vec{n}_{\text{data}} | r, \vec{n}_{\text{sig}}, \vec{n}_{\text{bkg}}, \vec{\theta} \right) = - \left(\sum_j^M \log \text{Poisson} \left(n_{\text{data},j} | r \cdot n_{\text{sig},j}(\vec{\theta}) + n_{\text{bkg},j}(\vec{\theta}) \right) \cdot \sum_i^N \log P(\theta_i) \right).$$

Especially the vanishing of the exponential functions which often occur in the PDFs of the nuisance parameters leads to a simpler computation.

The found likelihood estimate corresponds to the optimal set of free parameters $(\hat{r}, \hat{\vec{\theta}})$ for a given number of events in data $\vec{n}_{\text{data}}$ as well predicted events in the Monte Carlo signal samples $\vec{n}_{\text{sig}}$ and events in the background samples $\vec{n}_{\text{bkg}}$. These found parameters $(\hat{r}, \hat{\vec{\theta}})$ describe best the observed data for a given model.

In most of the cases only one or a few free parameters of the likelihood function are parameters of interest. In the case of this thesis only the best-fit value of the signal strength modifier $\hat{r}$ is of general interest, while the values of the nuisance parameters are still studied to check the assumed uncertainty model. Therefore, each nuisance parameter θ_i is expressed as a function of the signal strength modifier: $\theta_i = f(r)$.

The uncertainty on the best-fit value of the signal strength $\hat{r}$ can then be derived from a profile likelihood ratio (PLR):

$$q(r) = -2 \log \frac{\mathcal{L} \left(r, \hat{\vec{\theta}} \right)}{\mathcal{L} \left(\hat{r}, \hat{\vec{\theta}} \right)}.$$

Here, the parameter $\hat{\vec{\theta}}$ maximizes the likelihood function for a given signal strength r , while by $\mathcal{L} \left(\hat{r}, \hat{\vec{\theta}} \right)$ the likelihood estimate corresponding to the previously found global maximum of the likelihood function is used.

By scanning the PLR $q(r)$ the 1σ -uncertainty interval of the best-fit value of the signal strength $\hat{r}$ is determined. The 1σ -uncertainty interval corresponds to an increase in the parabolic shape of the PLR from the minimum obtained with the best-fit signal strength $\hat{r}$

by a factor of 1. In the case of a dependency of the likelihood on two signal strengths and of an associated two-dimensional likelihood contour the factor to derive the 1σ -uncertainty would be 2.3.

5.1.3 Properties of nuisance parameters

Nuisance parameters θ_i can be classified into rate-changing, shape-changing, and rate-and-shape-changing nuisance parameters. All of these three classes of nuisance parameters affect final discriminant output distributions which are used as inputs for a ML fit.

Rate-changing nuisance parameters A rate-changing nuisance parameter changes the expected rate of events in all bins of a final discriminant output distribution by the same factor. Such a nuisance parameter corresponds to a normalization uncertainty, which is often modeled by a log-normal distribution, so that the nuisance parameter itself is constrained by a normal distribution.

Shape-changing nuisance parameters Shape-changing nuisance parameters affect the number of expected events in each bin of a final discriminant distribution separately. Therefore, usually two additional Monte Carlo samples are produced to model the upwards and downwards shape variation due to such a shape uncertainty. For the production of these samples the values of some parameters are varied by $\pm 1\sigma$ compared to their nominal values used for the production of the nominal sample.

Rate-and-shape-changing nuisance parameters Rate-and-shape-changing nuisance parameters can be seen as a combination of both aforementioned nuisances. In general, such uncertainties are split into a rate-changing and shape-changing nuisance and corresponding uncertainties. The shape uncertainties are then normalized to the nominal prediction, while the rate uncertainty takes the necessary normalization into account.

5.1.4 Asimov dataset

The 1σ -uncertainty interval for the central value of the observed signal strength $r(\text{observed})$ is obtained from a PLR fit using the measurement data as described.

However, this observed signal strength and its associated uncertainty $\Delta r(\text{observed})$ strongly depends on the measurement data. Therefore, it is not a good measure of the experimental sensitivity of a physics analysis and the uncertainty on the signal strength r expected from the related uncertainty model. Furthermore, the measurement data can be blinded.

That is the reason why for a given signal strength r the expected uncertainty on the signal strength $\Delta r(\text{expected})$ is computed using a representative dataset, the so-called Asimov dataset [116]. Compared to measurement data, in such an Asimov dataset the observed data is replaced by fake data predicted from the sum of the Monte Carlo samples. Furthermore, it is defined in a way that the true parameter values are obtained if the dataset is used for the evaluation of all parameters in a ML fit. Hence, it can be applied to derive the expected uncertainties for a parameter of interest, in the case of this thesis the expected uncertainty on a signal strength, $\Delta r(\text{expected})$. By construction, the Asimov dataset is also free from statistical fluctuations for a given set of Monte Carlo samples it can be used to determine the maximum precision obtainable with a physics analysis.

Two types of Asimov datasets can be distinguished:

An Asimov dataset constructed for a signal strength expectation $r(\text{expected})$ of zero consists only of the background model in which the nuisance parameters θ_i and all parameters of interest except the signal strength r are set to their estimation values before the ML fit. The signal strength is set to zero, so that no signal contributions are considered. Such an Asimov dataset is only a good estimate of the expected uncertainty on a signal strength modifier r if the signal contribution is quite smaller than the background contribution. For example, this is often the case in searches. Otherwise, the uncertainty model would be also driven by signal contribution. Since in the analyses presented in this thesis signal contributions are in the same order as background contributions, this assumption holds not true. For this reason such an Asimov dataset is not used in this thesis.

An Asimov dataset can also be constructed considering a signal strength expectation $r(\text{expected})$ of one. In this way, signal and background is assumed according to the prediction by the Monte Carlo samples based on the Standard Model prediction. Again, the nuisance parameters θ_i and all parameters of interest are set to their estimation values before the fit. In addition, for the signal strength a value of one is assumed. Such Asimov datasets are used throughout this thesis to provide an estimate of the uncertainty expected for a signal strength measurement.

5.2 Multivariate data analysis

Multivariate analysis (MVA) methods [117] are employed for classification of data and dimensionality reduction. In the context of this thesis, MVA methods like linear discriminant analysis (section 5.2.2) or neural networks (section 5.2.3) are used to classify events into different categories.

In the next subsection the receiver-operator-characteristic (ROC) curve is shortly introduced since its integral is often used as a performance indicator for a classifier.

Afterwards a general overview of the two MVA methods mentioned before will be given. The setup of these methods for the physics analysis presented in this thesis will be discussed in chapter 9.

5.2.1 Performance of a classifier

MVA methods are often used to classify data into categories. In such a case it is of importance to estimate or quantify the performance of such classifier.

A suitable performance indicator of a classifier is the receiver-operating-characteristic (ROC) curve. Furthermore, if the integral of this curve is computed a quantitative assessment of the separation power of a classifier is possible. This integral value is also termed area-under-curve (AUC) score. To obtain such a ROC curve the true positive rate is generally plotted against the false positive rate. However, in particle physics it is more common to plot the background rejection versus the signal efficiency for a classifier. Both plotting styles lead to the same AUC score but the curves appear mirrored along their x- and y-axis.

The value of the ROC integral can vary between zero and one. In practice ROC integral

somewhere between 0.5 and one are obtained, while a larger value represents a better performing classifiers and a better separation between different categories. A classifier choosing randomly should lead to a mean ROC value of 0.5 in the case of binary classification. If a classifier would provide a perfect categorization of data the associated ROC integral would be one.

The concept of the ROC integral can be extended from the case of binary classification to multiclassification. This is achieved by considering a multiclass classifier as a combination of single binary classifiers. For each category the value of the ROC integral is determined based on its separation to the other categories. The common ROC integral of the multiclass classifier is then obtained by calculating the mean of the ROC integrals for each category.

5.2.2 Linear discriminant analysis

Linear discriminant (LD) analysis is a mathematically robust and rather simple MVA method. It is a generalization of the Fisher discriminant analysis [118] which was developed already in 1936. LD analysis is generally employed to reduce the dimensionality of data to a single variable and to separate two classes based on this obtained output variable. This is accomplished by finding of linear combination of n parameters β_i and associated input variables x_i for a given dataset with known classification, so that the computed single output variable $y(\vec{x})$ provides the best split between the two classes:

$$y(\vec{x}) = \sum_{i=1}^n \beta_i \cdot x_i.$$

In particle physics, usually, the two classes are designated as signal and background class and an event corresponds to single data. Furthermore, in many cases a bias parameter β_0 is introduced. By this additional bias parameter the value of the output variable $y(\vec{x})$ is adjusted:

$$y(\vec{x}) = \sum_{i=1}^n \beta_i \cdot x_i + \beta_0.$$

Often an adjustment is chosen in which the value of the output variable for data corresponding to the signal class is always above zero ($y(\vec{x}) > 0$) and for data corresponding to the background below zero ($y(\vec{x}) < 0$).

The finding of the parameters $\vec{\beta} = \sum \beta_i$ for a given set of input variables $\vec{x} = \sum x_i$ is called training. In this training a so-called training dataset, for which the classification of events into signal events ($y(\vec{x}) = 1$) and background events ($y(\vec{x}) = -1$) is known, is applied. If this training dataset contains m events, the following matrix representation can be used to compute the output variable $y_i(\vec{x})$ for each single data simultaneously:

$$\mathbf{Y} = \begin{pmatrix} y_1(\vec{x}) \\ y_2(\vec{x}) \\ \vdots \\ y_m(\vec{x}) \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1n} \\ 1 & x_{21} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{m1} & \dots & x_{mn} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{pmatrix} = \mathbf{X} \vec{\beta}.$$

Since finding an optimal set of the parameters $\vec{\beta}$ is of interest and only $\mathbf{X}$ and $\mathbf{Y}$ are known, a solution of the system of equations by the method of least squares is essential.

Therefore, the Moore-Penrose inverse of $\mathbf{X}$ transformation [119] $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ is applied to derive the parameters $\vec{\beta}$:

$$\vec{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}.$$

The obtained vector of parameters $\vec{\beta}$ can be geometrically interpreted as a vector which defines a perpendicular hyperplane: Depending on if an event is located below or above this hyperplane it is considered to belong to the class of signal or background events. The dot product between the set of input variables $\vec{x}$ and the parameter vector $\vec{\beta}$ is then the criterion to define if an event belongs to the signal (background) class:

$$\vec{x} \cdot \vec{\beta} \gtrless 0.$$

The LD analysis shows a good performance for Gaussian distributed variables with linear correlations. If the input variables are not Gaussian distributed, the LD analysis can often profit from suitable transformations of the input variables [120].

Therefore, transformations were used as part of the preparation of the training dataset for the physics analysis presented in this thesis. After application of these transformations the input variables were virtually Gaussian distributed with a mean of zero and a standard deviation of one.

Reference [121] demonstrates that adding input variables having low separation power to a set of input variables showing high separation power can lead to a decreased performance of the LD analysis in most cases. Furthermore, it is a requirement of each input variable that it is well modeled by Monte Carlo samples.

For this reason, only a small set of variables showing high separation power and good modeling by Monte Carlo samples were chosen in this thesis (more details will be given in section 9.2).

5.2.3 Neural networks

Neural networks (NNs), or in fact artificial neural networks, are MVA methods which are in general more complex than linear discriminants. Their name was inspired by the network of neurons in animal brains they are modeled after.

Basic principles of neural networks

Usually (artificial) NNs consist of different layers in which neurons (“nodes”) are arranged (figure 5.1). These nodes are connected to other nodes.

Two layers are special in an NN: The nodes of the first layer (also: input layer) are only connected to the subsequent layer. The nodes in this layer only have one inward connection. They get a single input value which corresponds to the value of an input variable. The last layer or output layer has only one outward connection which can be seen as the final output of an NN. The layers in between are referred to as hidden layers since their inputs or outputs are not directly accessible.

Different types of NNs exist which differ in the topology of the network, the connections made between the nodes, and the properties of the nodes.

In particle physics often and, in particular, in this thesis only feed-forward NNs are of interest for the classification of events. The reason is that the occurrence of an event for a given physics process can be described by Poisson statistics and is not time-dependent. This means that the occurrence of an event is not determined by the events which were

observed before. Therefore, no feedback loop which would take this into account is required.

This leads to feed-forward NNs having only connections between consecutive layers and no connections to other layers.

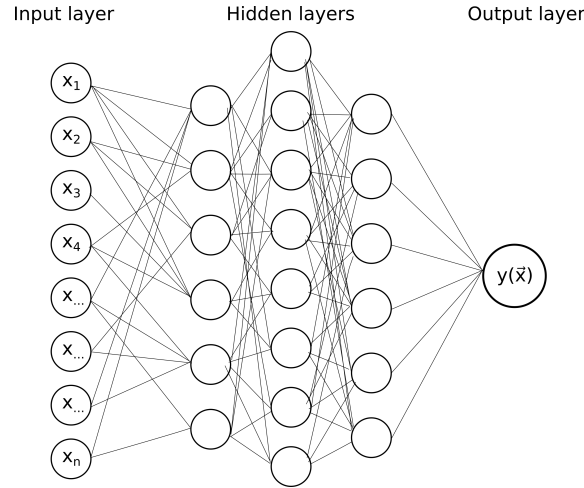


Figure 5.1: Schematic representation of a feed-forward neural network: The nodes in each layer are only connected to nodes in the subsequent layer. The input values x_i of nodes in the input layer correspond to values of input variables. After three hidden layers follows the output layer. This contains one single output node which provides the output value $y(\vec{x})$ of the neural network.

Nodes represent mathematical functions by which a single output $o(\vec{i})$ is derived from a set of inputs i_i provided by the inward connections of other nodes (figure 5.2). The obtained output $o(\vec{i})$ is then forwarded to other nodes by outward connections.

Each connection to another node is associated with a weight factor w_i . Therefore, first of all, the product of input times weight factor $i_i \cdot w_i$ is computed for each connection in a node. Afterwards, the products of all connections are added up, while a bias factor b is added to the sum. Finally, an activation function f is applied to derive an output $y(\vec{x})$:

$$o(\vec{i}) = f \left(\sum i_i \cdot w_i + b \right).$$

The bias factor is chosen to adjust the output $o(\vec{i})$ obtained by the function f . Often a bias factor is chosen so that for data which fulfills a categorization an output $o(\vec{i})$ close to one would be obtained, while for other data an output $o(\vec{i})$ close to zero would be observed.

For the activation function f mostly monotonically increasing and non-linear functions which are easily differentiable are chosen. By choosing monotonically increasing activation functions f it is taken into account that if two events have a large difference in their input values this generally corresponds to the events being rather different. Therefore, such a large difference in the initial input values x_i should also result in quite different inputs i_i to the single nodes. Furthermore, this should lead to different outputs $o(\vec{i})$ obtained by the single nodes and maybe a different overall output value $y(\vec{x})$ and categorization of the two events.

Usually non-linear activation functions are employed since a combination of linear function can be replaced by a single linear function by making use of simple arithmetic transformations. Therefore, in NNs containing more than one hidden layer linear activation functions f would not constitute an advantage.

The differentiability of the activation function f is mostly a requirement since the back-propagation-of-error algorithm [122] [123] [124] is used in the training of a neural network which presumes this.

For this reason, suitable activation functions would be e.g., the tangens hyperbolicus function or the exponential linear unit (ELU) function [125] which was applied throughout this thesis.

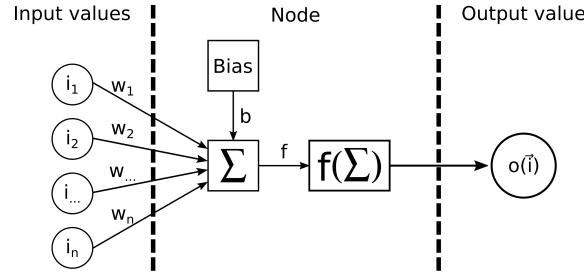


Figure 5.2: Schematic representation of a neural network node: The node obtains inputs i_i by connections to other nodes. A weight factor w_i is associated with each connection. The products of input times weight factor $i_i \cdot w_i$ are added up. Furthermore, a bias factor b is added to the sum. Afterwards by application of an activation function f the output $o(\vec{i})$ is obtained. This output $o(\vec{i})$ is forwarded to other nodes.

Training of neural networks

The finding of optimal weight factors w_i and bias factors $\vec{b}$ requires – in analogy to a linear discriminant – a training of the neural network. Again, this leads to the need for a training dataset in which the categorization of single data is known. For example, the predicted output value of a node in the output layer should be either one (single data belongs to category defined by this node) or zero (single does not belong to category).

Generally, as a first step of the processing of input variables in an NN a transformation step is introduced: For each input variable its distribution in the training dataset is determined and a suitable transfer function with an associated scaling factor is derived. By application of such transfer functions the values for each input variables are transformed, such that the resulting distributions are virtually Gaussian. This is done to make sure that each input variable has the same influence on the training. In this way, the data to be classified will also be properly normalized and possible differences in the input variable distribution of data and training data are mitigated. In strict terms this normalization is not necessary, but it can improve the performance of the training [126].

Usually NNs are then trained by making use of the backpropagation-of-error algorithm [122] [123] [124], while employing batch training: Out of the training dataset batches with a given size n are randomly constructed. All single data contained in a batch is fed one after the other into such an NN. The output value $y(\vec{x})$ of the NN is computed for each single data and is compared to the known classification. Therefore, the obtained output value $y(\vec{x})$ of an output node should be either one or zero. The deviation between the obtained output value $y(\vec{x})$ and the expected output value $\hat{y}(\vec{x})$ is then determined for all data in a batch by a loss function $L(y(\vec{x}), \hat{y}(\vec{x}))$. Possible loss functions $L(y(\vec{x}), \hat{y}(\vec{x}))$ are

the euclidean distance

$$L(y(\vec{x}), \hat{y}(\vec{x})) = \frac{1}{n} \sum_i^n (\hat{y}(\vec{x}) - y_i(\vec{x}))^2,$$

or the cross entropy

$$L(y(\vec{x}), \hat{y}(\vec{x})) = \frac{1}{n} \sum_i^n (\hat{y}(\vec{x}) \cdot \ln y_i(\vec{x}) + (1 - \hat{y}(\vec{x})) \cdot \ln (1 - y_i(\vec{x}))).$$

The obtained output value $y(\vec{x})$ for a given single data not only depends on the input variables but also on the applied weights w_i and bias factors $\vec{b}$ as parameters of an NN. Therefore, a optimization algorithm can be used to obtain the set of parameters leading to the smallest loss function $L(y(\vec{x}), \hat{y}(\vec{x}))$. The optimization algorithm evaluates the loss function for all data in a batch and determines new weights w_i and bias factors $\vec{b}$ by a gradient descent method. The weight change is then given by the partial derivative of the loss function $L(y(\vec{x}), \hat{y}(\vec{x}))$ with respect to the corresponding weight

$$\Delta w_i = -\eta \cdot \frac{\partial L(y(\vec{x}), \hat{y}(\vec{x}))}{\partial w_i}.$$

Here, the learning rate η was introduced. The learning rate η defines by which amount the weights are adjusted per optimization step and should be suitably chosen. If the learning rate η is too small the change of weight takes too long, if it is too large the weights can fluctuate and will never reach an optimal value. Similarly, the bias factors $\vec{b}$ are adjusted. Following this adjustment the next batch is employed for a further optimization of the NN parameters. After all batches were used, the complete dataset has been used once for the optimization. This is called an epoch.

Reference [127] showed that using the cross-entropy as loss function can yield an improved performance of NNs compared to NNs trained by making use of the euclidean distance as loss function. That is the reason why throughout this thesis the cross entropy is employed as a loss function.

Nowadays, mostly stochastic gradient descent methods instead of gradient descent methods are applied for the NN parameter optimization. The major difference is that stochastic gradient descent methods approximate a gradient by computing gradients from single data and update the parameter every time. An advantage of this class of optimizers is that they usually obtain an optimized set of parameters faster. However, during the training the value of the loss function $L(y(\vec{x}), \hat{y}(\vec{x}))$ can fluctuate heavily [128].

Among the most known of these methods is the Adam optimizer [129] which according to [128] might be the best choice out of the generally available optimizers. This finding was confirmed by own studies.

Therefore, in this thesis the Adam optimizer was employed.

Measures against overtraining

The training of an NN can lead to a so-called overtraining: By the training a NN should learn the features provided by the training dataset. However, an NN cannot only learn general features of the data but over time and after a few epochs it can also memorize specific characteristics of the training dataset due to its usually large number of free parameters. In this way, it loses its generalization abilities. Such an NN would still show

a continuously improving classification of the training dataset with each further epoch. However, its performance on data it was not trained on would get worse and worse.

As a countermeasure the initial (training) dataset for which the classification of single data is known is split into three datasets: A training dataset is still used to obtain an optimized set of NN parameters. An additional validation dataset is not employed for optimization but utilized to determine the performance of an NN after each epoch based on an performance indicator. Furthermore, after the training of an NN is finished a test dataset is employed for an unbiased check of the performance of the NN on an so far unused dataset. The ROC integral stated later in this thesis is always derived from such a test dataset.

In general the ROC integral mentioned before is used as a performance indicator to spot overtraining. For each epoch the ROC integral is determined for the training dataset and the validation dataset. In the beginning of the training the ROC integral should considerably increase with every epoch (figure 5.3). After a certain number of epochs the increase in the ROC integrals typically slows down. With further epochs the ROC integral of the training dataset will normally slightly improve. However, at some point the ROC integral of the validation dataset will have reached a maximum and will decrease again. This is considered as a sign that overtraining occurred and the NN has already partly lost its generalization abilities. At latest now the training of the NN should be stopped.

A more promising measure against overtraining is early stopping. By early stopping the training of an NN is stopped as soon as the improvement in proper classification and the associated increase in the ROC integral of the validation dataset stagnates. It is assumed that by a suitable choice of the early-stopping interval the training of a NN can already be stopped before overtraining occurs. The early-stopping interval defines how many epochs must have passed, in which the ROC integral of validation dataset has not improved, before the training is stopped.

An example of early stopping is shown in figure 5.3 for the reference neural network employed for the physics analysis in this thesis. Here, an early-stopping interval of 15 epochs was chosen. The training stopped after 214 epochs since no improvement in the ROC integral of the validation dataset was observed since the 199th epoch. The NN configuration of the 199th epoch is used later in the physics analysis.

Another measure against overtraining of NNs is regularization techniques. These are often used in ill-posed optimization problems. In general the training of a complex classifier, like an NN, is an underdetermined problem since such a classifier can have a large number of free parameters, while the size of the training dataset is finite. Therefore, it is quite challenging to constrain all free parameters properly. By regularization techniques additional information is provided during the training of classifier. In this way, the free parameters can be better constrained. Furthermore, this approach usually leads to more homogenous classification models since parameters having large values are suppressed. Therefore, large fluctuations in the obtained output values of a classifier are less frequent. In this manner, the classifier is also forced to keep its generalization abilities since single outliers of the output values and corresponding single data have fewer influence on the overall training. For NNs often weight decay, which is also termed L-regularization, is employed. Here, the L refers to adding a coefficient to the linear model of the loss function $L(y(\vec{x}), \hat{y}(\vec{x}))$, so

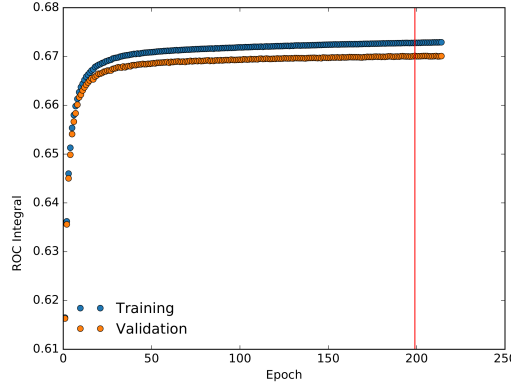


Figure 5.3: Example of the learning behaviour and of early stopping during the training of a neural network: The training of the neural network used for the reference physics analysis in this thesis is shown. In the beginning of the training the classification performance, shown by the receiver-operator-curve (ROC) integral, is increasing strongly, while with increasing number of training epochs the classification performance approaches a nearly constant value. The training of the neural network is stopped, after the ROC integral the validation sample has not increased for 15 epochs. The corresponding 199th epoch is shown by a red line.

that the new loss function $L'(y(\vec{x}), \hat{y}(\vec{x}))$ would be:

$$L'(y(\vec{x}), \hat{y}(\vec{x})) = L(y(\vec{x}), \hat{y}(\vec{x})) + \frac{\lambda}{2} \cdot \|w_i\|.$$

In this formula the regularization parameter λ and a norm for the weights w_i of the NN $\|w_i\|$ was introduced. A suitable choice of the regularization parameter λ is necessary. The regularization parameter λ should be large enough to enforce small individual weights w_i , so that overtraining will be avoided. Nonetheless, if the regularization parameter λ has been selected too large, the learning of an NN and its adoption of to data is suppressed. Furthermore, different kind of norms of the weights w_i are conceivable. However, in most cases the either L1- or L2-regularization is applied. The difference between the two schemes is that either the absolute value of the weights (L1) or the square of the weights (L2) is employed, so that the extended loss function becomes:

$$L'(y(\vec{x}), \hat{y}(\vec{x})) = L(y(\vec{x}), \hat{y}(\vec{x})) + \frac{\lambda}{2} \cdot \sum |w_i| \quad (\text{L1-regularization}),$$

or

$$L'(y(\vec{x}), \hat{y}(\vec{x})) = L(y(\vec{x}), \hat{y}(\vec{x})) + \frac{\lambda}{2} \cdot \sum w_i^2 \quad (\text{L2-regularization}).$$

Since the change of weights w_i is determined by the partial derivative of the loss function, by L1-regularization the change of single weights as a function of the regularization term is independent of the weight itself. Therefore, L1-regularization has the advantage that it pushes more weights close to zero. However, according to [130] L2-regularization leads to an overall better training performance, especially if pruning (will be discussed next) is applied during the training.

Therefore, for the NNs employed in this thesis L2-regularization is used.

A pruning of NN parameters by the dropout method can lead to simpler NN models. In

addition, it reduces the influence of single NN nodes on the overall output of a given NN. If dropout is applied in the training of an NN, the output of a node $o(\vec{i})$ is set to zero for a fraction d of randomly selected nodes. In this way, these nodes are virtually switched off for the corresponding training epoch. The outputs of the remaining nodes is rescaled by the factor $\frac{1}{1-d}$, so that still the same output value for each single data should be obtained as for an NN with all nodes being active. As before the loss function is evaluated and the backpropagation-of-error algorithm employed to adjust the weights batch by batch or per epoch. However, it is expected that an NN having better generalization abilities is obtained by this kind of training. The reason for this is that the emergence of nodes having a significant influence on the overall NN performance is suppressed. Furthermore, the obtained NN should be less prone to statistical fluctuations of the training data.

Part II

Motivation

Motivation

Even more than 20 years after the discovery of the top quark at the Tevatron [98] [99] in 1995 the study and measurement of the $t\bar{t}+\geq 1b$ -jet signal processes, which in the context of this thesis comprise of the $t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, and $t\bar{t}+b$ processes, is an important field of study. The reason for this is that the description, calculation, proper simulation, and determination of these processes is challenging. Furthermore, these processes enter many other measurements done at the Large Hadron Collider as background processes. A detailed motivation of the measurements and studies presented in this thesis will be given in the following.

Description of the $t\bar{t}+\geq 1b$ -jet processes by Monte Carlo event generators: Test of heavy-flavour production and Quantum Chromodynamics

Out of the various $t\bar{t}+X$ processes especially the description, calculation, and proper simulation of the $t\bar{t}+\geq 1b$ -jet processes as well as their measurement is of special interest. It can be seen as a crucial test of our understanding of the theory of Quantum Chromodynamics (QCD) and the analysis methods usually applied to predict and measure a QCD process: The $t\bar{t}+\geq 1b$ -jet processes provide a unique possibility to explore and study heavy-flavour quark production since at least two top quarks and one bottom quark are involved in such a process [131]. In the $t\bar{t}+\geq 1b$ -jet processes the associated dynamics of the heavy-flavour quark production are very pronounced due to the at least four QCD vertices occurring in the $t\bar{t}+b\bar{b}$ and $t\bar{t}+2b$ processes in leading-order of perturbation theory (LO). This also leads to an α_s^4 -dependency of the corresponding cross sections. That is also the reason why these processes are particularly sensitive to variations of the renormalization and factorization scale. The related uncertainty on the $t\bar{t}+b\bar{b}$ cross section is in the order of 70% – 80% at leading-order of perturbation theory [131]. According to [132, 133] the scale factor between the leading-order and next-to-leading order (NLO) predictions of the $t\bar{t}+b\bar{b}$ cross section is roughly $k(\frac{\text{NLO}}{\text{LO}}) \approx 1.8$ and, hence, quite large, while a rather large 30% – 35% uncertainty on the NLO $t\bar{t}+b\bar{b}$ cross section can be observed, too. This rather large jump in the cross section from LO to NLO and the large uncertainty of the NLO $t\bar{t}+b\bar{b}$ cross section can be seen as a sign that it is hard to describe the $t\bar{t}+b\bar{b}$ process and the two other related $t\bar{t}+\geq 1b$ -jet processes properly in lower orders of perturbation theory only. Proper simulations of the $t\bar{t}+\geq 1b$ -jet processes in higher orders of perturbation theory considering the full phase space seem not feasible up to now. Furthermore, it appears that more insight into suitable choices of the renormalization and factorization could be beneficial for the $t\bar{t}+\geq 1b$ -jet processes [131].

Also the decision on energy scale values other than the renormalization and factorization scale, like the scale of the matrix element and parton shower matching, represents a tough choice: For the description and prediction of these processes by application of Monte Carlo event generators multiple scales, spanning a wide energy range, have to be considered. As an example of these different energy scales, the typical energy of the emerging bottom quark system is roughly in the order of two to four times the bottom quark mass. Hence,

it is far below the energy of $pp \rightarrow t\bar{t}$ production taking place at energies in the order of two times the top quark mass [134]. As an effect of this scale dependency, in [135] it was found that the choice of the value of the parton shower matching scale, used to match the matrix element event generator output with a general-purpose event generator, can have a significant impact on $t\bar{t} + \geq 1b$ -jet production. The cause for this is that $gg \rightarrow b\bar{b}$ splittings can be either described by matrix elements or by the parton shower of a general-purpose event generator.

Based on all these reasons, various approaches to describe and simulate the $t\bar{t} + \geq 1b$ -jet production in NLO precision by a combination of a matrix element event generator and a general-purpose event generator are conceivable and were developed: In the first NLO simulation of the $t\bar{t} + b\bar{b}$ process conducted by using the POWHEL+Pythia6 event generator [37, 136, 137] a five-flavour scheme (5FS) was applied together with the POWHEG matching method [30, 31]. In contrast to that, a Sherpa+OpenLoops result [135] obtained with a four-flavour scheme (4FS) and the MC@NLO matching method [54, 138] proved soon afterwards that a 4FS computation can describe the full phase-space of the bottom quarks. For this reason it can deliver a more consistent description of the $t\bar{t} + \geq 1b$ -jet production, while a description in the 5FS has to restrict the phase-space to avoid collinear $gg \rightarrow b\bar{b}$ singularities due to the massless bottom quarks [131].

A comparison of the integrated $t\bar{t} + \geq 1b$ -jet cross section predicted by the three event generator combinations MG5aMC(NLO)+Pythia8 [29, 38, 139, 140], POWHEL+Pythia8 [38, 139, 141, 142], and Sherpa+OpenLoops [143–145], demonstrated excellent agreement between the POWHEL+Pythia8 and Sherpa+OpenLoops event generators. However, the MG5aMC(NLO)+Pythia8 event generator yielded a roughly 30% larger integrated $t\bar{t} + \geq 1b$ -jet cross section and showed a stronger dependency on the chosen parton shower matching scale. This result is shown in figure 5.4 and was published among further $t\bar{t} + \geq 1b$ -jet related studies as part of the fourth volume of the Handbook of LHC Higgs Cross Sections [47]. The observed difference is striking since both the MG5aMC(NLO) and Sherpa matrix element generator employ a 4FS for the computation, while POWHEL makes use of a 5FS computation. Nonetheless, in principle, good agreement between the different event generators using NLO perturbation theory would be expected. One can conclude from this that even today the simulation of the $t\bar{t} + \geq 1b$ -jet production is afflicted by uncertainties related to the choice of a specific event generator and the resulting uncertainty on its proper description of the $t\bar{t} + \geq 1b$ -jet production.

A recently published paper [131], which was used as the major reference for this part of the chapter, assesses the theoretical uncertainties in the $t\bar{t} + \geq 1b$ -jet production further from a theoretical point of view by providing new Monte Carlo comparison studies. It mentions also a new POWHEG generator for the 4FS $t\bar{t} + \geq 1b$ -jet production, which was not yet available for own tests at the time of writing this thesis.

However, the two NLO $t\bar{t} + \geq 1b$ -jet-production comparison-studies mentioned so far lacked a comparison with data, while the agreement between data and a Monte Carlo simulation of the $t\bar{t} + \geq 1b$ -jet processes should be considered the final test of the quality of a Monte Carlo event generator.

For this reason, in the measurement presented in part IV the POWHEG+Pythia8 event generator combination is applied to describe all $t\bar{t} + X$ processes including the $t\bar{t} + \geq 1b$ -jet ones at reconstruction level. In this way, the prediction of the POWHEG+Pythia8 event

generator for the $t\bar{t}+\geq 1\text{b-jet}$ signal processes can be compared to data at reconstruction level.

In part VI the POWHEG+Pythia8 prediction of the $t\bar{t}+\geq 1\text{b-jet}$ processes is compared to $t\bar{t}+\geq 1\text{b-jet}$ predictions by the MG5aMC(NLO)+Pythia8 and Sherpa+OpenLoops event generator to provide more insight on the differences between the $t\bar{t}+\geq 1\text{b-jet}$ modeling of different event generators. This comparison is also done at reconstruction level by making use of events which have passed the selection and object-reconstruction steps described in chapter 7.

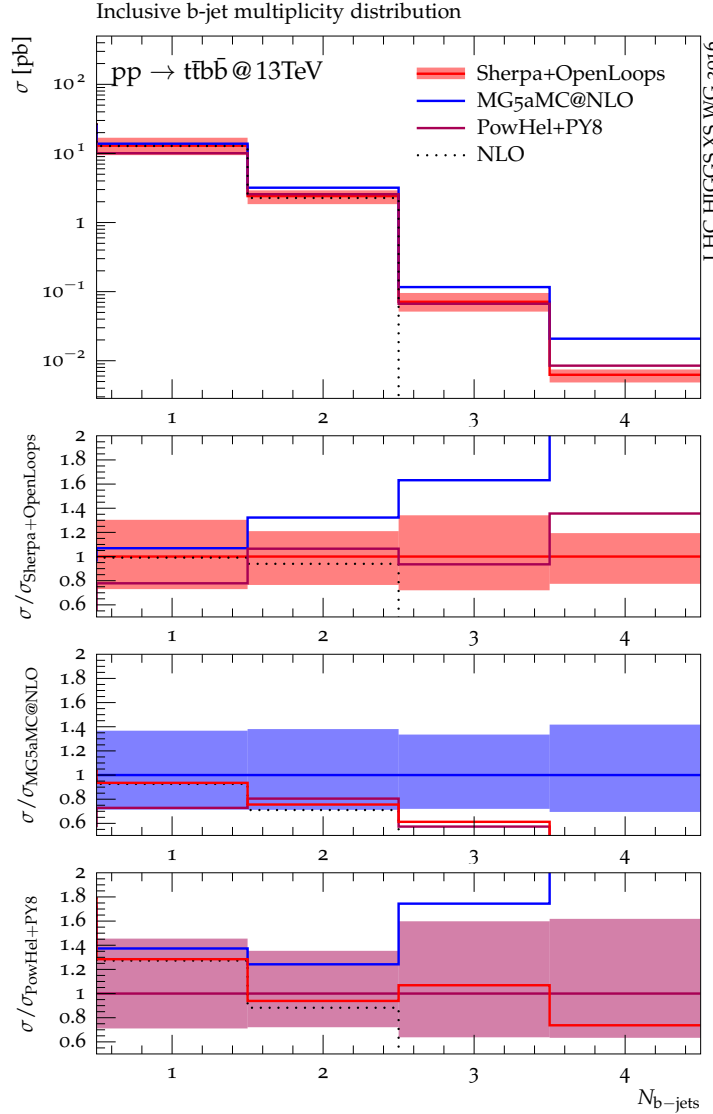


Figure 5.4: Predictions of the integrated NLO $t\bar{t}+\geq 1\text{b-jet}$ production cross section: The integrated NLO $t\bar{t}+\geq 1\text{b-jet}$ cross section is presented for the MG5aMC(NLO)+Pythia8, POWHEL+Pythia8, and Sherpa+OpenLoops event generators as well as for a fixed-order NLO computation in inclusive bins considering events with different numbers of b-tagged jets. In the three ratio plots one normalized event generator prediction is compared to the other ones and to the fixed-order NLO computation, in addition its scale variation band is shown. Taken from [47].

Distinction between the $t\bar{t}+\geq 1b$ -jet processes:

Dependency on the chosen b-tagging settings and the applied Monte Carlo event generators

The classification of events as belonging to one of the three processes $t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, and $t\bar{t}+b$, which together form the $t\bar{t}+\geq 1b$ -jet process, depends strongly on how heavy-flavour jets at particle level (“gen-jets”) are defined, identified and mapped to bottom hadrons. Therefore, different kinds of migrations of events between these three process classes can occur:

For example, by raising or lowering the threshold of the minimum transverse momentum p_T required for a gen-jet, an event can contain more or fewer b-tagged gen-jets. In this way, by increasing the p_T threshold an event considered a $t\bar{t}+b\bar{b}$ event can become a $t\bar{t}+b$ event. Similarly, the change of the value of the anti- k_T jet clustering algorithm [87] or even using another jet clustering algorithm could lead to a different categorization of an event. As an example, an event formerly classified as a $t\bar{t}+b$ event can mutate to a $t\bar{t}+2b$ event if the anti- k_T jet clustering value is raised from 0.4 (the value used in this thesis) to a slightly larger value because then an additional bottom hadron could become part of this gen-jet.

Some further modifications of the gen-jet definition or their mapping to bottom hadrons are conceivable which could lead to an alteration of the event classification (Ref. [146] contains a few examples). Therefore, the distinction between the $t\bar{t}+\geq 1b$ -jet processes must always be viewed in the context of the heavy-flavour identification applied in the corresponding analysis. The heavy-flavour identification employed in this thesis is described in section 6.4.

Besides the dependency of the $t\bar{t}+\geq 1b$ -jet classification on the heavy-flavour identification settings, also the choice of the Monte Carlo event generator applied for the $t\bar{t}+\geq 1b$ -jet production can lead to a change in the fractions of the $t\bar{t}+\geq 1b$ -jet process classes.

In the fourth version of the Handbook of LHC Higgs Cross Sections [47] it was shown, for example, that the ΔR distribution between the first and second b-tagged jet, which describes the distance between these two jets, can be quite different for a $t\bar{t}+\geq 1b$ -jet production using either one of the three aforementioned event generators MG5aMC(NLO)+Pythia8, POWHEL+Pythia8, and Sherpa+OpenLoops. Also here, MG5aMC(NLO)+Pythia8 predicts a quite different differential distribution, while comparing POWHEL+Pythia8 and Sherpa+OpenLoops still sizable differences can be observed, too, as apparent from figure 5.5. Similar findings were also presented in [146] and in addition it was found that already large differences in the ΔR distribution for bottom hadrons can occur if different Monte Carlo event generators are compared. The observed difference seems to be caused mainly by the chosen general-purpose event generator. Furthermore, the differences smear out if jet clustering is applied and the b-tagged jet distributions are compared, which is also expected.

Analogue to the direct influence of the b-tagging settings, such deviations in differential distributions related to the b-tagging of jets and caused by different Monte Carlo generator predictions can also result in an uncertain distinction between the three $t\bar{t}+\geq 1b$ -jet process classes. Similar to the earlier example, a change in the ΔR distance between two bottom hadrons and corresponding b-tagged jets of an imaginable event could lead to a mutation of an event considered as a $t\bar{t}+b\bar{b}$ event to a $t\bar{t}+2b$ event since one of the bottom hadrons and associated b-tagged jets then would become part of the other jet.

In figure 5.6 an exemplary differentiation between the three $t\bar{t}+\geq 1b$ -jet processes is shown by using the $t\bar{t}+X$ samples employed in this thesis (section 6.4): In this figure the ΔR

value of the b-tagged jet pair with the smallest ΔR value (DR of b-jets w/ min. DR) distribution is presented for the three processes. A detailed description of this event variable which is also employed as input variable of multivariate classifiers in this thesis can be found in section 9.2.

The observable difference in the DR of b-jets w/ min. DR distribution between the $t\bar{t}+b\bar{b}$ process and the other two processes can be attributed to the at least two b-tagged jets which occur in $t\bar{t}+b\bar{b}$ events, while in $t\bar{t}+2b$ events and $t\bar{t}+b$ events only one additional b-tagged jet can emerge per definition: Due to the additional b-tagged jets the minimal distance between b-tagged jets becomes smaller.

As mentioned previously, a change in the ΔR distance chosen for the jet clustering would alter the differentiation between the three processes and, hence, the distribution shown in figure 5.6.

Even though the distinction between the $t\bar{t}+\geq 1b$ -jet process classes is afflicted by the aforementioned imponderabilities a differentiation of them is of particular importance since the underlying Feynman diagrams and interactions describing the production of events of these three process classes can be quite different (section 6.4). Especially, in the case of $t\bar{t}+b$ events different origins are possible: $t\bar{t}+b$ events can originate from explicit $t\bar{t}+b$ Feynman diagrams, but they can also arise from $t\bar{t}+b\bar{b}$ or $t\bar{t}+2b$ Feynman diagrams and subsequent simulations, in which one additional bottom hadron was not identified at particle level for different reasons.

A detailed and individual study of the three process classes can, therefore, provide valuable insight into the $t\bar{t}+\geq 1b$ -jet production mechanisms and their description by Monte Carlo event generators. For this reason, by using neural networks a multivariate analysis method allowing a multiclassification and individual study of single $t\bar{t}+\geq 1b$ -jet process classes is introduced in this thesis.

In part V the three $t\bar{t}+\geq 1b$ -jet processes will be measured individually but simultaneously. The obtained signal strength $r(\text{observed})$ for each of these three process classes ($t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$) will be compared to the signal strength $r(\text{expected})$ predicted by the POWHEG+Pythia8 event generator. Furthermore, the cross section σ of these three process classes will also be determined, while the joint $t\bar{t}+\geq 1b$ -jet signal strength r and cross section σ will be already obtained by the first measurement (section 11.7 of part IV).

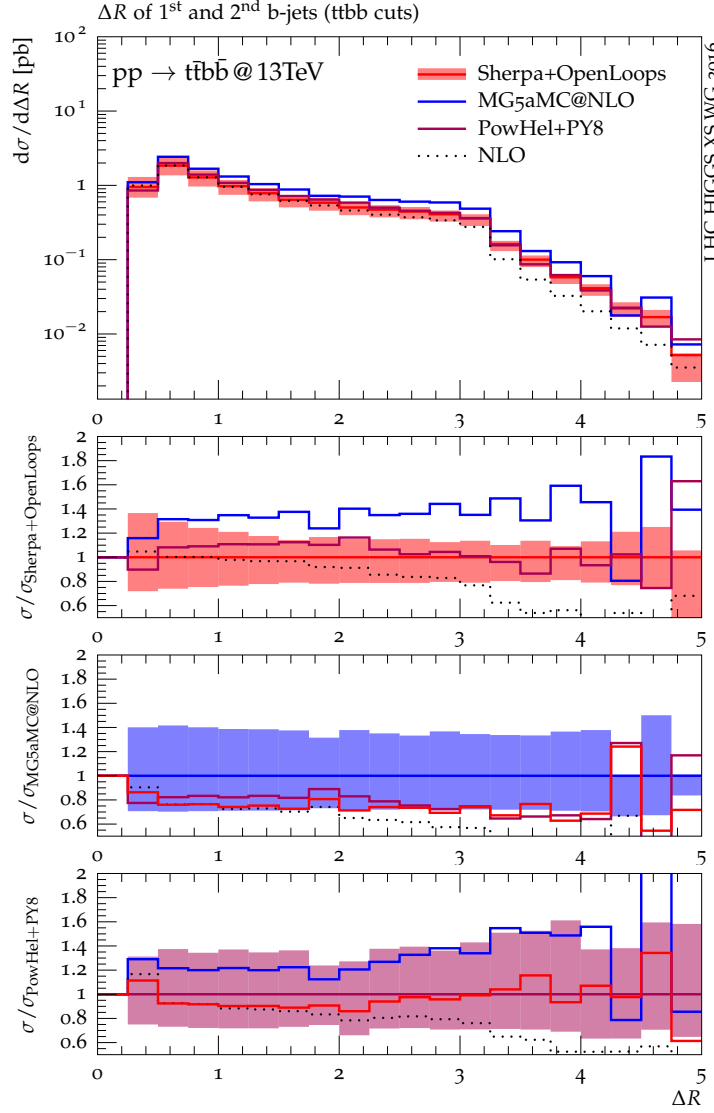


Figure 5.5: Predictions of the ΔR value of the first and second b-tagged jet: The ΔR value of the b-tagged jets having the largest and second largest p_T value is presented for the MG5aMC(NLO)+Pythia8, POWHEL+Pythia8, and Sherpa +OpenLoops event generator as well as for a fixed-order NLO computation. In the three ratio plots one normalized event generator prediction is compared to the other ones and to the fixed-order NLO computation, in addition its scale variation band is shown. Taken from [47].

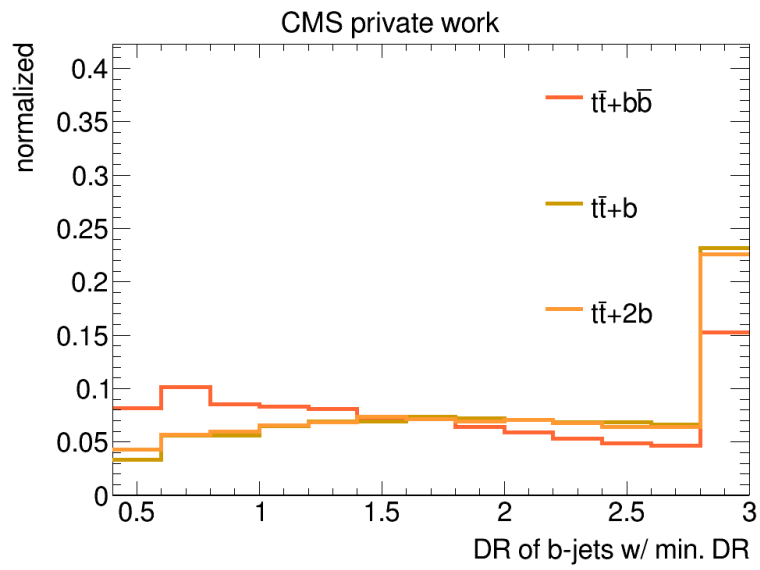


Figure 5.6: Example of the differentiation between the three $t\bar{t}+\geq 1b$ -jet processes: The ΔR value of the b-tagged jet pair with the smallest ΔR value (DR of b-jets w/ min. DR) distribution is presented for the three processes. A detailed description of this event variable can be found in section 9.2.

Importance of the $t\bar{t}+X$ and particularly the $t\bar{t}+\geq 1b\text{-jet}$ processes as background processes for various other LHC analyses

In many Standard Model (SM) and Beyond Standard Model (BSM) analyses conducted at the Large Hadron Collider (LHC) the $t\bar{t}+X$ processes can constitute a non-negligible and often important background – among others – to the actual signal process of interest. The major reason for this can be seen in the dominant production of many processes of interest by QCD at the LHC energies due to the hard partonic interactions of the protons in the initial state and since also the production of $t\bar{t}+X$ processes is dominated by QCD. Meanwhile, Quantum Electrodynamics (QED) corrections in next-to-leading order of perturbation theory contribute only less than 5 % to the overall $t\bar{t}$ production [147]. Furthermore, reference [148] found that the effects of electroweak corrections are very small as far as the total $t\bar{t}$ cross section is concerned.

An example of a signal search, which is plagued by the $t\bar{t}+X$ background is the search for the Higgs boson production in association with a top quark pair ($pp \rightarrow t\bar{t}H$). Meanwhile, the study of the $t\bar{t}H$ production presents an important remaining test of the SM predictions since these searches allow direct access to the Top-Higgs-Yukawa coupling. This coupling represents the strongest coupling of the Higgs boson to any of the SM fermions because the Higgs boson couples corresponding to the mass of the fermion. Up to now the Top-Higgs-Yukawa coupling could be only indirectly assessed and constrained from the study of the other Higgs production and decay modes ($H \rightarrow \gamma\gamma$) described by the SM.

Recently, a combination of the searches of the $t\bar{t}H$ production including various Higgs decay modes lead to 5.2σ observation of the $t\bar{t}H$ process by the CMS collaboration with a measured signal strength $r(\text{observed})$ of $1.27^{+0.32}_{-0.27}$ [111], while the ATLAS collaboration found 4.2σ evidence for the $t\bar{t}H$ production end of last year [113] (a short review of these measurements will be given in part III).

An increased $t\bar{t}H$ signal strength r and related Top-Higgs-Yukawa coupling strength could hint at new physics, while the $t\bar{t}H$ process can be also used to determine the CP-structure of the coupling [82]. Therefore, it is certain that even after its recent observation the $t\bar{t}H$ process will remain an interesting object of study and will attract further attention.

Out of the various Higgs boson decay modes, which can be considered as part of the $t\bar{t}H$ search and process study, the Higgs boson decay into a bottom quark pair ($H \rightarrow b\bar{b}$) is of special interest since it has the largest branching fraction of all Higgs decays with roughly 58 %. The next largest branching fraction is roughly 6.3 % for the Higgs boson decay to a tau lepton pair [47].

However, the larger statistics of the $H \rightarrow b\bar{b}$ decay mode come on a price: The $t\bar{t}H(b\bar{b})$ analysis suffers from large $t\bar{t}$ background, about 1600 times larger than the $t\bar{t}H$ signal, and especially an irreducible $t\bar{t}+b\bar{b}$ and $t\bar{t}+2b$ background. The cause of the irreducible background is that the $t\bar{t}H(b\bar{b})$ and the $t\bar{t}+b\bar{b}$ process as well as the $t\bar{t}+2b$ process can share the same final state with the only difference being that the additional bottom quarks and associated b-tagged jets either originate from the Higgs boson or from gluon radiation (cf. figures 6.1 and 4.3).

Due to the uncertainties on the proper simulation of the $t\bar{t}+\geq 1b\text{-jet}$ production by Monte Carlo event generators and the ambiguous distinction between the $t\bar{t}+\geq 1b\text{-jet}$ processes discussed before, the treatment of these irreducible backgrounds in the $t\bar{t}H(b\bar{b})$ analyses gets even more complicated, while the treatment of these uncertainties differs between the ATLAS and CMS collaboration: For an improved description of the $t\bar{t}+X$ background both the ATLAS collaboration in their $t\bar{t}H(b\bar{b})$ analysis, making use of the 2015 and 2016

data with a total luminosity of $\mathcal{L} = 36.1 \text{ fb}^{-1}$ [114], and the CMS collaboration in their corresponding $t\bar{t}H(b\bar{b})$ analysis, which only used the 2016 data corresponding to a luminosity of $\mathcal{L} = 35.9 \text{ fb}^{-1}$ [115], split the $t\bar{t}$ process into $t\bar{t}+X$ sub processes. In this way, smaller uncertainties for the better understood $t\bar{t}+1f$ process can be applied, while larger individual uncertainties can be attributed to the $t\bar{t}+hf$ processes (including the $t\bar{t}+c\bar{c}$ process).

However, even in the way of $t\bar{t}+X$ process splitting and the defined $t\bar{t}+X$ process classes both analyses differ. In the ATLAS $t\bar{t}H(b\bar{b})$ analysis a ΔR matching is applied with a ΔR value of $\Delta R = 0.4$ and a transverse momentum cut of $p_T = 5 \text{ GeV}$ on the jets to derive six $t\bar{t}+X$ process classes. These process classes are defined analogously to the process class definition outlined in section 4.5 with the major difference that an additional $t\bar{t}+\geq 3b\text{-jet}$ class is introduced. In comparison to that the CMS $t\bar{t}H(b\bar{b})$ analysis makes use of the ghost hadron matching (see [149]) and considers $t\bar{t}+\geq 3b\text{-jet}$ events in the $t\bar{t}+b\bar{b}$ process class. This split of the $t\bar{t}$ process into $t\bar{t}+X$ sub processes by the CMS collaboration is explained in more detail in section 6.4.

Furthermore, the two collaborations apply different systematic uncertainties related to the modeling of the $t\bar{t}+hf$ processes in their analysis, which cannot be discussed here in detail. One major difference between the two analysis is that in the ATLAS $t\bar{t}H(b\bar{b})$ analysis the events in the $t\bar{t}+\geq 1b\text{-jet}$ process classes, which were obtained quite similar to the CMS approach from a 5FS inclusive $t\bar{t}$ simulation using the POWHEG+Pythia8 event generator, are rescaled to match the predictions of an NLO $t\bar{t}+b\bar{b}$ sample originating from a 4FS $t\bar{t}+b\bar{b}$ Sherpa+OpenLoops simulation. A comparison of the relative fractions predicted by the POWHEG+Pythia8 and the Sherpa+OpenLoops event generator is shown in figure 5.7. The 4FS $t\bar{t}+b\bar{b}$ Sherpa+OpenLoops simulation should in principle provide $t\bar{t}+\geq 1b\text{-jet}$ events, and particularly $t\bar{t}+\geq 3b\text{-jet}$ events, with a higher precision. But it comes with the drawback that $t\bar{t}+1f$ events are not included and the simulation itself is computationally quite expensive in comparison to possible POWHEG+Pythia8 or MG5aMC (NLO)+Pythia8 simulations. Therefore, the sole use of such a 4FS $t\bar{t}+b\bar{b}$ Sherpa+OpenLoops sample was not an option for the CMS analysis. Moreover, the CMS collaboration decided against a rescaling of their 5FS POWHEG+Pythia8 $t\bar{t}+\geq 1b\text{-jet}$ events with such a sample since this kind of rescaling changes the relative fractions of the $t\bar{t}+\geq 1b\text{-jet}$ process classes described by the POWHEG+Pythia8 sample. For this reason it could introduce additional systematic uncertainties which are difficult to specify. However, in the CMS analysis an additional 50% uncertainty for each of the four $t\bar{t}+hf$ process class is applied, which is considered a conservative treatment of the related uncertainties and is assumingly too large.

In the first measurement presented in this thesis (part IV), the $t\bar{t}+\geq 1b\text{-jet}$ signal strength r and its associated overall uncertainty Δr is determined using an analysis setup similar to the CMS $t\bar{t}H(b\bar{b})$ analysis. The aim is to derive a better estimate of the uncertainty on $t\bar{t}+\geq 1b\text{-jet}$ production and to provide this as an input for a future CMS $t\bar{t}H(b\bar{b})$ analysis. Apart from this, in the linear discriminant based and neural network based reference analyses, delivering the major results presented in this thesis, besides the $t\bar{t}+X$ processes the $t\bar{t}H$ process is considered as an individual process class in the multiclassification. In this manner, not only the signal-to-background separation of the present analyses could be improved, but this developed approach makes it also possible to determine the signal strength of the $t\bar{t}+\geq 1b\text{-jet}$ processes and the $t\bar{t}H$ process simultaneously. This is demonstrated in part V.

In a future analysis this simultaneous measurement could reduce the dependency on a proper modeling of the $t\bar{t}+\geq 1b\text{-jet}$ processes for the determination of the $t\bar{t}H$ signal. The ATLAS $t\bar{t}H(b\bar{b})$ measurement already employs a rather similar approach: In this measure-

ment the normalization of the $t\bar{t}+\geq 1b$ -jet processes as well of the $t\bar{t}+c\bar{c}$ processes is left freely floating. To derive the uncertainty on the $t\bar{t}H$ signal strength, with the exception of these two normalization factors all nuisance parameters are fixed by a Maximum-Likelihood (ML) fit. However, the signal strength of the $t\bar{t}+\geq 1b$ -jet process or the $t\bar{t}+c\bar{c}$ process is not derived.

Meanwhile, in section 11.6 of part IV the influence of including or neglecting the $t\bar{t}H$ process as a process class in the neural network analysis will be discussed.

In part VI – as mentioned before – the $t\bar{t}+\geq 1b$ -jet predictions of the POWHEG+Pythia8 event generator are compared to predictions by the MG5aMC(NLO)+Pythia8 and Sherpa+OpenLoops event generator. These studies maybe can also provide helpful insight for the future choice of a Monte Carlo event generator applied to describe the $t\bar{t}+\geq 1b$ -jet processes in a future $t\bar{t}H(b\bar{b})$ analysis.

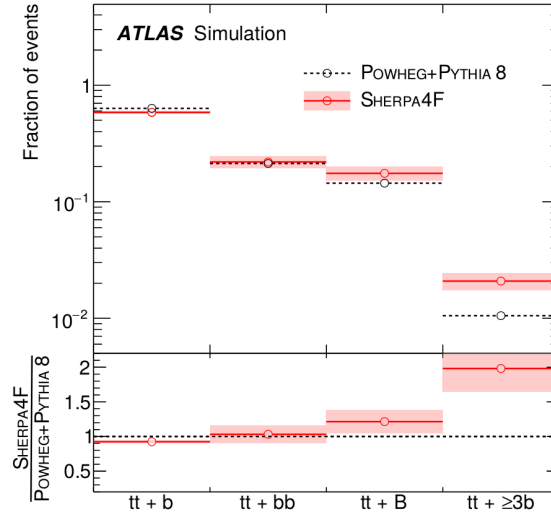


Figure 5.7: Prediction of the relative fractions of the $t\bar{t}+\geq 1b$ -jet process classes: The relative fractions of the $t\bar{t}+\geq 1b$ -jet process classes, according to their definition in the ATLAS $t\bar{t}H(b\bar{b})$ analysis, are compared between the predictions of the POWHEG+Pythia8 and the Sherpa+OpenLoops event generator. In addition, the systematic uncertainties considered for the Sherpa+OpenLoops event generator are shown as an uncertainty band. The shown $t\bar{t}+B$ process class corresponds to the $t\bar{t}+2b$ process class in this thesis. Taken from [150].

Application of neural network based classification considering different $t\bar{t}+X$ process classes

In experimental particle physics multivariate statistical analysis [117] and its related techniques are essential workhorses in many cases. By making use of multiple input variables they produce output variables or final discriminants with a better separation between signal and background than would be obtained from the application of single variables.

Multivariate analysis (MVA) methods and also neural networks were used in a significant number of particle physics publications since the 1980s and the early 1990s [151–158], while at first they were mostly applied to improve identification of physics objects and reconstruction of events and less to produce final discriminants. At the Large Electron-Positron Collider (LEP) MVA methods were already established as a tool set of heavy quark physics [159].

Later on, the application of neural networks as a final classifier for a particle physics analysis became more often. For example, in the Run 2 of the Tevatron the CDF collaboration used neural networks provided by the Neurobayes neural network package [160], while by the DØ collaboration at the Tevatron and in the first years of the LHC other MVA methods like (boosted) decision trees [161, 162] were still more commonly employed. In this context the “Measurement of the Cross Section for $t\bar{t}$ Production in $p\bar{p}$ Collisions using the Kinematics of Lepton+Jets Events” analysis [163] conducted by the CDF collaboration at the Tevatron and employing neural networks is worth mentioning because it can be seen as a predecessor of the physics analysis presented in this thesis.

Nonetheless, it can be correctly stated that only from roughly 2015 on the time was ripe for a rediscovery of neural networks as MVA techniques and, especially, a widespread application of deep learning techniques in particle physics and also in industry which is still picking up momentum. In 2015, the TensorFlow open-source software library [164] was released by Google, which provided a rather easy framework to adopt deep¹ neural networks in various software projects. In the same year, the NVidia company entered the deep learning market [165] by releasing the seventh version of the NVidia CUDA programming platform which allowed to run TensorFlow and other frameworks providing artificial neural networks on a graphical processing unit (GPU). The straightforward parallelization of the steps needed to train neural networks on a GPU, which is provided by TensorFlow and other frameworks by abstraction of more complex steps, can be seen as a breakthrough and the ignition of an upcoming new era of computing which is expected to provide artificial intelligence computing to a wide scope of application [166, 167].

In particle physics the adoption of new technologies and implementation of state-of-the-art techniques has a long history. It was generally driven by the quest for an improvement of the precision of physics measurements as demonstrated by outlining the adoption of the first multivariate analysis techniques above.

Another example is the replacement of the CSVv1 b-tagging algorithm [168], which is relying on a likelihood ratio to combine various basic event variables to a b-tagging value for a jet and which was applied in LHC Run 1 as the default algorithm by the CMS collaboration, with the CSVv2 b-tagging algorithm [169]. The CSVv2 algorithm uses rather simple neural networks instead and was replaced by the DeepCSV b-tagging algorithm [90] during the LHC Run 2, based on more complex neural networks and including more event

¹The term of a deep neural network is not generally defined. But usually feed-forward neural networks containing at least a few hidden layers or more complex convolutional neural networks are considered deep neural networks. In the context of this thesis the term “neural network” without the addition “deep” will be used and refers to feed-forward networks.

information. From figure 5.8 it is evident that the application of a more elaborate neural network and the possibility to include further event information improves both the b-tagging efficiency as well as the efficiency of tagging charm jets.

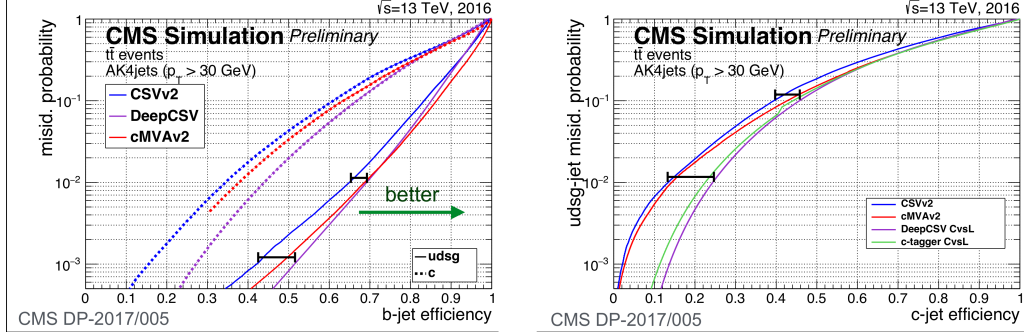


Figure 5.8: Comparison between the performance of different jet tagging algorithms applied by the CMS collaboration: In the left plot, the b-tagging efficiency of the three jet taggers CSVv2, DeepCSV, cMVA2 is compared, while in the right plot their c-tagging efficiency is shown. Curves tending to the right present a better jet tagging efficiency. Taken from [90].

In this thesis, a neural network based analysis, which relies on the TensorFlow open-source software library, is introduced to measure the $t\bar{t} + \geq 1b$ -jet signal strength r . As part of this, a neural network is used for a multiclassification of the $t\bar{t} + X$ processes as well as of the $t\bar{t}H$ background. To my knowledge it is one of the first analyses using multiclassification for a $t\bar{t} + \geq 1b$ -jet measurement.

The multiclassification approach is motivated by the expectation that a simultaneous classification of the individual $t\bar{t} + X$ processes leads to a better separation between the $t\bar{t} + \geq 1b$ -jet signal processes and the $t\bar{t} + c\bar{c}$, $t\bar{t} + lf$, $t\bar{t}H$ background processes and, hence, a better precision of a $t\bar{t} + \geq 1b$ -jet signal strength measurement. Moreover, the multiclassification and the associated determination of multiple process classes enables to consider either a single process class or a combination of quite a few process classes as signal process during a Maximum-Likelihood fit providing the final signal strength r .

Hence, a plurality of new signal strength measurements and further studies are conceivable. Three of these various possibilities are exploited in this thesis: In part IV a measurement of the $t\bar{t} + \geq 1b$ -jet signal strength is presented, while in part V signal strengths of single process classes are obtained by a simultaneous measurement.

In the context of the variety of new opportunities due to the multiclassification, these measurements and the related studies can also be seen as a trailblazer and part of a possibly upcoming flood of new physics analyses making use of multiclassifying neural networks.

Part III

Review of past $t\bar{t} + \geq 1b\text{-jet}$ and $t\bar{t} + b\bar{b}$ measurements

Review of past $t\bar{t}+\geq 1\text{b-jet}$ and $t\bar{t}+b\bar{b}$ measurements

In the following a selection of past $t\bar{t}+\geq 1\text{b-jet}$ and $t\bar{t}+b\bar{b}$ measurements will be reviewed. Only measurements which were conducted at the Large Hadron Collider and which contain a dedicated result for either the $t\bar{t}+\geq 1\text{b-jet}$ or the $t\bar{t}+b\bar{b}$ process were selected. After the presentation of the single measurements a summary and comparison of these measurements will be given. The measurements will be also compared to the measurements presented in this thesis.

Measurement of fiducial cross sections for $t\bar{t}$ production with one or two additional b-jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector

This measurement [170] was conducted by the ATLAS collaboration. It makes use of an integrated luminosity of $\mathcal{L} = 20.3\text{fb}^{-1}$ which was obtained at a center-of-mass energy of $\sqrt{s} = 8$ TeV. As part of this measurement two $t\bar{t}+\geq 1\text{b-jet}$ fiducial cross sections were extracted by a binned profile likelihood fit: One $t\bar{t}+\geq 1\text{b-jet}$ fiducial cross section is derived from a semileptonic measurement, in which exactly one lepton and at least five jets, two of which must fulfill a b-tagging criterion, are required. A set of isolation criteria with an efficiency of 90 % is used for electrons, while for muons a 97 % efficient set of selection criteria is used. The second $t\bar{t}+\geq 1\text{b-jet}$ fiducial cross section is obtained from a dileptonic measurement, in which exactly two leptons and at least three jets, two of which are b-tagged, are required. The summary of the main event selection is also shown in table 5.1.

Table 5.1: Summary of the main event selection criteria applied in the ATLAS $t\bar{t}+\geq 1\text{b-jet}$ measurement: The requirements of the semileptonic and dileptonic measurement of the $t\bar{t}+\geq 1\text{b-jet}$ fiducial cross section are presented. Values taken from [170].

Requirement	$t\bar{t}+\geq 1\text{b-jet}$ semileptonic	$t\bar{t}+\geq 1\text{b-jet}$ dileptonic
N_{leptons}	1	2
Electron iso. efficiency	90 %	90 %
Muon iso. efficiency	97 %	97 %
N_{jets}	≥ 5	≥ 2
$N_{\text{b-tagged jets}}$	≥ 2	≥ 2

For the simulation of the $t\bar{t}$ signal events the POWHEG Box Version 1 next-to-leading order (NLO) matrix event generator [30,31,171] interfaced to the Pythia6 general-purpose Monte Carlo generator [38] was applied. The $t\bar{t}$ events were renormalized to the inclusive cross section obtained from a next-to-next-to-leading order (NNLO) QCD calculation which included resummation of next-to-next-to-leading logarithmic (NNLL) soft gluon terms

and which was done with the TOP++ program [172–177].

From both measurements a $t\bar{t} + \geq 1b$ -jet fiducial cross section was observed which is higher than the predicted cross section of the POWHEG+Pythia6 sample. For the semileptonic measurement an observed signal strength of $r(\text{observed}) = 1.32$ was found and for the dileptonic an observed signal strength of $r(\text{observed}) = 1.30$. Due to the higher acceptance times branching ratio in the semileptonic decay channel, the total measurement uncertainty of $\Delta r \approx 25\%$ is smaller than the total measurement uncertainty of $\Delta r \approx 32\%$ obtained in the dileptonic decay channel. In the semileptonic measurement major uncertainties are uncertainties related to the b-tagging of charm jets (+16.2 %, –13.4 %), the total modelling of the $t\bar{t}$ event content (+13.1 %, –13.7 %), and to the modelling of the parton shower and hadronisation (+11.4 %, –12.1 %). In comparison to that, major uncertainties in the dileptonic measurement are the total modelling of the $t\bar{t}$ event content (+23.8 %, –16.1 %), generator uncertainties (+23.3 %, –15.1 %) as well as the statistical uncertainty (+19.2 %, –17.9 %).

Besides the determination of the $t\bar{t} + \geq 1b$ -jet signal strength and cross section, also an $t\bar{t} + b\bar{b}$ fiducial cross section was measured using the dileptonic decay channel and two different approaches. In the first cut-based approach tight selection criteria were applied, while in the second fit-based approach a looser selection was employed allowing the extraction of the background normalisation from data. In both analyses it was required that the mass of the two leptons is above 15 GeV and for opposite-sign, same-flavour leptons an additional cut of ± 10 GeV around the Z^0 boson mass was applied. A summary of the main event selection criteria of the two $t\bar{t} + b\bar{b}$ measurements is given in table 5.2.

Table 5.2: Summary of the main event selection criteria applied in the ATLAS $t\bar{t} + b\bar{b}$ measurement: The requirements of the cut-based and fit-based measurement of the $t\bar{t} + b\bar{b}$ fiducial cross section are presented. Values taken from [170].

Requirement	$t\bar{t} + b\bar{b}$ cut-based	$t\bar{t} + b\bar{b}$ fit-based
N_{leptons}	2	2
Electron iso. efficiency	98 %	98 %
Muon iso. efficiency	97 %	97 %
$m_{ll} > 15 \text{ GeV}$	Yes	Yes
$ m_{ee/\mu\mu} - 91 \text{ GeV} > 10 \text{ GeV}$	Yes	Yes
N_{jets}	≥ 4	≥ 4
$N_{b\text{-tagged jets}}$	≥ 4	≥ 2

The fiducial cross sections obtained from the two $t\bar{t} + \geq 1b$ -jet measurements as well from the $t\bar{t} + b\bar{b}$ measurements are presented in table 5.3.

Table 5.3: Summary of the measured fiducial cross sections in the ATLAS $t\bar{t}+\geq 1\text{b-jet}$ and $t\bar{t}+b\bar{b}$ measurement: The observed cross sections $\sigma(\text{obs.})$ and predicted cross sections $\sigma(\text{pred.})$ are taken from [170]. The measured signal strength $r(\text{observed})$ is calculated out of these two cross sections.

Analysis	$\sigma(\text{obs.})$ [fb]			$\sigma(\text{pred.})$ [fb]	$r(\text{observed})$
		stat.	syst.		
$t\bar{t}+\geq 1\text{b-jet}$ semileptonic	950	± 70	$^{+240}_{-190}$	720	$1.32^{+0.43}_{-0.36}$
$t\bar{t}+\geq 1\text{b-jet}$ dileptonic	50	± 10	$^{+15}_{-10}$	38	$1.32^{+0.65}_{-0.53}$
$t\bar{t}+b\bar{b}$ cut-based	19.3	± 3.5	± 5.7	12.3	$1.57^{+0.75}_{-0.75}$
$t\bar{t}+b\bar{b}$ fit-based	13.5	± 3.3	± 3.6	12.3	$1.10^{+0.56}_{-0.56}$

Measurement of $t\bar{t}$ production with additional jet activity, including b quark jets, in the dilepton decay channel using pp collisions at $\sqrt{s} = 8$ TeV

This measurement [178] was conducted by the CMS collaboration in the dileptonic decay channel. It makes use of an integrated luminosity of $\mathcal{L} = 19.7 \text{ fb}^{-1}$ which was obtained at a center-of-mass energy of $\sqrt{s} = 8$ TeV. As part of this measurement differential distributions of the $t\bar{t} + \geq 1 \text{ b-jet}$ and the $t\bar{t} + b\bar{b}$ cross section were derived in the visible phase space of the additional b-tagged jets and in the full phase space of the $t\bar{t}$ system. For the simulation of the nominal $t\bar{t}$ signal events the MG5aMC(LO) matrix event generator [179] interfaced to the MadSpin package [180] and the Pythia6 general-purpose Monte Carlo generator [38] was applied. The $t\bar{t}$ events were renormalized to the inclusive cross section from a next-to-next-to-leading order (NNLO) QCD calculation which included resummation of next-to-next-to-leading logarithmic (NNLL) soft gluon terms and which was done with the TOP++ program [172–177]. In addition, further Monte Carlo event generators were employed to produce additional $t\bar{t}$ samples, so that the predictions of these event generators could be compared to the obtained result at particle-level: MC@NLO [54], POWHEG Box Version 1 [31], and MG5aMC(NLO) [29] which were either interfaced with Pythia6 [38] or Herwig6 [181].

The measurement found that the observed $t\bar{t} + \geq 1 \text{ b-jet}$ signal strength $r(t\bar{t} + \geq 1 \text{ b-jet}, \text{observed})$ and associated $t\bar{t} + \geq 1 \text{ b-jet}$ cross section is roughly 1.3 times larger than the signal strength $r(\text{expected})$ predicted by the MG5aMC(LO) or POWHEG Box Version 1 event generator. The $t\bar{t} + b\bar{b}$ cross section is underestimated by a factor 1.8 by both event generators. No uncertainties on these estimates of the observed $t\bar{t} + \geq 1 \text{ b-jet}$ and of the observed $t\bar{t} + b\bar{b}$ signal strength are given in this publication.

The dominant systematic uncertainties in this measurement are the b-tagging efficiency (up to 20 %), the jet-energy scale (JES, up to 15%), and the variation of the renormalization and factorization scale (8 %).

In figure 5.9 the $t\bar{t} + \geq 1 \text{ b-jet}$ differential cross sections as a function of the transverse momentum p_T and the absolute pseudorapidity $|\eta|$ of the leading additional b-tagged jet are shown as an example.

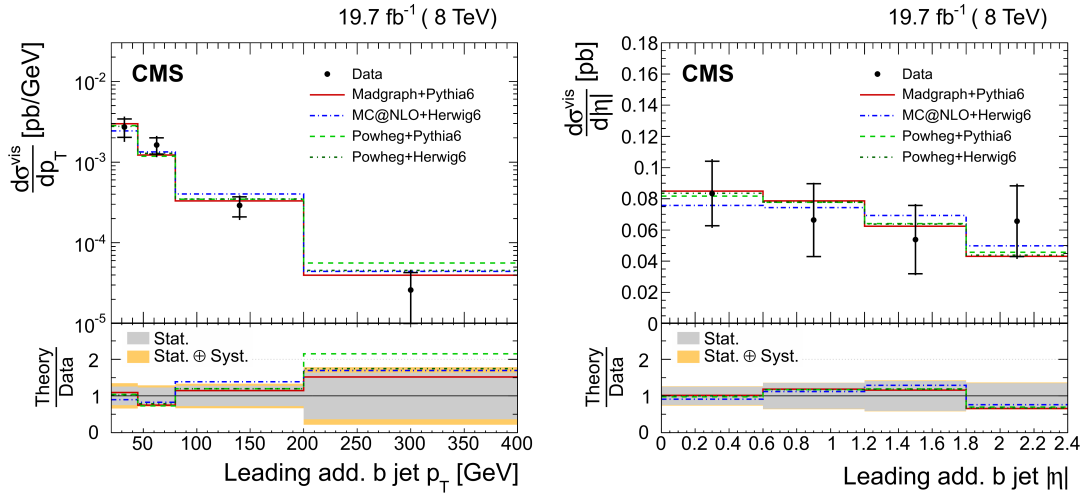


Figure 5.9: Differential cross section distributions of the CMS $t\bar{t} + \geq 1b$ -jet measurement: The differential $t\bar{t} + \geq 1b$ -jet cross section is given as a function of the transverse momentum p_T (left plot) and of the absolute pseudorapidity $|\eta|$ of the leading b-tagged jet. Data are compared with predictions from MG5aMC(LO) interfaced with Pythia6, MC@NLO interfaced with Herwig6, and POWHEG Box Version 1 interfaced with Pythia6 and Herwig6, normalized to the measured inclusive cross section. The inner (outer) vertical bars indicate the statistical (total) uncertainties. The lower part of each plot shows the ratio of the predictions to the data. Taken from [178].

Measurements of $t\bar{t}$ cross sections in association with b jets and inclusive jets and their ratio using dilepton final states in pp collisions at $\sqrt{s} = 13$ TeV

This measurement [182] was conducted by the CMS collaboration in the dileptonic decay channel. It makes use of an integrated luminosity of $\mathcal{L} = 2.3 \text{ fb}^{-1}$ which was obtained at a center-of-mass energy of $\sqrt{s} = 13$ TeV. As part of this measurement the $t\bar{t}+b\bar{b}$ and $t\bar{t}+j\bar{j}$ cross section were derived in the visible phase space, defined as events having two leptons with $p_T > 20$ GeV and $|\eta| < 2.4$, plus at least four jets, including at least two b-tagged jets. Furthermore, the two cross sections were measured in the full phase space, corrected for acceptance and branching fractions.

For the simulation of the nominal $t\bar{t}$ signal events the POWHEG Box Version 2 matrix event generator [30,31,171] interfaced to the Pythia8 general-purpose Monte Carlo generator [37,183] was applied. The $t\bar{t}$ events were renormalized to the inclusive cross section from a next-to-next-to-leading order (NNLO) QCD calculation which included resummation of next-to-next-to-leading logarithmic (NNLL) soft gluon terms and which was done with the TOP++ program [177]. In addition, further matrix element event generators were employed to produce additional $t\bar{t}$ samples for cross checks and studies of systematic uncertainties: MG5aMC(LO) [179] and MG5aMC(NLO) [29] which were also interfaced with Pythia8.

For the discrimination between the $t\bar{t}+b\bar{b}$ and $t\bar{t}+j\bar{j}$ events the CSVv2 b-tagging discriminators [168] of the third and fourth leading jets are applied, which are shown in figure 5.10.

The measurement found an inclusive $t\bar{t}+b\bar{b}$ cross section of $\sigma(t\bar{t}+b\bar{b}) = 0.088 \pm 0.012(\text{stat.}) \pm 0.029(\text{syst.})$ pb and a $t\bar{t}+j\bar{j}$ cross section of $\sigma(t\bar{t}+j\bar{j}) = 4.0 \pm (\text{stat.})0.6 \pm 1.3(\text{syst.})$ pb in the visible phase space. The individual cross sections were also determined for the full phase space: $\sigma(t\bar{t}+b\bar{b}) = 4.0 \pm 0.6(\text{stat.}) \pm 1.3(\text{syst.})$ pb and $\sigma(t\bar{t}+j\bar{j}) = 184 \pm 6(\text{stat.}) \pm 33(\text{syst.})$ pb. The measured cross sections $\sigma(t\bar{t}+b\bar{b})$ and $\sigma(t\bar{t}+j\bar{j})$ as well as the predictions by the POWHEG Box Version 2 event generator are also given in table 5.4.

By a comparison of the observed and predicted $t\bar{t}+b\bar{b}$ cross section $\sigma(t\bar{t}+b\bar{b})$ it can be concluded that the POWHEG event generator underestimates the $t\bar{t}+b\bar{b}$ cross section by a factor of ≈ 1.26 in the visible phase space and by a factor of ≈ 1.25 in the full phase space. Similarly to the other measurements, the dominant uncertainties of this measurement are related to the b-tagging of jets (up to 19 %), the dependency on the chosen scale in the parton shower (13 %), the dependency on the applied Monte Carlo event generator (9.4 %), and the jet-energy scale and jet-energy resolution (7.8 %).

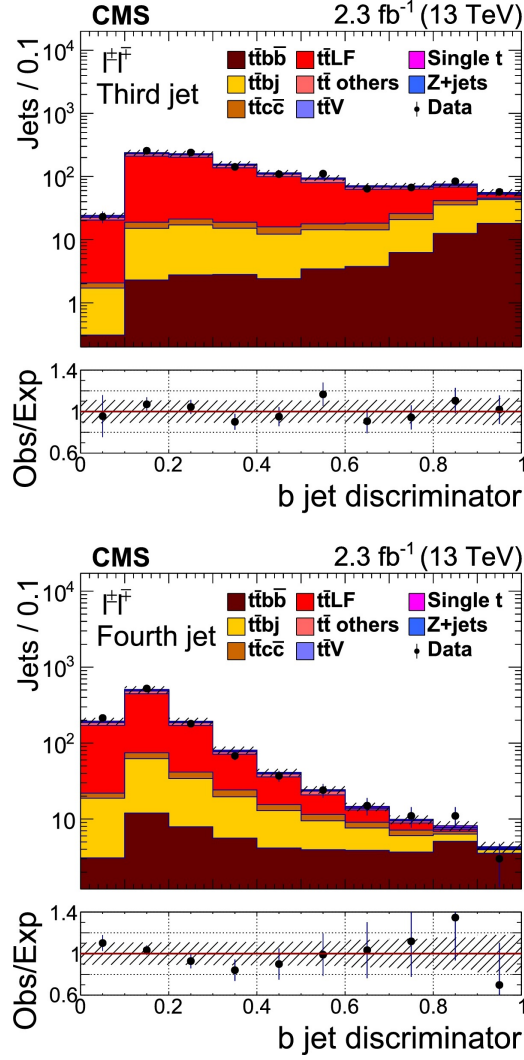


Figure 5.10: Distributions of the CSVv2 b-tagging discriminators for the third (top) and fourth (bottom) leading jets in decreasing order of b-tagging discriminator value. The points show the data and the stacked histograms are from simulated events, normalized by the results of the fit. The ratio of the number of data events to the expected number, as given by the stacked histograms, is shown in the lower panels. The hatched region indicates the modelling uncertainty in the MC simulation. Taken from [182].

Phase space	Obs./Pred.	$\sigma(t\bar{t}+b\bar{b})$ [pb]	$\sigma(t\bar{t}+j\bar{j})$ [pb]
Visible	Observed	$0.088 \pm 0.012 \pm 0.029$	$3.7 \pm 0.1 \pm 0.7$
Visible	POWHEG	0.070 ± 0.009	5.1 ± 0.5
Full	Observed	$4.0 \pm 0.6 \pm 1.3$	$184 \pm 6 \pm 33$
Full	POWHEG	3.2 ± 0.4	257 ± 26

Table 5.4: Summary of the measured $\sigma(t\bar{t}+b\bar{b})$ and $\sigma(t\bar{t}+j\bar{j})$ cross sections as well as their ratio $\sigma(t\bar{t}+b\bar{b})/\sigma(t\bar{t}+j\bar{j})$ in the CMS measurement: The cross sections and their ratio were determined for the visible and full phase space. The uncertainties on the measurements are separated into statistical and systematic components, while those are combined for the POWHEG Box Version 2 predictions. Values taken from [182].

Summary and comparison

An overview of the analysis setup of the three reviewed measurements is given in table 5.5, while the derived results are summarized in table 5.6.

The (older) ATLAS measurement and the older CMS measurement were conducted at a center-of-mass energy $\sqrt{s}$ of 8 TeV, while the newer CMS measurement was done at a center-of-mass energy $\sqrt{s}$ of 13 TeV. The measurements presented in this thesis are also conducted at this center-of-mass energy. Due to the increase in the center-of-mass energy an increase in the predicted $t\bar{t}+b\bar{b}$ cross section $\sigma(t\bar{t}+b\bar{b}, \text{pred.})$ of roughly a factor 4 is expected [184].

The two older measurements made use of a similar luminosity $\mathcal{L}$ of roughly 20 fb^{-1} . The newer CMS measurement used only a luminosity of 2.3 fb^{-1} . For the measurements presented in this thesis a luminosity of 35.9 fb^{-1} is employed.

The ATLAS measurement is the only measurement in which the $t\bar{t}+\geq 1\text{b-jet}$ and the $t\bar{t}+b\bar{b}$ process was measured in the semileptonic decay channel of the top quark – like in this thesis – and in the dileptonic decay channel as well. The two other measurements were conducted by using only the dileptonic decay channel. None of the three measurements gives a result for the $t\bar{t}+2\text{b}$ or $t\bar{t}+b$ process as it is done in this thesis.

The event selection applied in the three measurements is outlined by presenting the number of required leptons, jets and b-tagged jets. In contrast to these measurements and, in particular the ATLAS semileptonic measurement, in this thesis a tighter event selection was chosen by requiring at least six jets and two b-tagged jets besides a single lepton.

In this way, a phase space region is selected in which $t\bar{t}+\geq 1\text{b-jet}$ events are enriched and background is partly suppressed. Furthermore, the rather homogenous phase space made it possible to use only a single classifier. If a larger phase space would have been considered it would have been necessary to train and employ multiple multivariate classifiers covering different jet multiplicity and b-jet multiplicity categories which would have complicated a comparison between the reference analyses employed in this thesis (chapter 9).

The ATLAS measurement and the newer CMS measurement rely on b-tagging probability distributions as final discriminators or make use of a cut-and-count approach. The older CMS measurement employs a boosted decision tree (BDT) as classifier. Compared to that, the best classifier used in this thesis is a neural network, while also a b-jet multiplicity distribution and a distribution provided by a linear discriminant are employed. These two simpler classifiers are provided as a cross check of the elaborate neural network and allow the comparison with the past measurements.

For the generation of the nominal $t\bar{t}$ sample different event generators were used: The ATLAS measurement makes use of the older version of the next-to-leading order (NLO) POWHEG matrix element (ME) event generator (POWHEG Box V1) together with the older version of the Pythia general-purpose Monte Carlo (GPMC) event generator (Pythia6). The older CMS measurement employs a leading-order (LO) ME event generator by making use of MG5aMC(LO) and matches it to the Pythia6 event generator. For the newer CMS measurement with the combination of the NLO POWHEG Box Version 2 ME generator and the Pythia8 GPMC event generator the same combination is used as in this thesis. In principle, it is expected that these newer event generators provide a more accurate description of the $t\bar{t}+\geq 1\text{b-jet}$ and the $t\bar{t}+b\bar{b}$ processes as already mentioned in part II.

All measurements (including the measurements of this thesis) use the same program (TOP++) to calculate the cross section of the inclusive $t\bar{t}$ sample at next-to-next-leading order (NNLO) QCD accuracy with next-to-next-leading log (NNLL) soft gluon resummation.

The observed signal strengths $r(\text{observed})$ are summarized in table 5.6. The observed signal strengths $r(\text{observed})$ of the ATLAS measurement and of the newer CMS measurement conducted at $\sqrt{s} = 13$ TeV were calculated by using the published cross section results. In addition, the observed $t\bar{t}+b\bar{b}$ cross section for the full phase space $\sigma(t\bar{t}+b\bar{b})$ is listed which was only available for the newer CMS $\sqrt{s} = 13$ TeV measurement.

As can be seen from table 5.6 the measurements are for the observed $t\bar{t}+\geq 1\text{b-jet}$ signal strength $r(t\bar{t}+\geq 1\text{b-jet}, \text{observed})$ in good agreement with each other. According to the measurements the observed $t\bar{t}+\geq 1\text{b-jet}$ signal strength $r(t\bar{t}+\geq 1\text{b-jet}, \text{observed})$ is roughly 1.3.

However, for the observed $t\bar{t}+b\bar{b}$ signal strength $r(t\bar{t}+b\bar{b}, \text{observed})$ there is a rather strong tension between the measurements since the observed $t\bar{t}+b\bar{b}$ signal strength ranges from 1.1 to 1.8.

Table 5.5: Overview of the analysis setup in the three reviewed $t\bar{t}+\geq 1\text{b-jet}$ and $t\bar{t}+b\bar{b}$ measurements: For each measurement the center-of-mass energy $\sqrt{s}$, the luminosity $\mathcal{L}$, the measured processes and the decay channels are stated. Furthermore, the number of required leptons, required jets, and required b-tagged jets is listed. In addition, the classifier employed to derive the result is stated. The event generator combination to produce the nominal $t\bar{t}$ sample is also listed as well as the $t\bar{t}$ cross section $\sigma(t\bar{t})$ applied for the $t\bar{t}$ event sample normalization. Values taken from [170, 178, 182].

Measurement	ATLAS		CMS	CMS
$\sqrt{s}$	8 TeV		8 TeV	13 TeV
Luminosity $\mathcal{L}$	20.3 fb^{-1}		19.7 fb^{-1}	2.3 fb^{-1}
Processes	$t\bar{t}+\geq 1\text{b-jet}, t\bar{t}+b\bar{b}$		$t\bar{t}+\geq 1\text{b-jet}, t\bar{t}+b\bar{b}$	$t\bar{t}+b\bar{b}$
Channel	semilepton	dilepton	dilepton	dilepton
Required leptons	1	2	2	2
Required jets	≥ 5	≥ 4	-	≥ 4
Required b-tagged jets	≥ 2	≥ 4	≥ 3	≥ 2
Classifier	b-tag. prob.	b-tag. prob. / cut-and-count	Boosted Decision Tree	b-tag. prob.
Nominal $t\bar{t}$ sample:	POWHEG Box V1+		MG5aMC(LO)+	POWHEG Box V2+
Event generator	Pythia6		Pythia6	Pythia8
$\sigma(t\bar{t})$	NNLO+NNLL inclusive $t\bar{t}$ cross section from TOP++ program			

Table 5.6: Summary of the signal strengths and cross sections derived in the three reviewed $t\bar{t}+\geq 1\text{b-jet}$ and $t\bar{t}+b\bar{b}$ measurements: The observed signal strength $r(\text{observed})$ and the observed cross section for the full phase space is listed if available. In the case of the ATLAS measurement the observed signal strength $r(\text{observed})$ is calculated from fiducial cross sections. In the case of the CMS $\sqrt{s} = 13$ TeV measurement the observed signal strength $r(\text{observed})$ is calculated from full phase space cross sections. Values taken from [170, 178, 182].

Analysis	$r(\text{observed})$	$\sigma(\text{obs.})$ [pb]
ATLAS measurement		
$t\bar{t}+\geq 1\text{b-jet}$ semileptonic	$1.32^{+0.43}_{-0.36}$	-
$t\bar{t}+\geq 1\text{b-jet}$ dileptonic	$1.32^{+0.65}_{-0.53}$	-
$t\bar{t}+b\bar{b}$ cut-based	1.57 ± 0.75	-
$t\bar{t}+b\bar{b}$ fit-based	1.10 ± 0.56	-
CMS $\sqrt{s} = 8$ TeV measurement		
$t\bar{t}+\geq 1\text{b-jet}$	≈ 1.3	-
$t\bar{t}+b\bar{b}$	≈ 1.8	-
CMS $\sqrt{s} = 13$ TeV measurement		
$t\bar{t}+b\bar{b}$	1.25 ± 0.59	$4.0 \pm 0.6(\text{stat.}) \pm 1.3(\text{syst.})$

Part IV

Measurement of the $t\bar{t} + \geq 1\text{b-jet}$ signal strength and cross section

Measurement of the $t\bar{t}+\geq 1\text{b-jet}$ signal strength and cross section

In this part of the thesis, a measurement of the $t\bar{t}+\geq 1\text{b-jet}$ cross section, expressed by the $t\bar{t}+\geq 1\text{b-jet}$ signal strength r , using the CMS data from the Large Hadron Collider run in 2016 is conducted.

In the following chapters the analysis will be presented: In chapter 6 the Monte Carlo samples applied to describe the signal and background processes will be discussed in detail. This is followed by a few details about the CMS measurement data considered for this analysis and the selection of events, independent of their origin from Monte Carlo samples or measurement data (chapter 7). In chapter 8 an overview of the analysis framework is given. Furthermore, the agreement between the predictions by the Monte Carlo samples and the measurement data after the event selection is checked. The strategy of this analysis is explained in detail in chapter 9. In this chapter also the three different sub analyses, which are employed throughout this thesis, are introduced. A discussion of the systematic uncertainties considered in all analyses follows in chapter 10. The overall result of the $t\bar{t}+\geq 1\text{b-jet}$ signal strength r measurement is presented in the first part of chapter 11, while further aspects of the result are studied closer in the subsequent sections. In the last chapter of this part the previously obtained results are discussed and summarized as well as compared to the findings of the previously reviewed measurements (part III).

6. Signal and background processes

Event samples derived from Monte Carlo simulations are an important ingredient to a particle physics analysis, as explained in section 2.2. Each process which could contain events passing the selection criteria of an analysis has to be considered via the inclusion of associated event samples in the corresponding analysis.

However, since the preparation and production of samples of Monte Carlo simulated events (“MC samples”) can be a time-consuming, challenging, and computationally intensive task, Monte Carlo samples of general interest are produced centrally by the CMS collaboration. Favourable side effects of this sharing of Monte Carlo samples among analysis groups are also that the limited storage space is better exploited since the produced Monte Carlo samples can be in the order of Terabytes and that produced samples get cross checked by different groups.

For this analysis the Standard Model physics processes listed in table 6.1 are considered and their corresponding Monte Carlo event samples are applied. All these event samples were produced centrally in the course of the CMS “RunII Summer16 MiniAOD” campaign.

Not considered in this analysis is the so-called QCD background which corresponds to events originating from light-flavour multijet processes because it was shown that these processes only contribute with a tiny fraction of events to the selected phase space. Furthermore, since only a small number of QCD events fall in the phase space defined by the selection criteria of this analysis, the few remaining events have quite large uncertainties and some bins of the final discriminants have even negative event numbers. Thus, these events seem more of hindrance than beneficial for this analysis. For this reason, these kind of processes are neglected for the remainder of this thesis. The total number of neglected QCD events after the preselection would correspond to $N(\text{QCD}) = 4.7 \pm 31.2$. In comparison the total number of events is roughly 92 000.

In table 6.1, besides the nominal Monte Carlo samples, also eight $t\bar{t}$ systematic samples are stated which are applied to take into account systematic uncertainties due to a variation of the $t\bar{t}$ signal samples. They will be discussed in more detail in section 10.1.

6.1 General properties of the Monte Carlo samples

For the production of most of the event samples a dedicated matrix element event generator like MG5aMC [29] (version 5.2.2.2) or POWHEG Box Version 2 [30–34] (version from September 2015) was applied for the simulation of the hard sub process. If either MG5aMC(NLO) or POWHEG Box Version 2 is listed as event generator, the matrix element computation was done in next-to-leading (NLO) order of perturbation theory, while in the case of MG5aMC(LO) only a leading-order computation was conducted. In some rare cases like for the WW+jets, the WZ+jets, and the ZZ+jets processes only the internal matrix elements of the general-purpose Monte Carlo (GPMC) event generator Pythia8 [37, 38] (version 8.226) were exploited.

Pythia8 was also used for all event samples as a GPMC event generator to simulate the parton shower, hadronization, and hadron decay steps (cf. section 2.2.1). For the modelling of the underlying events by the Pythia8 GPMC event generator either one of the following two CMS in-house tunes “CUETP8M1” or “CUETP8M2” was applied. Both tunes are derived from the Pythia8 Monash tune [139] and “CUETP8M2” is a newer version of the “CUETP8M1” tune described in [185].

Moreover, all event samples derived from the Monte Carlo event generators underwent a CMS detector simulation (cf. section 2.2.1) for which the Geant 4 toolkit [51] was used.

In all samples the parameter of the top-quark mass is set to $m(\text{top}) = 172.5 \text{ GeV}$ and the parameter of the Higgs-boson mass to $m(\text{H}) = 125 \text{ GeV}$.

The NNPDF3.0 set [186] of parton distribution functions (PDFs) was used in all cases as nominal PDF set.

6.2 Cross sections of the Monte Carlo samples

For all event samples inclusive cross section predictions from dedicated theoretical calculations more precise than by the matrix element generator computations are available. Therefore, all event samples of a given physics process are scaled to the corresponding cross section derived from theoretical calculations. This is done by summing up the weights of

all events $\sum_i^N w_i$ contained in an event sample of a given physics process and scaling of this sum, so that the overall event weight matches the number of expected events in the measurement data. Hereby, the number of expected events $N(\text{expct.})$ is derived by multiplication of the cross section prediction $\sigma(\text{pred.})$ with the integrated luminosity $\int \mathcal{L} \text{ dt}$, which was recorded to obtain the measurement data and which corresponds in the case of this analysis to 36 fb^{-1} [187]:

$$\sum_i^N w_i := \sigma(\text{pred.}) \cdot \int \mathcal{L} \text{ dt} = N(\text{expct.})$$

In table 6.1 also the cross section predictions $\sigma(\text{pred.})$ are given. For the split $t\bar{t}$ and $t\bar{t}H$ samples the cross section prediction $\sigma(\text{pred.})$ is multiplied by the corresponding branching ratio $\mathcal{BR}$.

For the total cross section of the top-quark pair production ($t\bar{t}$) samples a prediction at next-to-next-leading order (NNLO) QCD accuracy with next-to-next-leading logarithmic (NNLL) soft gluon resummation using the top++2.0 program (see [177] and references therein) is used. Following the recommendations of the LHC Higgs Cross Section Working Group (LHCHXSWG) [47] an NLO QCD cross section which includes next-to-leading

logarithmic (NLL) soft-gluon radiation corrections is applied for the $t\bar{t}H$ samples. Also following the LHCHSWG recommendations NLO QCD cross sections including electroweak (EW) corrections are used for the $t\bar{t}W$ and $t\bar{t}Z$ processes. The cross section of the s-channel single-top quark production and the tW-channel production is obtained from an NNLO approximate calculation [105], while the t-channel cross section is derived from an NLO calculation using the Hathor v2.1 program [188, 189]. For the W+jets and Z+jets processes NNLO cross sections calculated with the program FEWZ 3.1 [190–193] are used. The NNLO QCD cross section of the WW+jets process is taken from [194], while the NLO QCD cross sections of the WZ+jets and the ZZ+jets processes are calculated with the MCFM 6.6 program [195, 196].

6.3 Choice of the $t\bar{t}$ signal samples

Three kinds of $t\bar{t}$ samples are centrally produced by the CMS collaboration using different matrix element generators which are considered in this thesis:

First of all, inclusive $t\bar{t}$ samples are produced by the MG5aMC(NLO) event generator in next-to-leading order perturbation theory, while making use of the so-called FxFx merging procedure [56] mentioned in section 2.2.3. These inclusive $t\bar{t}$ samples are obtained by merging three NLO MG5aMC(NLO) event samples: An exclusive $t\bar{t}+0$ jet sample is merged with an exclusive $t\bar{t}+1$ jet and an inclusive $t\bar{t}+2$ jets sample. The resulting inclusive $t\bar{t}+0/1/2$ jets sample has NLO accuracy in processes with up to two additional jets. However, this excellent description of the $t\bar{t}$ processes in higher jet multiplicities comes at a cost. NLO event samples produced with MG5aMC(NLO) have a non-negligible fraction of events which have a negative event weight, resulting in a larger statistical uncertainty of these samples. According to the weak law of large numbers the relative statistical uncertainty of a result obtained using N Monte Carlo events, each having a weight w_i , can be generally stated as:

$$\frac{\Delta N}{N} = \frac{\sqrt{\sum_i^N w_i^2}}{\sum_i^N w_i}.$$

In the case of events having only positive event weights with value $w_i = +1$ this relation simplifies to

$$\frac{\Delta N}{N} = \frac{1}{\sqrt{N}},$$

which corresponds to the well-known relative statistical uncertainty usually considered for Monte Carlo simulations. However, in the case of the inclusive, FxFx-merged $t\bar{t}+0/1/2$ jets samples roughly $f_n = 32.7\%$ of all events have a negative event weight of -1 , while the other 67.3% of events have a positive event weight of $+1$ ¹. Therefore, the following relation has to be applied to determine the relative statistical uncertainty

$$\frac{\Delta N}{N} = \frac{1}{\sqrt{n}} \cdot \frac{1}{(1 - 2 \cdot f_n)} = 2.89 \frac{1}{\sqrt{n}},$$

¹The case that the absolute value of the negative event weight corresponds to the value of the positive event weight is not universally valid, but in the case of NLO event generation with MG5aMC(NLO) holds true.

which shows that the statistical uncertainty is 2.89 times larger than originally expected due to the many events with negative event weight. Put differently, to obtain the same statistical precision an event sample with 2.89 times more events is required in comparison to an event sample which does not contain events with negative event weight. Besides the significantly lower statistical precision for a given number N of events, events with negative event weight can also cause problems in the training and evaluation of multivariate analysis methods. Hence, one often tries to avoid them.

The number of ($t\bar{t}$) events which can be produced by the CMS collaboration in a Monte Carlo campaign is limited. In addition, for the training of the multivariate analysis methods applied in this analysis a sufficient statistical precision of the $t\bar{t}$ signal samples is vital. For this reason, the MG5aMC(NLO) $t\bar{t}$ samples are not considered a suitable choice and are not applied in this analysis.

The second kind of $t\bar{t}$ samples, which are centrally available and are also produced by application of a merging method, are MG5aMC(LO) $t\bar{t}$ samples. Here, the matrix element computation is conducted in leading-order perturbation theory. Furthermore, MLM merging [55] is applied to derive inclusive $t\bar{t}+0/1/2/3$ jets samples having leading-order accuracy in processes with up to three additional jets. Due to the leading-order matrix element computation no events with negative event weight occur. Therefore, in principle these kind of samples would be a good choice to produce large event samples. However, because of the only leading-order matrix elements applied these samples lack a bit in the description of the hard sub process, the final-state of the top-quark system and, hence, the kinematic properties of the top quark. For this reason, only rather small MG5aMC(LO) event samples are produced by the CMS collaboration.

In comparison to the aforementioned two kinds of samples, the third kind of $t\bar{t}$ samples are computed in next-to-leading order by using the POWHEG Box Version 2 event generator. They have only a tiny fraction of events with negative event weight. Furthermore, since the fraction of events with negative event weight f_n is smaller than 0.5 % and it can be shown that the corresponding events have a negligible impact on the overall description of the $t\bar{t}$ processes, these events having negative weights can safely be neglected in CMS analyses. Nonetheless, the POWHEG Box Version 2 $t\bar{t}$ samples have some drawbacks. Therefore, they are considered only as a good compromise between accuracy and feasibility of producing event samples with sufficient statistical precision. The reason for this is that only the $t\bar{t}+0$ jet process is simulated in next-to-leading order and no merging technique is applied. That is the reason why at most one additional parton and associated jet is described in leading-order accuracy by the matrix element generator. Compared to that all the further partons leading to additional jets have to be produced by the parton shower of the GPMC event generator. In the context of this analysis, in which the $t\bar{t}+\geq 1$ b-jet processes are of particular interest, this can be problematic since the parton shower generally does not describe emissions of high-energetic partons well. Nevertheless, large POWHEG Box Version 2 $t\bar{t}$ samples are centrally produced by the CMS collaboration and are used as default samples for the $t\bar{t}$ process in this thesis.

6.4 Split of the $t\bar{t}$ samples into $t\bar{t}+X$ sub processes according to the flavour of the additional hadrons and jets

Of greatest interest are $t\bar{t}$ processes where the top-quark pair is created in association with additional quarks. These $t\bar{t}+X$ processes are not uniformly defined and their proper definition as well as their differentiation to one another is still part of an ongoing discussion [47, 131]. This is detailed in section II.

Nonetheless, one can define five types of $t\bar{t}+X$ processes based on the particle-level information in the following way (illustrated in figure 6.1) that is rather universally accepted. Still, the problem lies in the detail of the individual implementation. In particular, the definition of bottom hadrons and the applied jet reconstruction algorithm can differ between experiments and analyses.

The previously discussed $t\bar{t}$ samples can be classified and split into $t\bar{t}+X$ sub samples according to the flavour of additional quarks and hadrons occurring in the final-state of the hard sub processes by a heavy-flavour identification based on particle-level information. Such a split of the $t\bar{t}$ samples into five distinct $t\bar{t}+X$ sub samples ($t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$, $t\bar{t}+c\bar{c}$, and $t\bar{t}+lf$) is done in this analysis by using the GenHFBadronMatcher [146, 149] provided in the CMS software framework (CMSSW) [197]. Consequently, the different properties of the individual $t\bar{t}+X$ sub processes can be exploited by multivariate analysis techniques to separate the sub processes in the measurement data.

For the classification of $t\bar{t}$ events according to the flavour of the additional bottom and charm hadrons, which do not originate from the top-quark decay, the GenHFBadronMatcher method requires the matching of selected hadrons to jets at particle-level (“gen-jets”) and the subsequent matching of the selected hadrons to quarks.

Matching of selected hadrons to jets

For the matching of selected hadrons to gen-jets the CMS hadron-based JetFlavour tool [198] is employed. This tool allows an identification of the flavour of a gen-jet based on the flavour of the hadron it contains. It makes usage of so-called “ghost” hadrons to decide if a bottom or charm hadron gets reclustered into a gen-jet which would then be referred to either as a bottom or charm gen-jet: For selected bottom or charm hadrons the four-momentum of the corresponding hadron and its partons gets rescaled by a very small number. In this manner, they become soft (having low energy and low momentum) and are referred to as “ghost particles”. These ghost particles are then added to the parton collection of an event instead of the original hadrons and associated partons, before the gen-jets are reclustered using the default jet clustering parameters. In this way, the selected soft bottom and charm hadrons as well as their partons practically do not affect the overall gen-jet clustering. Nonetheless, they still get clustered into gen-jets by the direction of their three-momentum vector.

Depending on if at least one bottom ghost hadron is clustered into gen-jet, such a gen-jet is considered a bottom gen-jet. A gen-jet is referred to as a charm gen-jet if no bottom ghost hadrons are clustered into the corresponding gen-jet, but at least one charm ghost hadron. The remaining gen-jets which do not contain bottom or charm ghost hadrons are declared as light-flavour gen-jets.

Matching of selected hadrons to quarks

The GenHfHadronMatcher tool makes use of the derived bottom and charm hadrons as well of the matched gen-jets. For each bottom- and charm-hadron the particle chain, which is stored for every event and contains the event history of a simulated event up to the hard scattering process, is scanned to identify the last mother quark of the selected hadron. By this means and in the case of using the Pythia8 event generator as parton shower, it may be decided with a high degree of certainty if the selected hadron and its predecessors originates from the decay of a top quark or if it has been radiated before. The information if such a selected hadron is an additional hadron or emerging from a top-quark decay is then used to classify the associated event into one of five distinct $t\bar{t}+X$ processes (cf. figure 6.1):

- $t\bar{t}+b\bar{b}$ events: Events featuring at least two bottom gen-jets of which each contains at least one bottom hadron not originating from a top-quark decay.
- $t\bar{t}+2b$ events: Events featuring a single bottom gen-jet in which two additional bottom hadrons are clustered.
- $t\bar{t}+b$ events: Events featuring a single bottom gen-jet which contains exactly one bottom hadron not emerging from a top-quark decay.
- $t\bar{t}+c\bar{c}$ events: Events featuring no bottom gen-jets, but at least one charm gen-jet containing an additional charm hadron.
- $t\bar{t}+lf$ events: All remaining events which either contain no bottom and charm gen-jets or only such gen-jets with bottom and charm hadrons coming out of a top-quark decay.

A clear separation between the $t\bar{t}+b\bar{b}$ and the $t\bar{t}+2b$ processes cannot be made. Particularly, a separation between the two processes depends, among others, on the jet algorithm and the parameters chosen for the jet algorithm. For similar reasons, $t\bar{t}+2b$ events can also be classified as $t\bar{t}+b$ events by mistake, but in this case the imprecise identification and definition of the underlying bottom hadrons is mostly the cause.

The softened requirement of only one charm gen-jet, even though the process class is called $t\bar{t}+c\bar{c}$, is motivated by the difficult identification of charm hadrons and charm-tagged jets at reconstruction level. It is assumed that in most of the cases a charm-quark pair occurs, but that the identification of only one charm hadron and its successors is quite likely.

This matching of selected hadrons to quarks and, especially, the decision if such a hadron emerges from a top-quark decay or not relies strongly on the event history provided by the GPMC event generator. Some GPMC event generators like Herwig7 [35, 36] and Sherpa [143] consider the matching between a top quark occurring in the final state and its subsequent decay products as unphysical. Hence, they only provide a truncated event history. Therefore, if these event generators are used to produce a $t\bar{t}$ sample, the aforementioned splitting procedure based on additional bottom and charm hadrons does not work reliably, if at all.

6.5 Choice of the $t\bar{t}H$ background samples

For the same reasons as in the case of the $t\bar{t}$ samples, the POWHEG Box Version 2 $t\bar{t}H$ event samples are chosen for the analyses conducted in this thesis. Besides the reasons already given before, the same choice of event generator seems reasonable. This choice is especially made because the $t\bar{t}+b\bar{b}$ and $t\bar{t}H(b\bar{b})$ processes can share the same final-state

Table 6.1: Summary of the Monte Carlo event samples considered in this analysis: For the respective simulated process the applied Monte Carlo matrix element event generator as well as the assigned cross section σ is stated, while, if applicable, the corresponding branching ratio $\mathcal{BR}$ was taken into account. In addition to the nominal MC event samples also the samples used to evaluate the $t\bar{t}$ systematic uncertainties are listed.

Simulated process	Applied MC event generator	σ [pb]
$t\bar{t}$ inclusive	POWHEG Box Version 2	831.76
$t\bar{t}$ semileptonic	POWHEG Box Version 2	365.9744
$t\bar{t}$ dileptonic	POWHEG Box Version 2	91.4936
$t\bar{t}H(b\bar{b})$	POWHEG Box Version 2	0.2953
$t\bar{t}H(\text{non}b\bar{b})$	POWHEG Box Version 2	0.2118
$t\bar{t}W$ hadronic	MG5aMC(NLO) +FxFx merg.	0.4062
$t\bar{t}W$ leptonic	MG5aMC(NLO) +FxFx merg.	0.2043
$t\bar{t}Z$ hadronic	MG5aMC(NLO) +FxFx merg.	0.5297
$t\bar{t}Z$ leptonic	MG5aMC(NLO) +FxFx merg.	0.2529
Single top s-channel	MG5aMC(NLO) +FxFx merg.	3.70
Single top t-channel (t)	POWHEG Box Version 2	136.02
Single top t-channel ($\bar{t}$)	POWHEG Box Version 2	80.95
Single top tW-channel	POWHEG Box Version 2	35.85
W+jets	MG5aMC(LO) +MLM merg.	3091.522
Z+jets	MG5aMC(LO) +MLM merg.	939.825
WW+jets	Pythia8	118.7
WZ+jets	Pythia8	47.12
ZZ+jets	Pythia8	31.73
$t\bar{t}$ inclusive, initial-state radiation (ISR) scale up	POWHEG Box Version 2	831.76
$t\bar{t}$ inclusive, initial-state radiation (ISR) scale down	POWHEG Box Version 2	831.76
$t\bar{t}$ inclusive, final-state radiation (FSR) scale up	POWHEG Box Version 2	831.76
$t\bar{t}$ inclusive, final-state radiation (FSR) scale down	POWHEG Box Version 2	831.76
$t\bar{t}$ inclusive, underlying event tune (UE) up	POWHEG Box Version 2	831.76
$t\bar{t}$ inclusive, underlying event tune (UE) down	POWHEG Box Version 2	831.76
$t\bar{t}$ inclusive, hdamp variation up	POWHEG Box Version 2	831.76
$t\bar{t}$ inclusive, hdamp variation down	POWHEG Box Version 2	831.76

7. Measurement data and event selection

In this section the measurement data used and the selection of the events applied in this analysis will be shortly summarized. Please refer to sections 3.2 and 3.3 for a general description of the CMS experiment and correspondingly of the object identification and event reconstruction.

7.1 Measurement data

For this analysis measurement data from runs of the Large Hadron Collider (LHC) taking place in the year 2016 is used which was recorded by the CMS experiment. More particularly, the data was taken from LHC runs with proton-proton collisions, at which the beam energy of each proton corresponded to 6.5 GeV. In this time period roughly a total integrated luminosity of $\int \mathcal{L} dt = 38 \text{ fb}^{-1}$ was delivered to the CMS experiment [199]. Of this amount of measurement data an integrated luminosity of $\int \mathcal{L} dt = 36.459 \text{ fb}^{-1}$ was recorded and certified by the data certification group of the CMS Collaboration (data sets internally referred to as “Golden JSON 2016”) [199]. After the application of luminosity corrections [200] using Van-der-Meer scans [201, 202] the total integrated luminosity available as measurement data corresponds to $\int \mathcal{L} dt \approx 35.918 \text{ fb}^{-1}$ [203].

7.2 Trigger

Since in this analysis only semileptonic events are of interest, two types of exclusive triggers based on leptons and corresponding trigger paths are applied.

By the semileptonic electron trigger path (internal name: HLT Ele27 WPTight Gsf ν^*) events which have exactly one leading electron are preselected. Furthermore, the electron of the event must have a transverse momentum of $p_T > 30 \text{ GeV}$, have an electron ID according to the tight working point of the electron identification algorithm, and must occur in a range of $|\eta| < 2.1$. In this case, the event is considered as a candidate for this analysis.

Similar to the electron trigger, events which have exactly one leading muon are preselected by two orthogonal semileptonic muon trigger path (internal names: HLT IsoMu24 ν^* , HLT IsoTkMu24 ν^*). Here, the muon of the event must have a transverse momentum of $p_T > 26 \text{ GeV}$ and an muon ID according to the tight working point. In addition, it

must occur in a range of $|\eta| < 2.1$. Furthermore, it must fulfill an isolation criterion of $\text{iso} < 0.15$. The cut on the transverse momentum p_T of the muon can be lower due to the excellent muon identification and tracking capabilities of the CMS detector.

However, such event candidates are still vetoed if one or more additional leptons occur. For the veto additional electrons are taken into account which have a transverse momentum of $p_T > 15 \text{ GeV}$, have an electron ID according to the tight working point, and occur in a range of $|\eta| < 2.4$. In the case of additional muons as veto leptons, the following requirements must be fulfilled: An additional muon must have a transverse momentum of $p_T > 15 \text{ GeV}$ and have a muon ID according to the tight working point. Additionally, it must occur in a range of $|\eta| < 2.4$ and must obey an isolation criterion of $\text{iso} < 0.25$.

7.3 Jets

In general events originating from the $t\bar{t}+X$ processes or the $t\bar{t}H(b\bar{b})$ process should contain at least six jets (cf. figures 6.1 and 4.3), while many of the background processes like the $V(V)+\text{jets}$ processes containing only vector bosons in the final-state usually feature events with fewer jets.

For a reduction of the number of background events, hence, events must contain at least six jets in this analysis. The possible jet candidates are derived from an anti- k_T -jet-clustering algorithm [87] with a jet-cone size of 0.4 and must have a jet ID in accordance with the loose working point. Furthermore, the transverse momentum of the leading jet candidate must be larger than 30 GeV . In contrast to that, for the subleading jets a transverse momentum of at least 20 GeV is sufficient. Moreover, all possible jets must occur in a range of $|\eta| < 2.4$. Nevertheless, jets are vetoed if their distance to leptons is smaller than $\Delta R(\text{jet}, \text{lepton}) < 0.4$. For the reduction of pile-up jet candidates are also vetoed if they would pass a pile-up jet ID corresponding to the loose working point.

7.4 B-tagging

For a further reduction of expendable background events and to increase the number of signal events in the overall analysis sample, a b-tagging requirement is applied in this analysis. Such a requirement reduces particularly the number of events from boson and diboson processes which generally feature fewer b-tagged jets. Therefore, possible event candidates must also possess at least two b-tagged jets out of the six or more selected jets in this analysis. These b-tagged jets must fulfill requirements consistent with the medium working point of the CMS CSVv2 b-tagging algorithm [89].

7.5 MET

Finally, event candidates have to pass a missing energy in the transverse plane $\cancel{E}_T$ requirement of $\cancel{E}_T > 20 \text{ GeV}$, too. Such a requirement would allow a possible measurement of the QCD background in a dedicated control region. However, as explained before, the number of QCD events in the phase space of this analysis is negligible. Therefore, such an additional measurement was not necessary in the end.

8. Analysis setup

In the following, an overview of the framework which is used to conduct the following analyses will be given. Furthermore, by a check of the Monte Carlo samples and the measurement data after the event selection it will be shown that the agreement between both is good.

8.1 Analysis framework

For this analysis a self-developed analysis framework is applied: It is based on the CMS offline software framework [197] and is also used to produce the results of the CMS $t\bar{t}H(b\bar{b})$ analyses [204, 205]. Sharing the analysis framework with the $t\bar{t}H(b\bar{b})$ analysis group is beneficial since by this means the framework is checked by more persons and the appendant efforts to implement the recommended settings for CMS physics analyses, to keep the framework up to date, and to add new features are shared.

The steering code of the analysis framework is mostly written in Python for simplicity. Also, the analysis itself is mostly defined by Python code. However, for performance reasons a code generator writing C++ code is invoked to generate the final code used for such an analysis. Furthermore, most of the analysis parts are fully automated and a continuous integration, delivery and deployment approach is applied. This approach mitigates the risks of introducing bugs in the code, provides confidence and trust in the analysis by being a completely transparent analysis. Moreover, it allows others the repetition of the the analysis and leads to the production of consistent results by an automated execution of the analysis code [206–208].

For the neural network based analyses (see section 9.5) the C++ version of the TensorFlow application programming interface [164] was compiled as a library and shipped as part of the final analysis code. In this way, previously trained neural networks can be used by the final analysis code. For the training of the neural networks the self-developed wrapper framework NNFlow [209] together with the TensorFlow open source library for machine learning was used. In the case of the linear discriminant based analyses (cf. section 9.4) such an inclusion of an additional program library was not necessary. The linear discriminants are trained by application of the TMVA multivariate data analysis tool kit [120] which provides standalone C++ code to evaluate the obtained linear discriminants in an analysis.

8.2 Check of Monte Carlo samples and measurement data agreement

For a physics analysis it is of importance that the data is well-modeled by the applied Monte Carlo (MC) samples and that the phase space of the measurement is properly described by the MC samples. Otherwise, it would be impossible to decide whether observable differences between the measurement data and the sum of all MC samples in a final discriminant output are due to physical reasons or just the outcome of an insufficient modeling of the considered signal and background processes by the MC event generators or if maybe a MC sample of some important physics process was unintentionally neglected. For this reason, a few basic and robust event variables are used to check the agreement between the measurement data and the sum of all MC samples after the application of the event selection (cf. section 7) using control distributions.

A short description of these event variables is given in table 8.1, while the corresponding control distributions are shown in figure 8.1 and 8.2. As can be seen from the distributions, the agreement between the measurement data and the sum of all MC samples is good. Therefore, the assumption can be made that this will hold also true for the final discriminator distributions.

Table 8.1: Description of the event variables used in the control plots

Variable name	Description
Eta of all electrons	η distribution considering all electrons
p_T of all electrons	p_T distribution considering all electrons
Eta of all muons	η distribution considering all muons
p_T of all muons	p_T distribution considering all muons
Number of jets	Distribution of the jet multiplicity
CSV	CSVv2 b-tagging values of all jets
Energy of all jets	Energy distribution considering all jets
Mass of all jets	Mass distribution considering all jets
Eta of all jets	η distribution considering all jets
p_T of all jets	p_T distribution considering all jets

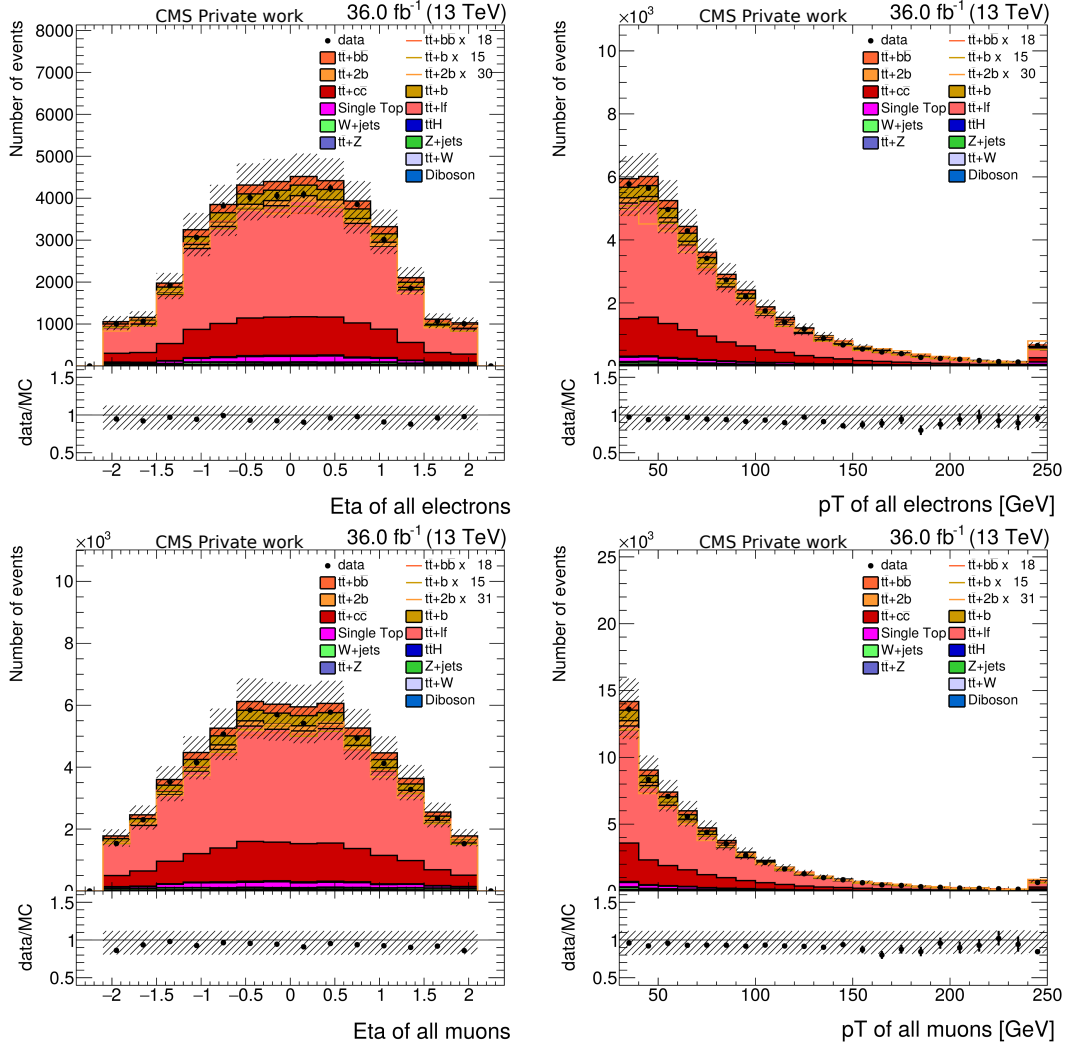


Figure 8.1: Control distributions to test the agreement between measurement data and the sum of all Monte Carlo samples: The measurement data (black dots) is compared to the stacked Monte Carlo histograms (filled blocks). In addition, the contributions of the three signal processes ($t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$) are individually scaled to the overall integral of the stacked histograms and are shown as solid lines. The uncertainty band includes all statistical and systematic uncertainties.

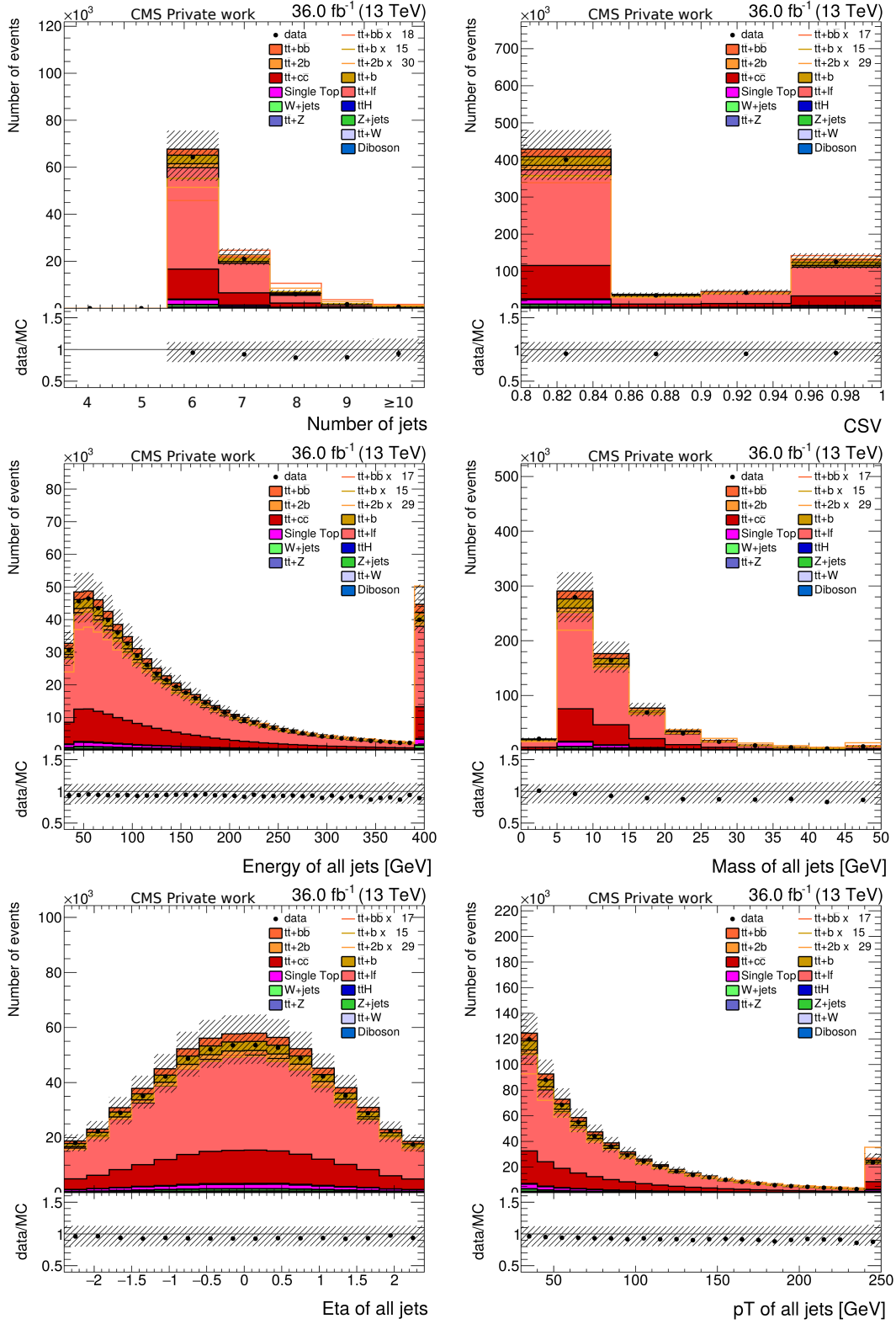


Figure 8.2: Control distributions to test the agreement between measurement data and the sum of all Monte Carlo samples: The measurement data (black dots) is compared to the stacked Monte Carlo histograms (filled blocks). In addition, the contributions of the three signal processes ($t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$) are individually scaled to the overall integral of the stacked histograms and are shown as solid lines. The uncertainty band includes all statistical and systematic uncertainties.

9. Analysis strategy

For the measurement of the $t\bar{t}+\geq 1\text{b-jet}$ ($t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$) signal strength r various types of analysis are conceivable. More basic and simple analyses come with the advantage that they are robust and easy to set up. These are transparent and verifiable to a higher degree as well. However, usually such analyses have the drawback that they cannot take full advantage of the information hidden in the data. Hence, they do not provide an optimum result regarding the achievable uncertainty of such a measurement. In contrast to that, multivariate analysis methods like neural networks tend to exploit more of the information available in data. But, they are also more difficult to comprehend and to check due to the highly sophisticated methods they make use of. Between these two extremes, there exist a variety of other analysis types. For example, a linear discriminant based analysis can be seen as a middle ground between both worlds and an entry step to a more complex multivariate analysis. As a multivariate analysis it allows to make use of different event variables and features, while its underlying mathematical model and analysis procedure can still be understood rather well.

The present analysis is divided into three reference sub analyses which cover the full range of possible analysis types:

The B-jet multiplicity based (B-jet mult.) analysis, discussed in the next section, is a simple and robust analysis. It is considered a baseline analysis for the two other analyses. This analysis makes use of the number of b-tagged jets distribution as a final discriminant to classify events in one out of four categories according to the number of b-tagged jets an event passing the event selection has.

A linear discriminant based (LD) analysis is applied in this thesis as a cross check of the more elaborate neural network based analysis employed later. Furthermore, it allows a check for correlations between the input variables (section 9.2) used in both multivariate analyses. By the linear discriminant events are classified into an output distribution which can be seen as the probability that a single event originates more likely from a $t\bar{t}+\geq 1\text{b-jet}$ ($t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$) signal process or from a background process. Its setup is presented in section 9.4.

In comparison with the binary classifying LD analysis, the reference neural network based (NN) analysis represents a multiclassification. Here, events are first assigned to one out

of six event classes based on the probability of emerging from a given process. Each event class and associated output distribution corresponds to either one of the three $t\bar{t} + \geq 1b$ -jet signal processes ($t\bar{t} + b\bar{b}$, $t\bar{t} + 2b$, $t\bar{t} + b$) or to three background processes ($t\bar{t} + c\bar{c}$, $t\bar{t} + lf$, $t\bar{t}H$) which are of special interest in this thesis. For each event class an output distribution is obtained. This distribution represents a kind of probability that a single, assigned event is really originating from the process associated with this event class.

This analysis is discussed in more detail in the section `refsec:analysis-1-nn-analysis`.

For both multivariate analyses (LD and NN) the event samples of various processes used in the training had to be reweighted, this reweighting is shortly discussed in section 9.3. Furthermore, a binning optimization algorithm was employed for both analyses, which is explained in section 9.6. This chapter ends with an overview of the setup of the Maximum-Likelihood (ML) fit (section 9.7), which was used to derive the final results for all three analyses.

9.1 B-jet multiplicity based analysis

The B-jet multiplicity based (B-jet mult.) analysis is a simple and robust analysis. It allows, as a baseline analysis, to estimate a conservative uncertainty on the signal strength result r for the $t\bar{t} + \geq 1b$ -jet ($t\bar{t} + b\bar{b}$, $t\bar{t} + 2b$, $t\bar{t} + b$) signal. The number of b-tagged jets distribution is chosen for this analysis. Subsequently, it will be shown that event variables related to b-tagging and b-tagged jets have the largest separation power between the six signal and background processes of interest ($t\bar{t} + b\bar{b}$, $t\bar{t} + 2b$, $t\bar{t} + b$, $t\bar{t} + c\bar{c}$, $t\bar{t} + lf$, $t\bar{t}H$).

In more detail, events which are passing the event selection criteria (cf. section 7) are sorted in a number of b-tagged jets final discriminant distribution (see figure 9.1 for the prefit distribution using the Asimov dataset and figure 11.1 for the distribution using the measurement data). The sorting is done in accordance with the number of b-tagged jets the event contains. Since the previously applied event selection requires that events have at least two b-tagged jets the final discriminant distribution binning starts with 2 b-tagged jets, too. For the final discriminant a binning with four bins was chosen, which means that the distribution spreads from two to five b-tagged jets, so that also the last bin contains a sufficient number of signal and background events. By using this final discriminant distribution together with a Maximum-Likelihood fit (discussed in section 9.7) the signal strength result for the $t\bar{t} + \geq 1b$ -jet processes is finally obtained which will be discussed in detail in chapter 11.

9.2 Input variables for the multivariate analyses

As multivariate analysis methods the linear discriminant and neural network based analyses explained below rely on well-described input variables which should also provide a good separation power between signal and background. The self-developed analysis framework which is shared with the $t\bar{t}H(b\bar{b})$ analysis groups provides in total 253 variables, from which finally 11 variables are chosen as input variables for the multivariate analyses. The method used to select these 11 input variables is discussed in the next section.

A description of these 11 input variables chosen for the subsequent multivariate analyses is given in table 9.1 and a discussion of these variables follows in section 9.2.2. Besides featuring a rather large separation power, also a good agreement between the stacked Monte Carlo histograms and the measurement data can be demonstrated for the 11 selected input variables, as can be seen from figures 9.2 and 9.3.

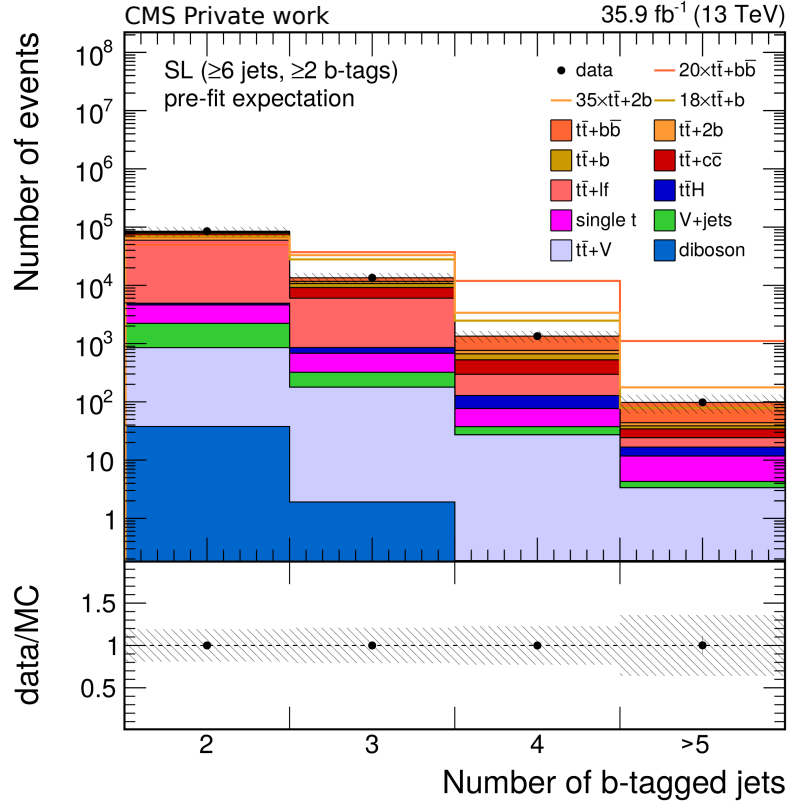


Figure 9.1: Prefit final discriminant distribution of the B-jet multiplicity based analysis using the Asimov dataset: As the final discriminant of the B-jet multiplicity based analysis the number of b-tagged jets distribution is chosen. The stacked Monte Carlo histograms (filled blocks) are compared to the Asimov dataset (black dots). In addition, the contributions of the three signal processes ($t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$) are individually scaled to the overall integral of the stacked histograms and are shown as solid lines. The hashed uncertainty band considers all systematic uncertainties, which are added in quadrature. The distribution is shown before the Maximum-Likelihood fit.

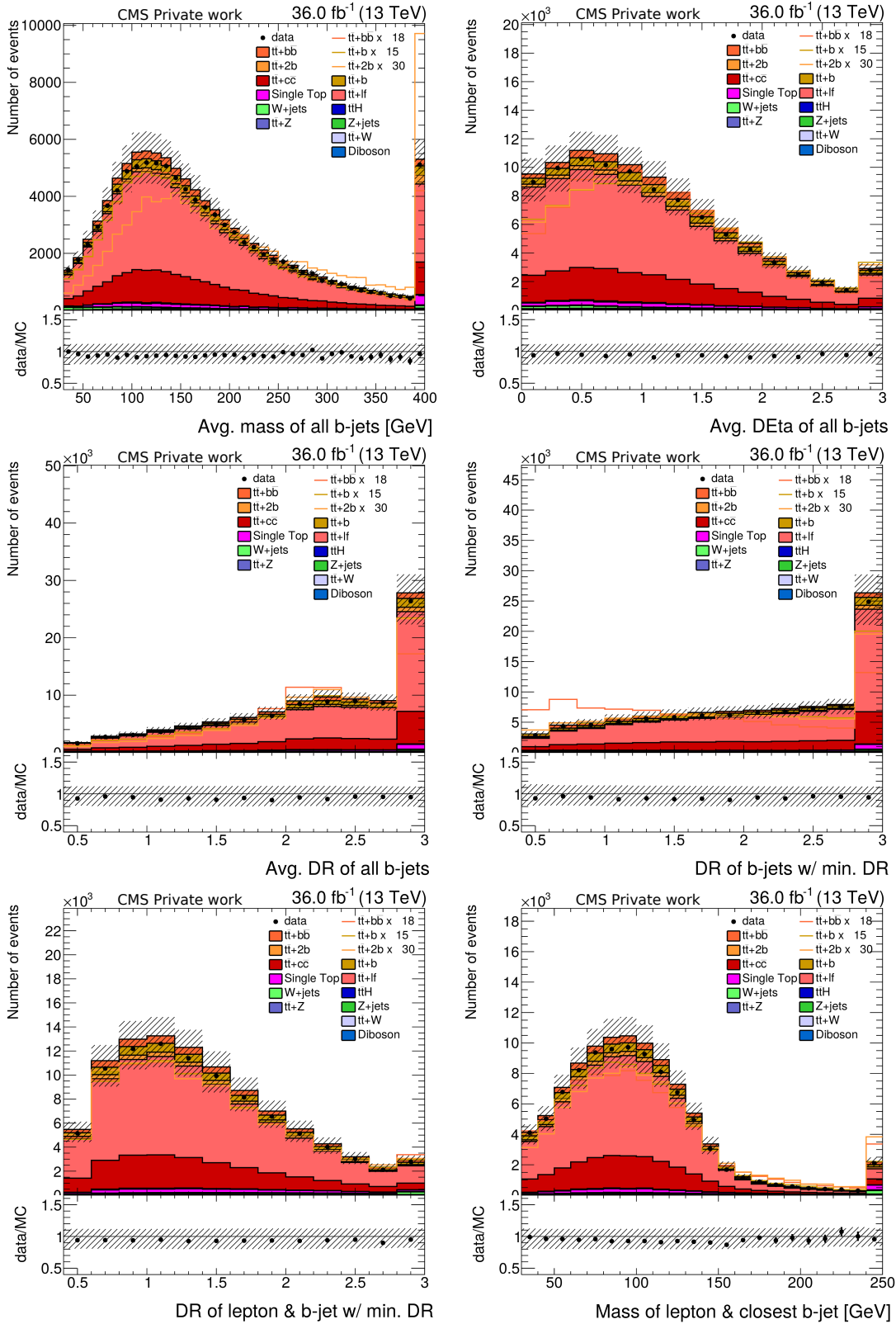


Figure 9.2: Input variables applied for the multivariate analyses: The stacked Monte Carlo histograms (filled blocks) are compared to the measurement data (black dots). In addition, the contributions of the three signal processes ($t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$) are individually scaled to the overall integral of the stacked histograms and are shown as solid lines. The uncertainty band includes all statistical and systematic uncertainties.

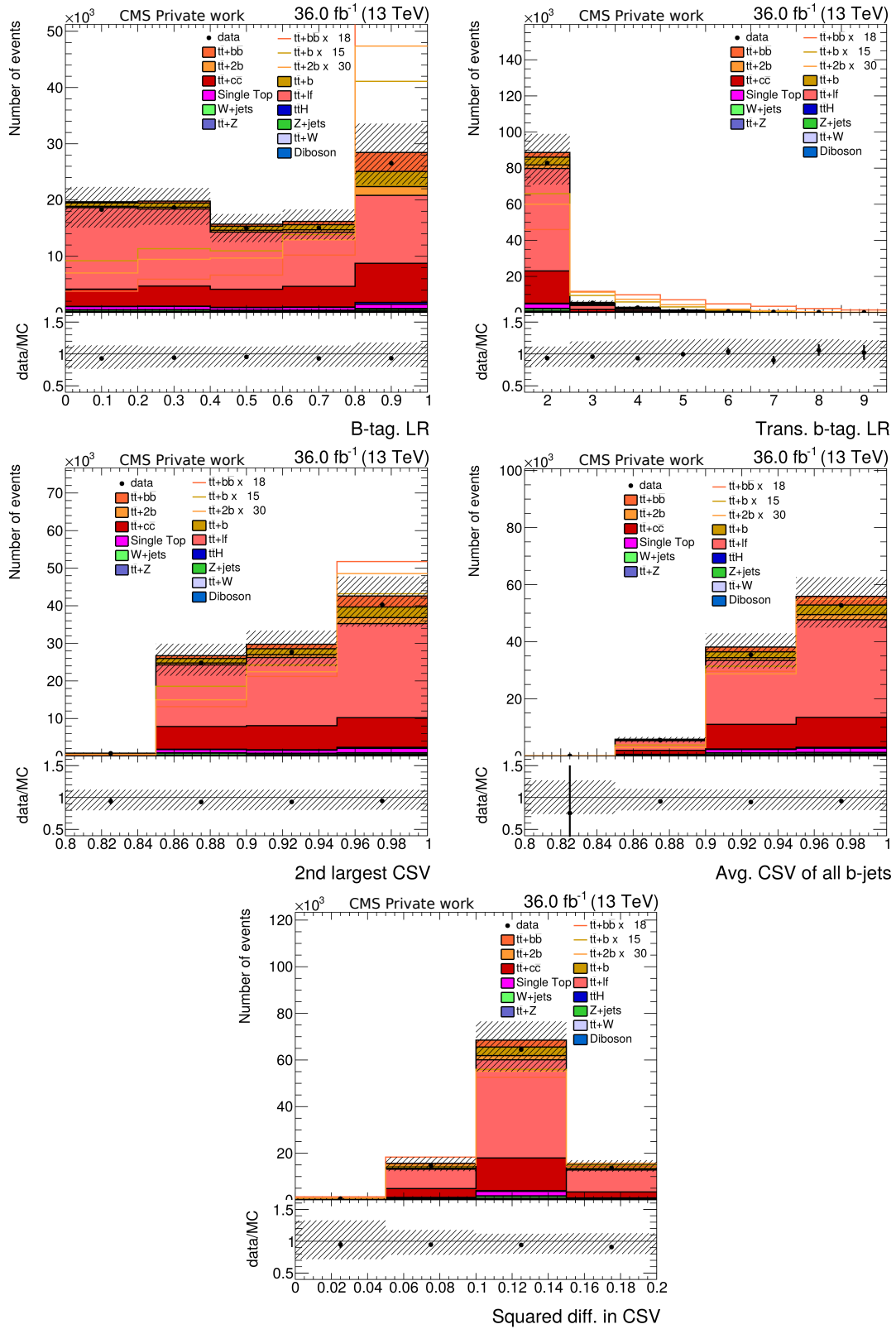


Figure 9.3: Input variables applied for the multivariate analyses: The stacked Monte Carlo histograms (filled blocks) are compared to the measurement data (black dots). In addition, the contributions of the three signal processes ($t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$) are individually scaled to the overall integral of the stacked histograms and are shown as solid lines. The uncertainty band includes all statistical and systematic uncertainties.

Table 9.1: Description of the input variables applied for the multivariate analyses: Beside the shortened variable name of the corresponding input variable a short description of the input variable is given.

Variable name	Description
Kinematic variables, considering b-tagged jets (and leptons)	
Avg. mass of all b-jets	Average mass of all b-tagged jets
Avg. DEta of all b-jets	Average $\Delta\eta$ value of all b-tagged jets
Avg. DR of all b-jets	Average ΔR value of all b-tagged jets
DR of b-jets w/ min. DR	ΔR value of the b-tagged jet pair with the smallest ΔR value
DR of lepton & b-jet w/ min. DR	ΔR value of the lepton and b-tagged jet with the smallest ΔR value
Mass of lepton & closest b-jet	Inverse mass of the lepton and closest b-tagged jet
B-tagging variables	
B-tag. LR	B-tagging likelihood ratio
Trans. b-tag. LR	Transformed b-tagging likelihood ratio
2nd largest CSV	Second-largest CSVv2 tagger discriminant value of all b-tagged jets
Avg. CSV of all b-jets	Average CSVv2 discriminant value of all b-tagged jets
Squared diff. in CSV	Sum of the squared differences between the CSVv2 discriminant value of a given b-tagged jet and the average CSVv2 discriminant value of all b-tagged jets

9.2.1 Choice of the input variables for the multivariate analyses

As explained previously, from 253 available event variables as inputs to the multivariate analyses 11 input variables are chosen. These 11 input variables are derived from preceding studies using dedicated neural network training: Various multiclassification neural networks were trained to differentiate between the following six processes $t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$, $t\bar{t}+c\bar{c}$, $t\bar{t}+lf$, and $t\bar{t}H$. Therefore, these networks contained six dedicated output nodes and associated event classes. As part of the training different network topologies (e.g. different numbers of hidden layers and nodes in hidden layers) and network parameters (e.g. different dropout rates) were tried out. Every neural network could exploit all of the 253 available event variables as input variables. After the training of each neural network the 253 input variables were ranked according to their contribution to the neural network and to the overall separation power of the neural network.

For this purpose, the connections between the input layer of the neural network and its output layer were evaluated. Each of the 253 input variables corresponds to a node of the input layer. Since only feed-forward neural networks were applied, each node in the input layer is only connected with nodes in the subsequent layer. The nodes in the subsequent layer are also only connected to nodes in the following layer. This kind of connection scheme goes on until the output layer is reached which contains six nodes, each representing one of the six considered event classes. All of these forward connections have an assigned weight which is derived by the training of the neural network. This weight can be considered a measure of the importance of a connection for the overall output of the neural network.

Under the assumption that an input variable having a rather large separation power should have an important influence on the overall neural network output, it should also have connections with either large positive or negative assigned weights due to positive or negative correlation with the neural network output. For this reason, the absolute sum of the weights of all connections going from the corresponding single input node to the six output nodes was calculated for each input variable. This absolute sum of associated weights was then used to rank the individual input variables.

Even though various neural network with different network topologies and parameters were tested, it was shown that in general the 11 input variables which were finally selected were among the input variables having the largest absolute sum of weights. Furthermore, all of them show good agreement between the Monte Carlo samples and measurement data which was an exclusion criteria for further promising input variables.

9.2.2 Discussion of input variables applied for the multivariate analyses

The 11 input variables applied for the multivariate analyses are listed in table 9.1 and are shown in figures 9.2 and 9.3.

The method-unspecific separation power between signal processes ($t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$) and background processes ($t\bar{t}+c\bar{c}$, $t\bar{t}+lf$, $t\bar{t}H$) of these variables was determined by using TMVA toolkit for Multivariate Analysis [120] and is stated for each input variable in table 9.2.

As can be seen from the description in the table 9.1 the 11 input variables can be roughly classified into two types of variables: Kinematic variables describing b-tagged jets, maybe in connection with leptons, and rather pure b-tagging variables.

However, it is noteworthy that all of these 11 input variables with rather large separation power are related to b-tagging and b-tagged jets. This is hardly surprising since the signal and the $t\bar{t}$ background processes ($t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$, $t\bar{t}+c\bar{c}$, $t\bar{t}+lf$) mostly differ in the number of b-tagged jets contained in an event. The kinematic variables like the ΔR value of the b-tagged jet pair with the smallest ΔR value (DR of b-jets w/ min. DR) or the mass of the lepton and closest b-tagged jet (Mass of lepton & closest b-jet) can be viewed as input variables exploited by the multivariate analyses to increase the separation power for processes which have the same number of b-tagged jets. For example, it can be safely assumed that especially these two kinematic variables can be used to better differentiate events of the $t\bar{t}+b\bar{b}$ and the $t\bar{t}H$ process as well as events of the $t\bar{t}+c\bar{c}$ and the $t\bar{t}+lf$ process. The first variable has a quite larger separation power than the second variable which indicates that the difference between the $t\bar{t}+b\bar{b}$ signal process and the $t\bar{t}H$ process could be more pronounced between two b-tagged jets than between a b-tagged jet and its closest lepton as one would expect.

Moreover, it was expected that the b-tagging likelihood ratio (B-tag. LR) and transformed b-tagging likelihood ratio (Trans. b-tag. LR) variables will be among the input variables having the largest separation power due to their construction: Both variables, which differ only by the non-linear transformation

$$\text{Trans. b-tag. LR} = \ln\left(\frac{\text{B-tag. LR}}{1 - \text{B-tag. LR}}\right),$$

are based on a likelihood method and will be now briefly described (Ref. [210] contains a more detailed explanation).

For the computation of the b-tagging likelihood ratio (B-tag. LR), first of all, the values of the CSVv2 b-tagging algorithm are derived for all jets in an event. Afterwards, these CSVv2 values are used to construct two hypotheses:

1. Hypothesis that 4 jets in the event are originating from b-quarks and the remaining jets from light quarks (up, down, strange quarks).
2. Hypothesis that only 2 jets in the event emerge from b-quarks and the remaining jets from light quarks (charm quarks not considered).

Finally, the B-tagging likelihood ratio (B-tag. LR) between the two hypotheses is evaluated by comparing with the expected CSVv2 probabilities derived from dedicated Monte Carlo simulations.

For this reason, the B-tagging likelihood ratio (B-tag. LR) and its transformed version (Trans. b-tag. LR) can especially be exploited by the multivariate analyses to distinguish between processes featuring different number of b-tagged jets like $t\bar{t}+b\bar{b}$ or $t\bar{t}H(b\bar{b})$ vs. $t\bar{t}+2b$ or $t\bar{t}+b$.

It will be shown later in section 11.3.1 that including both b-tagging likelihood ratios (B-tag. LR, Trans. b-tag. LR) helps to simplify the topology of the neural network used in the neural network based (NN) analysis, even though these two input variables are highly correlated, as can be seen from the linear discriminant based (LD) analysis.

Table 9.2: Separation power of input variables applied for the multivariate analyses: The separation power between signal processes ($t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$) and background processes ($t\bar{t}+c\bar{c}$, $t\bar{t}+lf$, $t\bar{t}H$) determined by the TMVA toolkit for Multivariate Analysis [120] is stated for each input variable applied by the linear discriminant and neural network based analyses. The variables are ordered from the variables having the largest separation power to the variables having the lowest.

Variable name	Separation power
Trans. b-tag. LR	1.553×10^{-1}
B-tag. LR	1.539×10^{-1}
2nd largest CSV	2.520×10^{-2}
DR of b-jets w/ min. DR	2.513×10^{-2}
Mass of lepton	1.209×10^{-2}
Squared diff. in CSV	8.933×10^{-3}
Avg. DR of all b-jets	8.309×10^{-3}
Avg. DEta of all b-jets	7.371×10^{-3}
Avg. CSV of all b-jets	7.060×10^{-3}
Avg. mass of all b-jets	3.617×10^{-3}
DR of lepton & b-jet w/ min. DR	2.453×10^{-3}

9.3 Reweighting of process classes for the training of the multivariate analyses

For the training of classifiers employed in this thesis all process classes ($t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$, $t\bar{t}+c\bar{c}$, $t\bar{t}+lf$, $t\bar{t}H$) are treated equally. This is achieved by an adjustment of the weights of single events, so that the sum of the weights of all events of a given process class corresponds to the sum of the weights of the other process classes.

Therefore, events emerging from the $t\bar{t}H$ process have the same influence in the training and on the final shape of a multivariate classifier as events from other processes like $t\bar{t}+b\bar{b}$. However, as an example, the cross section of the $t\bar{t}H$ process is quite smaller than the $t\bar{t}+b\bar{b}$ cross section (0.5 pb / 4.2 pb), while their events can have similar features. For this reason a classifier may not learn to properly distinguish $t\bar{t}+b\bar{b}$ events from other events since it could lead to a reduction of the identification of $t\bar{t}H$ events at the same time. In comparison to that, in the later measurement more $t\bar{t}+b\bar{b}$ events than $t\bar{t}H$ events are expected and, hence, the proper assignment of $t\bar{t}+b\bar{b}$ events is generally more important for the analysis.

In principle, this could be avoided by training a multivariate classifier with processes weighted according to their importance or cross section. However, in the case of the measurements presented in this thesis, it is challenging to do this weighting of processes and events in a reasonable way. The major reason is that the range of the cross sections of the signal and background processes of interest is rather wide. For example, the $t\bar{t}+1f$ cross section is approximately 730 pb, while the $t\bar{t}+2b$ cross section is roughly 4.6 pb. Even this cross section is rather large compared to the $t\bar{t}H$ cross section stated above at a center-of-mass energy of 13 TeV. By using cross sections as a basis for the reweighting of processes in the training $t\bar{t}+1f$ events and their separation from the other events would have the major influence on the training. In this way, the classifier would provide a rather good separation of the $t\bar{t}+1f$ processes from the other processes, but most likely processes having a smaller cross section and associated lower influence on the training would not be properly classified.

For the measurements presented in this thesis the proper assignment of $t\bar{t}+\geq 1b$ -jet events and later also $t\bar{t}H$ events (part V) as well as good separation of these events from background events is fundamental. As previously outlined all of these processes of interest have rather small cross sections and would have only a small influence on the classifier training if the process classes would be reweighted by their cross section. That is the reason why multivariate classifiers employed in this thesis were always trained with processes having equal overall weight instead.

9.4 Linear discriminant based analysis

The linear discriminant based (LD) analysis can be viewed as one of the simplest multivariate analysis methods. Therefore, it is a good compromise between the B-jet multiplicity based (B-jet mult.) baseline analysis and the highly-sophisticated neural network based (NN) analysis. The LD analysis is also used to exploit the linear correlations of the input variables in a straight forward way (see section 11.3.2), which would be a more cumbersome task with the NN analysis. In both analyses suitable transformations were applied, so that these input variables were virtually Gaussian distributed with a mean of zero and a standard deviation of one.

The LD analysis makes use of the “LD”-type of linear discriminant provided by the TMVA multivariate data analysis tool kit [120] (section 5.2.2 for a more general discussion). Nonetheless, also Fisher’s classical linear discriminant [118] was tried, but it was found that it leads to the same results. This was expected since it can be shown that both linear discriminant types are equivalent [120].

The linear discriminant is trained to distinguish between the $t\bar{t}+\geq 1b$ -jet signal processes ($t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$) and the background processes ($t\bar{t}+c\bar{c}$, $t\bar{t}+1f$, $t\bar{t}H$) using the aforementioned input variables for a binary classification. Later on, for each event the linear discriminant is evaluated and its output distribution is used as final discriminant in this analysis. The binning of the linear discriminant output distribution, shown in figure 9.4 for the Asimov dataset and in figure 11.2 for the measurement data, is optimized by a dedicated binning algorithm which is explained in section 9.6.

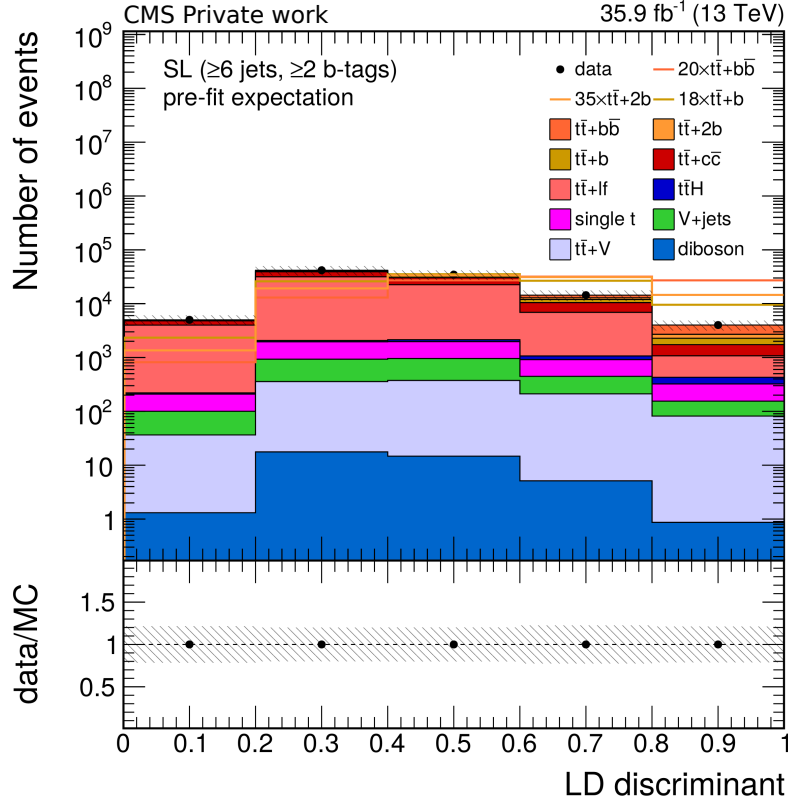


Figure 9.4: Prefit final discriminant distribution of the linear discriminant based analysis using the Asimov dataset: The output of the linear discriminant is chosen as the final discriminant. The stacked Monte Carlo histograms (filled blocks) are compared to the Asimov dataset (black dots). In addition, the contributions of the three signal processes ($t\bar{t} + b\bar{b}$, $t\bar{t} + 2b$, $t\bar{t} + b$) are individually scaled to the overall integral of the stacked histograms and are shown as solid lines. The hashed uncertainty band considers all systematic uncertainties which are added in quadrature. The distribution is shown before the Maximum-Likelihood fit.

9.5 Neural network based analysis

The neural network based (NN) reference analysis performs a multiclassification. Nonetheless, as part of a cross check with the LD reference analysis also a binary classifying neural network is tested. The multiclassifying neural network of the reference analysis differentiates between each of the three $t\bar{t}+\geq 1b$ -jet signal processes ($t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$) as well as the background processes ($t\bar{t}+c\bar{c}$, $t\bar{t}+lf$, $t\bar{t}H$) individually by defining six corresponding event classes and associated output distributions (see figure 9.5). These output distributions are then used as final discriminants in a Maximum-Likelihood (ML) fit to obtain the final $t\bar{t}+\geq 1b$ -jet signal strength result r .

Even though in this measurement only the signal strength r result of the combined $t\bar{t}+\geq 1b$ -jet signal process is derived, using an event class for each considered process is beneficial since it enables the ML fit to constrain the individual processes and related uncertainties better. This advantage of multiclassification will be confirmed in the measurements of part V. In these measurements signal strengths r of two or three processes are determined simultaneously.

For the implementation of the neural networks used in this thesis the TensorFlow open source library for machine learning [164] is applied, while for the training of the neural networks the self-developed wrapper framework NNFlow [209] is exploited in addition. A neural network containing 11 input nodes – corresponding to the 11 input variables discussed before –, two hidden layers with 100 nodes per hidden layer, and six output nodes associated with the six event classes ($t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$, $t\bar{t}+c\bar{c}$, $t\bar{t}+lf$, $t\bar{t}H$) is chosen for the NN reference analysis. But also other binary or multiclassifying neural networks featuring different network topologies are examined as part of this analysis, these networks will be explained in chapter 11.

However, in all cases the exponential linear unit (ELU) function [125] was employed as an activation function (section 5.2.3).

The application of neural networks as multivariate analysis tools requires to train the neural networks using Monte Carlo samples, before they can be used for an evaluation of events. In the case of the multiclassifying neural network six individual Monte Carlo samples corresponding to the $t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$, $t\bar{t}+c\bar{c}$, $t\bar{t}+lf$, and $t\bar{t}H$ processes are considered for the training, while in the binary case only two combined samples are applied ($t\bar{t}+\geq 1b$ -jet; $t\bar{t}+c\bar{c}$, $t\bar{t}+lf$, $t\bar{t}H$).

The samples consisting of events with known classification are split into three subsets: A training sample containing 40% of the total number of events, a validation sample containing 10% (both corresponding to events with an odd event number), and an evaluation sample containing 50 % of the events (corresponding to events with an even event number).

The 11 input variables previously discussed are used as inputs to the neural network. However, the scale of the input variables and their original distribution can differ substantially and this could lead to input variables having different influence in the training. Thus, a scaling function is applied for each input variable. This scaling function makes sure that the overall distribution of the input variable is Gaussian with a mean value of zero and a standard deviation of one. The derived scaling functions are not only applied during the training of the neural network, but also during the evaluation of events to obtain the final output distributions.

For the training of the various neural networks (section 5.2.3) batch training was employed. Each batch contained 500 randomly selected events. The learning rate η was set to $\eta = 3 \times 10^{-5}$. For the minimization of the loss function, for which the cross entropy was chosen, the adam optimizer [129] was utilized with the following parameters: $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 1e - 08$. As countermeasures against overtraining of the neural networks early-stopping, L2-regularization, and dropout were employed: An early-stopping interval of 15 epochs was chosen. The regularization parameter λ for the L2-regularization was set to $\lambda = 1 \times 10^{-12}$. For the dropout 30 % of the nodes were randomly selected and their output set to zero in each batch training. The general properties of the neural networks applied throughout this thesis are also stated in table 9.3.

Besides the scaling functions and the common activation functions of a neural network (section 5.2.3) a softmax function is applied as part of the output layer. In this way, the values of the six output nodes are transformed, so that the overall sum of the six output nodes is equal to one. By this means, the individual obtained value of an output node can be seen as a likelihood that an event was originating from the associated event class and corresponding process in the beginning. For each event, independent of being part of the training or evaluation step, the neural network is evaluated and the transformed values of the six output nodes are obtained. Afterwards, the event class an event most likely belongs to is chosen by picking the output node and corresponding event class with the largest value. Finally, the value of this output node is then used to fill the event with its event weight into the associated distribution of the selected event class. In this manner, events are always attributed to only one event class, which increases the purity of the event classes. This attribution of an event to an event class is also depicted in figure 9.6. Here, for a given event the output node associated with the $t\bar{t}+c\bar{c}$ event class has the largest value. Therefore, this event is classified as a $t\bar{t}+c\bar{c}$ event and its event weight is filled into the $t\bar{t}+c\bar{c}$ event class distribution.

In theory, the bin range for each of these six event class distributions could span from zero to one. However, in almost all cases the neural networks would not assign an event to only one event class by producing a single output node with value of one. Instead, in most cases one or two output nodes have rather large values for a given event and the remaining four output nodes still have non-zero values. Hence, the distribution of output node values of the event classes does not generally range from zero to one, but more likely from zero to roughly 0.7. The maximum of the distribution tends to be on the left side of the histogram (compare the prefit final discriminant distributions for the Asimov dataset in figure 9.5). As an improvement of the event class output distributions, which are used to derive the $t\bar{t}+\geq 1b\text{-jet}$ signal strength r result by a final Maximum-Likelihood fit, an binning optimization algorithm is applied. It will be introduced in the next section. Since it purges empty outer bins in an distribution, the prefit final discriminant distributions of the NN analysis will not span over the full range from zero to one.

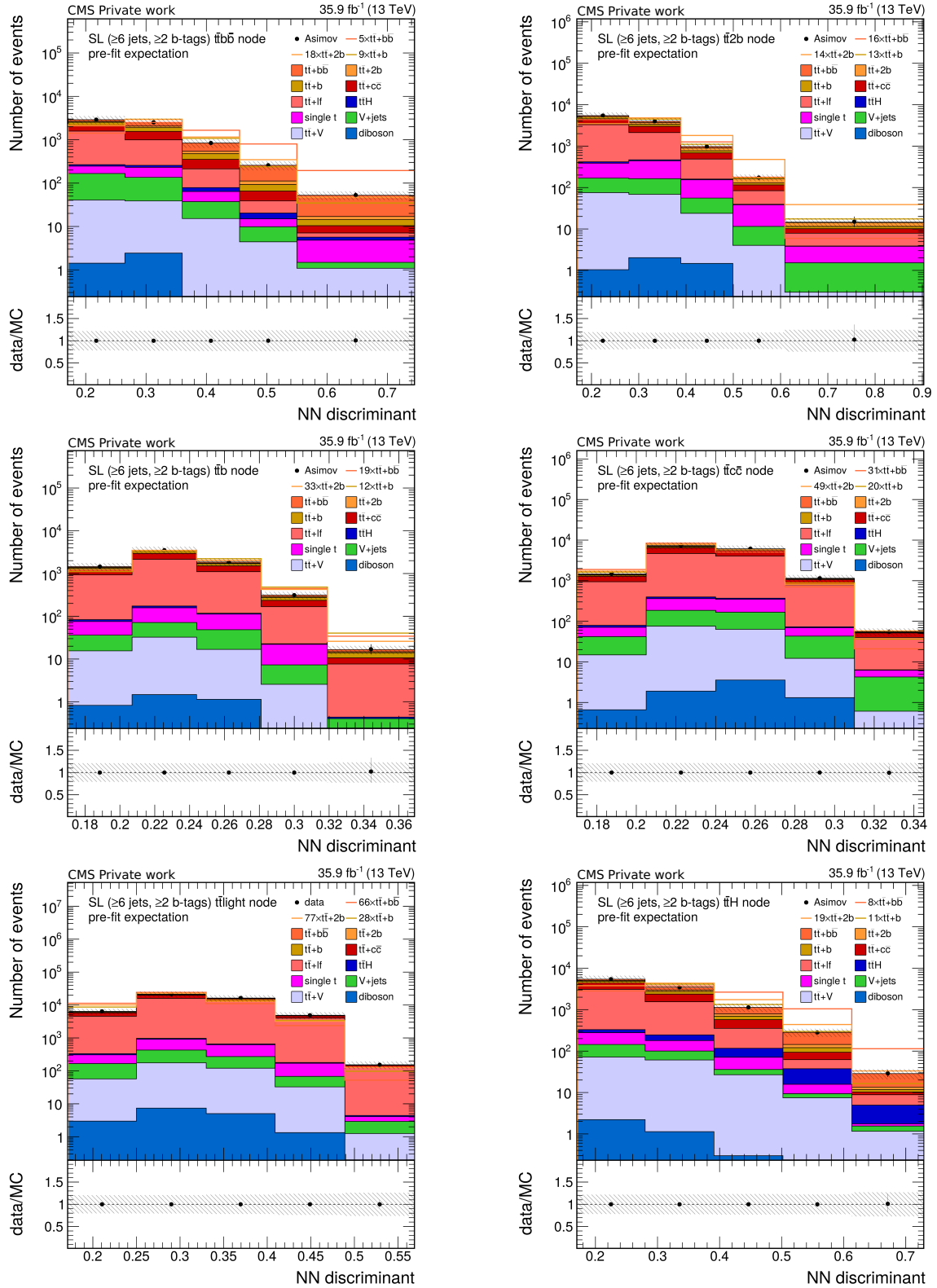


Figure 9.5: Prefit final discriminant distributions of the neural network based analysis using the Asimov dataset: The output distributions of the six output nodes of the neural network, which correspond to the six event classes ($t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$, $t\bar{t}+c\bar{c}$, $t\bar{t}+lf$, $t\bar{t}H$) are chosen as the final discriminant. The stacked Monte Carlo histograms (filled blocks) are compared to the Asimov dataset (black dots). In addition, the contributions of the three signal processes ($t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$) are individually scaled to the overall integral of the stacked histograms and are shown as solid lines. The hashed uncertainty band considers all systematic uncertainties which are added in quadrature. The distribution is shown before the Maximum-Likelihood fit.

Table 9.3: General properties of the neural networks employed in this thesis: The ELU function was used as an activation function. For the training of a neural network event samples with known classification were split into a training, validation, and test dataset. From these dataset batches of 500 randomly selected events were constructed. As part of the training a learning rate η and an optimizer to minimize the loss function was chosen. Furthermore, early-stopping, L2-regularization, and dropout were used to prevent overtraining of a neural network.

Property	Value
Activation function	Exponential linear unit (ELU)
Fraction of training events	40 % (80 % of events with odd number)
Fraction of validation events	10 % (20 % of events with odd number)
Fraction of evaluation events	50 % (100 % of events with even number)
Batch size	500 events
Learning rate η	3e-05
Loss function	cross entropy
Optimizer	Adam
Parameter β_1 of Adam optimizer	0.9
Parameter β_2 of Adam optimizer	0.999
Parameter ϵ of Adam optimizer	1e-08
Early-stopping interval	15 epochs
Early stopping of training after	199 epochs
L2-regularization parameter λ	1e-12
Dropout rate d	0.3

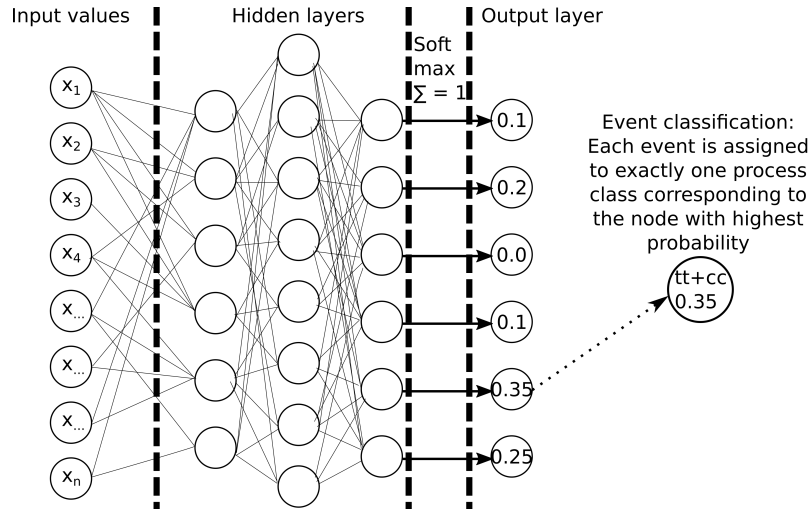


Figure 9.6: Overview of the neural network used for the neural network based analysis: The neural network expects for each input variable an input value (x_1 to x_n), which is normalized by using a scaling function (not shown). The values of the input layer are processed further by the hidden layers of the neural network, before the last hidden layer consisting of six nodes (one node for each of the considered sub-processes / event classes: $t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$, $t\bar{t}H$, $t\bar{t}+c\bar{c}$, and $t\bar{t}+lf$) is reached. The values of these six nodes get transformed by a so-called soft-max function, so that the sum over all six nodes in the output layer is one. Based on the node in the output layer with the largest value the neural network decides on the event class.

9.6 Application of a binning optimization algorithm

For the LD and NN analyses a dedicated binning optimization algorithm is applied since the number of background events in the outer bins can be rather small or even negligible. This can lead to large statistical fluctuations for the background processes and also to bins having negative numbers of expected background events if the corresponding background process was simulated in next-to-leading order of perturbation theory and, thus, its events can have negative event weight. For this reason, a fixed bin size would be hardly suitable and a better solution is to require a certain number of bins and a minimum number of background event in each bin. This requirement is met by the binning optimization algorithm: By using this algorithm a final discriminant has at least 5 bins and at maximum 20 bins. The bins usually have equal bin widths, but the width of the outer bins is increased if necessary, so that every bin contains at least 10 background events (events from all background processes counted together).

To obtain the optimized bins, final discriminant output distributions containing 1000 bins with a total range from zero to one are produced initially. Subsequently, the binning optimization algorithm merges bins starting from the left and right side of the total bin range, so that the number of weighted background events in each remaining bin is at least 10. Finally, the remaining bins are combined to bins with an equal bin width until the criteria of having at least 5 bins and at maximum 20 bins is met.

The binning optimization algorithm purges empty bins, if they correspond to the lowest or highest remaining bin of an output distribution. That is the reason why output distributions as in the case of the NN analysis can have a smaller bin range than zero to one.

9.7 Maximum-Likelihood fit

For the three reference analyses (B-jet mult., LD, NN) and all other analyses a Maximum-Likelihood (ML) fit (section 5.1) is applied to derive the signal strength r of the $t\bar{t} + \geq 1\text{b-jet}$ signal processes ($t\bar{t} + b\bar{b}$, $t\bar{t} + 2b$, $t\bar{t} + b$) and its associated uncertainties as a final analysis result. All systematic uncertainties (cf. next section) are taken into account as nuisance parameters θ during the minimization of the corresponding negative log-likelihood function. A global minimum of the likelihood function is found using the Minuit algorithm [211] by scanning the signal strength r as the parameter of interest, while the nuisance parameters θ are varied one after the other to find a local minimum of the likelihood function. This Maximum-Likelihood fit and further fit tests are conducted by making use of the CMS Higgs Combine tool [212] together with the CombineHarvester CMS software framework (CMSSW) package [213]. These tools are chosen and are widely used in the CMS collaboration since the proper implementation of a multi-dimensional Maximum-Likelihood fit, the corresponding handling of the convergence of the fit, and so on is a formidable challenge in itself. By employing the two tools, which rely on the ROOT data analysis framework [214] as well as the CMS software framework [197] themselves, trust can be placed in the results of the Maximum-Likelihood fits. Nevertheless, various tests are conducted to cross check the results produced with the three reference analysis and to increase the confidence in the results presented in chapter 11.

10. Systematic uncertainties

In this analysis different kinds of systematic uncertainties are considered during the Maximum-Likelihood fit discussed previously to obtain the final signal strength r result for the $t\bar{t}+\geq 1b$ -jet signal processes. All of the following uncertainties, if not stated otherwise, apply equally to the $t\bar{t}+\geq 1b$ -jet signal processes ($t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$) and all the background processes (including the background processes neglected by the classification via the multivariate analysis methods). Generally, these systematic uncertainties are treated as fully correlated between the signal and background processes.

The systematic uncertainties will be divided into theory uncertainties related to the applied Monte Carlo samples and into experimental uncertainties related to the measurement process. Furthermore, systematic uncertainties can also be classified according to their influence on the rate and / or shape of the obtained final discriminant output distributions into:

Rate uncertainties, which refer to systematic uncertainties affecting only the rates of physics processes in a final discriminant output distribution and are constrained by a log-normal distribution.

Shape uncertainties, which have an influence on the shape of a final discriminant output distribution and maybe (indirectly) also on the rate.

Of course, systematic uncertainties can also be an uncertainty affecting both the rate and the shape of an output distribution.

In the summary table 10.3, besides the classification of uncertainties according to their origin, these two types of systematic uncertainties will also be used to provide a complete overview of all systematic uncertainties applied.

10.1 Systematic uncertainties related to the applied Monte Carlo samples

In this section a short overview of the systematic uncertainties related to the usage of Monte Carlo samples will be given.

10.1.1 Cross section normalization rate uncertainties

In the most cases the Monte Carlo samples are generated using a combination of a dedicated matrix element (ME) event generator and a subsequent general-purpose Monte Carlo (GPMC) event generator in next-to-leading order of perturbation theory. Please refer to chapter 6 for more details about the chosen samples. However, the expectation for the signal and background cross sections are derived from theoretical calculations since in all cases these provide a cross section estimation with a higher accuracy and lower uncertainties (cf. section 6.2) due to inclusion of higher-order corrections. The rate uncertainties which are related to these cross section normalizations are given in table 10.1. Uncertainties that are common to multiple processes are considered as fully correlated, where applicable.

The cross section normalization rate uncertainties are split into systematic uncertainties related to the variation of the parton distribution function (PDF) and to the variation of the QCD scale. More details will be given below.

10.1.2 Extra rate uncertainty due to the cross section normalization of the $t\bar{t}+c\bar{c}$ process

The $t\bar{t}+c\bar{c}$ process due to its similarities to the $t\bar{t}+\geq 1b$ -jet signal processes represents an important source of irreducible background. Since up to now the cross section normalization of this process cannot be constrained to better than roughly fifty percent using either control region studies or higher-order theory calculations, an additional and uncorrelated fifty percent rate uncertainty is assigned to the cross section normalization of this process.

Table 10.1: Systematic uncertainties due to the normalization of cross sections: Each column in the table is an independent source of a rate uncertainty. If uncertainties are listed for more than one process (same column, but different rows), the uncertainties for these processes are fully correlated. The rate uncertainties are split into systematic uncertainties related to the variation of the parton distribution function (PDF) and of the QCD scale. In the case of the $t\bar{t}+c\bar{c}$ process an extra 50 % rate uncertainty is assigned which is not listed here.

Process	pdf				QCD Scale				
	gg $t\bar{t}H$	gg	q $\bar{q}$	qg	$t\bar{t}$	t	V+jets	VV+jets	$t\bar{t}H$
$t\bar{t}H$	3.6%								-9.2%/+5.8%
$t\bar{t}+\text{jets}$		4%			-4%/+2%				
$t\bar{t}W$			2%		-12%/+13%				
$t\bar{t}Z$		3%			-12%/+10%				
Single Top				3%		-2%/+3%			
V+jets			4%				1%		
DiBoson			2%					2%	

10.1.3 PDF variation uncertainties

For all Monte Carlo samples used in this thesis the NNPDF3.0 set [186] is used as the default parton distribution function (PDF) set. Variations of the final discriminant output distributions due to the intrinsic uncertainty of this PDF set are considered by making use

of the 100 sub pdfs contained in this PDF set to reweight the output distributions and to derive the associated shape uncertainties. In the case of the $t\bar{t}$ samples the obtained shape uncertainty is applied, while for all the other samples a rate uncertainty is used. These PDF uncertainties are listed in table 10.1 and PDF uncertainties are correlated for the same initial-state.

10.1.4 Extra shape uncertainties due to the chosen renormalization and factorization scale for the $t\bar{t}+X$ processes

For all $t\bar{t}+X$ processes additional, uncorrelated shape uncertainties are considered due to the choice of the renormalization scale μ_R and the factorization scale μ_F in the ME generator to produce these event samples. The shape uncertainties are derived from the changes to the final discriminant output distributions, after these scales are individually varied by a factor of either 0.5 or 2. Since the aforementioned cross section rate uncertainties should already completely cover any rate changes due to the parameters chosen in the matrix element generator, the scale variation uncertainties are getting scaled to retain the original overall cross section normalization and are only considered as shape uncertainties.

10.1.5 Extra rate uncertainties due to the matrix element generator and parton shower matching for the $t\bar{t}+X$ processes

The $t\bar{t}$ samples are produced using the POWHEG ME event generator matched to the Pythia8 GPMC event generator. For this event generation a hdamp value of $\text{hdamp} = 1.58m_t$ is applied which represents the choice of a resummation scale for the damping function limiting the resummation of higher-order effects by the Sudakov form factor (Ref. [215] provides detailed information about this parameter). Since the resummation scale is a physically unmotivated parameter and only an artifact of the approximation of higher-order calculations in leading or next-to-leading order event generation, it should be considered as an additional theory uncertainty. Therefore, by producing extra $t\bar{t}$ samples with hdamp values of $\text{hdamp} = m_t$ and $\text{hdamp} = 2.24m_t$ the hdamp resummation scale value is varied within reasonable bounds. In this way, corresponding uncorrelated shape variations for each $t\bar{t}+X$ process are obtained. However, due to the limited number of events in the extra $t\bar{t}$ samples large statistical fluctuations occur, resulting in unrealistically large variations of the final discriminant templates. It is for this reason that this shape variations are converted to an individual and uncorrelated rate uncertainty for each $t\bar{t}+X$ process by comparing the total event yield of the nominal and the varied samples.

10.1.6 Extra rate uncertainties due to the value of the strong coupling constant α_S chosen in the parton shower for the $t\bar{t}+X$ processes

As mentioned before, the $t\bar{t}$ samples are produced by employing the POWHEG ME event generator matched to the Pythia8 GPMC event generator. For these samples the impact of the value for the strong coupling constant α_S in the Pythia8 parton shower is evaluated independently for the initial-state radiation (ISR) and the final-state radiation (FSR) parton shower by including dedicated samples. These samples are produced with a variation of the α_S value in either the ISR or FSR shower by a factor of two up or down and lead to a shape variation in the final discriminant output distributions. However, as in the case of the hdamp $t\bar{t}$ samples large statistical fluctuations occur due to the limited number of events. Hence, for the same reasons as before the shape variations are converted to an individual and uncorrelated rate uncertainty for each $t\bar{t}+X$ process.

10.1.7 Extra rate uncertainties due to the underlying event tune of the parton shower for the $t\bar{t}+X$ processes

In the same manner as for the two previously discussed extra uncertainties, also rate uncertainties considering the choice of the underlying tune of the Pythia8 parton shower are obtained and applied. In table 10.2 the settings chosen for the underlying event tune to produce the nominal and varied $t\bar{t}$ samples are listed. Again, the limited size of the varied samples made it reasonable to convert the shape variations into uncorrelated rate uncertainties for the single $t\bar{t}+X$ processes.

Table 10.2: Settings to obtain the rate uncertainties due to the underlying event tune of the parton shower for the $t\bar{t}+X$ processes: The underlying event tune settings for the nominal POWHEG $t\bar{t}$ samples and the up- or down-varied samples are given in the table.

Parameter	Nominal	Up-varied	Down-varied
MultipartonInteractions:pT0Ref	2.1971	2.1279	2.2686
MultipartonInteractions:expPow	1.6085	1.7106	1.5619
ColourReconnection:range	6.5932	6.5000	8.7140

10.1.8 Shape uncertainties due to the limited size of the Monte Carlo samples

Due to the limited size of the $t\bar{t}$ Monte Carlo samples and all the background samples statistical fluctuations, affecting the event yield predictions for the nominal samples, occur. By introducing a shape uncertainty, which is constructed by varying each bin of a final discriminant output distribution individually by its statistical uncertainty, this is considered. Since all bins are statistically independent, each single shape variation can be viewed as an uncorrelated shape uncertainty.

10.2 Systematic uncertainties related to the measurement process

In this section a short overview of the systematic uncertainties related to the measurement process will be given.

10.2.1 Luminosity rate uncertainty

The integrated luminosity $\int \mathcal{L} dt$ delivered and recorded by the CMS detector is measured using five detectors: The silicon pixel detector, the drift tubes in the barrel (DT), the forward hadronic calorimeter (HF), the fast beam conditions monitor (BCM1F) and the pixel luminosity telescope (PLT). The recommended rate uncertainty on the luminosity for the 2016 measurement period is 2.5 % [187]. It is an important uncertainty since it affects all other rates.

10.2.2 Uncertainties due to pileup interactions

The occurrence of additional pileup interactions and the corresponding uncertainty on the number of pileup interactions is taken into account by including a rate and shape

uncertainty which is applied to all processes and considered fully correlated among them. These uncertainty is obtained by a variation of the cross section applied for the pileup interactions in the Monte Carlo samples by a factor of $\pm 4.6\%$ as recommended by [216].

10.2.3 Lepton ID, tracking, and isolation uncertainties

For the electrons and muons rate and shape uncertainties of the lepton ID, tracking, and isolation scale factors are applied and propagated to the final discriminant output distributions. In this way, the efficiency of reproducing a lepton by the Monte Carlo event generation and subsequent detector simulations is taken into account, which is the approach recommended by the CMS electron-gamma physics-object group [217] and by the CMS muon physics-object group [218]. For each lepton occurring in an event, the uncertainties are typically of the order of 1-2%.

10.2.4 Lepton trigger efficiency uncertainties

Similar to the other lepton uncertainties, variations in the final discriminant output distributions due to the modelling of the lepton trigger efficiency by Monte Carlo samples occur, which are considered as rate and shape uncertainties. These uncertainties are obtained by a variation of the lepton scale factors for each lepton emerging in an event.

10.2.5 Jet energy scale uncertainties

26 individual rate and shape uncertainties are applied in this analysis to take into account systematic uncertainties related to the jet energy scale (JES). They are obtained by a 1σ -variation of the jet energy scale for all jets at reconstruction level, so that after a new reconstruction of the jets and related kinematic variables varied final discriminant output distributions are derived. More information about these uncertainties can be found in [219].

In the case of the b-tagged jets for the 10 most important JES uncertainties dedicated JES uncertainties are determined, while for the other 16 JES uncertainties the same uncertainties as in the case of the non-b-tagged jets are used. All 26 systematic uncertainties are considered fully correlated.

10.2.6 Jet energy resolution uncertainties

Similar to the jet energy scale uncertainties, the jet energy resolution (JER) rate and shape uncertainties are derived by a variation of the difference in the jet energy between jets at particle level and jets at reconstruction level as suggested by [220].

10.2.7 B-tagging uncertainties

For the b-tagging of jets differential CSVv2 scale factors are applied which are derived from studies using heavy-flavour (hf) or light-flavour (lf) enriched control samples as described in [221] and [222]. The uncertainties related both to using the heavy-flavour and light-flavour scale factors can be divided into three categories which are considered separately:

Uncertainty on the heavy-flavour and light-flavour purity in the control samples

The uncertainty on the purity of the heavy-flavour and light-flavour control samples is taken into account by introducing a dedicated uncertainty which is obtained by shifting either the heavy-flavour or light-flavour scale factor by a factor of 1σ . This corresponds to a change in the contamination of a heavy-flavour (or light-flavour) enriched control region with light-flavour (or heavy-flavour) jets.

Uncertainties due to the size of the control samples

Four additional uncertainties, two for the heavy-flavour and two for the light-flavour b-tagging scale factors which are called b-Tag HF stats and b-Tag LF stats, respectively, are employed. In this manner, the limited size of the control samples and the consequential statistical uncertainties is taken into account. The first uncertainty of each class controls distortions in the CSV distribution leading to an overall tilt of the distribution (“linear”), while the second uncertainty controls distortions leading to a change in the lower and upper ends of the CSV distribution with respect to the relative center of the distribution (“quadratic”).

Uncertainties related to the scale factors of charm jets

Two separate additional uncertainties are used to control the systematic uncertainties on the distribution of the scale factors for charm jets. The first uncertainty considers an uncertainty on the linear shape of the charm jet scale factor distribution, while the second uncertainty refers to a quadratic shape deviation.

These uncertainties of the charm jet scale factors are constructed by taking the relative uncertainty from the heavy flavour scale factors and doubling it in size. The two uncertainties related to the c-tagging scale factors are then treated as independent to the uncertainties of the heavy-flavour and light-flavour scale factors. More details on these uncertainties can be found in [222].

10.3 Summary of systematic uncertainties

Table 10.3 summarizes the systematic uncertainties applied on the $t\bar{t}+\geq 1\text{b-jet}$ signal processes and all the background processes for this analysis. Furthermore, it contains a distinction between rate and shape uncertainties.

Table 10.3: Summary of the systematic uncertainties considered in this analysis

Source	Type	Notes
Systematic uncertainties related to the Monte Carlo samples		
QCD Scale ($t\bar{t}$)	Rate	Scale uncertainty for $t\bar{t}$ prediction
QCD Scale ($t\bar{t}+c\bar{c}$)	Rate	Additional scale uncertainty for $t\bar{t}+c\bar{c}$ prediction
QCD Scale ($t\bar{t}H$)	Rate	Scale uncertainty for $t\bar{t}H$ prediction
QCD Scale (t)	Rate	Scale uncertainty for single-top prediction
QCD Scale ($V+\text{jets}$)	Rate	Scale uncertainty for $W+\text{jets}$ and $Z+\text{jets}$ prediction
QCD Scale ($VV+\text{jets}$)	Rate	Scale uncertainty for diboson prediction
PDF (gg)	Rate	PDF uncertainty for gg initiated processes ($t\bar{t}$, $t\bar{t}Z$) except $t\bar{t}H$
PDF ($gg_{t\bar{t}H}$)		PDF uncertainty for $t\bar{t}H$
PDF ($q\bar{q}$)	Rate	PDF uncertainty for $q\bar{q}$ initiated processes ($t\bar{t}W$, W , Z).
PDF (qg)	Rate	PDF uncertainty for qg initiated processes (single top)
PDF ($t\bar{t}$)	Shape	PDF uncertainty for the $t\bar{t}$ samples
Q2 Scale ($t\bar{t}$)	Shape	Renormalization and factorization scale uncertainties for the $t\bar{t}+X$ processes
ME-PS matching ($t\bar{t}$)	Rate	Matrix element generator and parton shower matching uncertainties for the $t\bar{t}+X$ processes
PS Scale: ISR ($t\bar{t}$)	Rate	Initial-state radiation parton shower scale uncertainties for the $t\bar{t}+X$ processes
PS Scale: FSR ($t\bar{t}$)	Rate	Final-state radiation parton shower scale uncertainties for the $t\bar{t}+X$ processes
Underlying Event ($t\bar{t}$)	Rate	Underlying event tune uncertainties for the $t\bar{t}+X$ processes
Bin-by-bin statistics	Shape	Statistical uncertainty of the signal and background predictions due to the limited size of the Monte Carlo samples
Systematic uncertainties related to the measurement process		
Luminosity	Rate	Uncertainty on the luminosity measurement
Pileup	Rate, Shape	Uncertainties on the number of pileup interactions
Lepton ID, tracking, and isolation	Rate, Shape	Uncertainty on the lepton identification, tracking, and isolation
Lepton trigger efficiency	Rate, Shape	Uncertainty on the lepton triggering
JES [26 individual uncertainties]	Rate, Shape	Jet energy scale uncertainties
JER	Rate, Shape	Jet energy resolution uncertainties
b-Tag HF purity	Shape	Purity uncertainty of the heavy-flavour control region used to determine the b-tagging scale factors
b-Tag HF stats (linear)	Shape	Statistical uncertainty related to the determination of the b-tagging scale factors
b-Tag HF stats (quadratic)	Shape	Statistical uncertainty related to the determination of the b-tagging scale factors
b-Tag LF purity	Shape	Purity uncertainty of the light-flavour control region used to determine the b-tagging scale factors
b-Tag LF stats (linear)	Shape	Statistical uncertainty related to the determination of the b-tagging scale factors
b-Tag LF stats (quadratic)	Shape	Statistical uncertainty related to the determination of the b-tagging scale factors
b-Tag Charm (linear)	Shape	Uncertainties related to the scale factors of charm jets
b-Tag Charm (quadratic)	Shape	Uncertainties related to the scale factors of charm jets

11. Results

In this chapter – based on the foundations layed in the previous chapters – the results of the measurement of the $t\bar{t}+\geq 1\text{b-jet}$ ($t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$) signal strength and corresponding cross section using the reference analyses will be presented. Furthermore, some additional results and tests helping to assess the obtained results better will be discussed.

11.1 $t\bar{t}+\geq 1\text{b-jet}$ signal strength result

For the reference analyses the $t\bar{t}+\geq 1\text{b-jet}$ signal strength r result obtained with a Maximum-Likelihood fit is stated in table 11.1.

As a short reminder, the B-jet multiplicity (B-jet mult.) and linear discriminant (LD) based analyses perform a binary classification between the $t\bar{t}+\geq 1\text{b-jet}$ signal processes ($t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$) and the background processes ($t\bar{t}+c\bar{c}$, $t\bar{t}+lf$, $t\bar{t}H$). In comparison, the neural network (NN) based analysis does a multiclassification between the six associated event classes. Both the LD and the NN analyses make use of 11 input variables, discussed in sections 9.2 to 9.2.2. The B-jet mult. analyses, however, just uses the distribution of the number of b-tagged jets as final discriminant.

In table 11.1, the expected $t\bar{t}+\geq 1\text{b-jet}$ signal strength $r(\text{expected})$ result is given using an Asimov dataset under signal expectation ($S=1$) and the observed $t\bar{t}+\geq 1\text{b-jet}$ signal strength $r(\text{observed})$ result applying the measurement data after the event selection (cf. sections 7.1 and 7) is listed.

In this manner, the systematic uncertainties which correspond to a 1σ -uncertainty and which are related to the determination of the signal strength r result can be compared between both data sets. It is generally expected that the signal strength result of the Asimov data set $r(\text{expected})$ has slightly larger uncertainties, since the nuisance parameters do not get as much constrained as by real data, while only small differences between the obtained uncertainties of the Asimov data set and the measurement data indicate that the underlying model of the systematic uncertainties is already quite well described by the Asimov data set.

Some analyses also show the expected signal strength $r(\text{expected}, S=0)$ result for an Asimov data set with only background expectation ($S=0$). However, in the case of the analyses presented in this thesis showing a background-only expectation does not seem reasonable. The reason for this is that the signal contributions of the $t\bar{t}+\geq 1\text{b-jet}$ processes are rather large and a background-only model cannot be properly fitted to the measurement data.

Therefore, this kind of signal strength $r(\text{expected}, S=0)$ result is always neglected.

As can be seen from table 11.1 all analyses observe a signal strength $r(\text{observed})$ larger than expected from the predictions of the Standard Model of particle physics (SM). Furthermore, especially the two multivariate analyses (LD, NN) are in very good agreement with each other, both deriving a $t\bar{t}+\geq 1\text{b-jet}$ signal strength of $r(\text{observed}) = 1.2$. On the other hand, the smaller signal strength $r(\text{observed})$ result of the B-jet mult. analysis can be attributed to the rather basic number of b-tagged jets distribution, which is exploited as final discriminant (see figure 9.1 for the prefit version): Due to the smaller separation between signal and background processes and lower number of bins in its distribution, it is more likely that during the fit of this final discriminant background processes are pulled upwards reducing the amount of signal. The rather large reduction of the expected uncertainties in comparison to the observed uncertainties for the B-jet mult. analysis also can be seen as an indication that the separation between signal and background processes may be not as good as in the two other analysis.

Overall, one can safely say that the expected and observed systematic uncertainties are smaller for the NN analysis than for the other two analysis and that also the difference between the expected and observed uncertainties of this analysis is smallest. This is not surprising since the NN analysis utilizes rather complex multivariate analysis methods by making use of a multiclassifying neural network with 2 hidden layers and 100 nodes per hidden layer. However, it is striking that the much simpler linear discriminant based multivariate analysis comes to a similar result with compatible uncertainties. Moreover, if one neglects the different observed signal strength $r(\text{observed})$ result of the B-jet mult. analysis and only compares its expected uncertainties with the other two analyses, these uncertainties are also in the same order of magnitude. The reason for these rather similar and still quite large uncertainties is that the uncertainties of the analyses are mostly limited by large systematic uncertainties and rather less by statistical uncertainties which will be studied in the next section in more detail.

At this point one can conclude that the three reference analyses lead to an increased, but rather consistent $t\bar{t}+\geq 1\text{b-jet}$ signal strength of approximately $r(\text{observed}) = 1.20^{+23\%}_{-17\%}$, where the result of the NN analysis is used to estimate the uncertainty.

Table 11.1: $t\bar{t}+\geq 1\text{b-jet}$ signal strength r results for the reference analyses: For the three reference analyses the $t\bar{t}+\geq 1\text{b-jet}$ signal strength result for an Asimov data set under signal expectation $r(\text{expected})$ and the $t\bar{t}+\geq 1\text{b-jet}$ signal strength result for the measurement data after the event selection $r(\text{observed})$ is shown. The systematic uncertainties (listed in percent) correspond to a 1σ -uncertainty. Besides of that the type of classification and, if applicable, the chosen parameters are given.

Analysis	Classification	Parameters	$r(\text{expected})$	$r(\text{observed})$
B-jet mult.	binary	-	$1^{+32.7\%}_{-22.7\%}$	$1.09^{+25.0\%}_{-22.1\%}$
LD	binary	11 vars	$1^{+24.3\%}_{-17.6\%}$	$1.21^{+23.4\%}_{-19.5\%}$
NN	multiclass	11 vars, 2 HL, 100 nodes	$1^{+22.9\%}_{-17.1\%}$	$1.20^{+22.8\%}_{-17.4\%}$

11.2 Influence of systematic uncertainties on the $t\bar{t}+\geq 1b$ -jet signal strength result

As can be seen from table 11.1 all three reference analyses derive an uncertainty on the observed signal strength $\Delta r(\text{observed})$ in the order of $\Delta r(\text{observed}) \approx \pm 20\% - 24\%$, if one transforms the slightly different upwards and downwards uncertainties into a common two-sided uncertainty. It is of special interest that the three reference analyses derive a $t\bar{t}+\geq 1b$ -jet signal strength with almost the same uncertainty. In general, it would be expected that the multivariate analyses (LD, NN), by using more elaborate final discriminant output distributions, can constrain the result and its corresponding uncertainties better. It is assumed that at least the multivariate analyses are affected by large systematic uncertainties and, hence, cannot derive a more precise result. Therefore, in this section the influence of systematic uncertainties on the measurement of the $t\bar{t}+\geq 1b$ -jet signal strength will be discussed.

11.2.1 Prefit and postfit distributions of the three reference analyses

The prefit and postfit distributions of the three reference analysis using the selected measurement data are presented in figures 11.1 to 11.4. In the case of the B-jet mult. and LD analyses, which only use one final output distribution (either the number of b-tagged jets or the linear discriminant output), the prefit and postfit distributions are shown side-by-side in the corresponding figures 11.1 and 11.2. Due to the six event classes and associated final discriminant output distributions of the NN analyses the prefit distributions are shown in figure 11.3 and the postfit distributions in the subsequent figure 11.4.

It is important to note that the uncertainty bands in the prefit and postfit distributions are obtained differently. For the prefit distributions all systematic uncertainties are considered uncorrelated and the corresponding nuisance parameters θ are assumed to follow a Gaussian distribution. Therefore, the uncertainties of the nuisance parameters θ are added in quadrature. This results in rather large uncertainty bands compared to the uncertainty bands of the postfit distributions.

In the case of the postfit distributions, the nuisance parameters are getting constrained by the Maximum-Likelihood fit (see section 9.7) and correlations between them are taken into account by computing the covariance matrix by the fit. For the determination of the uncertainty bands the covariance matrix is used to construct probability distributions of the nuisance parameters. By sampling from multivariate normal distributions the rate changes of the final discriminant output distributions due to the constrained nuisance parameters can be calculated for every bin of the respective final discriminant output distribution.

From the prefit and postfit output distributions of the three reference analyses it can be observed that before the Maximum-Likelihood fit and also after the fit the agreement between the sum of Monte Carlo predictions and the measurement data is very good for all output distributions. Furthermore, the effect of the systematic uncertainties rarely leads to a systematic bin uncertainty (shown by the hashed uncertainty band) larger than 20 % of the bin expectation before the fit. After constraining the systematic uncertainties by the fit, these bin uncertainties do not exceed 10 % for the most part. Moreover, all bin-by-bin systematic uncertainties show consistent behaviour between the prefit and postfit distributions. In most cases, the bin uncertainties get drastically smaller after the fit. Only the right-most bins of the $t\bar{t}+2b$, $t\bar{t}+b$, and $t\bar{t}H$ output distributions of the NN analysis show only a slightly decrease of their uncertainties after the fit. This can be explained by the relative large uncertainty of the few measured events falling into these bins.

Due to the similar bin-by-bin systematic uncertainties observed in the postfit distributions, the similar observed uncertainty on the $t\bar{t}+\geq 1\text{b-jet}$ signal strength $r(\text{observed})$ seems plausible. However, a direct conclusion which systematic uncertainties and related nuisance parameters cause the rather large observed uncertainty of $\Delta r(\text{observed}) \approx \pm 20\% - 24\%$ cannot be drawn from the prefit and postfit plots. For such studies pull and impact distributions are more suitable, which will be discussed in the next section.

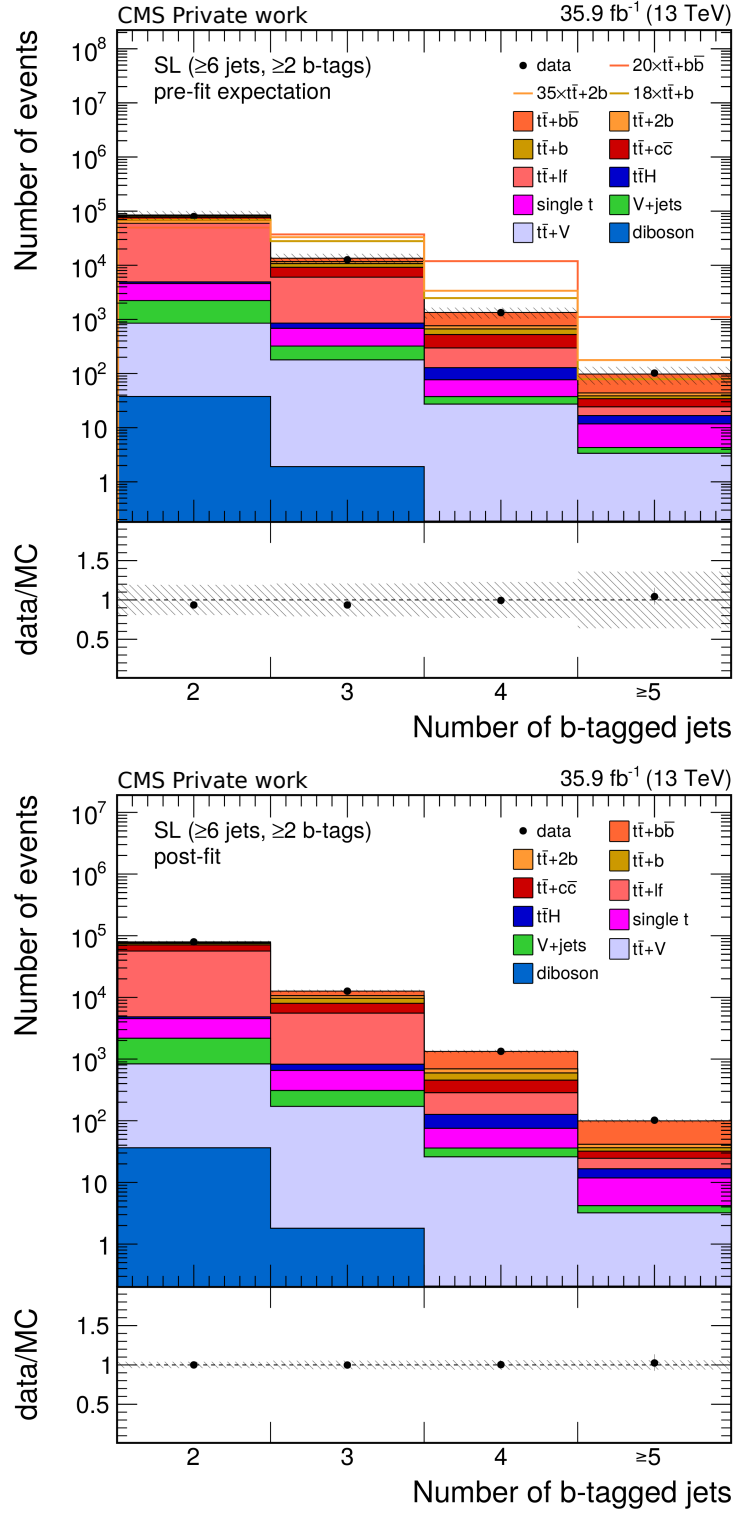


Figure 11.1: Final discriminant distributions of the B-jet multiplicity based analysis using the measurement data: As the final discriminant of the B-jet multiplicity based analysis the number of b-tagged jets distribution is chosen. The stacked Monte Carlo histograms (filled blocks) are compared to the measurement data (black dots). In addition, the contributions of the three signal processes ($t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$) are individually scaled to the overall integral of the stacked histograms and are shown as solid lines for the prefit distribution (upper plot). In the prefit distribution, the hashed uncertainty band considers all systematic uncertainties added in quadrature. In the postfit distribution (lower plot), the hashed uncertainty band considers all systematic uncertainties constrained by the Maximum-Likelihood fit via covariance matrix sampling.

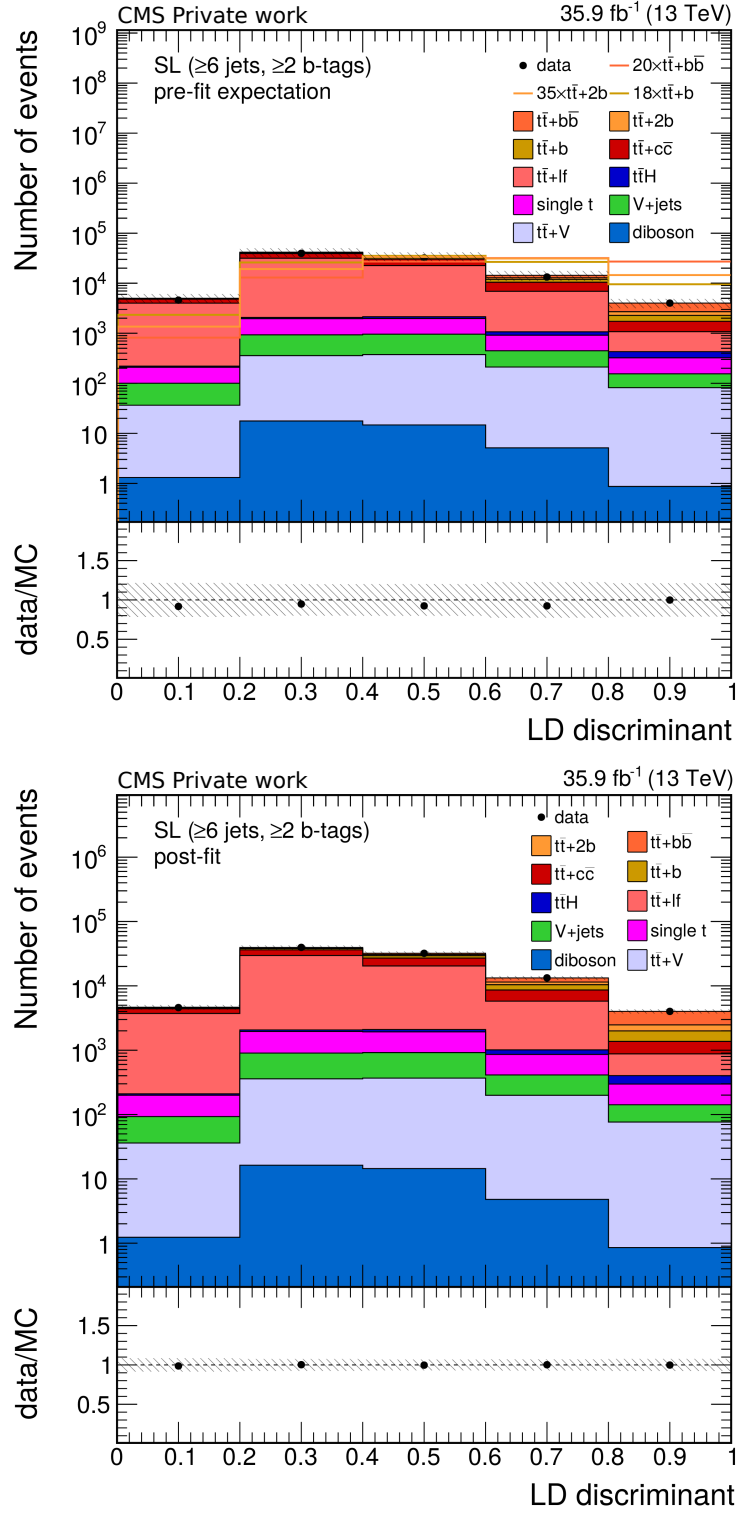


Figure 11.2: Final discriminant distributions of the linear discriminant based analysis using the measurement data: The output of the linear discriminant is chosen as the final discriminant. The stacked Monte Carlo histograms (filled blocks) are compared to the measurement data (black dots). In addition, the contributions of the three signal processes ($t\bar{t} + b\bar{b}$, $t\bar{t} + 2b$, $t\bar{t} + b$) are individually scaled to the overall integral of the stacked histograms and are shown as solid lines for the prefit distribution (upper plot). In the prefit distribution, the hashed uncertainty band considers all systematic uncertainties added in quadrature. In the postfit distribution (lower plot), the hashed uncertainty band considers all systematic uncertainties constrained by the Maximum-Likelihood fit via covariance matrix sampling.

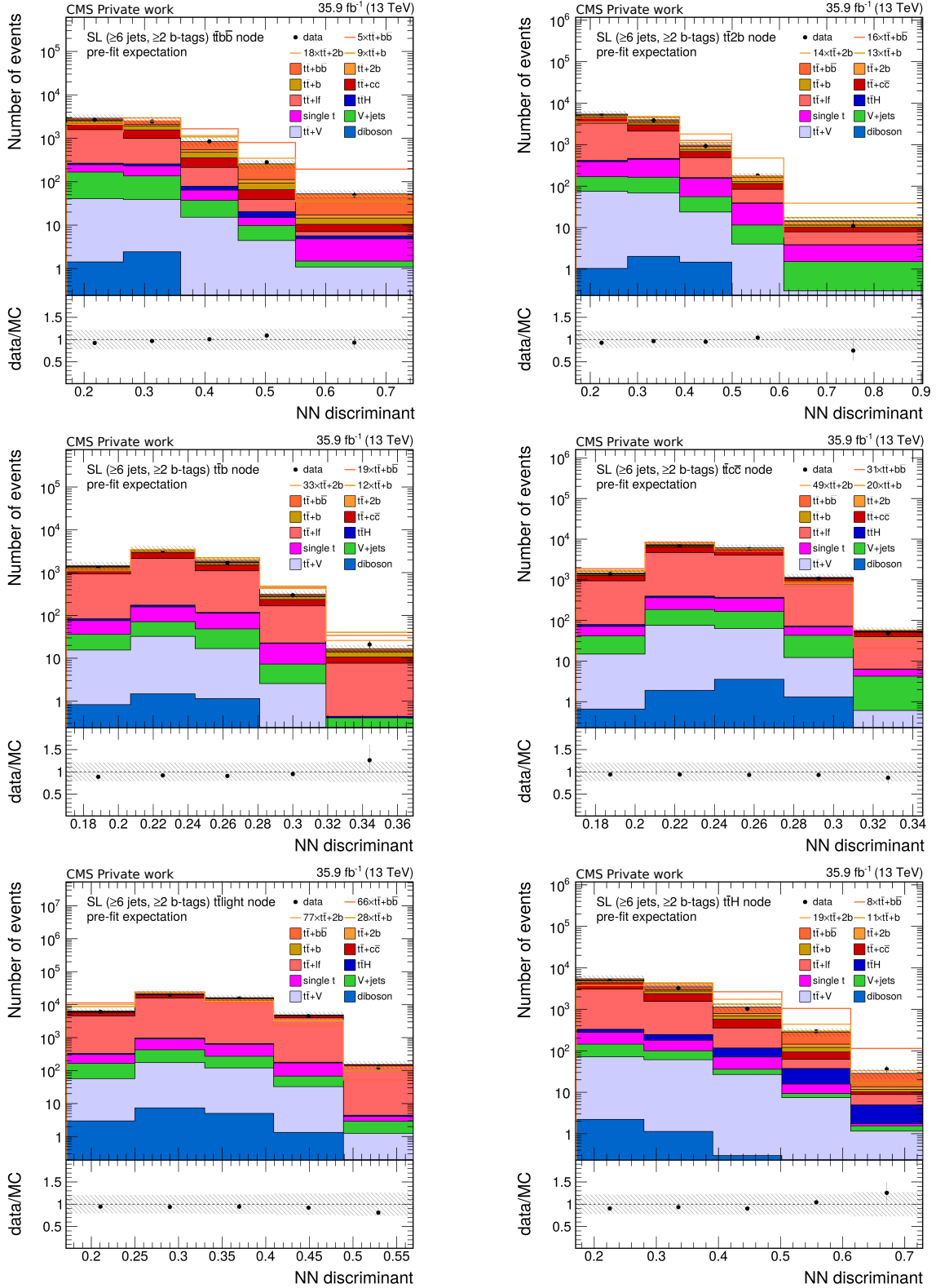


Figure 11.3: Prefit final discriminant distributions of the neural network based analysis using the measurement data: The output distributions of the six output nodes of the neural network, which correspond to the six event classes ($t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$, $t\bar{t}+c\bar{c}$, $t\bar{t}+lf$, $t\bar{t}H$) are chosen as the final discriminant. The stacked Monte Carlo histograms (filled blocks) are compared to the measurement data (black dots). In addition, the contributions of the three signal processes ($t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$) are individually scaled to the overall integral of the stacked histograms and are shown as solid lines. The hashed uncertainty band considers all systematic uncertainties added in quadrature. The distribution is shown before the Maximum-Likelihood fit.

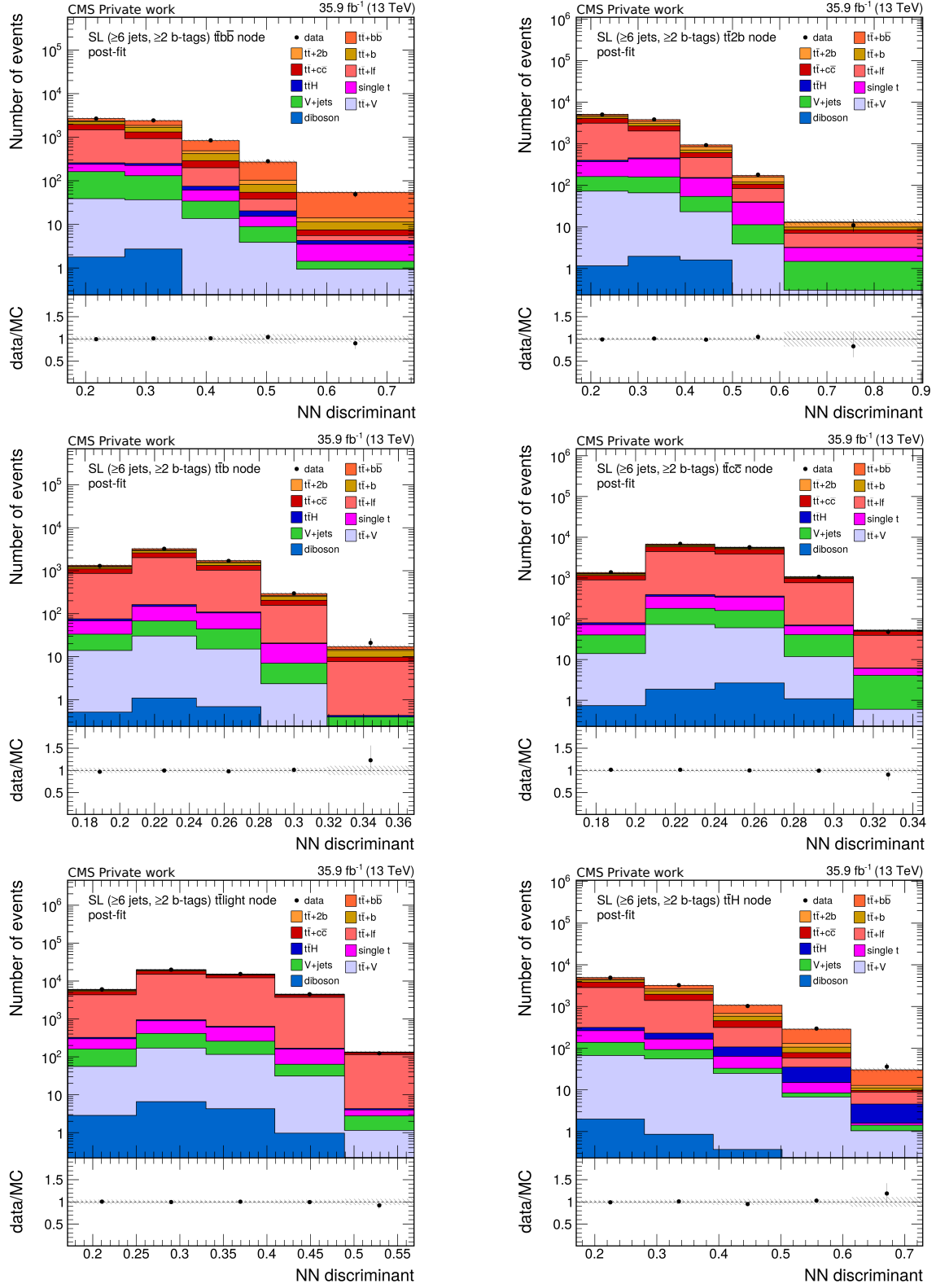


Figure 11.4: Postfit final discriminant distributions of the neural network based analysis using the measurement data: The output distributions of the six output nodes of the neural network, which correspond to the six event classes ($\text{tt}+\text{bb}$, $\text{tt}+2\text{b}$, $\text{tt}+\text{b}$, $\text{tt}+\text{c}\bar{\text{c}}$, $\text{tt}+\text{l}\bar{\text{l}}$, ttH) are chosen as the final discriminant. The stacked Monte Carlo histograms (filled blocks) are compared to the measurement data (black dots). The hashed uncertainty band considers all systematic uncertainties constrained by the Maximum-Likelihood fit via covariance matrix sampling.

11.2.2 Pull and impact distributions of the three reference analyses

Pull and impact distributions are quite useful to determine which systematic uncertainties and related nuisance parameters θ have the largest effect on the observed uncertainty of the $t\bar{t}+\geq 1b$ -jet signal strength $r(\text{observed})$ result. In pull distributions, for each systematic uncertainty the difference between the initial value of its nuisance parameter θ_0 and the best-fit value after the Maximum-Likelihood fit $\hat{\theta}$ divided by the uncertainty on this parameter before the fit $\Delta\theta$

$$\frac{\hat{\theta} - \theta_0}{\Delta\theta}$$

is shown. The constraints on the nuisance parameter correspond to a 1σ -confidence interval around the prefit value which is also presented. Large deviations of $\frac{\hat{\theta}-\theta_0}{\Delta\theta}$ from zero, or, in other words, pulls, indicate that the initially assumed value of a nuisance parameter θ and the corresponding model of a systematic uncertainty is not properly chosen and that the fit finds another value of this nuisance parameter more suitable. This often happens for systematic uncertainties, which proper estimation and modeling presents a challenge, and has to be understood. It can also indicate a problem with the fit model. Therefore, if occurring, the cause of such large pulls has to be checked.

Another feature occurring in pull distributions is that nuisance parameters can have rather strong constraints. This is usually caused by nuisance parameters and related systematic uncertainties for which rather large uncertainties are assumed in the beginning and which can be constrained quite well by the fit due to the knowledge obtained through the overall systematic uncertainty model. If systematic uncertainties related to rates of background processes are tightly constrained, in particular if the upper uncertainty interval is small, the sensitivity of the analysis is improved. The reason for the improvement is that the fit model cannot accommodate a signal-like excess by an upward fluctuation of the background model easily due to the constraints.

For the production of impact distributions each nuisance parameter is fixed either on the upper or lower border of its 1σ -confidence interval after the fit. The Maximum-Likelihood fit is then conducted again with the remaining nuisance parameters being freely floating. Afterwards, the change in the observed signal strength result Δr from the nominal observed signal strength result $r(\text{observed})$ is computed and shown as impact of the corresponding nuisance parameter and associated systematic uncertainty in the plot. In this manner, the influence of a systematic uncertainty on the overall fit result can be estimated. Furthermore, impact distributions are scanned for systematic uncertainties having single-sided impacts. Similar to systematic uncertainties with large pulls, systematic uncertainties with single-sided impacts suggest that these uncertainties are not properly modeled and should be understood better. Positive and negative correlations between the systematic uncertainties and their impact on the $t\bar{t}+\geq 1b$ -jet signal strength result can also be read from the corresponding impact distributions.

The combined pull and impact distributions listing the 13 most important systematic uncertainties of the neural network based (NN) analysis is presented in figure 11.5. This analysis is the most sensitive of the reference analyses. The combined pull and impact distributions of the other two reference analyses lead to similar findings and are part of appendix A together with a plot containing the 30 most important systematic uncertainties of the NN analysis.

From the combined pull and impact distributions it can be observed that the following

systematic uncertainties are most dominant in all three analyses:

B-tagging uncertainties

The uncertainty on the heavy-flavour and light-flavour purity in the control samples (b-Tag hf purity, b-Tag lf purity) to derive the b-tagging scale factors as well as the uncertainties on the scale factors of charm jets (b-Tag Charm (linear), b-Tag Charm (quadratic)) (cf. section 10.2.7) are among the most dominant uncertainties.

The strong impact of b-tagging uncertainties is not unexpected since as explained during the discussion of the input variables for the two multivariate analyses (section 9.2.2) b-tagging information is mostly exploited to separate the $t\bar{t} + \geq 1$ b-jet signal processes from the background processes, while the B-jet mult. analysis makes explicitly use of the number of b-tagged jets. Moreover, the proper b-tagging of bottom jets and even more the proper tagging of charm jets is still experimentally challenging and a continuing field of further study. For example, the CSVv2 b-tagging algorithm [89], which is used in this analysis and was used by default in the CMS Collaboration from the beginning of the LHC Run 2 up to now, is superseded for the CMS 2017 measurement data by the newer DeepCSV b-tagging algorithm [90] promising a more precise tagging of jets.

Uncertainties related to the parton shower

In all three analyses uncertainties related to the parton shower and to the description of either the $t\bar{t} + b\bar{b}$ or $t\bar{t} + lf$ process have significant impact on the overall result. Among them the uncertainty due to the chosen value of the strong coupling constant α_S for the description of the final-state radiation (FSR) by the parton shower (PS Scale: FSR ($t\bar{t} + b\bar{b}$), PS Scale: FSR ($t\bar{t} + lf$)) is the most important (section 10.1.6). It is followed by the impact of the uncertainty on the initial-state radiation (ISR) α_S value (PS Scale: ISR ($t\bar{t} + b\bar{b}$), PS Scale: ISR ($t\bar{t} + lf$)) and the uncertainty due to the matrix element generator and parton shower matching (ME-PS: ($t\bar{t} + b\bar{b}$), ME-PS: ($t\bar{t} + lf$)) (section 10.1.5).

Also this finding is expected since, as mentioned in the corresponding systematic uncertainty sections, the statistics of the Monte Carlo samples applied to derive and estimate these uncertainties is rather limited. For this reason rather large rate uncertainties have to be used to cover these systematic effects.

Looking at the constraints of the pull and the impact distribution of the neural network (NN) based analysis, it is interesting to note that the uncertainty related to the linear shape of the scale factor distribution of charm jets (B-Tag Charm (linear)) gets pulled and shows a deviation from its prefit value. Likewise its corresponding impact is single-sided. The pull and impact distributions of the other two analyses (available in appendix A) do not show single-sided impacts, but also a tendency that this nuisance parameter gets pulled to larger values. It can be assumed that the linear component of the shape of the scale factor distribution for the charm jets was mismodeled in the beginning and led to an underestimation of the associated initial value of the B-Tag Charm (linear) nuisance parameter. However, this should not be cause for excessive concern since it seems that this upward pull of the nuisance parameter gets balanced by a downward pull of the B-Tag Charm (quadratic) nuisance parameter, which reflects the uncertainty on the quadratic component of the shape of the charm jet scale factor distribution and which supports the above assumption of a mismodeling of the scale factor distribution. Furthermore, possibly the extra rate uncertainty due to the cross section normalization of the $t\bar{t} + c\bar{c}$ process (QCD Scale ($t\bar{t} + c\bar{c}$) nuisance parameter, cf. section 10.1.2) makes up for this mismodeling by getting pulled to negative values as well.

Besides of this all pulls and impacts look reasonable.

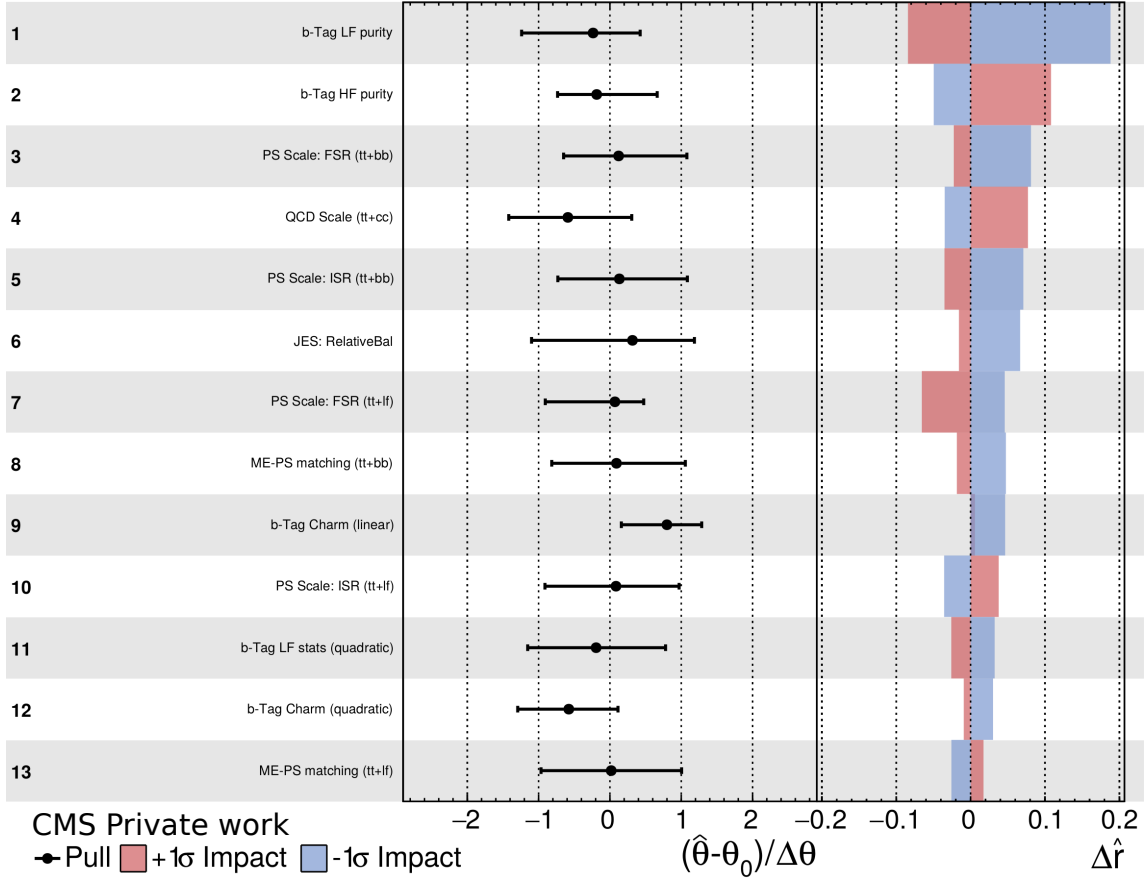


Figure 11.5: Pull and impact distribution of the neural network based analysis: The pull of a nuisance parameter θ (black point) corresponds to the difference between its initial value and its best-fit value after the fit divided by the uncertainty on it before the fit $\frac{\hat{\theta} - \theta_0}{\Delta\theta}$. The 1-sigma confidence interval according to its prefit value is shown as a black line. The impact of a nuisance parameter on the $t\bar{t}+\geq 1b$ -jet signal strength result Δr is obtained by repeating the fit while fixing this parameter to the upper or lower bound of its nominal postfit 1-sigma confidence interval. Positive (blue/red) or negative correlations (red/blue) between the nuisance parameter and the signal strength result can be concluded from the coloured bars. In the plot the 13 systematic uncertainties having the largest impact are presented.

11.2.3 Influence of specific systematic uncertainties on the $t\bar{t}+\geq 1\text{b-jet}$ signal strength result

In the following, the effects of specific systematic uncertainties on the $t\bar{t}+\geq 1\text{b-jet}$ signal strength result will be studied in detail, since – as explained in the previous section – especially systematic uncertainties related to the b-tagging of jets and to the parameters chosen for the parton shower during the production of the $t\bar{t}+X$ Monte Carlo samples seem to have a large impact.

The subsequent outcomes are presented using only the results produced with the neural network (NN) based reference analysis since this analysis has the smallest expected systematic uncertainty on the $t\bar{t}+\geq 1\text{b-jet}$ signal strength result and, hence, seems most suitable. Nonetheless, the findings of the NN analysis are consistent with the findings of the other two analyses, which will not be presented to avoid duplicate information. Moreover, subsequently, results produced with the Asimov dataset will be presented since due to the expected and observed $t\bar{t}+\geq 1\text{b-jet}$ signal strength $r(\text{expected/observed}) = 1$ the relative uncertainties can easily be determined and compared. But as a part of a cross check the subsequent findings were also reproduced for the measurement data.

In a first study none of the systematic uncertainties usually applied in the reference analyses and described in section 10 is considered during the Maximum-Likelihood fit. This approach is frequently chosen to determine for the overall uncertainty Δr on a signal strength result r which proportion can be attributed to the statistical uncertainties due to the Poisson distribution of the measurement data $\Delta r(\text{stats})$. Because the overall uncertainty Δr is defined as a quadratic sum of the statistical uncertainties $\Delta r(\text{stats})$ and the systematic uncertainties $\Delta r(\text{sys})$, knowing the overall and statistical uncertainties, the systematic uncertainties of a signal strength result can be easily computed via the following relation:

$$\Delta r(\text{sys}) = \sqrt{\Delta r^2 - \Delta r(\text{stats})^2}$$

For all three reference analyses it is found that the statistical uncertainty on the $t\bar{t}+\geq 1\text{b-jet}$ signal strength result is roughly $\Delta r(\text{stats}) = \pm 1.8\%$. From this finding it is clear that all three analyses are driven and limited by the systematic uncertainties, which are in the order of $\Delta r(\text{sys}) \approx \Delta r \approx \pm 20 - 24\%$.

For future analyses reducing these systematic uncertainties will be of particular interest. For this reason, it is vitally important to know which specific systematic uncertainties are the largest ones and should be reduced. Of course, due to the conclusions from the previously discussed impact plots the b-tagging and parton shower related systematic uncertainties are the natural candidates causing the overall large systematic uncertainty. Therefore, in further studies the Maximum-Likelihood fit is conducted without considering these two kinds of uncertainties. The results are given in table 11.2 and are explained below.

The assumption that the b-tagging and parton shower related systematic uncertainties affect the overall uncertainty of the $t\bar{t}+\geq 1\text{b-jet}$ signal strength result Δr most of all proves true. Without these kinds of uncertainties the overall uncertainty would be roughly $\Delta r \approx \pm 4.4\%$ which leads to the conclusion that the remaining systematic uncertainties contribute only to the overall uncertainty in the order of roughly $\Delta r(\text{sys}) \approx \pm 4.0\%$. Moreover, a few further things can be observed:

Table 11.2: Influence of specific systematic uncertainties on the overall uncertainty Δr of the $t\bar{t}+\geq 1b$ -jet signal strength r result for the neural network based reference analysis: The overall uncertainty is determined by redoing the Maximum-Likelihood fit, while some systematic uncertainties are not considered. The resulting downwards overall uncertainty ($-\Delta r$) and the upwards overall uncertainty ($+\Delta r$) corresponding to a 1σ -confidence interval is stated using the Asimov data set and the NN reference analyses. Similar results can be found for the measurement data and the other two reference analyses.

Systematic uncertainties	$-\Delta r$	$+\Delta r$
All systematic uncertainties	17.13%	22.88%
Without any systematic uncertainties	1.81 %	1.82 %
Without b-tagging related systematic uncertainties	10.84 %	17.00 %
Without parton shower related systematic uncertainties	13.23%	14.26 %
Without b-tagging and parton shower related systematic uncertainties	4.35 %	4.50 %
Without a subset of b-tagging and parton shower related systematic uncertainties	6.91 %	7.91 %

First of all, the effects of the b-tagging and parton shower related systematic uncertainties on the overall uncertainty are mostly uncorrelated between each other. That is the reason why the Maximum-Likelihood fit does only slightly exploit correlations between them to constrain the uncertainties of the fit. This can be seen if one combines the obtained uncertainty result without the b-tagging related systematic uncertainties with the uncertainty result without the parton shower related systematic uncertainties by quadratic sum. A comparison to the overall uncertainty result considering all systematic uncertainties leads to:

$$\begin{aligned}
-\sqrt{(10.84\%)^2 + (13.23\%)^2} &= -17.10\% && \approx -\Delta r = -17.13\% \\
+\sqrt{(17.00\%)^2 + (14.26\%)^2} &= +22.19\% && \approx +\Delta r = +22.88\%
\end{aligned}$$

It can be safely said that only upwards correlations of both kinds of systematic uncertainties are slightly exploited by the fit. The reason for this could be that the two types of uncertainties have – at least for downward variations – not a direct effect on each other.

Secondly, the parton shower related systematic uncertainties seem to be the major cause of the inequality of the upwards and downwards overall uncertainty Δr since the overall uncertainty after their removal becomes almost uniform ($-\Delta r : -13.23\%$, $+\Delta r : +14.26\%$). In comparison a removal of the b-tagging related systematic uncertainties leads to a quite unequal overall uncertainty ($-\Delta r : -10.84\%$, $+\Delta r : +17.00\%$). The unequal overall uncertainty due to the parton shower related uncertainties is most likely caused by converting the parton shower related shape uncertainties suffering from low statistics to rate uncertainties. Especially, in the case of an upwards variation of the α_S values in the parton shower (PS: FSR and PS: ISR systematic uncertainties) a change of the $t\bar{t}+X$ process shapes is expected. This not fully covered by using only rate uncertainties based on the nominal shapes due to a higher selection efficiency of events with more parton splittings and corresponding higher number of additional jets. Hence, it is likely that the associated upward rate uncertainties do not get constrained as much in the Maximum-Likelihood fit as their downward companions.

Thirdly, the b-tagging and parton shower related systematic uncertainty drive the overall uncertainty Δr roughly in the same order of magnitude, especially if one compares their effects on the downward overall uncertainty $-\Delta r$.

Finally, this common set of systematic uncertainties causing the large overall uncertainty at most can be further restricted by determination of a sub set of b-tagging and parton shower related systematic uncertainties. In a future analysis studies to reduce the systematic uncertainties contained in this sub set can then be made, which hopefully leads to a reduction of the overall uncertainty Δr . It can be found that the following systematic uncertainties are the major cause of a large overall uncertainty: b-Tag LF purity, b-Tag HF purity, b-tag Charm (linear), b-Tag Charm (quadratic), b-Tag LF stats (quadratic); ME-PS matching ($t\bar{t}+b\bar{b}$), ME-PS matching ($t\bar{t}+lf$); PS: FSR ($t\bar{t}+b\bar{b}$), PS: FSR ($t\bar{t}+lf$), PS: ISR ($t\bar{t}+b\bar{b}$), PS: ISR($t\bar{t}+lf$). If these systematic uncertainties are neglected during the Maximum-Likelihood fit the overall uncertainty gets reduced to $-\Delta r : -6.91, +\Delta r : +7.91$. Better b-tagging scale factors or a new b-tagging algorithm as explained previously could lead to a decrease of these systematic uncertainties as an improvement of the Monte Carlo samples applied to describe the $t\bar{t}+X$ processes.

11.3 Influence of correlations between input variables on the $t\bar{t}+\geq 1b$ -jet signal strength result

Some of the input variables applied in the linear discriminant and neural network based multivariate analyses are strongly correlated as can be seen from the correlation matrix presented in figure 11.6. Since correlations between input variables can have a negative impact on the accuracy of the $t\bar{t}+\geq 1b$ -jet signal strength result r , the influence of correlations will be investigated more closely in this section.

11.3.1 Strong correlations between pairs of input variables

It can be observed from the correlation matrix (cf. figure 11.6), which is produced as part of the LD based analysis using TMVA [120], that no huge negative correlations between input variables occur, while the following pairs of input variables are strongly positively correlated:

Correlation between b-tagging likelihood ratio and transformed b-tagging likelihood ratio

A strong linear correlation of 87 % between the b-tagging likelihood ratio (B-tag. LR) and the transformed b-tagging likelihood ratio (Trans. b-tag. LR) is expected since the only difference between these two variables is the following non-linear transformation:

$$\text{Trans. b-tag. LR} = \ln\left(\frac{\text{B-tag. LR}}{1 - \text{B-tag. LR}}\right)$$

Nonetheless, the application of both variables in the multivariate analyses is beneficial for two reasons: Both variables have the largest separation power between signal ($t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$) and background ($t\bar{t}+c\bar{c}$, $t\bar{t}+lf$, $t\bar{t}H$) events as stated in table 9.2. Furthermore, the application of both variables helps to simplify the network topology of the neural network used in the NN analysis as will be discussed later in this section.

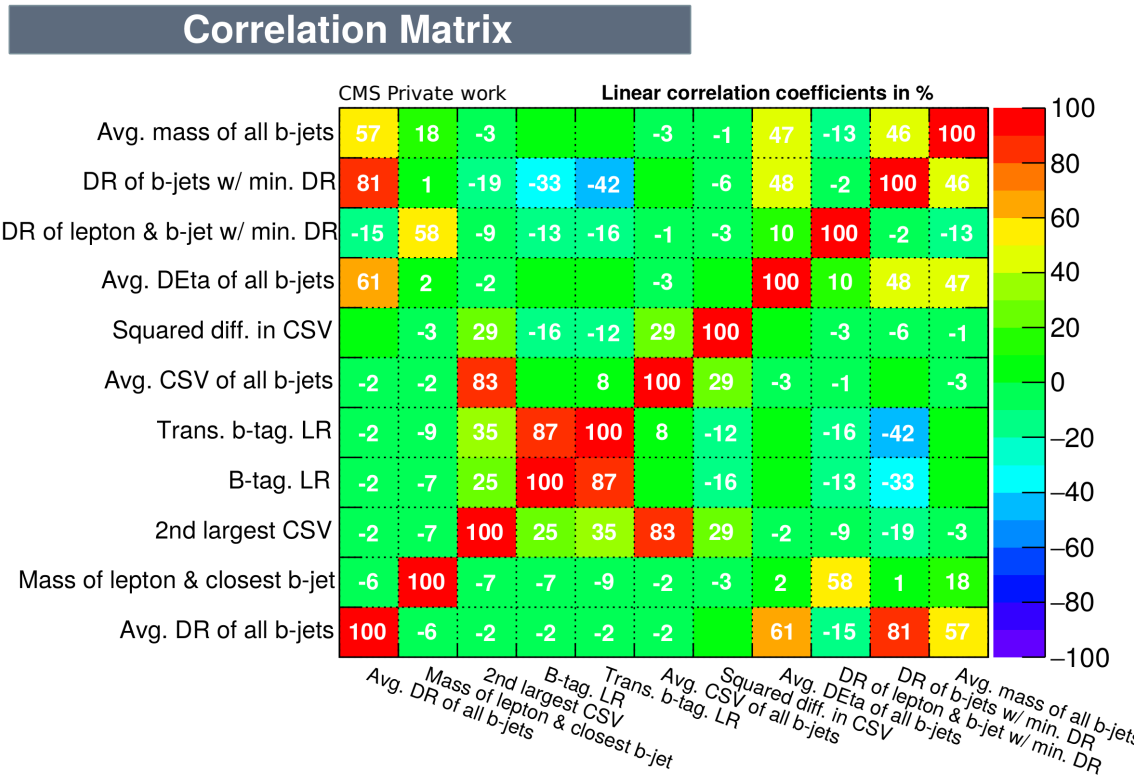


Figure 11.6: Correlation matrix of the input variables applied in the multivariate analyses: The correlation matrix shows linear correlation coefficients between the 11 input variables applied in the linear discriminant and neural network based reference analyses.

Correlation between second-largest CSV and average CSV of all b-tagged jets

The pair of the second-largest CSVv2 tagger discriminant value of all b-tagged jets (2nd largest CSV) and the average CSVv2 discriminant value of all b-tagged jets (Avg. CSV of all b-jets) input variables represents the second-largest linear correlation of input variables with 83 %. Also here a large correlation is rather expected: Due to the event selection requirement of six jets and at least two b-tagged jets, each event contains at least two b-tagged jets, while increasingly large multiplicities of b-tagged jets are less likely as it is evident from the number of b-tagged jets distribution applied by the B-jet. mult. based analysis (figure 9.1). Hence, the Avg. CSV of all b-jets input variables shares roughly 46 % of its information with the 2nd largest CSV input variable since it contains the CSVv2 tagger discriminant values of all b-tagged jets including the second-largest CSVv2 value and on average an event passing the event selection contains 2.17 b-tagged jets.

In principle, one could construct an input variable which calculates the average CSVv2 discriminant value of all b-tagged jets without considering the second-largest CSVv2 value. Including the second-largest CSVv2 value in the computation of the average value should provide more information about the overall system of b-tagged jets in an event. The reason for this is that a larger value of the second-largest b-tagged jet should also in general be linked with larger CSV values of the other b-tagged jets and more b-tagged jets occurring in an event. Leaving out the second-largest CSV value in the averaging process would diminish this common link between all b-tagged jets and most likely reduce the intrinsic separation power exploitable by a multivariate analysis technique.

Correlation between ΔR value of the b-jet pair with the smallest ΔR value and average ΔR value of all b-jets

The input variable ΔR value of the b-tagged jet pair with the smallest ΔR value (DR of b-jets w/ min. DR) and the input variable average ΔR value of all b-tagged jets (Avg. DR of all b-jets) also show a relatively large linear correlation with a correlation coefficient of 81 %.

Similar considerations as in the case of the previous input variable pair apply since the one input variable can be seen as a subset of the other one. Again, the occurrence of a correlation between these two input variables is hardly surprising. Again, it is still reasonable to consider both variables in a multivariate analysis since by using only the ΔR average variable less information could be exploitable than by using both variables.

11.3.2 Influence of strongly correlated input variables on the $t\bar{t} + \geq 1\text{b-jet}$ signal strength

To study which influence strongly correlated input variables could have on the $t\bar{t} + \geq 1\text{b-jet}$ signal strength result r , the b-tag. LR and trans. b-tag. LR input variables are chosen. They show the largest correlation of all 11 input variables and also have the largest separation power between signal and background of all input variables according to the input variable ranking (table 9.2). Therefore, it is expected that the correlation of this input-variable pair will affect the overall $t\bar{t} + \geq 1\text{b-jet}$ signal strength result the most.

In a first step the b-tag. LR input variable was removed from the set of input variables used in the LD and NN based analyses, so that these two analyses made use of only 10 input variables instead of the 11 input variables applied for their corresponding reference analyses. The comparison of the reference analyses with the analyses not containing the b-tag. LR input variable is given in table 11.3. Besides the expected $t\bar{t} + \geq 1\text{b-jet}$ signal strength

$r(\text{expected})$ this table also states the ROC integral values derived by the associated analysis using a validation dataset. The signal strength result r for the Asimov dataset is presented since it provides a better comparability between the individual analyses. Moreover, the ROC integrals of the binary classifying LD analysis and the multiclassifying NN analysis cannot be directly compared (a discussion of these differences is part of the next section 11.4). Only a comparison of the ROC values of the same type of analysis is sensible.

Table 11.3: Influence of strongly correlated input variables on the $t\bar{t}+\geq 1b$ -jet signal strength using the b-tagging likelihood variables as an example: The expected $t\bar{t}+\geq 1b$ -jet signal strength $r(\text{expected})$ is stated for the Asimov dataset as well as the ROC integral value. The comparison includes the baseline linear discriminant (LD) and neural network (NN) based analyses and different LD and NN based analyses, in which the b-tag. LR variable was not considered as an input variable (10 vars). Furthermore, NN analyses with one (3HL) or two (4HL) additional hidden layers are presented.

Analysis	Parameters	$r(\text{expected})$	ROC integral value
LD reference	11 vars	$1^{+24.30\%}_{-17.64\%}$	0.7352
NN reference	11 vars, 2 HL, 100 nodes	$1^{+22.88\%}_{-17.13\%}$	0.6701
LD	10 vars	$1^{+23.81\%}_{-17.77\%}$	0.7354
NN	10 vars, 2HL, 100 nodes	$1^{+26.10\%}_{-18.61\%}$	0.6698
NN	10 vars, 3HL, 100 nodes	$1^{+22.85\%}_{-17.15\%}$	0.6698
NN	10 vars, 4HL, 100 nodes	$1^{+67.05\%}_{-20.55\%}$	0.6696

As can be seen from table 11.3, the ROC integral values for the two LD analyses as well as for the four NN analyses are nearly identical.

In the case of the NN analysis the relation between ROC integral value and expected signal strength $r(\text{expected})$ is not so clear: Even though the reference neural network considering all 11 input variables and consisting of two hidden layers with 100 nodes per layer has the largest ROC integral of 0.6701, which is one of the reasons why this network is chosen as the reference network in the first place. The neural network considering only 10 variables, but having three hidden layers shows a slightly better expected signal strength $r(\text{expected})$, while having a lower ROC integral of 0.6698. It will be shown later in section 11.5 that these difference in the expected signal strength $r(\text{expected})$ is fully covered by statistical fluctuations due to the chosen random seed for the training of a neural network with a given network topology. Nonetheless, the difference between the ROC integrals and still compatible performance of both neural networks is interesting. However, a proper explanation is hard to give. Most likely, the neural network with three hidden layers performs a little bit better in a phase-space region which allows the Maximum-Likelihood to constrain the signal strength result better than the reference network, while being worse in a not so important phase-space region. This could be a possible reason why it has a decreased ROC integral.

Moreover, one can conclude from the above comparison that the neural network needs an additional hidden layer, if it is only supplied with 10 input variables, to obtain the same accuracy as the reference neural network. This can be explained by the strong correlation between the b-tag. LR variable, which is removed in this study, and the transformed b-tag. LR variable as well as the difference of this two input variables due to a non-linear transformation. By adding an additional hidden layer the neural network has the possibility to do further non-linear transformations of input variables, since the activation function in the neurons is non-linear itself. In this special case it seems highly likely that the neural network uses the additional hidden layer to reconstruct the missing b-tag. LR variable in some way, so that it also can exploit its information. For this reason, the neural network containing of 10 input variables and having only two hidden layers performs worse than the reference network and the neural network with three hidden layers as well as the LD analyses.

In this context it is interesting that the neural network considering 10 input variables and consisting of two additional hidden layers performs also worse compared to the NN reference analysis. However, the major cause of this behaviour can be seen by the more challenging training of a neural network with more hidden layers and also most likely be attributed to the lack of Monte Carlo sample statistics to sufficiently train a neural network with four hidden layers. Especially, the second point gets supported by the slightly worse ROC integral of this neural network.

The lack of Monte Carlo statistics could also be the cause why the neural network considering 10 variables and consisting of only 2 hidden layers performs not as good as the reference neural network or the much simpler LD analyses. In general, one would expect that a neural network with at least one hidden layer and one node in this layer performs as good as a linear discriminant. But the picture here is more complicated since multiclassifying neural networks are compared to binary classifying linear discriminants. Due to the multiclassification the neural networks are suffering more from limited statistics of the Monte Carlo samples than the linear discriminants. Therefore, it seems highly likely that the lack of the b-tag. LR variable as input variable cannot be compensated by a better trained neural network consisting of two hidden layers alone due to the training samples not being large enough. In comparison to that, the neural network with three hidden layers should require smaller training samples since it can far easier reconstruct the missing information of the b-tag. LR variable due to its additional hidden layer.

In the next section the difference between a binary and multiclassifying neural network will be studied in more depth to underpin these statements.

11.4 Influence of binary or multiclassification on the $t\bar{t}+\geq 1b$ -jet signal strength result

In the previous section and in table 11.3 it was observed that the binary classifying linear discriminant based (LD) reference analysis and the neural network (NN) based reference analysis making use of a multiclassification have quite different ROC integral values of 0.7352 and 0.6701, respectively. In this section the assumption will be checked that this difference in the ROC integral is due to the fact that multiclassification is more challenging than binary classification.

Since the linear discriminant method implemented in the TMVA toolkit for Multivariate Analysis [120] does not allow to conduct a multiclassification, binary classifying neural networks similar to the reference neural network and considering all 11 input variables are constructed for a comparison with the LD reference analysis. As in the case of the LD analyses these neural networks are trained to provide a binary classification between the $t\bar{t}+\geq 1b$ -jet ($t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$) signal processes and the $t\bar{t}+c\bar{c}$, $t\bar{t}+lf$, $t\bar{t}H$ background processes. The obtained expected signal strength result $r(\text{expected})$ as well as the corresponding ROC integral value of these neural networks is given in table 11.4 and can be compared to the LD and NN reference results.

It is expected that in the case of binary classification the presented NN analysis with one node in one hidden layer has a ROC integral value compatible to the LD reference analyses. Furthermore, it should derive a similar or slightly better expected signal strength result $r(\text{expected})$. The result is expected to improve for the NN analysis due to the usage of a non-linear activation function in the single node, while the linear discrimination only exploits linear relations.

Results obtained with binary neural networks having more nodes or more hidden layers should always be better, but are not the major scope of this study. Nonetheless, a result from a neural network consisting of 100 nodes in one hidden layer is given as an example.

The expectations stated above prove true as can be seen from table 11.4. If a binary classifying neural network with a single node in one hidden layer is used, a value of the ROC integral equivalent to the ROC integral of the LD reference analysis can be obtained. The small difference in the ROC integral value most likely comes from numerical differences between the two different multivariate analysis frameworks TensorFlow and TMVA. Furthermore, the result of the expected signal strength $r(\text{expected})$ of this neural network has slightly smaller uncertainties, which can be explained by the exploitation of non-linear relations of the input variables by the neural network. But overall one can say that both analyses expect the same uncertainties on the signal strength r .

Moreover, as also expected, the binary classifying neural network with 100 nodes in a single hidden layer performs better than the simpler neural network and the LD reference analyses: The better ROC integral value leads to an expected signal strength result having slightly smaller uncertainties.

The rather large difference in the ROC integral value between the multiclassifying neural network of the reference analysis and the binary classifying neural networks can be ex-

Table 11.4: Comparison of binary classifying neural networks with the linear discriminant and neural network based reference analyses: The expected $t\bar{t}+\geq 1b$ -jet signal strength $r(\text{expected})$ is stated as well as the ROC integral value for two different neural network (NN) based analyses, in which the number of nodes in the one hidden layer is varied. The binary neural networks are trained to classify between the $t\bar{t}+\geq 1b$ -jet ($t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$) signal processes and the $t\bar{t}+c\bar{c}$, $t\bar{t}+lf$, $t\bar{t}H$ background processes. The results are compared to the binary classifying LD reference analysis and the multiclassifying NN reference analysis.

Analysis	Parameters	$r(\text{expected})$	ROC integral value
LD reference	11 vars	$1^{+24.30\%}_{-17.64\%}$	0.7352
NN reference	11 vars, 2 HL, 100 nodes	$1^{+22.88\%}_{-17.13\%}$	0.6701
Binary neural networks: $t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$ vs. $t\bar{t}+c\bar{c}$, $t\bar{t}+lf$, $t\bar{t}H$			
NN	11 vars, 1 HL, 1 nodes	$1^{+23.52\%}_{-17.44\%}$	0.7349
NN	11 vars, 1 HL, 100 nodes	$1^{+22.56\%}_{-17.34\%}$	0.7413

plained by stronger requirements in the case of the multiclassification to obtain the same ROC integral value. Here, a signal (or background) event not only needs to be properly attributed to the signal (background) event class, but it is also required that a signal event like a $t\bar{t}+b\bar{b}$ event is put in the proper corresponding $t\bar{t}+b\bar{b}$ signal event class. In the likely case that a $t\bar{t}+b\bar{b}$ event is wrongly assigned to the $t\bar{t}+2b$ or $t\bar{t}+b$ event class, the ROC integral of the multiclassification analysis is reduced, while it stays the same for the binary classification analysis, in which such an event would still be properly assigned.

11.5 Influence of neural network training on the $t\bar{t}+\geq 1b$ -jet signal strength result

For an estimation of how much the training of a neural network (NN) influences the obtained $t\bar{t}+\geq 1b$ -jet signal strength result and its related uncertainty, four additional neural networks, which had exactly the same configuration and network topology as the neural network used in the NN reference analysis, are trained. The only difference between these neural networks as well as the original reference neural network is the chosen random seed at the beginning of the training.

In table 11.5 the expected and observed signal strength r as well as the value of the ROC integral is given for each of the neural network based analyses. In addition, the arithmetic mean is applied to calculate the average expected and observed signal strength as well as the related uncertainties and the average ROC integral value. Moreover, by using a linear regression implemented in the ROOT data analysis framework [214] the five signal strength results are combined to common expected and observed signal strength results $r(\text{expected})$ and $r(\text{observed})$, respectively. However, since no correlations between the individual results are considered the uncertainties of these two combined signal strength results are certainly underestimated.

Nevertheless, the linear regression result can serve as a reference point to estimate the influence of the training on the obtained $t\bar{t}+\geq 1b$ -jet signal strength result.

Table 11.5: Influence of neural network training on the $t\bar{t}+\geq 1b$ -jet signal strength result: The expected $t\bar{t}+\geq 1b$ -jet signal strength $r(\text{expected})$ and the observed signal strength $r(\text{observed})$ is stated as well as the ROC integral value for a set of neural network (NN) based analyses. The only difference between these neural network analyses is the random seed chosen for the training of the network. In addition, the arithmetic mean is applied to calculate the average signal strengths and an average ROC integral. Furthermore, common results obtained from a linear regression are presented.

Analysis	$r(\text{expected})$	$r(\text{observed})$	ROC integral value
NN reference	$1^{+22.88\%}_{-17.13\%}$	$1.20^{+22.75\%}_{-17.35\%}$	0.6701
NN 2nd trial	$1^{+22.09\%}_{-17.49\%}$	$1.21^{+19.29\%}_{-15.67\%}$	0.6703
NN 3rd trial	$1^{+22.88\%}_{-17.39\%}$	$1.21^{+18.80\%}_{-15.49\%}$	0.6695
NN 4nd trial	$1^{+22.43\%}_{-17.16\%}$	$1.27^{+20.34\%}_{-15.06\%}$	0.6692
NN 5nd trial	$1^{+22.99\%}_{-17.54\%}$	$1.28^{+25.09\%}_{-15.09\%}$	0.6698
Arithmic mean	$1^{+22.7\%}_{-17.3}$	$1.23^{+21.3\%}_{-15.7\%}$	0.6698
Linear regression	$1^{+10.1\%}_{-7.8\%}$	$1.25^{+8.4\%}_{-7.5\%}$	-

From the results stated in table 11.5 one can conclude that the chosen random seed in the training of a neural network and the training itself can have a non-negligible influence on the obtained signal strength results r . After the comparison with the other

NN results the NN reference analysis appears to predict a rather low observed signal strength $r(\text{observed}) = 1.20$, while the common observed signal strength seems to be rather $r(\text{observed}) = 1.25$ according to the result of the linear regression.

Furthermore, it is striking that the value of the ROC integral does neither indicate if a larger or smaller signal strength result r will be obtained at the end of the full analysis chain nor can be used to estimate the uncertainties related to the final signal strength results r . For example, a comparison between the 2nd trial and the 4th trial shows a rather large difference in the ROC integral of 0.6703 and 0.6692, respectively. However, the neural network of the 4th trial obtains a signal strength result with smaller relative uncertainties than the reference neural network. Also a comparison between the 3rd trial and the 5th trial which differ in the absolute value of the observed signal strength $r(\text{observed})$ and a comparison of their ROC integral to other trials does not provide further insight.

Most likely both effects can be ascribed to the finding of local minima by the particular neural network instead of a global minimum in the phase-space of the separation problem. Since the phase-space of the selected events itself and also the corresponding multiclassification is a multidimensional and highly complex problem it is very unlikely that such a global minimum is found and quite expected that nearby minima are found and lead to slightly different signal strength results r .

According to the findings presented in table 11.5 and the difference in the expected and observed signal strength r it seems reasonable to assume an intrinsic uncertainty on the $t\bar{t} + \geq 1\text{b-jet}$ signal strength results of roughly 7% – 10% which is caused by the influence of the neural network training and which cannot be considered as a systematic uncertainty in the Maximum-Likelihood fit. Here, for this intrinsic uncertainty the results from the linear regression are used as a first estimate. But as mentioned briefly before, the intrinsic uncertainty could be larger since in the linear regression no correlations between the particular results are considered, while in principle the results should be highly correlated. For a cross check, this intrinsic uncertainty can also be estimated by calculation of the standard deviation σ for the arithmetic mean of all five NN analyses. This standard deviation σ corresponds to $\sigma = 3.78\%$. Therefore, the assumption that the intrinsic uncertainty on the $t\bar{t} + \geq 1\text{b-jet}$ signal strength is smaller than 10% appears plausible. Meanwhile, since the overall uncertainties on the $t\bar{t} + \geq 1\text{b-jet}$ signal strength result r are quite larger this intrinsic uncertainty is fully covered by the overall uncertainty.

11.6 Influence of including the $t\bar{t}H$ process as a background process in the training of neural networks

For the linear discriminant and neural network based reference analyses the $t\bar{t}H$ process was always included as a background process in the training of the corresponding classifier. It was also considered as a background process in all other analyses shown previously. Furthermore, each multiclassifying neural network contained a $t\bar{t}H$ output node until now.

This inclusion of the $t\bar{t}H$ process is motivated by the possibility to conduct a simultaneous measurement of the $t\bar{t} + \geq 1\text{b-jet}$ signal strength and the $t\bar{t}H$ signal strength r (part V). So far, in the present $t\bar{t}H(b\bar{b})$ measurements [114,115] large uncertainties on the $t\bar{t} + \geq 1\text{b-jet}$ contributions have to be considered. The reason for this is that the exact size of the $t\bar{t} + \geq 1\text{b-jet}$ contribution to the phase space of the $t\bar{t}H(b\bar{b})$ measurement is difficult to predict by Monte Carlo samples: As explained in the motivation of this measurement

(part II) different $t\bar{t}+\geq 1b\text{-jet}$ as well as $t\bar{t}+b\bar{b}$ cross sections σ are predicted by different Monte Carlo event generators. Furthermore, the measurement of the $t\bar{t}H(b\bar{b})$ process as well as this measurement is conducted in a phase space which contains events with rather large jet multiplicities and a few b -tagged jets. Usually the proper description of such a phase space represents a challenge for Monte Carlo event generators. The reason for this is that besides a proper simulation of the hard collision process by a matrix element generator the radiation of a few additional partons must be performed correctly by the subsequent parton shower of a general-purpose Monte Carlo event generator. This necessary interplay between the two classes of event generators places hard demands on their proper setup and tune which are difficult to fulfill in practice.

Therefore, from the perspective of a $t\bar{t}H(b\bar{b})$ analysis a simultaneous measurement of both signal strengths ($t\bar{t}+\geq 1b\text{-jet}$ and $t\bar{t}H$) can be beneficial: Since the $t\bar{t}+\geq 1b\text{-jet}$ background and the $t\bar{t}H$ signal would be determined simultaneously the uncertainty assumed for the $t\bar{t}+\geq 1b\text{-jet}$ background in the $t\bar{t}H(b\bar{b})$ analysis could be smaller. The reason for this is that by leaving the $t\bar{t}+\geq 1b\text{-jet}$ signal strength freely floating the uncertainty on the $t\bar{t}+\geq 1b\text{-jet}$ contribution would most likely be stronger constrained than if a fixed prediction of the $t\bar{t}+\geq 1b\text{-jet}$ background is employed. Therefore, it can be assumed that the uncertainty of the $t\bar{t}H$ signal strength is reduced if the $t\bar{t}+\geq 1b\text{-jet}$ signal strength is measured in parallel.

In contrast to that, it is assumed that a simultaneous measurement only leads to a slight improvement of the uncertainty of the measured $t\bar{t}+\geq 1b\text{-jet}$ signal strength if at all. The reason is that $t\bar{t}H$ process is only a minor background to a $t\bar{t}+\geq 1b\text{-jet}$ measurement. Moreover, one can also argue that compared to other background processes like the $t\bar{t}V$ processes the $t\bar{t}H$ process has an insignificant impact on the $t\bar{t}+\geq 1b\text{-jet}$ measurement. Therefore, constraining the $t\bar{t}H$ process better most likely will have no effect on the uncertainty of the $t\bar{t}+\geq 1b\text{-jet}$ signal strength. For this reason it may be considered that the $t\bar{t}H$ process is not included as a background process during the training of the classifier of a multivariate analyses. In this line of thought, a separate process class in a multiclassifying neural network for the $t\bar{t}H$ process does also not appear to be strictly necessary.

Furthermore, the inclusion of the $t\bar{t}H$ process in the training of a multivariate classifier can even have a negative impact on the performance of a classifier as explained in section 9.3: For the training of classifiers employed in this thesis all process classes are treated equally. Therefore, the neural network may not learn to properly distinguish $t\bar{t}+\geq 1b\text{-jet}$ signal events from other events since it could lead to a reduction of the identification of $t\bar{t}H$ events at the same time.

In a similar context, it can be said that a multiclassification consisting of fewer classes can be beneficial. The proper assignment of events to a process class becomes more likely if fewer process classes and output nodes have to be considered.

This is motivated by the following considerations: Under the assumption that the neural network does not know anything about the process classes and just randomly assigns an event to an process class, it will assign one out of six events correctly if the neural network contains six output nodes. However, if such a neural network would contain only five process classes and corresponding output nodes 20% of all events should be properly assigned but sensitivity may not depend (strongly) on this.

On the other hand, an additional process class and corresponding final output distribution can provide extra information which can be employed by the Maximum-Likelihood fit to

obtain the $t\bar{t}+\geq 1\text{b-jet}$ signal strength.

For the reasons outlined above, it may be relevant to study the influence of including the $t\bar{t}H$ process as a background process in the training of neural networks. Therefore, it is now studied in detail:

A multiclassifying neural network analogous to the neural network of the NN reference analysis was defined and trained. The only difference to the NN reference analysis and network is that the $t\bar{t}H$ process is not included as a background process in the training and the neural network does not contain a $t\bar{t}H$ output node.

In addition, a binary classifying neural network was trained, in which the $t\bar{t}H$ process was neglected during the training. Apart from that, this neural network corresponds to the binary classifying neural network which was used for a comparison with the LD reference analysis (section 11.4).

In table 11.6 the expected $t\bar{t}+\geq 1\text{b-jet}$ signal strength $r(\text{expected})$ values obtained from these analyses are stated. Furthermore, the results of the LD and NN reference analyses are given for comparison. In the case of the stated NN reference analysis result the arithmetic mean was applied to combine the results of five NN analyses. All of these analyses consisted of neural networks having the same configuration and topology but different random seeds were chosen in the training of the neural network (section 11.5).

The ROC integral values of the analysis including the $t\bar{t}H$ process cannot be compared to the ROC integral values of the analyses in which the $t\bar{t}H$ process was neglected. The reason for this is that the normalization of the input variables by scaling factors can be different: Before the training this scaling factor is determined by considering the distribution of input variables of each process employed in the training. The distribution of an input variable can e.g. differ between the $t\bar{t}H$ process and the other processes. The change in the joint input variable normalization by excluding a process like the $t\bar{t}H$ process, hence, could also change the response of the neural network and lead to slight changes of the ROC integral values. Therefore, in this case ROC integral values are not a good metric for the sensitivity of an analysis and not stated here.

For the two neural networks in which the $t\bar{t}H$ process was not included, a more precise measurement of the $t\bar{t}+\geq 1\text{b-jet}$ signal strength $r(\text{expected})$ is obtained than for the reference neural networks they are compared to. For this reason, the clear conclusion can be drawn that the inclusion of the $t\bar{t}H$ process as a background process in the training and also as an output node in a multiclassifying neural network impairs the measurement of the $t\bar{t}+\geq 1\text{b-jet}$ signal strength in this particular instance. Furthermore, the binary and the multiclassifying NN analyses predict an almost identical uncertainty on the $t\bar{t}+\geq 1\text{b-jet}$ signal strength of $\Delta r(\text{expected}) \approx {}^{+21.6\%}_{-17.2\%}$, in which the arithmetic mean of the uncertainty of these two measurements is given. Therefore, it can be concluded that including the $t\bar{t}H$ process in the training reduces the sensitivity and not the additional output node in the multiclassifying neural network.

The most likely reason is that by inclusion of the $t\bar{t}H$ process in the training neural networks, will learn less about the proper assignment of $t\bar{t}+X$ events. It can be assumed that some features which would be employed by the neural network to separate $t\bar{t}+X$ process cannot be exploited since some $t\bar{t}H$ events would show the same features. This assumption is supported by the following observation: In table 11.7 the fraction of properly assigned $t\bar{t}+X$ events is compared for the two multiclassifying neural networks, the one without

Table 11.6: Influence of including the $t\bar{t}H$ process in the training of neural networks: The expected $t\bar{t}+\geq 1b$ -jet signal strength $r(\text{expected})$ is stated for different neural network (NN) based analyses. A multiclassifying neural network was trained, in which the $t\bar{t}H$ process was not included as a background process in the training and which does not contain a $t\bar{t}H$ process class. Apart from that, it corresponds to the neural network of the NN reference analysis, whose result is also given. This result was obtained by calculating the arithmetic mean of five measurements obtained with neural networks, in which training different random seeds were used. Furthermore, the $t\bar{t}+\geq 1b$ -jet signal strength result from a binary classifying neural network, in which the $t\bar{t}H$ process was not included, is presented. Except for the inclusion of the $t\bar{t}H$ process this neural network corresponds to the binary classifying neural network which was employed for a comparison with the LD reference analysis. The results of both analyses are also given.

Analysis	Parameters	$r(\text{expected})$
Reference analyses		
LD reference	11 vars	$1^{+24.3\%}_{-17.6\%}$
NN reference (arithmetic mean)	11 vars, 2 HL, 100 nodes	$1^{+22.7\%}_{-17.3\%}$
Multiclass neural network excl. $t\bar{t}H$: $t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$, $t\bar{t}+c\bar{c}$, $t\bar{t}+lf$		
NN	11 vars, 2HL, 100 nodes	$1^{+21.7\%}_{-17.3\%}$
Binary neural network incl. $t\bar{t}H$: $t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$ vs. $t\bar{t}+c\bar{c}$, $t\bar{t}+lf$, $t\bar{t}H$		
NN	11 vars, 1 HL, 1 nodes	$1^{+23.5\%}_{-17.4\%}$
Binary neural network excl. $t\bar{t}H$: $t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$ vs. $t\bar{t}+c\bar{c}$, $t\bar{t}+lf$		
NN	11 vars, 1 HL, 1 nodes	$1^{+21.5\%}_{-17.0\%}$

inclusion of the $t\bar{t}H$ process and the neural network of the NN reference analysis. These fractions were obtained by applying the neural networks to the validation dataset.

Table 11.7: Comparison of properly assigned $t\bar{t}+X$ events: The fraction of properly assigned $t\bar{t}+X$ events is compared for two multiclassifying neural networks. For the multiclassifying neural network (NN without $t\bar{t}H$) the $t\bar{t}H$ process was not included as a background process in the training and it does not contain a $t\bar{t}H$ process class, while the other neural network (reference NN) belongs to the NN reference analysis and includes the $t\bar{t}H$ process.

$t\bar{t}+X$ process	NN without $t\bar{t}H$	reference NN
$t\bar{t}+b\bar{b}$	44.3 %	24.5 %
$t\bar{t}+2b$	30.6 %	26.1 %
$t\bar{t}+b$	12.7 %	10.5 %
$t\bar{t}+c\bar{c}$	21.3 %	19.3 %
$t\bar{t}+lf$	59.1 %	59.4 %

As expected the inclusion of the $t\bar{t}H$ process in the training drastically reduces the number of properly assigned $t\bar{t}+b\bar{b}$ events. The reason can be seen by the two processes having similar features as discussed previously in detail. For the same reason, the proper assignment of $t\bar{t}+2b$ events is also significantly affected by the inclusion of the $t\bar{t}H$ process. The assignment of $t\bar{t}+b$ and $t\bar{t}+c\bar{c}$ events is only slightly affected, while the $t\bar{t}+lf$ assignment is not affected at all. It even seems that the $t\bar{t}+lf$ assignment gets slightly improved, but this could be just an effect of statistical fluctuations and would need further study.

It can be summarized that by including the $t\bar{t}H$ process in the training of neural network classifiers the expected precision of the $t\bar{t}+\geq 1b$ -jet signal strength decreases. An NN analysis without inclusion of the $t\bar{t}H$ process would most likely yield a $t\bar{t}+\geq 1b$ -jet signal strength with smaller uncertainties. Nevertheless, for the NN reference analysis and corresponding neural network the $t\bar{t}H$ process was still considered. The reason is that this NN reference analysis was also employed to conduct a simultaneous measurement of the $t\bar{t}+\geq 1b$ -jet and $t\bar{t}H$ signal strength in a further analysis (part V). In this way, the $t\bar{t}+\geq 1b$ -jet signal strength of both analysis can be easily compared.

11.7 $t\bar{t}+\geq 1b$ -jet cross section

So far, the result of this analysis was always stated as the $t\bar{t}+\geq 1b$ -jet signal strength r . The signal strength is a very useful quantity if a prediction based on Monte Carlo samples should be compared to measurement data. Therefore, it is often employed in experimental particle physics. However, for a comparison between different measurements which may rely on different Monte Carlo predictions the specification of a cross section σ is more appropriate. In addition, cross sections σ are frequently used in theoretical particle physics, while signal strengths r are seldomly employed and are only meaningful relative to the cross section they are normalized to.

To provide easy access to the $t\bar{t}+\geq 1b$ -jet measurement, the $t\bar{t}+\geq 1b$ -jet signal strength r result of the NN reference analysis is now employed to determine the associated $t\bar{t}+\geq 1b$ -jet cross section $\sigma(t\bar{t}+\geq 1b\text{-jet})$ in the full phase space and at a center-of-mass energy of 13 TeV.

A visible cross section is not derived since a reproduction of the visible phase space would be a challenging task for others. The reason for this is that the definition of the $t\bar{t} + \geq 1b$ -jet processes strongly depends on the GenHfHadronMatcher which is employed to split the inclusive POWHEG Box Version 2 $t\bar{t}$ sample into $t\bar{t} + X$ sub samples (section 6.4). Therefore, interested parties would principally need to implement the GenHfHadronMatcher to derive the same visible phase space. It seems highly unlikely that someone is willing to take this trouble. Especially since the signal strength result is provided which could be easily compared to an result obtained in a similar phase space.

For the purpose of determining the $t\bar{t} + \geq 1b$ -jet cross section in the full phase space $\sigma(t\bar{t} + \geq 1b\text{-jet})$ first of all the $t\bar{t} + \geq 1b$ -jet cross section $\sigma(t\bar{t} + \geq 1b\text{-jet, pred.})$ predicted by the $t\bar{t}$ samples was determined: The $t\bar{t}$ samples employed in this analysis were produced by a combination of the POWHEG Box Version 2 [30–34] matrix element generator and the Pythia8 [37, 38] general-purpose Monte Carlo event generator (section 6.3). By application of the GenHfHadronMatcher [146, 149] these $t\bar{t}$ samples were split into $t\bar{t} + X$ sub samples (section 6.4). A total cross section at next-to-next-leading order (NNLO) QCD accuracy with next-to-next-leading logarithmic (NNLL) soft gluon resummation was only available for the inclusive $t\bar{t}$ production process (section 6.2). It is for this reason that this cross section of 831.76 pb had to be used to obtain the $t\bar{t} + \geq 1b$ -jet cross section $\sigma(t\bar{t} + \geq 1b\text{-jet, pred.})$. Therefore, the number of events per $t\bar{t} + X$ category in the inclusive POWHEG Box Version 2 $t\bar{t}$ sample was determined. In total roughly 77 million events were analyzed and classified by the GenHfHadronMatcher, while no further cuts were applied. In table 11.8 the number of events per $t\bar{t} + X$ category is listed. All samples are inclusive in the $t\bar{t}$ decay channels.

The number of $t\bar{t} + \geq 1b$ -jet events $n_{t\bar{t} + \geq 1b\text{-jet}}$ was calculated by summing the number of $t\bar{t} + b\bar{b}$ events $n_{t\bar{t} + b\bar{b}}$, $t\bar{t} + 2b$ events $n_{t\bar{t} + 2b}$, and $t\bar{t} + b$ events $n_{t\bar{t} + b}$:

$$n_{t\bar{t} + \geq 1b\text{-jet}} = n_{t\bar{t} + b\bar{b}} + n_{t\bar{t} + 2b} + n_{t\bar{t} + b} = 2\,218\,801$$

The cross section of the $t\bar{t} + \geq 1b$ -jet processes $\sigma(t\bar{t} + \geq 1b\text{-jet, pred.})$ predicted by the POWHEG Box Version 2 event generator is then calculated by making use of the number of $t\bar{t} + \geq 1b$ -jet events $n_{t\bar{t} + \geq 1b\text{-jet}}$, the total number of $t\bar{t}$ events $n_{t\bar{t}} = 76\,707\,098$, and the aforementioned inclusive $t\bar{t}$ cross section $\sigma(t\bar{t}) = 831.76$ pb:

$$\sigma(t\bar{t} + \geq 1b\text{-jet, pred.}) = \frac{n_{t\bar{t} + \geq 1b\text{-jet}}}{n_{t\bar{t}}} \cdot \sigma(t\bar{t}) \approx 24.06 \text{ pb}$$

In the same fashion, the predicted cross sections $\sigma(\text{pred.})$ of the single $t\bar{t} + b\bar{b}$, $t\bar{t} + 2b$, and $t\bar{t} + b$ process are derived. All four predicted cross sections $\sigma(\text{pred.})$ are also listed in table 11.8.

The $t\bar{t} + \geq 1b$ -jet cross section in the full phase space and at a center-of-mass energy of 13 TeV, $\sigma(t\bar{t} + \geq 1b\text{-jet, obs.})$, is then determined by multiplying the $t\bar{t} + \geq 1b$ -jet signal strength $r(\text{observed})$ measured by the NN reference analysis and its uncertainties by the predicted $t\bar{t} + \geq 1b$ -jet cross section, $\sigma(t\bar{t} + \geq 1b\text{-jet, pred.})$:

$$\begin{aligned} \sigma(t\bar{t} + \geq 1b\text{-jet, obs.}) &= r(\text{observed}) \cdot \sigma(t\bar{t} + \geq 1b\text{-jet, pred.}) \\ &= 1.23^{+21.3\%}_{-15.7\%} \cdot 24.06 \text{ pb} \\ &\approx 29.6^{+6.3}_{-4.6} \text{ pb} \end{aligned}$$

Here, for the value of the $t\bar{t} + \geq 1b$ -jet signal strength $r(\text{observed})$ the arithmetic mean of five measurements was employed. For these measurements neural networks were used, which had the same setup and topology as the NN reference network, but had different

Table 11.8: Number of $t\bar{t}+X$ events in the inclusive POWHEG Box Version 2 $t\bar{t}$ sample and predicted cross sections $\sigma(\text{pred.})$: The number of events per $t\bar{t}+X$ category is stated in this tables. It was determined by analyzing the inclusive POWHEG Box Version 2 $t\bar{t}$ sample. The classification of the events was made by the GenHfHadronMatcher, while no further cuts were applied. In addition, the total number of events in the $t\bar{t}$ sample is given. Furthermore, the cross section $\sigma(\text{pred.})$ of the $t\bar{t}+\geq 1\text{b-jet}$ processes as well as the cross section of the single $t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, and $t\bar{t}+b$ process predicted by the POWHEG Box Version 2 event generator is stated. For the $t\bar{t}$ process the cross section at NNLO QCD and NNLL accuracy is listed.

$t\bar{t}+X$ category	Number of events	$\sigma(\text{pred.})$
$t\bar{t}+b\bar{b}$	384 253	4.16 pb
$t\bar{t}+2b$	427 108	4.63 pb
$t\bar{t}+b$	1 407 440	15.26 pb
$t\bar{t}+\geq 1\text{b-jet}$	2 218 801	24.06 pb
$t\bar{t}+c\bar{c}$	7 211 374	-
$t\bar{t}+lf$	67 276 914	-
$t\bar{t}$	76 707 089	831.76 pb

random seeds in the training (section 11.5).

Unfortunately, the obtained $t\bar{t}+\geq 1\text{b-jet}$ cross section $\sigma(t\bar{t}+\geq 1\text{b-jet, obs.})$ of $29.6^{+6.3}_{-4.6}$ pb cannot be compared to another $t\bar{t}+\geq 1\text{b-jet}$ measurement since no such measurement conducted at a center-of-mass energy of 13 TeV and stating a $t\bar{t}+\geq 1\text{b-jet}$ cross section $\sigma(t\bar{t}+\geq 1\text{b-jet})$ exists so far.

12. Discussion of results

In the previous chapter the results of the measurement of the $t\bar{t}+\geq 1\text{b-jet}$ signal strength r using the 2016 CMS data together with further detailed studies were presented.

It was found that the three reference analyses (B-jet multiplicity based, linear discriminant based, neural network based) are in good agreement with each other if one takes the uncertainty on the signal strength result Δr into account. The uncertainty on the expected signal strength $r(\text{expected})$ ranges between $\Delta r(\text{expected}) = {}^{+33\%}_{-23\%}$ in the case of the B-jet multiplicity based (B-jet mult.) analysis having the largest uncertainties and $\Delta r(\text{expected}) = {}^{+23\%}_{-17\%}$ in the case of the neural network based (NN) analysis being the most precise one.

Furthermore, it can be observed that for all three analyses the observed $t\bar{t}+\geq 1\text{b-jet}$ signal strength $r(\text{observed})$ is larger than the signal strength predicted by the combination of the POWHEG Box Version 2 matrix element generator and the Pythia8 general-purpose Monte Carlo event generator. The linear discriminant based (LD) and NN analysis observe a similar signal strength of $r(\text{observed}) \approx 1.2$, while the B-jet mult. analysis observes a smaller signal strength of $r(\text{observed}) \approx 1.1$. It is assumed that the smaller observed signal strength of the B-jet mult. based analysis can be attributed to the simpler final discriminant and the reduced separation power between $t\bar{t}+\geq 1\text{b-jet}$ signal processes and $t\bar{t}+X$ background processes.

Nonetheless, all results are consistent. In the following the signal strength $r(\text{observed})$ of $1.20^{+23\%}_{-17\%}$ measured by the NN reference analysis is taken as main result. This result is strongly supported by the two other analyses acting as baseline analysis (B-jet mult.) and cross check analysis (LD).

The observed signal strength of $r(\text{observed}) \approx 1.2$ is also in good agreement with other $t\bar{t}+\geq 1\text{b-jet}$ measurements: As reviewed in chapter III, two measurements were conducted at the Large Hadron Collider at a center-of-mass energy of 8 TeV, which contained a measurement of the $t\bar{t}+\geq 1\text{b-jet}$ signal strength r , too. Both measurements, one from the ATLAS collaboration using dileptonic and semileptonic events and one from the CMS collaboration using only dileptonic events, observed a signal strength $r(\text{observed})$ which was roughly 1.3 times larger than the signal strength predicted by either the POWHEG Box Version 1 matrix element generator or the MG5aMC(LO) matrix element generator.

The overall uncertainty of these two $t\bar{t}+\geq 1\text{b-jet}$ signal strength measurements is roughly $\Delta r(\text{expected}) \approx 25\% - 30\%$ and, therefore, only slightly larger than the overall uncertainty expected for the analyses presented in this thesis. The reason for the similar precision, despite more luminosity was employed in this measurement, is that the past measurements were conducted using the dileptonic channel (CMS measurement) or using both the dileptonic and semileptonic channel (ATLAS measurement). The systematic uncertainties in the case of the dileptonic measurements are smaller. Therefore, these measurements are less driven by systematic uncertainties and can provide a similarly precise result, while making use of less luminosity.

As the measurement presented in this thesis the ATLAS measurement conducted in the semileptonic channel is plagued by rather large systematic uncertainties compared to the dileptonic measurements. The dominant systematic uncertainties of the semileptonic measurements are related to the b-tagging of jets and to the usage of the parton shower which is quite expected as explained previously. But also the two reviewed dileptonic measurements suffer from similar further major systematic uncertainties like uncertainties related to the usage of Monte Carlo event generators and the modelling of the $t\bar{t}$ processes or the uncertainty on the jet-energy scale.

For this reason, considering the overall uncertainty on the signal strength Δr , it can be concluded that the results of the measurement presented here are consistent with the past two measurements.

This finding is of particular importance since for the two reviewed measurements the older version 1 of the next-to-leading order POWHEG matrix element generator interfaced with the Pythia6 general-purpose Monte Carlo event generator was applied to produce the nominal $t\bar{t}$ samples. Additionally, in the past CMS $t\bar{t}+\geq 1\text{b-jet}$ measurement further $t\bar{t}$ samples were considered for a comparison to the nominal one. For example, among other event generators, the MG5aMC(LO) matrix element generator was used to produced a $t\bar{t}$ event sample at leading-order of perturbation accuracy, in which merging was employed to include matrix elements with up to three additional partons.

Compared to the past measurements, for this measurement $t\bar{t}$ samples are produced in next-to-leading order of perturbation theory by the newer version 2 of the POWHEG generator interfaced to the Pythia8 event generator. In principle, it would be expected that the newer Monte Carlo event generators and associated $t\bar{t}$ samples provide a more accurate modeling of the $t\bar{t}+X$ and, particularly, the $t\bar{t}+\geq 1\text{b-jet}$ processes.

However, independent of the different matrix element generators and parton showers applied, all three measurements find that the deviation of $t\bar{t}+\geq 1\text{b-jet}$ signal strength from one is sizeable and underestimated by (roughly) the same factor by the event generators. This could be seen as a sign that the modelling of the heavy-flavour production by Monte Carlo event generators is inadequate up to now.

Besides the maybe delicate modeling of the heavy-flavour production by the Monte Carlo event generators all three $t\bar{t}+\geq 1\text{b-jet}$ measurements would gain from reduced systematic uncertainties related to the parton shower and the Monte Carlo event generators in general as well as from smaller systematic uncertainties due to the b-tagging of jets. At least the measurement presented here (and the ATLAS semileptonic measurement) are driven by these two kinds of dominant systematic uncertainties, as shown in section 11.2.

For the reduction of the b-tagging systematic uncertainties newer b-tagging algorithms like the DeepCSV b-tagging algorithm, established by the CMS collaboration during the 2017 data taking period, [90] seem promising.

In the case of the parton shower systematic uncertainties it is harder to predict how fast and by how much these uncertainties can be further reduced. Their reduction would either

require larger additional $t\bar{t}$ samples allowing for the in-depth study of shape changes in the final discriminants due to the chosen parton shower parameters or another treatment and propagation of these uncertainties by the applied Monte Carlo event generator: The production of even larger additional samples than the already existing ones, from which each contains about 100 million events and which are employed to take parton shower variations into account (section 10.1.6), seems not very feasible due to the constraints on the overall event sample production in a large collaboration like ATLAS or CMS.

Another possible option would be to rewrite general-purpose Monte Carlo event generators, so that for considering parton shower variations no additional samples have to be produced. Instead parton shower variations could be included in the nominal $t\bar{t}$ samples. In this way, only large nominal $t\bar{t}$ samples must be produced. However, including parton shower uncertainties as part of nominal $t\bar{t}$ samples could require some difficult changes in the established general-purpose Monte Carlo event generators.

Hence, it appears quite likely that for the near future further measurements of the $t\bar{t} + \geq 1b$ -jet processes will mostly be conducted using the dileptonic decay channel, in which the uncertainties related to the b-tagging of jets and to the parton shower are less important. That is also the reason why $t\bar{t} + \geq 1b$ -jet measurements in the dileptonic decay channel are less driven by systematic uncertainties so far.

Correlations between input variables could per se have an influence on the separation power of a multivariate analysis and, hence, on the overall $t\bar{t} + \geq 1b$ -jet signal strength r measurement. For this reason, the influence of correlations between input variables for the LD and NN analysis was studied as part of this measurement. As an example the rather strong correlation of 87 % between the b-tagging likelihood ratio (B-tag. LR) and the transformed b-tagging likelihood ratio (Trans. b-tag. LR) input variable was presented and examined.

It was found that for the LD analysis the removal of the B-tag. LR input variable had a rather negligible influence on the expected uncertainty of the $t\bar{t} + \geq 1b$ -jet signal strength $\Delta r(\text{expected})$. The slightly better performance after removing the B-tag. LR variable can most likely be explained by the better numerical stability of the matrix transformation necessary to obtain the linear discriminant if one variable less has to be considered.

At the same time, the removal of the B-tag. LR input variable leads to a worse performance of the neural network which consisted of 100 nodes in two hidden layers and is constructed analogously to the reference network. In the case of just 10 input variables, only by adding a third hidden layer consisting of 100 nodes the same expected uncertainties for the $t\bar{t} + \geq 1b$ -jet signal strength $\Delta r(\text{expected})$ could be regained as for the reference neural network which exploited also the B-tag. LR input variable. This necessity of a third hidden layer in the case of 10 input variables can most likely be attributed to the intrinsic reconstruction of the missing B-tag. LR variable from the correlated Trans. b-tag. LR variable by the neural network.

Nonetheless, one can conclude from the correlation studies that the influence of correlated input variables has a rather innocuous influence on the overall $t\bar{t} + \geq 1b$ -jet signal strength result. However, the addition or removal of correlated input variables can lead to a change of the overall expected uncertainty $\Delta r(\text{expected})$ in the case of a neural network based analysis, which should be kept in mind. Furthermore, as done as part of this measurement, various neural network topologies should be tried for a given set of input variables to exploit the dependency of the overall $t\bar{t} + \geq 1b$ -jet signal strength uncertainty on the chosen network architecture.

As part of these studies on the influence of the chosen neural network, the performance of binary classifying neural networks was compared to the performance of multiclassify-

ing neural networks. The major difference between both classes of neural networks is the obtained value of the ROC integral. The value of the ROC integral is smaller for the multiclassifying neural networks, but this can be explained by stronger requirements in the case of the multiclassification. Besides of that, both approaches lead to $t\bar{t}+\geq 1\text{b-jet}$ signal strength results with a similar expected overall uncertainty $\Delta r(\text{expected})$. For this reason, the usage of multiclassifying neural networks should be favoured since in principle these networks provide more information about the possible origin of an event by sorting these into different process classes.

In another study the effect of the training of a neural network to derive an expected and observed $t\bar{t}+\geq 1\text{b-jet}$ signal strength r result was examined. It was found that the choice of the random seed for the training of the network has only a rather negligible effect on the expected uncertainty of the signal strength $\Delta r(\text{expected})$. But still sizable differences in the observed signal strength $r(\text{observed})$ and the related observed overall uncertainty $\Delta r(\text{observed})$ can occur. Furthermore, the training can lead to quite different values of the ROC integral. Most likely both effects can be ascribed to the finding of a local minimum by the corresponding neural network instead of the global minimum.

The difference in the observed signal strength results $r(\text{observed})$ is fully covered by the associated overall uncertainty $\Delta r(\text{observed})$. Nevertheless, one can assume an intrinsic uncertainty on the absolute value of the measured $t\bar{t}+\geq 1\text{b-jet}$ signal strength of roughly 7% – 10% due to influence of the neural network training.

However, this intrinsic uncertainty on the absolute value of the observed signal strength $r(\text{observed})$ can be mitigated by conducting a $t\bar{t}+\geq 1\text{b-jet}$ measurement multiple times, while using identical neural networks with different random seeds applied in the training. Out of this multiple measurements the arithmetic mean of the value of the observed $t\bar{t}+\geq 1\text{b-jet}$ signal strength $r(\text{observed})$ is then determined. At the same time, the arithmetic mean can also be applied to derive the average relative uncertainties of the observed $t\bar{t}+\geq 1\text{b-jet}$ signal strength $\Delta r(\text{observed})$.

On that basis, an expected $t\bar{t}+\geq 1\text{b-jet}$ signal strength of $r(\text{expected}) = 1_{-17\%}^{+23\%}$ and an observed $t\bar{t}+\geq 1\text{b-jet}$ signal strength of $r(\text{observed}) = 1.23_{-16\%}^{+21\%}$ is derived by the neural network based reference analysis and used from now on.

Further on, the influence of including the $t\bar{t}H$ process as a background process in the training of neural networks was studied. It was shown that if the $t\bar{t}H$ process is included in the training the expected precision of the $t\bar{t}+\geq 1\text{b-jet}$ signal strength $r(\text{expected})$ decreases. The reason for this is that by including the $t\bar{t}H$ process a neural network will learn less about the proper assignment of $t\bar{t}+X$ events since $t\bar{t}H$ events can show similar features as $t\bar{t}+b\bar{b}$ and $t\bar{t}+2b$ events. Therefore, in particular, fewer $t\bar{t}+b\bar{b}$ and $t\bar{t}+2b$ events are properly assigned if the $t\bar{t}H$ process was included in the training.

For this reason, the $t\bar{t}H$ process should be neglected in the training of a neural network if only the measurement of the $t\bar{t}+\geq 1\text{b-jet}$ signal strength is of interest. However, in the training of the neural networks of the NN reference analysis the $t\bar{t}H$ process was included since in this way a simultaneous measurement of the $t\bar{t}+\geq 1\text{b-jet}$ and $t\bar{t}H$ signal strength can be conducted later on (part V).

In the last section based on the observed $t\bar{t}+\geq 1\text{b-jet}$ signal strength $r(\text{observed})$ of $1.23_{-16\%}^{+21\%}$, measured by the NN reference analysis, the corresponding $t\bar{t}+\geq 1\text{b-jet}$ cross section $\sigma(t\bar{t}+\geq 1\text{b-jet, obs.})$ was determined. Considering the full phase space the signal strength $r(\text{observed})$ corresponds to $29.6_{-4.6}^{+6.3}$ pb. So far it is the first $t\bar{t}+\geq 1\text{b-jet}$ cross section result at a center-of-mass energy of 13 TeV.

Part V

Simultaneous measurement of individual signal strengths and cross sections

Simultaneous measurement of individual signal strengths and cross sections

In this part of the thesis, two $t\bar{t}+\geq 1\text{b-jet}$ measurements using the CMS data from the Large Hadron Collider run in 2016 will be presented.

In these measurements cross sections σ , expressed by signal strengths r , of more than one signal process are simultaneously determined. For this determination, the classifiers of the reference analyses, which were already employed for the measurement of the joint $t\bar{t}+\geq 1\text{b-jet}$ signal strength $r(t\bar{t}+\geq 1\text{b-jet})$ and cross section $\sigma(t\bar{t}+\geq 1\text{b-jet})$ (part IV), are used again.

In the first of the two measurements, the cross section of the $t\bar{t}+b\bar{b}$ process $\sigma(t\bar{t}+b\bar{b})$, the $t\bar{t}+2\text{b}$ process $\sigma(t\bar{t}+2\text{b})$, and the $t\bar{t}+b$ process $\sigma(t\bar{t}+b)$ as well as the corresponding signal strengths ($r(t\bar{t}+b\bar{b})$, $r(t\bar{t}+2\text{b})$, $r(t\bar{t}+b)$) are measured simultaneously. From now on this will be referred to as the $t\bar{t}+b\bar{b}/t\bar{t}+2\text{b}/t\bar{t}+b$ measurement.

In the second measurement, the signal strength $r(t\bar{t}+\geq 1\text{b-jet})$ and cross section $\sigma(t\bar{t}+\geq 1\text{b-jet})$ of the $t\bar{t}+\geq 1\text{b-jet}$ process is determined together with the signal strength $r(t\bar{t}H)$ and cross section $\sigma(t\bar{t}H)$ of the $t\bar{t}H$ process by a simultaneous measurement. In short form it will be designated as the $t\bar{t}+\geq 1\text{b-jet}/t\bar{t}H$ measurement.

The analyses conducted by these two measurements ($t\bar{t}+b\bar{b}/t\bar{t}+2\text{b}/t\bar{t}+b$ and $t\bar{t}+\geq 1\text{b-jet}/t\bar{t}H$) are largely identical to one another and also to the reference analyses employed for the joint $t\bar{t}+\geq 1\text{b-jet}$ measurement. Therefore, to avoid a repetition, in chapter 13 only the differences in the three measurements will be discussed.

The results of the $t\bar{t}+b\bar{b}/t\bar{t}+2\text{b}/t\bar{t}+b$ and the $t\bar{t}+\geq 1\text{b-jet}/t\bar{t}H$ measurement will be presented in chapter 14. This part closes with a discussion of the results obtained by the two measurements in chapter 15. Here, the results will also be compared to the findings of the reference analyses (part IV) and the measurements reviewed previously (part III).

13. Differences between the $t\bar{t}+\geq 1\text{b-jet}$, $t\bar{t}+b\bar{b}/t\bar{t}+2b/t\bar{t}+b$, and $t\bar{t}+\geq 1\text{b-jet}/t\bar{t}H$ signal strength and cross section measurements

All three measurements ($t\bar{t}+\geq 1\text{b-jet}$, $t\bar{t}+b\bar{b}/t\bar{t}+2b/t\bar{t}+b$, and $t\bar{t}+\geq 1\text{b-jet}/t\bar{t}H$) employ the same three analysis strategies, a B-jet multiplicity based (B-jet mult.) one, a linear discriminant based (LD) one, and an analysis based on a neural network (NN). These three analysis strategies are already discussed at length in chapter 9.

The only difference between the three measurements is the type of Maximum-Likelihood (ML) fit (section 5.1) and the signal and background definition, which is used to derive the respective signal strength r . From the signal strength r the cross section σ can easily be determined as described in section 11.7.

Due to taking only one signal into consideration, a one-dimensional ML fit is employed in the joint $t\bar{t}+\geq 1\text{b-jet}$ measurement, while for the two other measurements a multi-dimensional ML fit is used.

For all measurements the prefit final discriminant distributions (figure 9.1, fig. 9.4, and fig. 9.5) obtained from each of the three aforementioned analyses are employed in either a one-dimensional or a multi-dimensional ML fit:

In comparison to the joint $t\bar{t}+\geq 1\text{b-jet}$ signal strength $r(t\bar{t}+\geq 1\text{b-jet})$ measurement by a one-dimensional ML fit, in which the $t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, and $t\bar{t}+b$ process are considered as a combined signal; the $t\bar{t}+b\bar{b}$ process, $t\bar{t}+2b$ process, and $t\bar{t}+b$ process are defined as individual signals in the multi-dimensional ML fit of the $t\bar{t}+b\bar{b}/t\bar{t}+2b/t\bar{t}+b$ measurement. Therefore, for the three signal processes, ($t\bar{t}+b\bar{b}/t\bar{t}+2b/t\bar{t}+b$) signal strengths, $r(t\bar{t}+b\bar{b})$, $r(t\bar{t}+2b)$, and $r(t\bar{t}+b)$, are assigned.

Each of these signal strengths is set to an initial value of one. In this way, the SM prediction of the $t\bar{t}+\geq 1\text{b-jet}$ processes, which is given by the POWHEG Box Version 2 $t\bar{t}$ sample and the NNLO+NNLL $t\bar{t}$ cross section (chapter 6), is assumed before the fit. However,

the signal strengths are left freely floating during the fit.

All other processes (chapter 6) are included as background and single templates for each of the background processes are used, while the individual backgrounds are set to their Standard Model (SM) expectation.

The corresponding profile likelihood ratio (section 5.1) is then evaluated, while all systematic uncertainties (chapter 10) are taken into account as constrained nuisance parameters θ . Here, all three signal strengths ($r(t\bar{t}+b\bar{b})$, $r(t\bar{t}+2b)$, $r(t\bar{t}+b)$) as parameters of interest are scanned simultaneously by defining a three-dimensional grid of sets of possible signal strength values, while the nuisance parameters θ are varied to obtain a local minimum of the likelihood function. The best-fit signal strengths $r(t\bar{t}+b\bar{b})$, $r(t\bar{t}+2b)$, and $r(t\bar{t}+b)$ are then given by the global minimum of the profile likelihood function.

Analogous to the $t\bar{t}+b\bar{b}/t\bar{t}+2b/t\bar{t}+b$ measurement the $t\bar{t}+\geq 1b\text{-jet}$ and $t\bar{t}H$ signal strengths ($r(t\bar{t}+\geq 1b\text{-jet})$, $r(t\bar{t}H)$) and cross sections ($\sigma(t\bar{t}+\geq 1b\text{-jet})$, $\sigma(t\bar{t}H)$) are determined in the $t\bar{t}+\geq 1b\text{-jet}/t\bar{t}H$ measurement. The only difference to the previously explained measurement is that the three processes $t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, and $t\bar{t}+b$ are taken as a combined signal, while the $t\bar{t}H$ process is included as a further signal in the multi-dimensional ML fit.

14. Results

$t\bar{t}+b\bar{b}/t\bar{t}+2b/t\bar{t}+b$ measurement

In the $t\bar{t}+b\bar{b}/t\bar{t}+2b/t\bar{t}+b$ measurement the signal strength of the $t\bar{t}+b\bar{b}$ process $r(t\bar{t}+b\bar{b})$, of the $t\bar{t}+2b$ process $r(t\bar{t}+2b)$, and of the $t\bar{t}+b$ process $r(t\bar{t}+b)$ are determined simultaneously.

However, only in the case of the NN reference analysis the multi-dimensional ML fit showed a good convergence behaviour. In the case of the B-jet mult. and LD analysis the ML fit had difficulties to determine the signal strength $r(t\bar{t}+2b)$ of the $t\bar{t}+2b$ process.

The initial assumption that this is caused by a smaller number of degrees of freedom provided to the ML fit by the B-jet mult. and LD analysis could be excluded: A LD analysis containing a final discriminant distribution divided in 30 bins, the same number of bins as provided by the six final discriminant distributions of the NN analysis, still resulted in a non-properly converging ML fit.

Instead it seems that the B-jet mult. and LD cannot provide sufficient separation of the $t\bar{t}+2b$ process to the two other signal processes ($t\bar{t}+b\bar{b}$, $t\bar{t}+b$). The multiclassification done by the neural network apparently gives an advantage to the NN reference analysis. Comparing the final discriminant distributions of the three analyses (figure 9.1, fig. 9.4, and fig. 9.5), it can be observed that in the single distributions of the B-jet mult. and LD analyses events corresponding to all three processes are enriched in the rightmost bins. In the same figures the individual contributions of the three signals are shown by solid lines which are scaled to the overall integral of the stacked histograms. It is rather obvious that events of all three processes will be similarly classified. In comparison, the dedicated $t\bar{t}+2b$ final discriminant distribution of the NN analysis demonstrates that the neural network can sufficiently separate events of the $t\bar{t}+2b$ process and the two other signal processes.

Accordingly, only results obtained with the NN reference analysis will be given for the $t\bar{t}+b\bar{b}/t\bar{t}+2b/t\bar{t}+b$ measurement in the following:

The expected signal strengths $r(\text{expected})$ and observed signal strengths $r(\text{observed})$ of the three processes are presented in table 14.1. From the observed signal strength $r(\text{observed})$ the cross section σ was derived as described in section 11.7. The results obtained with the NN reference analysis are presented additionally for a comparison with the sum of the

$t\bar{t}+b\bar{b}/t\bar{t}+2b/t\bar{t}+b$ cross sections. Since correlations between the three $t\bar{t}+b\bar{b}/t\bar{t}+2b/t\bar{t}+b$ cross section results are difficult to estimate and the cross section sum is only used for a comparison with the $t\bar{t}+\geq 1b$ -jet cross section of the NN reference analysis, no uncertainty on this summed $t\bar{t}+b\bar{b}/t\bar{t}+2b/t\bar{t}+b$ cross section is given.

Furthermore, in figure 14.1 two-dimensional likelihood contour plots are shown in which two observed signal strengths $r(\text{observed})$ measured by the NN reference analysis are drawn.

Table 14.1: Results of the $t\bar{t}+b\bar{b}/t\bar{t}+2b/t\bar{t}+b$ measurement: The expected signal strength $r(\text{expected})$ measured by the NN reference analysis is stated for the $t\bar{t}+b\bar{b}$ process, the $t\bar{t}+2b$ process and the $t\bar{t}+b$ process. Furthermore, the observed signal strength $r(\text{observed})$ is given for these processes. From this observed signal strength $r(\text{observed})$ the cross section σ was derived. For a comparison, the sum of the $t\bar{t}+b\bar{b}/t\bar{t}+2b/t\bar{t}+b$ cross sections and the results obtained with the NN reference analysis are presented in addition. The systematic uncertainties correspond to a 1σ -uncertainty.

Process	$r(\text{expected})$	$r(\text{observed})$	σ
$t\bar{t}+b\bar{b}$	1 ± 0.13	1.28 ± 0.15	5.3 ± 0.6 pb
$t\bar{t}+2b$	$1^{+0.65}_{-0.55}$	$0.98^{+0.71}_{-0.62}$	$4.5^{+3.3}_{-2.9}$ pb
$t\bar{t}+b$	1 ± 0.21	1.14 ± 0.21	17.4 ± 3.2 pb
$t\bar{t}+b\bar{b}/t\bar{t}+2b/t\bar{t}+b$			27.2 pb
$t\bar{t}+\geq 1b$ -jet	$1^{+0.23}_{-0.17}$	$1.23^{+0.26}_{-0.19}$	$29.6^{+6.3}_{-4.6}$ pb

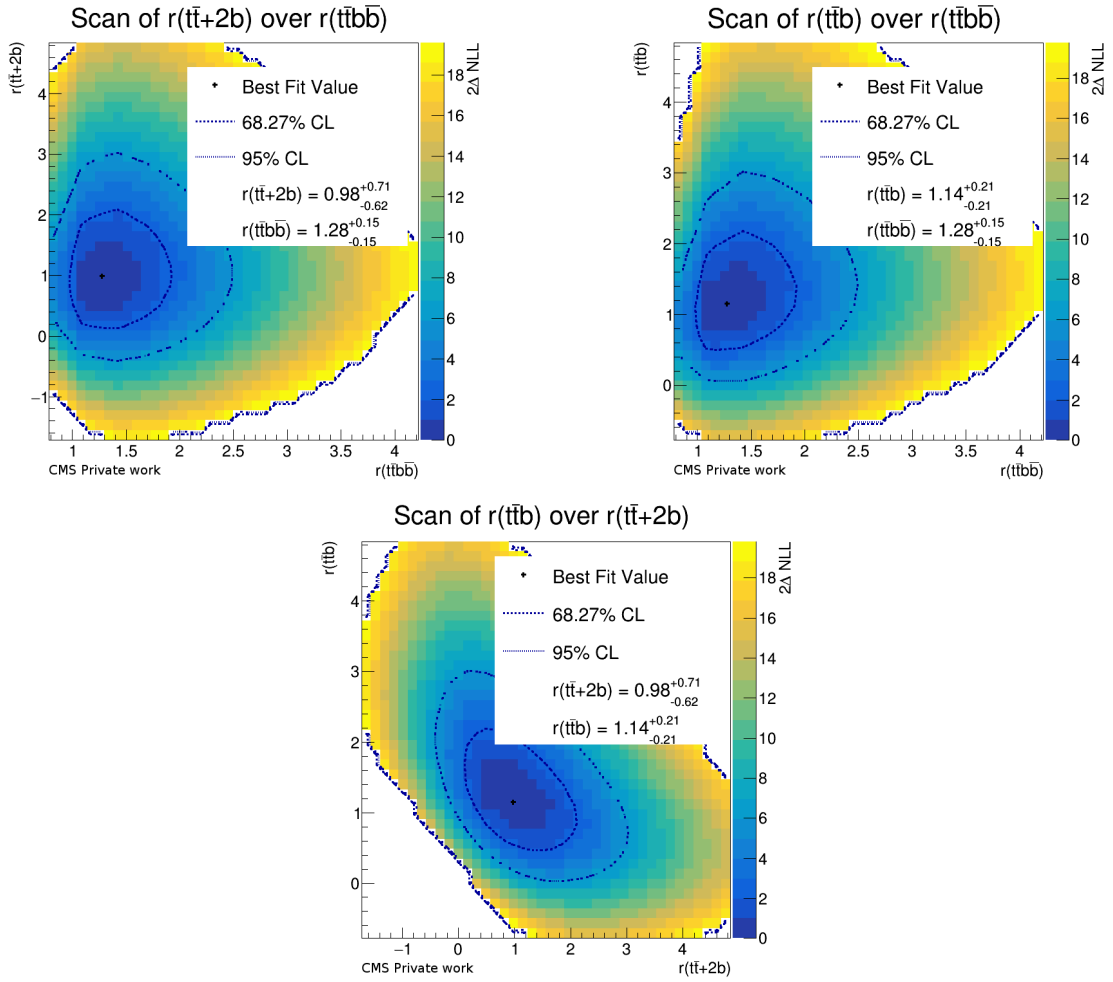


Figure 14.1: Likelihood contour plots of the $\bar{t}\bar{t}+b\bar{b}/\bar{t}\bar{t}+2b/\bar{t}\bar{t}+b$ measurement: Two observed signal strenghts $r(\text{observed})$ are plotted against each other. In addition, their best-fit value is drawn and stated. Furthermore, the 68 % and 95 % confidence level of the two-dimensional likelihood contour is shown.

$t\bar{t} + \geq 1\text{b-jet}/t\bar{t}H$ measurement

In the $t\bar{t} + \geq 1\text{b-jet}/t\bar{t}H$ measurement the signal strength of the $t\bar{t} + \geq 1\text{b-jet}$ process $r(t\bar{t} + \geq 1\text{b-jet})$ and of the $t\bar{t}H$ process $r(t\bar{t}H)$ are determined by a multi-dimensional ML fit.

In contrast to the $t\bar{t} + b\bar{b}/t\bar{t} + 2b/t\bar{t} + b$ measurement, the ML fit converged for all three analyses (B-jet mult., LD, NN). Nonetheless, only the NN analysis has good sensitivity. For the other two analyses large uncertainties on the $t\bar{t}H$ signal strength $r(t\bar{t}H)$ are expected, which is confirmed by even larger uncertainties on the observed signal strength $r(t\bar{t}H, \text{observed})$. Therefore, they will only be listed in table 14.2 for completeness.

As before, an additional linear discriminant based analysis in which the final discriminant distribution had the same number of bins as the six final discriminant distributions of the NN reference analysis, was tested. However, increasing the number of bins in the LD analysis did not significantly improve the precision of the expected and observed signal strengths r . For this reason, this result is not presented. It can once more be assumed that the lower precision obtained with the B-jet mult. and LD analysis is caused by the smaller separation between the two signal processes.

The expected and observed signal strength r of the $t\bar{t} + \geq 1\text{b-jet}$ process and the $t\bar{t}H$ process measured by the three reference analyses is stated in table 14.2. For the determination of the $t\bar{t}H$ cross section $\sigma(t\bar{t}H)$ the $t\bar{t}H(b\bar{b})$ cross section prediction of 0.2953 pb (section 6.2) is employed. The $t\bar{t} + \geq 1\text{b-jet}$ results of the dedicated $t\bar{t} + \geq 1\text{b-jet}$ measurement (part IV) are stated additionally for comparison.

Furthermore, the likelihood contour of the $t\bar{t} + \geq 1\text{b-jet}$ signal strength $r(t\bar{t} + \geq 1\text{b-jet}, \text{observed})$ and the $t\bar{t}H$ signal strength $r(t\bar{t}H, \text{observed})$ is shown in figure 14.2. Again, the results of the NN reference analysis are used.

Table 14.2: Results of the $t\bar{t} + \geq 1\text{b-jet}/t\bar{t}H$ measurement: The expected signal strength $r(\text{exp.})$ measured by the reference analyses (B-jet mult., LD, NN) is stated for the $t\bar{t} + \geq 1\text{b-jet}$ process and the $t\bar{t}H$ process. Furthermore, the observed signal strength $r(\text{obs.})$ is given for these processes. From the observed signal strength of the NN analysis the cross section σ was derived. For a comparison, the $t\bar{t} + \geq 1\text{b-jet}$ signal strength results measured by the dedicated $t\bar{t} + \geq 1\text{b-jet}$ measurement are presented in addition. The systematic uncertainties correspond to a 1σ -uncertainty.

Process	B-jet mult.		LD		NN		σ
	$r(\text{exp.})$	$r(\text{obs.})$	$r(\text{exp.})$	$r(\text{obs.})$	$r(\text{exp.})$	$r(\text{obs.})$	
$t\bar{t} + \geq 1\text{b-jet}$	1 ± 0.50	$0.86^{+0.40}_{-0.30}$	$1^{+0.57}_{-0.37}$	$1.20^{+0.65}_{-0.40}$	$1^{+0.23}_{-0.18}$	$1.21^{+0.27}_{-0.22}$	$29.1^{+6.5}_{-5.3} \text{ pb}$
$t\bar{t}H$	1 ± 1.53	4.30 ± 8.95	1 ± 1.67	-0.01 ± 6.39	1 ± 0.81	0.80 ± 0.83	$0.2 \pm 0.2 \text{ pb}$
$t\bar{t} + \geq 1\text{b-jet}$	$1^{+0.33}_{-0.23}$	$1.09^{+0.27}_{-0.24}$	$1^{+0.24}_{-0.18}$	$1.21^{+0.28}_{-0.24}$	$1^{+0.23}_{-0.17}$	$1.23^{+0.26}_{-0.19}$	$29.6^{+6.3}_{-4.6} \text{ pb}$

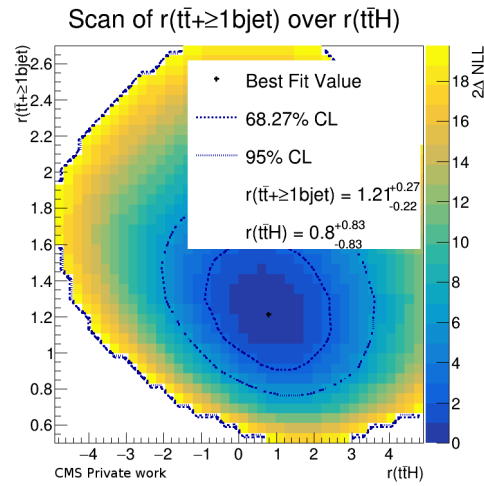


Figure 14.2: Likelihood contour plot of the $\bar{t}\bar{t}+\geq 1\text{b-jet}/\bar{t}\bar{t}\text{H}$ measurement: The observed signal strength of the $\bar{t}\bar{t}\text{H}$ process $r(\bar{t}\bar{t}\text{H}, \text{observed})$ is plotted over the observed signal strength of the $\bar{t}\bar{t}+\geq 1\text{b-jet}$ process $r(\bar{t}\bar{t}+\geq 1\text{b-jet}, \text{observed})$. In addition, their best-fit value is drawn and stated. Furthermore, the 68 % and 95 % confidence level of the two-dimensional likelihood contour is shown.

15. Discussion of results

As already discussed in the results chapter (chapter 14), out of the three analyses (B-jet mult., LD, NN) only the NN analysis showed a good convergence behaviour and leads to results with sufficient precision for the two measurements, the $t\bar{t}+b\bar{b}/t\bar{t}+2b/t\bar{t}+b$ measurement and the $t\bar{t}+\geq 1b\text{-jet}/t\bar{t}H$ measurement. The shortcomings of the other two analyses can be attributed to their lower separation power. In particular, these had difficulties to separate the $t\bar{t}+2b$ process from the $t\bar{t}+b\bar{b}$ and $t\bar{t}+b$ process in the $t\bar{t}+b\bar{b}/t\bar{t}+2b/t\bar{t}+b$ measurement.

For this reason, only results obtained by using the NN reference analysis will be discussed in the following.

$t\bar{t}+b\bar{b}/t\bar{t}+2b/t\bar{t}+b$ measurement

A comparison of the sum of the observed $t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, and $t\bar{t}+b$ cross section of $\sigma(\text{obs.}) = 27.2$ pb obtained with the $t\bar{t}+b\bar{b}/t\bar{t}+2b/t\bar{t}+b$ measurement to the $t\bar{t}+\geq 1b\text{-jet}$ cross section ($\sigma(t\bar{t}+\geq 1b\text{-jet}, \text{obs.}) = 29.6^{+6.3}_{-4.6}$ pb) derived from the $t\bar{t}+\geq 1b\text{-jet}$ measurement shows the good agreement between the two measurements. Here, due to the difficult to estimate correlations between the $t\bar{t}+b\bar{b}/t\bar{t}+2b/t\bar{t}+b$ cross section results no uncertainties of the $t\bar{t}+b\bar{b}/t\bar{t}+2b/t\bar{t}+b$ cross section sum are given. Nonetheless, it can be assumed that these uncertainties will be in similar order of magnitude as the uncertainties of the $t\bar{t}+\geq 1b\text{-jet}$ cross section.

Interestingly enough, the $t\bar{t}+b\bar{b}/t\bar{t}+2b/t\bar{t}+b$ measurement leads to signal strengths and cross sections with rather symmetric uncertainties, while the uncertainties on the $t\bar{t}+\geq 1b\text{-jet}$ signal strength and cross section of the $t\bar{t}+\geq 1b\text{-jet}$ measurement are rather asymmetric. An explanation for this is hard to give, but it is assumed that it is caused by the fewer degrees of freedom in the $t\bar{t}+\geq 1b\text{-jet}$ measurement: Since only one signal is fitted in this measurement, the Maximum-Likelihood (ML) fit has more difficulties to adjust the signal template, so that, together with the background, the shape of the measurement data is matched. In the case of three signals as in the $t\bar{t}+b\bar{b}/t\bar{t}+2b/t\bar{t}+b$ measurement the signal templates can be individually adjusted, so that the agreement of signals plus backgrounds with the measurement data should be better.

The $t\bar{t}+2b$ process has by far the largest relative uncertainties, even though the $t\bar{t}+2b$

process out of the $t\bar{t} + \geq 1b$ -jet processes is the process with the smallest expected and observed cross section. Again, one can conclude that the most challenging separation is the separation of the $t\bar{t} + 2b$ process from the other $t\bar{t} + \geq 1b$ -jet processes, which is expected as explained earlier.

For Monte Carlo event generator experts the different signal strengths obtained in the $t\bar{t} + b\bar{b}/t\bar{t} + 2b/t\bar{t} + b$ measurement maybe also of interest: The observed signal strength of the $t\bar{t} + b\bar{b}$ process of 1.28 ± 0.15 clearly indicates that the amount of $t\bar{t} + b\bar{b}$ signal observed in measurement data is underpredicted by the combination of the POWHEG Box Version 2 matrix element generator [30–34] and the Pythia8 general-purpose Monte Carlo event generator [37, 38] employed in this thesis to describe all $t\bar{t}$ processes. Meanwhile, the observed signal strength of the $t\bar{t} + b$ process is only 1.14 ± 0.21 . Therefore, it seems that the prediction of this process by the aforementioned combination of Monte Carlo event generators is better. The signal strength of the $t\bar{t} + 2b$ process is – as discussed previously – afflicted with large uncertainties; for this reason, no conclusions will be drawn for this process.

A comparison of the observed signal strength ($r(t\bar{t} + b\bar{b}, \text{observed}) = 1.28 \pm 0.15$) and cross section ($\sigma(t\bar{t} + b\bar{b}, \text{obs.}) = 5.3 \pm 0.6$ pb) of the $t\bar{t} + b\bar{b}$ process to the results of the reviewed $t\bar{t} + b\bar{b}$ measurements (part III) shows a good agreement with the newest measurement, the CMS $\sqrt{s} = 13$ TeV measurement, which observes a $t\bar{t} + b\bar{b}$ signal strength of 1.25 ± 0.59 and a $t\bar{t} + b\bar{b}$ cross section of $4.0 \pm 0.6(\text{stat.}) \pm 1.3(\text{syst.})$ pb.

As mentioned in the review part (part III) the other reviewed $t\bar{t} + b\bar{b}$ measurements are in strong tension to the CMS $\sqrt{s} = 13$ TeV measurement and, hence, also to the measurement presented in this thesis.

However, the uncertainties on the observed $t\bar{t} + b\bar{b}$ signal strength $r(t\bar{t} + b\bar{b}, \text{observed})$ and the observed $t\bar{t} + b\bar{b}$ cross section $\sigma(t\bar{t} + b\bar{b}, \text{obs.})$ are quite smaller in the measurement presented in this thesis than in the CMS $\sqrt{s} = 13$ TeV measurement.

Since in section 11.2.3 it was shown that systematic uncertainties are dominating the overall uncertainties of a signal strength measurement presented in this thesis and statistical uncertainties are rather negligible the uncertainties are not split in statistical uncertainties (stat.) and systematic uncertainties (syst.) as it is done for the CMS measurement.

The reason for the different statistical uncertainties is that the measurement presented in this thesis made use of an integrated luminosity of 35.9 fb^{-1} , while the CMS $\sqrt{s} = 13$ TeV measurement could only use a luminosity of 2.3 fb^{-1} .

Furthermore, it can be observed that the precision of CMS $\sqrt{s} = 13$ TeV measurement is limited by the rather large systematic uncertainty of ± 1.3 pb on the $t\bar{t} + b\bar{b}$ cross section result. In comparison to that, the systematic uncertainty on the $t\bar{t} + b\bar{b}$ cross section of the measurement presented in this thesis is quite smaller, roughly 0.6 pb.

In principle, it would be expected that a $t\bar{t} + b\bar{b}$ measurement conducted in the dileptonic channel has smaller systematic uncertainties than a semileptonic $t\bar{t} + b\bar{b}$ measurement. The reason for this is that, for example, large dominant systematic uncertainties related to the b -tagging of jets do not need to be considered.

This contradiction can be explained by the fact that the CMS $\sqrt{s} = 13$ TeV $t\bar{t} + b\bar{b}$ measurement was conducted at the beginning of the LHC Run 2 which is also why the employed integrated luminosity is rather small. Commonly, after an increase of the center-of-mass energy $\sqrt{s}$ the estimation of systematic uncertainties must be renewed, while in

general this estimation improves with more event statistics or rather luminosity. Furthermore, the reduction of systematic uncertainties is part of continuous effort to improve the object identification and event reconstruction by the CMS detector as well as physics analyses. Therefore, it can be safely assumed that the systematic uncertainties employed in the $t\bar{t}+b\bar{b}$ measurement presented in this thesis are smaller not least because this measurement was conducted at a later time.

For the reasons stated above, it can be concluded that the $t\bar{t}+b\bar{b}$ measurement presented in this thesis is the most precise measurement of the $t\bar{t}+b\bar{b}$ signal strength and cross section so far.

For the $t\bar{t}+2b$ and the $t\bar{t}+b$ process no other measurement conducted at the Large Hadron Collider and suitable for a comparison exists. Therefore, no further conclusions can be drawn.

$t\bar{t}+\geq 1b\text{-jet}/t\bar{t}H$ measurement

A comparison of the observed $t\bar{t}+\geq 1b\text{-jet}$ signal strength $r(t\bar{t}+\geq 1b\text{-jet}, \text{observed})$ and cross section $\sigma(t\bar{t}+\geq 1b\text{-jet}, \text{obs.})$, obtained from the NN reference analysis, between the $t\bar{t}+\geq 1b\text{-jet}/t\bar{t}H$ measurement

$$\begin{aligned} r(t\bar{t}+\geq 1b\text{-jet}, \text{observed}) &= 1.21_{-0.22}^{+0.27} , \\ \sigma(t\bar{t}+\geq 1b\text{-jet}, \text{obs.}) &= 29.1_{-5.3}^{+6.5} \text{ pb} , \end{aligned}$$

and the $t\bar{t}+\geq 1b\text{-jet}$ measurement

$$\begin{aligned} r(t\bar{t}+\geq 1b\text{-jet}, \text{observed}) &= 1.23_{-0.19}^{+0.26} , \\ \sigma(t\bar{t}+\geq 1b\text{-jet}, \text{obs.}) &= 29.6_{-4.6}^{+6.3} \text{ pb} , \end{aligned}$$

shows good agreement.

It can be concluded that the additional fit of the $t\bar{t}H$ signal in the $t\bar{t}+\geq 1b\text{-jet}/t\bar{t}H$ measurement does not significantly affect the determination of the $t\bar{t}+\geq 1b\text{-jet}$ signal strength and cross section.

The observed $t\bar{t}H$ signal strength $r(t\bar{t}H, \text{observed})$ is now compared to a recent ATLAS $t\bar{t}H(b\bar{b})$ measurement [114] and a recent CMS $t\bar{t}H(b\bar{b})$ measurement [115]:

Both measurements employ a similar luminosity (ATLAS: $\mathcal{L} = 36.1 \text{ fb}^{-1}$) or even the same luminosity (CMS: $\mathcal{L} = 35.9 \text{ fb}^{-1}$) as the measurement presented in this thesis.

Furthermore, both measurements conduct the measurement of the $t\bar{t}H(b\bar{b})$ signal strength in the semileptonic decay channel of the top quark as well as the dileptonic decay channel, while the measurement presented in this thesis only exploits the semileptonic decay channel. Therefore, in the following only the semileptonic measurements and results will be discussed.

For the semileptonic decay channel, the ATLAS measurement requires at least five jets, from which two must fulfill a b-tagging requirement, besides a single lepton. Semileptonic events considered in the CMS measurement must have one single lepton and at least four jets, from which two must fulfill a b-tagging requirement. For the measurement presented in this thesis events have to fulfill a harder requirement: Besides one lepton at least six jets, from which two are b-tagged, are required.

In all three measurements (including the one presented in this thesis) the $t\bar{t}$ samples was

produced using a combination of the POWHEG Box Version 2 matrix element generator [30–34] and the Pythia8 general-purpose Monte Carlo event generator [37, 38]. In the case of the CMS measurement and this measurement this event generator combination was also employed for the modeling of the $t\bar{t}H$ process. In contrast to that, in the ATLAS measurement a combination of the MG5aMC(NLO)+Pythia8 event generators [29, 38, 139, 140] is used to model the $t\bar{t}H$ process. Furthermore, the $t\bar{t} + \geq 1b\text{-jet}$ processes contained in the $t\bar{t}$ sample are scaled to match the predictions of a next-to-leading order (NLO) $t\bar{t} + b\bar{b}$ sample generated with the Sherpa+OpenLoops event generator [143–145] in the ATLAS measurement.

For the total cross section of the $t\bar{t}$ samples a prediction at next-to-next-leading order (NNLO) QCD accuracy with next-to-next-leading logarithmic (NNLL) soft gluon resummation using the top++2.0 program (see [177] and references therein) is used in all three measurements.

In the ATLAS measurement a boosted decision tree (BDT) is employed as classifier, which can exploit among common event variables b -tagging probabilities and information provided by a matrix element method. In the CMS dileptonic measurement also a BDT using common event variables, b -tagging probabilities, and information provided by a matrix element method is used. However, for the CMS semileptonic measurement a multiclassifying neural network – similar to the neural network used in this thesis – is employed. The major difference between the two neural networks is that the neural network of the CMS measurement considers other event variables and makes use of additional information provided by a matrix element method.

All three (semileptonic) measurements consider similar (ATLAS measurement) or the same (CMS measurement and measurement of this thesis) systematic uncertainties. However, in the ATLAS measurement the normalizations of the $t\bar{t} + b\bar{b}$ and $t\bar{t} + c\bar{c}$ background is left freely floating, while the CMS measurement considers them as nuisance parameters.

Due to leaving the $t\bar{t} + b\bar{b}$ and $t\bar{t} + c\bar{c}$ normalization freely floating the ATLAS measurement determines, besides the observed $t\bar{t}H$ signal strength $r(t\bar{t}H, \text{observed})$

$$r(t\bar{t}H, \text{observed}) = 0.95^{+0.65}_{-0.62} \quad ,$$

also the $t\bar{t} + \geq 1b\text{-jet}$ signal strength $r(t\bar{t} + b\bar{b}, \text{observed})$ and the $t\bar{t} + c\bar{c}$ signal strength $r(t\bar{t} + c\bar{c}, \text{observed})$

$$\begin{aligned} r(t\bar{t} + b\bar{b}, \text{observed}) &= 1.24 \pm 0.10 \quad , \\ r(t\bar{t} + c\bar{c}, \text{observed}) &= 1.63 \pm 0.23 \quad . \end{aligned}$$

Here, the observed $t\bar{t}H$ signal strength $r(t\bar{t}H, \text{observed})$ is stated which is derived from the semileptonic measurement only, so that it is better comparable to the measurement of this thesis. In contrast to that, the $t\bar{t} + b\bar{b}$ and $t\bar{t} + c\bar{c}$ signal strengths are given for a combined measurement using the semileptonic and dileptonic decay channel since no semileptonic result was available for a comparison.

By the semileptonic measurement of the CMS collaboration a $t\bar{t}H$ signal strength of

$$r(t\bar{t}H, \text{observed}) = 0.84^{+0.52}_{-0.50}$$

is observed. Again, only the result of the semileptonic measurement is given for a better comparison with the measurement presented in this thesis, even though a combined semileptonic and dileptonic result would have been available.

Compared to the two reviewed measurements, the $t\bar{t}+\geq 1\text{b-jet}/t\bar{t}H$ measurement presented in this thesis observes a $t\bar{t}H$ signal strength $r(t\bar{t}H, \text{observed})$ of

$$r(t\bar{t}H, \text{observed}) = 0.80 \pm 0.83 \quad ,$$

and a $t\bar{t}+\geq 1\text{b-jet}$ signal strength $r(t\bar{t}+\geq 1\text{b-jet}, \text{observed})$ of

$$r(t\bar{t}+\geq 1\text{b-jet}, \text{observed}) = 1.21^{+0.27}_{-0.22} \quad .$$

It can be concluded that all three measurements are in good agreement with each other.

However, the two reviewed measurements feature $t\bar{t}H$ signal strength results with smaller uncertainties than this measurement, even though a similar luminosity was used and only the results of the semileptonic measurements were compared.

It is plausible that the better performance of the two reviewed measurements compared to this one comes from including additional information provided by a matrix element method in the classifier. Furthermore, the two reviewed measurements employ a looser jet selection and, hence, can include more categories in the final Maximum-Likelihood (ML) fit.

Besides the good agreement in the $t\bar{t}H$ signal strength the ATLAS measurement and the measurement presented in this thesis also predict a rather similar $t\bar{t}+\geq 1\text{b-jet}$ signal strength, while the $t\bar{t}+\geq 1\text{b-jet}$ definition is slightly different:

In the ATLAS $t\bar{t}H(b\bar{b})$ measurement in addition to the $t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, and $t\bar{t}+b$ process classes two further $t\bar{t}+\geq 1\text{b-jet}$ process classes are introduced. Events having more than two additional b-tagged jets fall into the $t\bar{t}+\geq 3\text{b-jet}$ process class and are not included in the $t\bar{t}+b\bar{b}$ process class like in the measurements presented in this thesis. Events with additional b-tagged jets entirely originating from final-state radiation or multi-parton interactions are considered in a separate process class. Furthermore, the relative contributions of the $t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$, and the $t\bar{t}+\geq 3\text{b-jet}$ process class are scaled to match the predictions of a Sherpa+OpenLoops [143–145] $t\bar{t}+b\bar{b}$ sample. This rescaled and combined process scales are employed in the ML fit to derive the $t\bar{t}+\geq 1\text{b-jet}$ signal strength.

Part VI

Study of a possible improvement of the $t\bar{t} + \geq 1b$ -jet modeling by Monte Carlo event generators

Study of a possible improvement of the $t\bar{t}+\geq 1\text{b-jet}$ modeling by Monte Carlo event generators

In the measurements presented in this thesis and the reviewed measurements signal strengths of the $t\bar{t}+\geq 1\text{b-jet}$, the $t\bar{t}+b\bar{b}$, and the $t\bar{t}+b$ process larger than expected from available Monte Carlo event generators are observed. This leads to the assumption that an improvement of the modeling of the $t\bar{t}+\geq 1\text{b-jet}$ processes by Monte Carlo event generators would be beneficial. Therefore, in the following, possible approaches to improve the $t\bar{t}+\geq 1\text{b-jet}$ modeling by Monte Carlo event generators will be discussed.

The concept of combining a five-flavour scheme (5FS) $t\bar{t}$ sample with a four-flavour scheme (4FS) $t\bar{t}+b\bar{b}$ sample will be introduced and motivated (chapter 16). In chapter 17 follows a description how such a combined 5FS $t\bar{t}$ / 4FS $t\bar{t}+b\bar{b}$ sample can be constructed, before the focus will be on the proper normalization which is required for the combination of the two samples (chapter 18). Since this normalization is still afflicted with shortcomings only preliminary conclusions can be drawn in chapter 19.

16. Motivation of a combined 5FS $t\bar{t}$ / 4FS $t\bar{t}+b\bar{b}$ sample

In the measurements presented previously in this thesis the combination of the POWHEG Box Version 2 matrix element (ME) event generator [30, 31, 171] and the Pythia8 general-purpose Monte Carlo (GPMC) event generator [37, 183] is employed to produce next-to-leading order (NLO) $t\bar{t}$ samples by making use of the five-flavour scheme (5FS). As explained in section 2.1.4 the 5FS seems a suitable choice for the generation of the $t\bar{t}$ samples since in measurements of signal strengths ($r(t\bar{t}+\geq 1b\text{-jet})$, $r(t\bar{t}+b\bar{b})$, $r(t\bar{t}+2b)$, $r(t\bar{t}+b)$, $r(t\bar{t}H)$) rather inclusive observables are of particular interest.

These $t\bar{t}$ samples (chapter 6) are then split into $t\bar{t}+X$ sub processes by using the GenHFWadronMatcher tool [146, 149] (section 6.4). In this way, separate Monte Carlo event samples of the $t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, and $t\bar{t}+b$ signal process are obtained and also event samples of the $t\bar{t}+c\bar{c}$ and the $t\bar{t}+lf$ background process.

All of these processes are later considered as single processes in the final discriminant distributions of the three reference analyses, the B-jet multiplicity based (B-jet mult.) analysis, the linear discriminant based (LD) analysis, and the neural network based (NN) analysis (chapter 9). Furthermore, these single processes are used for the training of the multivariate classifiers of the LD and NN analysis (section 5.2).

As can be deduced from the measurements presented in this thesis (part IV and part V) and the reviewed measurements (part III), the $t\bar{t}+\geq 1b\text{-jet}$ process is underpredicted by Monte Carlo event generators so far. All measurements observe signal strengths $r(t\bar{t}+\geq 1b\text{-jet}, \text{observed})$ of roughly 1.3. This also applies to the $t\bar{t}+b\bar{b}$ and $t\bar{t}+b$ processes individually, for which a signal strength of $r(t\bar{t}+b\bar{b}, \text{observed}) = 1.28 \pm 0.15$ and $r(t\bar{t}+b, \text{observed}) = 1.14 \pm 0.21$, respectively, is derived in this thesis. The $t\bar{t}+2b$ process could be the only exception since the observed signal strength of this process is $0.98^{+0.71}_{-0.62}$. However, the uncertainties on the $t\bar{t}+2b$ signal strength $r(t\bar{t}+2b, \text{observed})$ are rather large, so that no definitive statement can be made.

The $t\bar{t}+\geq 1b\text{-jet}$ production processes are predicted by the Standard model of particle physics (SM). Until now, no hint of physics beyond the Standard Model (BSM) is observed for these processes. Therefore, the significant deviation of the measured $t\bar{t}+\geq 1b\text{-jet}$ signal strengths from one leads to the conclusion that these processes are not properly

modeled by the Monte Carlo event generators and methods currently employed.

This issue is well-known among large parts of the particle physics community and active research is undertaken in this context as already mentioned in the motivation (part II): References [47, 131, 223] are examples of rather recent publications describing this issue and suggesting possible solutions.

Among the possible solutions which are discussed is a further tuning of the currently employed Monte Carlo event generators. However, large efforts to tune the event generators used by default by the CMS collaboration have already been made in the recent past (see reference [185]). By this tuning, for example, the modeling of the $t\bar{t}+1f$ processes was improved, so that the agreement between measurement data and Monte Carlo prediction has particularly gotten better in high jet-multiplicity and large transverse momentum categories [224].

Nonetheless, in the near future it is not expected that tuning the Monte Carlo event generators currently employed by the CMS collaboration will improve the $t\bar{t}+\geq 1b$ -jet modeling.

Another proposal for solution, made e.g. by [47, 131, 223], involves using dedicated $t\bar{t}+b\bar{b}$ event samples which are produced by employing a four-flavour scheme (4FS).

In contrast to the 5FS, in the 4FS the assumption is made that bottom quarks cannot be part of the composite proton below the factorization scale μ_F due to their large mass and that they only appear as quark-antiquark pairs in scattering processes. Therefore, bottom quarks are not included in the parton distribution function of the proton. The advantage of the 4FS is the presumably better description of the kinematics of the $t\bar{t}+b\bar{b}$ process, while the prediction of the total cross section of a process or other inclusive observables can be inaccurate because of large logarithms possibly occurring in computation of initial-state and final-state radiation, which are not resummed.

For example, the ATLAS collaboration makes use of such a 4FS $t\bar{t}+b\bar{b}$ sample produced by the Sherpa+OpenLoops event generator [143–145] in their $t\bar{t}H(b\bar{b})$ analysis [114] as outlined in the motivation (part II): $t\bar{t}+\geq 1b$ -jet events derived from a 5FS $t\bar{t}$ sample which was generated by using the same event generator combination as in this thesis (POWHEG+Pythia8) are rescaled to match the predictions of this 4FS $t\bar{t}+b\bar{b}$ sample.

However, the CMS collaboration is skeptical of this rescaling approach. The reason is that the rescaling changes the relative fractions of the $t\bar{t}+\geq 1b$ -jet process classes described by the POWHEG+Pythia8 sample. Therefore, it could introduce additional systematic uncertainties which are difficult to specify.

Furthermore, in this way, the main advantage of 4FS $t\bar{t}+b\bar{b}$ event generation, namely providing $t\bar{t}+\geq 1b$ -jet events, and particularly $t\bar{t}+b\bar{b}$ and $t\bar{t}+\geq 3b$ -jet events, with higher precision, is not exploited.

For this reason, this approach is not discussed further here.

The best solution so far, which is also advocated by [47, 131, 223], seems to be the combination of a 5FS $t\bar{t}$ sample with a 4FS $t\bar{t}+b\bar{b}$ sample: The 5FS $t\bar{t}$ sample would provide a modeling of $t\bar{t}+lf$ and $t\bar{t}+c\bar{c}$ events, while the 4FS $t\bar{t}+b\bar{b}$ sample would be employed for an improved modeling of $t\bar{t}+\geq 1b\text{-jet}$ events.

This is motivated by different limitations of the two sample types: The 5FS can only poorly describe $t\bar{t}+b\bar{b}$ events in which the two additional bottom quarks of the $t\bar{t}+b\bar{b}$ matrix element have a small transverse momentum and angular separation. The reason for this is that in this collinear region of the phase space the approximation that bottom quarks are massless breaks down. Compared to that, the 4FS has the advantage that one out of the two $g \rightarrow b\bar{b}$ splittings required to produce the $t\bar{t}+b\bar{b}$ process can be described as part of a NLO $t\bar{t}+b\bar{b}$ matrix element. However, a 4FS $t\bar{t}+b\bar{b}$ sample does not include contributions of the $t\bar{t}+lf$ or $t\bar{t}+c\bar{c}$ processes [47].

This approach will be followed in the further course of this thesis.

17. Construction of a combined 5FS $t\bar{t}$ / 4FS $t\bar{t}+b\bar{b}$ sample

For the reasons stated in chapter 16 a combined 5FS $t\bar{t}$ / 4FS $t\bar{t}+b\bar{b}$ sample would be highly desirable, at least for a cross check with the default 5FS $t\bar{t}$ sample.

According to [223], for a combination of a 5FS $t\bar{t}$ sample with a 4FS $t\bar{t}+b\bar{b}$ sample a veto on events involving bottom quarks in the $t\bar{t}$ sample would be sufficient. However, the veto has to take place at particle level, after the parton shower, but the veto should not be applied on bottom quarks originating from the decay of the involved top quarks or from the underlying event.

Applying such an event veto is not possible for analyses conducted by the CMS collaboration at the moment. The reason is that due to the large number of $t\bar{t}$ events (> 150 million) required by analyses to obtain enough statistics the 5FS $t\bar{t}$ samples are produced globally by a dedicated event sample production system called Monte Carlo Request Management (McM). The technical realization of such a veto would require non-trivial changes to the McM system. Furthermore, it still needs to be confirmed by a comparison with measurement data that such a combined sample leads to a better modeling of the $t\bar{t}$, and in particular the $t\bar{t}+\geq 1b$ -jet, processes, before the default 5FS $t\bar{t}$ can be replaced.

For this reason a combined 5FS $t\bar{t}$ / 4FS $t\bar{t}+b\bar{b}$ sample has to be obtained by a slightly different replacement of 5FS $t\bar{t}+\geq 1b$ -jet events with 4FS $t\bar{t}+\geq 1b$ -jet events, which takes into account the available means and event samples provided by the CMS collaboration.

In addition to the 5FS $t\bar{t}$ samples produced by using POWHEG+Pythia8 and employed in this thesis or other 5FS samples produced by using either MG5aMC(LO) [179] or MG5aMC(NLO) [29, 140] as matrix element generator, two 4FS samples were produced which should match the properties of the 2016 measurement data: A 4FS $t\bar{t}+b\bar{b}$ sample was produced using a combination of the MG5aMC(NLO) matrix element event generator and the Pythia8 GPMC event generator. By using the Sherpa+OpenLoops event generator a further 4FS $t\bar{t}+b\bar{b}$ sample was produced.

Both 4FS samples will be used in the following together with 5FS POWHEG $t\bar{t}$ sample.

The 5FS POWHEG $t\bar{t}$ sample is split into $t\bar{t}+X$ sub samples by using the GenHfHadronMatcher tool as mentioned earlier. The GenHfHadronMatcher (section 6.4) acts in a similar way as the suggested event veto based on additional bottom quarks by providing a possibility to derive a $t\bar{t}+b\bar{b}$, a $t\bar{t}+2b$, and a $t\bar{t}+b$ sub sample in addition to a $t\bar{t}+c\bar{c}$ and a $t\bar{t}+lf$ sub sample. The major difference between the proposed event veto and the approach of the GenHfHadronMatcher is that the GenHfHadronMatcher identifies $t\bar{t}+\geq 1b$ -jet processes by additionally occurring bottom hadrons which are not originating from the decay of a top quark, while the suggested event veto relies directly on additional bottom quarks.

By using the process classes determined by the GenHfHadronMatcher two combined 5FS $t\bar{t}$ / 4FS $t\bar{t}+b\bar{b}$ sample can be easily constructed: From the 5FS POWHEG sample only the $t\bar{t}+c\bar{c}$ and the $t\bar{t}+lf$ sub sample is used for a combination with either the 4FS MG5aMC-(NLO) $t\bar{t}+b\bar{b}$ sample or the 4FS Sherpa+OpenLoops $t\bar{t}+b\bar{b}$ sample, while the POWHEG $t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, and $t\bar{t}+b$ sub samples are neglected.

18. Normalizing 4FS $t\bar{t} + \geq 1b$ -jet events

Before 5FS $t\bar{t} + \geq 1b$ -jet events are replaced by 4FS $t\bar{t} + \geq 1b$ -jet events during the construction of a combined sample, $t\bar{t} + \geq 1b$ -jet events obtained from the 4FS $t\bar{t} + b\bar{b}$ sample must be properly normalized, so that their sum of weights matches the weight of the replaced 5FS $t\bar{t} + \geq 1b$ -jet events. Otherwise, the exchange of the $t\bar{t} + \geq 1b$ -jet events would lead to a change in the (differential) $t\bar{t} + \geq 1b$ -jet cross section. For this reason, the stacked Monte Carlo histograms of all processes ($t\bar{t} + X$, $t\bar{t}H$, and background processes) would no longer show good agreement with measurement data.

Currently, this proper normalization of the integrated 4FS $t\bar{t} + \geq 1b$ -jet events represents a challenge: Various normalization schemes have been tried, but after the replacement of 5FS $t\bar{t} + \geq 1b$ -jet events by 4FS $t\bar{t} + \geq 1b$ -jet events a change in the number of events which pass the event selection (chapter 7) of the three reference analyses is always observed. The exchange of the $t\bar{t} + \geq 1b$ -jet events also leads to a poor modeling of the measurement data by the Monte Carlo samples, so that it does not seem reasonable at this point to use the obtained combined 5FS $t\bar{t} / 4FS t\bar{t} + b\bar{b}$ event sample for an analysis or to show a derived final discriminant distribution.

Especially, the observed significant change in the number of events passing the event selection after replacing the $t\bar{t} + \geq 1b$ -jet events is astonishing. As an example the b -tagged jets final discriminant distribution of the B -jet mult. analysis is shown in figure 18.1. Here, the events of the $t\bar{t} + b\bar{b}$ process, the $t\bar{t} + 2b$ process, and the $t\bar{t} + b$ process of the 5FS POWHEG $t\bar{t}b\bar{b}$ samples are combined to a common 5FS POWHEG $t\bar{t} + \geq 1b$ -jet sample and compared to the two 4FS $t\bar{t} + \geq 1b$ -jet samples produced by using either the MG5aMC (NLO)+Pythia8 or the Sherpa+OpenLoops event generator.

For the construction of the combined 5FS $t\bar{t} / 4FS t\bar{t} + b\bar{b}$ samples the sum of the event weight of all 5FS $t\bar{t} + \geq 1b$ -jet events was determined and the weight of the inserted 4FS $t\bar{t} + \geq 1b$ -jet events adjusted, so that their sum matches the sum of the weights of the 5FS $t\bar{t} + \geq 1b$ -jet events. Both weight sums were determined using particle-level information, meaning that no event selection or cuts were applied.

This normalization scheme is motivated by the different jet multiplicity distributions ob-

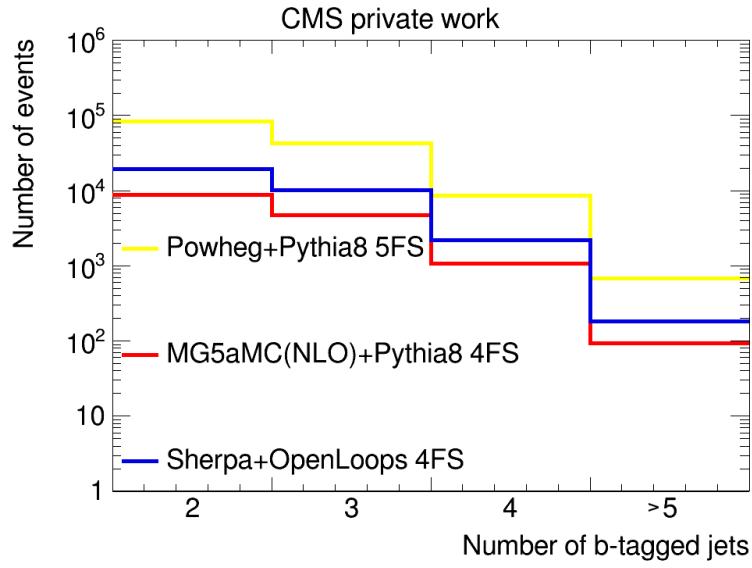
served for events in the three samples as shown in figure 18.2. For the determination of this number of jets distribution a transverse momentum cut of 20 GeV is imposed on gen-level jets obtained with the anti- k_T jet clustering algorithm and a distance parameter of 0.4. In principle, by considering the full phase (no cuts applied) for the normalization of the $t\bar{t}+\geq 1b$ -jet events a difference in the jet multiplicity or other observables should have no influence on the overall normalization. Furthermore, by this normalization scheme the $t\bar{t}+\geq 1b$ -jet cross section of all three $t\bar{t}+\geq 1b$ -jet event samples should be equal and should correspond to the $t\bar{t}+\geq 1b$ -jet cross section $\sigma(t\bar{t}+\geq 1b\text{-jet, pred.})$ of 24.06 pb predicted by the POWHEG event generator as described in section 11.7.

In figure 18.1a the number of $t\bar{t}+\geq 1b$ -jet events predicted by using each of the three $t\bar{t}+\geq 1b$ -jet samples is shown, while in figure 18.1b the number of predicted $t\bar{t}+\geq 1b$ -jet events is normalized to one for each $t\bar{t}+\geq 1b$ -jet sample.

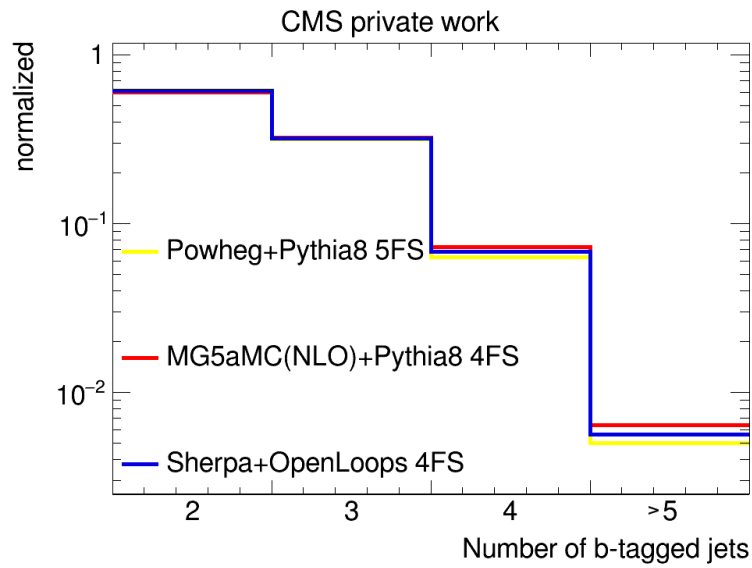
As seen in figure 18.1 the total number of $t\bar{t}+\geq 1b$ -jet events as well as the number of $t\bar{t}+\geq 1b$ -jet events per bin is significantly smaller if the two combined 5FS $t\bar{t}/4\text{FS } t\bar{t}+b\bar{b}$ samples are used instead of the 5FS POWHEG $t\bar{t}+\geq 1b$ -jet sample. Furthermore, it seems that the 5FS POWHEG $t\bar{t}+\geq 1b$ -jet sample has a smaller fraction of events with four or more b-tagged jets than the other two combined samples.

At the same time the 5FS POWHEG $t\bar{t}+\geq 1b$ -jet sample has in comparison with the combined 5FS $t\bar{t}/4\text{FS } t\bar{t}+b\bar{b}$ samples more events with higher jet multiplicity, while the other two samples have more events with less than six jets.

Due to having more events with fewer than six jets the two combined 5FS $t\bar{t}/4\text{FS } t\bar{t}+b\bar{b}$ samples have a lower number of events in the selected phase space region of events containing at least six jets, from which are two b-tagged. For this reason also a poorer modeling of the measurement data by the stack of Monte Carlo histograms is observed if either one of the two combined 5FS $t\bar{t}/4\text{FS } t\bar{t}+b\bar{b}$ samples is included instead of the default 5FS POWHEG $t\bar{t}$ sample which already shows a good agreement with measurement data.



(a) Number of b-tagged jets (not normalized)



(b) Number of b-tagged jets (normalized to one)

Figure 18.1: Comparison of the number of b-tagged jets predicted by the three $t\bar{t} + \geq 1b$ -jet samples: The number of b-tagged jets after the event selection (section 7) predicted by using the 5FS POWHEG+Pythia8 $t\bar{t} + \geq 1b$ -jet sample, the 4FS MG5aMC(NLO)+Pythia8 $t\bar{t} + \geq 1b$ -jet sample, and the 4FS Sherpa+Open Loops $t\bar{t} + \geq 1b$ -jet sample is compared. In figure (a) the number of events predicted by using the three samples is shown, while in figure (b) the distribution is normalized to one.

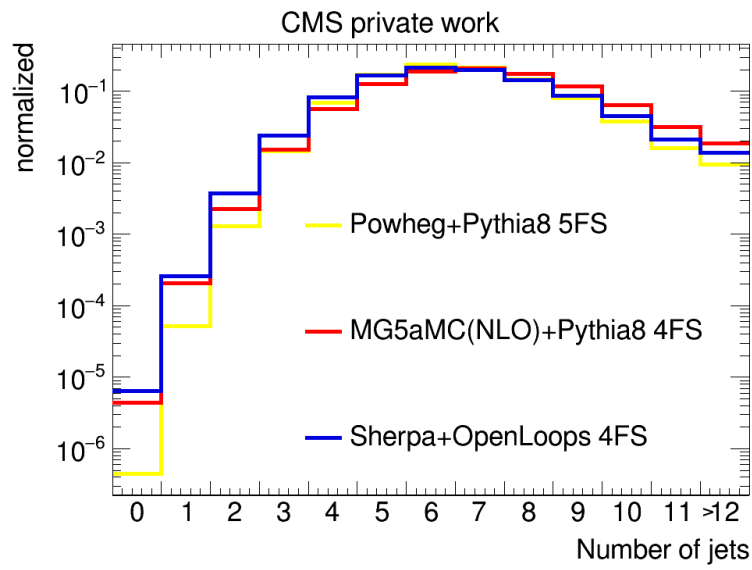


Figure 18.2: Comparison of the number of jets predicted by the three $t\bar{t} + \geq 1b$ -jet samples: The number of jets predicted by using the 5FS POWHEG+Pythia8 $t\bar{t} + \geq 1b$ -jet sample, the 4FS MG5aMC(NLO)+Pythia8 $t\bar{t} + \geq 1b$ -jet sample, and the 4FS Sherpa+OpenLoops $t\bar{t} + \geq 1b$ -jet sample is compared.

19. Preliminary conclusions

Even after a few studies and different normalization schemes tried out a well-founded explanation of the difference between the 5FS POWHEG $t\bar{t}+\geq 1\text{b-jet}$ sample and the two combined 5FS $t\bar{t}+\geq 1\text{b-jet}$ / 4FS $t\bar{t}+b\bar{b}$ samples is hard to give at present:

First of all, it can be speculated that the 4FS does not properly describe the full range of $t\bar{t}+\geq 1\text{b-jet}$ processes encountered in measurement data and that a modeling of $t\bar{t}+\geq 1\text{b-jet}$ processes by the 5FS approximates measurement data better.

Furthermore, it can be assumed that the two Monte Carlo event generators MG5aMC (NLO)+Pythia8 and Sherpa+OpenLoops employed for the production of the underlying 4FS $t\bar{t}+b\bar{b}$ samples need some further tuning. This is plausible for the Sherpa+OpenLoops event generator, which has not yet been used much in the CMS collaboration. However, the MG5aMC(NLO)+Pythia8 is used by default for the production of many process samples other than the $t\bar{t}$ samples. An improper tuning of this event generator, therefore, seems unlikely but a non-optimal tuning for processes involving top and bottom quarks cannot be fully excluded.

More insight in this issue may also be gained by applying the GenHFWadronMatcher on the 4FS MG5aMC(NLO)+Pythia8 $t\bar{t}+b\bar{b}$ sample before it is used to replace the 5FS POWHEG+Pythia8 $t\bar{t}+\geq 1\text{b-jet}$ events. In this way, similar events should be classified as $t\bar{t}+\geq 1\text{b-jet}$ events from the 4FS MG5aMC(NLO)+Pythia8 $t\bar{t}+b\bar{b}$ sample as from the 5FS POWHEG+Pythia8 $t\bar{t}$ sample. Therefore, it could be that a better normalization is reached by these means.

It can be argued that the GenHFWadronMatcher should have been applied also on the 4FS $t\bar{t}+b\bar{b}$ samples from the beginning. However, since the GenHFWadronMatcher relies partly on parton-level information it cannot be used for Sherpa+OpenLoops samples. The reason is that for samples produced with the Sherpa event generator such information is deliberately not available.

That is the reason why a more general replacement and normalization scheme was aimed for from the beginning, which will allow the combination of 5FS $t\bar{t}$ samples with 4FS $t\bar{t}+b\bar{b}$ samples independent of the event generator they are produced with.

As for now, it can be concluded that more studies are necessary before a combined 5FS $t\bar{t}+\geq 1b\text{-jet}$ / 4FS $t\bar{t}+b\bar{b}$ sample can be employed in a physics analysis.

In the same context, it cannot be said for the time being if a combination of a 5FS $t\bar{t}+\geq 1b\text{-jet}$ sample with a 4FS $t\bar{t}+b\bar{b}$ sample would lead to an improved modeling of the $t\bar{t}+\geq 1b\text{-jet}$ process and if it would resolve the difference between the predicted $t\bar{t}+\geq 1b\text{-jet}$, $t\bar{t}+b\bar{b}$, and $t\bar{t}+b$ signal strengths $r(\text{expected})$ and the corresponding observed signal strengths $r(\text{observed})$.

Part VII

Summary and outlook

Summary and outlook

After the discovery of the Higgs boson in 2012, which was the final missing particle predicted by the Standard Model of particle physics (SM), and at latest with the Run 2 of the Large Hadron Collider (LHC) starting in 2015, a large-scale transition process began: While searches for new particles, physics beyond the Standard Model (BSM), or the discovery of previously undiscovered SM physics processes like the $t\bar{t}H$ process were of major importance in the early days of the LHC, precision measurements of SM processes are becoming more and more important nowadays.

Since no new physics has been found at the LHC up to now, an early discovery of new physics becomes less and less likely every day. The reason for this lies in the scaling of the discovery potential at a fixed center-of-mass energy $\sqrt{s}$ with the root of the number of analyzed events. The LHC has delivered a rather large luminosity since its start, while the CMS detector is performing well and has already recorded an integrated luminosity of proton-proton collisions of roughly 114 fb^{-1} by the end of 2017 [225]. It can be assumed that a jump in luminosity as expected by the high-luminosity LHC (HL-LHC) projected for 2026 is essential before an extensive search for new physics will be worthwhile again.

Meanwhile, many physics analyses conducting precision measurements are no longer limited by the size of the dataset as in the beginning of the LHC data taking, but more and more systematic uncertainties dictate the accuracy of the outcome of such analyses.

Dominant systematic uncertainties in precision measurements are often related to the event reconstruction and the identification of more complex objects, for example, the identification of b-tagged jets, or to the modeling of major signal and background processes by Monte Carlo event generators.

An example of the latter are the $t\bar{t}+\geq 1\text{b-jet}$ processes which are defined by a top-quark pair production physics process accompanied by additionally occurring bottom quarks. The measurement of these $t\bar{t}+\geq 1\text{b-jet}$ processes suffers from large systematic uncertainties on the b-tagging of jets if the semileptonic decay channel of top quarks is exploited. Furthermore, in measurements of the $t\bar{t}+\geq 1\text{b-jet}$ and, in particular, $t\bar{t}+b\bar{b}$ production conducted so far, signal strengths well above one are observed, so that it can be concluded that these processes are underestimated by the Monte Carlo event generators currently employed.

At the same time, these $t\bar{t}+\geq 1\text{b-jet}$ processes are important background processes for many other LHC physics analyses. Therefore, considering the non-optimal modeling of $t\bar{t}+\geq 1\text{b-jet}$ processes, large systematic uncertainties have to be dealt within such analyses. For example, the $t\bar{t}H(b\bar{b})$ measurement conducted by the CMS collaboration [115], which recently led to a 5.2σ discovery of the $t\bar{t}H$ process together with other $t\bar{t}H$ channels [111], assigned an additional 50% uncertainty to the $t\bar{t}+\geq 1\text{b-jet}$ processes.

Therefore, the boundaries of precision measurements restricted by systematic uncertainties can often be pushed in two ways:

On the one hand, the employment of more elaborate methods, e.g. in the object identification and event reconstruction, can reduce systematic uncertainties. The replacement of a likelihood ratio used by the CMS collaboration for the b-tagging of jets in the LHC Run 1 by methods based on neural networks in LHC Run 2 may serve as an example [90, 169]. On the other hand, a better understanding and modeling of physics processes can also lead to measurement results with higher precision.

With the measurements of the $t\bar{t}+\geq 1b$ -jet signal strength and cross section using the CMS 2016 measurement data recorded at a center-of-mass energy of 13 TeV and presented in this thesis, both possibilities to push the boundaries were exploited: The determination of a more precise $t\bar{t}+\geq 1b$ -jet signal strength and cross section itself is worth striving for, but it can also lead to decreased systematic uncertainties assumed for the $t\bar{t}+\geq 1b$ -jet processes in other measurements. Moreover, a promising multivariate analysis method was introduced by using multiclassifying neural networks in the measurements presented in this thesis, which to the best of my knowledge was never used for a measurement of the $t\bar{t}+\geq 1b$ -jet processes until now.

In addition to the neural network based analysis, a straightforward analysis based on the B-jet multiplicity was employed as a cross check for these measurements. Furthermore, a linear discriminant based analysis was used to compare a binary classifying multivariate analysis method with the multiclassifying multivariate analysis method implemented by using neural networks.

In the first measurement part (part IV), the signal strength and cross section of the combined $t\bar{t}+\geq 1b$ -jet processes were determined. In contrast, the single signal strengths and cross sections of either the $t\bar{t}+b\bar{b}$, the $t\bar{t}+2b$, and the $t\bar{t}b$ process or the single signal strengths and cross sections of the $t\bar{t}+\geq 1b$ -jet and the $t\bar{t}H$ process were simultaneously measured in part V.

Here, the advantages of a multiclassification provided by the neural network employed became clearly evident: For the simultaneous measurement of the $t\bar{t}+b\bar{b}/t\bar{t}+2b/t\bar{t}+b$ processes, the multi-dimensional Maximum-Likelihood fit showed good convergence behaviour only in the case of the neural network based analysis and was used to determine signal strength and cross section results. By contrast, the fit did not converge for the simpler B-jet multiplicity and linear discriminant based analyses. In similar fashion, for the simultaneous measurement of the $t\bar{t}+\geq 1b$ -jet and $t\bar{t}H$ signal strength and cross section rather large uncertainties were expected and observed if the two more straightforward analyses were used, while the neural network based analysis reached results and associated uncertainties comparable to reviewed measurements.

In table 19.1, the various signal strengths r and cross sections σ determined by employing the neural network based analysis are summarized.

All results are in good agreement with previous measurements.

The results obtained for the $t\bar{t}+\geq 1b$ -jet and $t\bar{t}+b\bar{b}$ processes represent the most precise results up to now, while for the $t\bar{t}+2b$ and the $t\bar{t}+b$ processes no measurement conducted at the LHC at a center-of-mass energy of 13 TeV and suitable for a comparison exists.

By specific $t\bar{t}H(b\bar{b})$ measurements realized by the ATLAS and CMS collaboration just recently, a $t\bar{t}H$ signal strength with smaller uncertainties was determined [114, 115]. However, the difference in the uncertainties can be explained by a tighter event selection employed in the measurements of this thesis and resulting fewer categories. In contrast to the two

$t\bar{t}H(b\bar{b})$ measurements, the analyses presented in this thesis were also not specifically optimized for measuring the $t\bar{t}H$ signal strength. Moreover, only one classifier was used in this thesis always, while the two $t\bar{t}H(b\bar{b})$ measurements employ various classifiers trained for single categories.

The underestimation of the $t\bar{t}+\geq 1b$ -jet signal strengths (except maybe for the $t\bar{t}+2b$ signal strength, which has rather large uncertainties), which is also observed by other measurements [170, 178, 182], confirms the assumption that the $t\bar{t}+\geq 1b$ -jet processes could benefit from improved modeling by Monte Carlo event generators.

For this reason, in part VI, feasible approaches for such an improvement were discussed and studied by introducing the possibility to combine an inclusive $t\bar{t}$ sample produced by using the five-flavour scheme with a $t\bar{t}+b\bar{b}$ sample produced by making use of the four-flavour scheme. Regrettably, the necessary normalization required to employ such combined samples is still proving problematic. Therefore, no final conclusions can be drawn if a combined sample would provide an improved modeling of $t\bar{t}+X$ processes and, in particular, $t\bar{t}+\geq 1b$ -jet processes.

Table 19.1: Summary of the signal strengths and cross sections determined by the measurements presented in this thesis: The expected signal strength $r(\text{expected})$ of the $t\bar{t}+\geq 1b$ -jet process, the $t\bar{t}+b\bar{b}$ process, the $t\bar{t}+2b$ process, the $t\bar{t}+b$, and the $t\bar{t}H$ process is stated. Furthermore, the observed signal strength $r(\text{observed})$ is given for these processes. From the observed signal strength $r(\text{observed})$ the cross section σ was derived. All results were determined using the neural network based analysis. The results of processes marked by an asterisk (*) were obtained in a simultaneous measurement with other processes. The systematic uncertainties correspond to a 1σ -uncertainty.

Process	$r(\text{expected})$	$r(\text{observed})$	σ
$t\bar{t}+\geq 1b\text{-jet}$	$1^{+0.23}_{-0.17}$	$1.23^{+0.26}_{-0.19}$	$29.6^{+6.3}_{-4.6}$ pb
$t\bar{t}+b\bar{b}$ (*)	1 ± 0.13	1.28 ± 0.15	5.3 ± 0.6 pb
$t\bar{t}+2b$ (*)	$1^{+0.65}_{-0.55}$	$0.98^{+0.71}_{-0.62}$	$4.5^{+3.3}_{-2.9}$ pb
$t\bar{t}+b$ (*)	1 ± 0.21	1.14 ± 0.21	17.4 ± 3.2 pb
$t\bar{t}+\geq 1b\text{-jet}$ (*)	$1^{+0.23}_{-0.18}$	$1.21^{+0.27}_{-0.22}$	$29.1^{+6.5}_{-5.3}$ pb
$t\bar{t}H$ (*)	1 ± 0.81	0.80 ± 0.83	0.2 ± 0.2 pb

Outlook

The measurements presented in this thesis would benefit from an improved b-tagging of jets which should reduce the related dominant systematic uncertainties.

In the 2016/2017 year-end shutdown the old CMS pixel detector consisting of three detector layers was replaced by a new pixel detector as part of the CMS Phase-1 upgrade [69]. Due to the increased number of 124 million pixels and an additional fourth detector layer an improvement in the b-tagging capabilities of the CMS detector is rather certain.

In addition, an improved b-tagging algorithm [90, 169] is already available for the CMS 2017 measurement data, but has not been available by default for the CMS 2016 measurement data yet.

Furthermore, the measurements would in general gain from an improved modeling of the $t\bar{t}+X$ processes and, in particular the $t\bar{t}+\geq 1b$ -jet processes, by Monte Carlo event generators. Systematic uncertainties related to the modeling by Monte Carlo event generators are among the most important systematic uncertainties. Such an improvement can either come from using a new or better tuned Monte Carlo event generator or from making use of a combined $t\bar{t}/t\bar{t}+b\bar{b}$ sample as mentioned previously.

However, if no further reduction of systematic uncertainties can be reached, future $t\bar{t}+\geq 1b$ -jet measurements will most likely be conducted in the dileptonic decay channel of top quarks. In comparison to the semileptonic decay channel, this channel is less affected by systematic uncertainties until now and hence will allow more precise results.

The usage of multiclassifying neural networks clearly showed advantages in contrast to binary classification. This is also observed in other measurements like the recent CMS $t\bar{t}H(b\bar{b})$ measurement [115].

That is the reason why a wide-spread use of multiclassifying neural networks in physics analysis – similar to the rapidly growing adoption of neural networks in many fields of industry – can be expected.

For example, in a future $t\bar{t}H(b\bar{b})$ analysis, a multiclassifying neural network could be employed to measure the $t\bar{t}H$ signal strength and the $t\bar{t}+\geq 1b$ -jet signal strength simultaneously.

Of course, such a simultaneous measurement is also feasible with traditional multivariate analysis methods like binary classifying boosted decision trees. However, the multiclassifying neural networks have the advantage that they can easily provide individual categories for multiple signals (e.g. $t\bar{t}H$ and $t\bar{t}+\geq 1b$ -jet) and background. Therefore, a better separation between multiple signal processes as well an improved background estimation can be expected which should lead to an improved final result.

A simultaneous measurement of the $t\bar{t}H$ process and the $t\bar{t}+\geq 1b$ -jet process seems to be beneficial independent of the multivariate classifier applied. The reason is that the uncertainty on the dominant $t\bar{t}+\geq 1b$ -jet background to the $t\bar{t}H$ signal can be stronger constrained if the $t\bar{t}+\geq 1b$ -jet normalization is left freely floating in the ML fit which was already demonstrated in the ATLAS $t\bar{t}H(b\bar{b})$ measurement [114].

Thinking further ahead, the neural networks employed in this thesis or in the CMS $t\bar{t}H(b\bar{b})$ measurement, which both consider six process classes ($t\bar{t}+b\bar{b}$, $t\bar{t}+2b$, $t\bar{t}+b$, $t\bar{t}+c\bar{c}$, $t\bar{t}+lf$,

$t\bar{t}H$) in the multiclassification, could be extended by process classes for the $t\bar{t}W$ and the $t\bar{t}Z$ process. In this way, a simultaneous measurement of all three processes involving top-quark pair production associated with an additional boson seems conceivable. First steps have already been taken in this direction, but particularly the separation of the $t\bar{t}Z$ process from the other two $t\bar{t}+\text{Boson}$ processes still needs improvement, before this approach could be exploited in a measurement.

Danksagung

Ein afrikanisches Sprichwort besagt: “Es braucht zwei Menschen, um ein Kind zu zeugen. Um ein Kind aber großzuziehen, braucht es ein ganzes Dorf”.

In meinem Fall würde ich davon sprechen, dass es zwei Dörfer waren: Das Dorf der Familie und der Freunde sowie das Dorf der Arbeitsgruppe von Prof. Dr. Ulrich Husemann bzw. des Instituts für Experimentelle Teilchenphysik (ETP) am Karlsruher Institut für Technologie (KIT).

Zunächst einmal möchte ich allen danken, die zum Erfolg dieser Arbeit beigetragen haben. Es sind zu viele, die mich in fachlicher, beruflicher und persönlicher Hinsicht geprägt, inspiriert und durch ihr Vorbild geleitet haben, um alle hier zu nennen.

Getreu dem Spruch “Menschen treten in unser Leben und begleiten uns eine Weile. Einige bleiben für immer, denn sie hinterlassen ihre Spuren in unseren Herzen.” kann ich aber festhalten, dass es in meinem Verstand und Herzen nur so von (Teilchen-)Spuren wimmelt.

Im Rahmen dieser Arbeit möchte ich ganz herzlich meinem Doktorvater Herrn Prof. Dr. Ulrich Husemann danken, der meinen Weg durch die Teilchenphysik seit meiner Masterarbeit begleitet hat. Er hat mir vielfältige wertvolle Erfahrungen ermöglicht, indem er mir sehr viele Freiheiten gelassen und auch jederzeit ein offenes Ohr hatte, wenn ich mit neuen kreativen Einfällen und technischen Weiterentwicklungen die Teilchenphysik und die Arbeitsweise in der Arbeitsgruppe vorantreiben wollte. Gleichzeitig habe ich gerade zu Anfang sehr stark von seinem bewunderswert großen Wissen zur Teilchenphysik, Datenanalyse und verwandten Fachgebieten profitieren können.

Im Zuge der Doktorarbeit möchte ich auch meinem Korreferenten Herrn PD Dr. Stefan Gieseke herzlich danken. Zu Beginn meiner Promotionszeit durfte ich in seiner Arbeitsgruppe an der Entwicklung des Herwig7-Ereignisgenerators mitarbeiten, wodurch ich dann auch ein MCNet-Stipendium erhielt, um diese Tätigkeit für 4 Monate an der University of Durham fortzusetzen. Diese Arbeit an und mit Monte-Carlo-Ereignisgeneratoren begann bereits in meiner Masterarbeit und hat sich bis zum Ende meiner Promotion größtenteils fortgesetzt, aber einen tieferen Einblick habe ich erst in PD Dr. Stefan Giesekes Gruppe am Institut für Theoretische Physik (ITP) des KIT sowie in der Gruppe von Prof. Peter Richardson und Dr. Simon Plätzer am Institute for Particle Physics Phenomology (IP³) der University of Durham erlangt.

Krankheitsbedingt konnte PD Dr. Stefan Gieseke leider nicht an der mündlichen Verteidigung meiner Doktorarbeit teilnehmen, sodass ich Herrn Prof. Dr. Dieter Zeppenfeld für die kurzfristige Übernahme des Korreferats sehr dankbar bin.

Hiermit möchte ich mich auch äußerst herzlich bei meinen Kollegen aus der Arbeitsgruppe Husemann bedanken:

Zuallerst sind Dr. Matthias Schröder und Karim El Morabit zu nennen, mit denen ich zu jeder Zeit Ideen zur eigenen Analyse diskutieren konnte und Fragen zum gemeinsamen Analyse-Framework, welches auf Arbeiten von Dr. Hannes Mildner und Dr. Shawn Williamson basiert, klären konnte. Dr. Matthias Schröder übernahm auch die inhaltliche und sprachliche Korrektur der Grundlagenteile dieser Arbeit sowie der CMS-internen Analyse-Note. Dafür beiden meinen tiefen Dank.

Weiterhin möchte ich Simon Kudella, der kurz nach mir promoviert hat, für den regen Austausch über alle möglichen Belange betreffend Promotion danken. Wie heißt es so schön, “zu zweit ist man weniger allein”.

Ebenfalls danken möchte ich Michael Waßmer und Philip Keicher, die mir als Experten für die NTuple-Produktion und für das Higgs-Combine-Tool bei der Ausarbeitung der Analyse dazu auftretende Fragen sehr detailliert und umfangreich beantworten konnten.

Besonderen Dank möchte ich auch unserem ehemaligen Masterstudenten Lukas Hilser zollen, der das von mir erdachte NNFlow-Framework zum Training neuronaler Netze basierend auf den Bachelorarbeiten von Maximilian Welsch, Martin Lang und Jan van der Linden weitestgehend eigenständig entwickelt hat und einen großen Teil der von mir verwendeten neuronalen Netze bereitgestellt hat.

Zuletzt gilt mein Dank der Vielzahl an Masterstudenten und Bachelorstudenten, die ich direkt und eigenverantwortlich oder als sekundärer Ansprechpartner betreuen durfte. Es war mir immer eine große Freude Euch am Anfang die Teilchenphysik näher zu bringen, dann mit Euch zu lernen und am Schluss von Euch zu lernen, wenn ihr die Experten auf Eurem eigenen (kleinen) Fachgebiet geworden seid.

Nicht nur in fachlicher Hinsicht sondern auch in Hinblick auf meine Führungskompetenz war diese gemeinsame Arbeit an einem Thema immer eine Bereicherung für mich.

Besondere Teilchenspuren hinterlassen haben dabei in umgekehrter zeitlicher Reihenfolge: Andrej Saibel, Korbinian Schweiger (beide Masterstudenten), Lars Sowa, Jan van der Linden, Maximilian Welsch, Martin Lang, Dominik Beutel, Julia Hunt, Denise Müller (alle Bachelorstudenten).

Vielen Dank Euch und allen weiteren Mitarbeitern des Instituts!

Ganz am Schluss möchte ich eine Einrichtung sowie drei Personen noch ausdrücklich erwähnen, die meinen akademischen Werdegang sehr stark geprägt und überaus positiv beeinflusst haben, das Heidelberger Life-Science Lab, welches eine Einrichtung am Deutschen Krebsforschungszentrums (DKFZ) in Heidelberg ist, und deren Leiterin Frau Dr. Katrin Platzer sowie Herrn Dr. Frank Becker, der meine erste Bachelorarbeit im Rahmen meines DHBW-Studiums betreut hat, und Herrn Dr. Thomas Keck als Studienkollege im Physik-Studium.

Bibliography

- [1] D. Brown, “Origin”. Transworld, 2017.
- [2] D. Lincoln, “The Standard Model: The most successful theory ever.” <http://news.fnal.gov/2011/11/the-standard-model-the-most-successful-theory-ever/>. Date viewed: 08.01.2018.
- [3] R. Merton, “On the Shoulders of Giants: The Post-Italianate Edition”. University of Chicago Press, 1965.
- [4] J. J. Aubert et al., “Experimental Observation of a Heavy Particle J ”, Phys. Rev. Lett. 33, 1404–1406 Dec, 1974.
- [5] J. E. Augustin et al., “Discovery of a Narrow Resonance in e^+e^- Annihilation”, Phys. Rev. Lett. 33, 1406–1408 Dec, 1974.
- [6] F. Hasert et al., “Search for elastic muon-neutrino electron scattering”, Physics Letters B 46, 121 – 124, no. 1, 1973.
- [7] F. Hasert et al., “Observation of neutrino-like interactions without muon or electron in the gargamelle neutrino experiment”, Physics Letters B 46, 138 – 140, no. 1, 1973.
- [8] F. Hasert et al., “Observation of neutrino-like interactions without muon or electron in the Gargamelle neutrino experiment”, Nuclear Physics B 73, 1 – 22, no. 1, 1974.
- [9] G. Arnison et al., “Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s} = 540$ GeV”, Physics Letters B 122, 103 – 116, no. 1, 1983.
- [10] M. Banner et al., “Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider”, Physics Letters B 122, 476 – 485, no. 5, 1983.
- [11] G. Arnison et al., “Experimental observation of lepton pairs of invariant mass around $95 \text{ GeV}/c^2$ at the CERN SPS collider”, Physics Letters B 126, 398 – 410, no. 5, 1983.
- [12] P. Bagnaia et al., “Evidence for $Z^0 \rightarrow e^+e^-$ at the CERN pp collider”, Physics Letters B 129, 130 – 140, no. 1, 1983.
- [13] F. Englert and R. Brout, “Broken Symmetry and the Mass of Gauge Vector Mesons”, Phys. Rev. Lett. 13, 321–323 Aug, 1964.
- [14] P. Higgs, “Broken symmetries, massless particles and gauge fields”, Physics Letters 12, 132 – 133, no. 2, 1964.
- [15] P. W. Higgs, “Broken Symmetries and the Masses of Gauge Bosons”, Phys. Rev. Lett. 13, 508–509 Oct, 1964.

- [16] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, “Global Conservation Laws and Massless Particles”, *Phys. Rev. Lett.* 13, 585–587 Nov, 1964.
- [17] **ATLAS Collaboration**, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”, *Phys. Lett. B* 716, 1–29 2012.
- [18] **CMS Collaboration**, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”, *Phys. Lett. B* 716, 30–61 2012.
- [19] Cush, “File:Standard Model of Elementary Particles.svg — Wikimedia Commons, the free media repository.”
https://commons.wikimedia.org/w/index.php?title=File:Standard_Model_of_Elementary_Particles.svg&oldid=269046622. Date viewed: 23.01.2018.
- [20] S. Tomonaga, “On a relativistically invariant formulation of the quantum theory of wave fields”, *Prog. Theor. Phys.* 1, 27–42 1946.
- [21] J. Schwinger, “Quantum Electrodynamics. I. A Covariant Formulation”, *Phys. Rev.* 74, 1439–1461 Nov, 1948.
- [22] H. Fritzsch, “Quantum Flavordynamics”, in: “Deeper Pathways in High-Energy Physics”, ed. A. Perlmutter and L.F. Scott, pp. 275–322. Springer US, Boston, MA, 1977.
- [23] S. Bilenky and J. Hošek, “Glashow-Weinberg-Salam theory of electroweak interactions and the neutral currents”, *Physics Reports* 90, 73 – 157, no. 2, 1982.
- [24] **Particle Data Group**, “Review of Particle Physics”, *Chin. Phys. C* 40, 100001, no. 10, 2016.
- [25] M. Kobayashi and T. Maskawa, “CP-Violation in the Renormalizable Theory of Weak Interaction”, *Progress of Theoretical Physics* 49, 652–657, no. 2, 1973.
- [26] A. Buckley et al., “General-purpose event generators for LHC physics”, *Phys. Rept.* 504, 145–233 2011.
- [27] A. Saibel, “Study of $t\bar{t}+b\bar{b}$ Modelling and Uncertainties of Monte Carlo Generators for the $t\bar{t}H(b\bar{b})$ -Analysis at the CMS Experiment”. Master thesis, Karlsruher Institut für Technologie (KIT), 2017.
- [28] M. Harrendorf, “Comparative studies of Monte Carlo event generation in leading order and next-to-leading order for the CMS collaboration”. Master thesis, Karlsruher Institut für Technologie (KIT), 2014.
- [29] J. Alwall et al., “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”, *Journal of High Energy Physics* 2014, 79 Jul, 2014.
- [30] P. Nason, “A New method for combining NLO QCD with shower Monte Carlo algorithms”, *JHEP* 11, 040 2004.
- [31] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with Parton Shower simulations: the POWHEG method”, *JHEP* 11, 070 2007.
- [32] S. Alioli et al., “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX”, *JHEP* 06, 043 2010.
- [33] S. Frixione, P. Nason, and G. Ridolfi, “A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction”, *JHEP* 09, 126 2007.

- [34] H. B. Hartanto et al., “Higgs boson production in association with top quarks in the POWHEG BOX”, *Phys. Rev. D* 91, 094003, no. 9, 2015.
- [35] M. Bahr et al., “Herwig++ Physics and Manual”, *Eur. Phys. J. C* 58, 639–707 2008.
- [36] J. Bellm et al., “Herwig 7.0/Herwig++ 3.0 release note”, *The European Physical Journal C* 76, 196, no. 4, 2016.
- [37] T. Sjöstrand, S. Mrenna, and P. Z. Skands, “PYTHIA 6.4 Physics and Manual”, *JHEP* 05, 026 2006.
- [38] T. Sjöstrand et al., “An Introduction to PYTHIA 8.2”, *Comput. Phys. Commun.* 191, 159–177 2015.
- [39] A. Gehrmann-DeRidder, “Perturbative QCD (1/3).” <https://indico.cern.ch/event/230448/attachments/381911/531283/CERN1.pdf>.
Date viewed: 23.01.2018.
- [40] R. P. Feynman, “Space-Time Approach to Quantum Electrodynamics”, *Phys. Rev.* 76, 769–789 Sep, 1949.
- [41] G. Quast and A. Raspereza, “Master course 2013 Higgs Physics lecture no. 4 - Calculation of physics processes.” http://www-ekp.physik.uni-karlsruhe.de/~quast/vorlesung/TP2HiggsSS13/V04_Calculation-ofPhysicsProcesses.pdf.
Date viewed: 20.10.2014, Original document not any longer online available.
- [42] G. Ossola, C. G. Papadopoulos, and R. Pittau, “CutTools : a program implementing the OPP reduction method to compute one-loop amplitudes”, *Journal of High Energy Physics* 2008, 042, no. 03, 2008.
- [43] Y. L. Dokshitzer, “Calculation of the Structure Functions for Deep Inelastic Scattering and e^+e^- Annihilation by Perturbation Theory in Quantum Chromodynamics.”, *Sov. Phys. JETP* 46, 641–653 1977.
- [44] V. N. Gribov and L. N. Lipatov, “ e^+e^- pair annihilation and deep inelastic e p scattering in perturbation theory”, *Sov. J. Nucl. Phys.* 15, 675–684 1972.
- [45] G. Altarelli and G. Parisi, “Asymptotic Freedom in Parton Language”, *Nucl. Phys.* B126, 298–318 1977.
- [46] F. Maltoni, G. Ridolfi, and M. Ubiali, “b-initiated processes at the LHC: a reappraisal”, *JHEP* 07, 022 2012.
- [47] **LHC Higgs Cross Section Working Group**, “Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector”. CERN Reports (ISSN 2519-8076), 2016.
- [48] M. Czakon, P. Fiedler, and A. Mitov, “Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_s^4)$ ”, *Phys. Rev. Lett.* 110, 252004 2013.
- [49] M. Czakon et al., “Top-pair production at the LHC through NNLO QCD and NLO EW”, *JHEP* 10, 186 2017.
- [50] B. Webber, “Parton shower Monte Carlo event generators”, *Scholarpedia* 6, 10662, no. 12, 2011.
- [51] S. Agostinelli et al., “Geant4—a simulation toolkit”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506, 250 – 303, no. 3, 2003.

- [52] S. Williamson, “Search for Higgs-Boson Production in Association with a Top-Quark Pair in the Boosted Regime with the CMS Experiment”. PhD thesis, Karlsruher Institut für Technologie (KIT), 2016.
- [53] A. Buckley, “Tools for event generator tuning and validation” in *Proceedings, HERA and the LHC Workshop Series on the implications of HERA for LHC physics: 2006-2008*, pp. 768–773. 2008. [arXiv:0809.4638](https://arxiv.org/abs/0809.4638).
- [54] S. Frixione and B. R. Webber, “Matching NLO QCD computations and parton shower simulations”, JHEP 06, 029 2002.
- [55] M. Mangano, “The so-called MLM prescription for ME/PS matching.” <http://www-cpd.fnal.gov/personal/mrenna/tuning/nov2002/mlm.pdf.gz>. Date viewed: 25.01.2018.
- [56] R. Frederix and S. Frixione, “Merging meets matching in MC@NLO”, JHEP 12, 061 2012.
- [57] S. Catani et al., “QCD matrix elements + parton showers”, JHEP 11, 063 2001.
- [58] L. Lönnblad, “Correcting the color dipole cascade model with fixed order matrix elements”, JHEP 05, 046 2002.
- [59] F. Krauss, “Matrix elements and parton showers in hadronic interactions”, JHEP 08, 015 2002.
- [60] N. Lavesson and L. Lönnblad, “W+jets matrix elements and the dipole cascade”, JHEP 07, 054 2005.
- [61] L. Rossi and O. Brüning, “Introduction to the HL-LHC Project”, in: “The High Luminosity Large Hadron Collider”, ed. Oliver Brüning and Lucio Rossi, ch. Chapter 1, pp. 1–17. WORLD SCIENTIFIC, 2015.
- [62] A. G. et al., “High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1”. CERN Yellow Reports: Monographs. CERN, Geneva, 2017.
- [63] Y. Ohnishi et al., “Accelerator design at SuperKEKB”, Progress of Theoretical and Experimental Physics 2013, 03A011, no. 3, 2013.
- [64] L. Evans and P. Bryant, “LHC Machine”, Journal of Instrumentation 3, S08001, no. 08, 2008.
- [65] J. Haffner, “The CERN accelerator complex. Complexe des accélérateurs du CERN” Oct, 2013.
- [66] **CMS Collaboration**, “CMS Physics: Technical Design Report Volume 1: Detector Performance and Software”. Technical Design Report CMS. CERN, Geneva, 2006.
- [67] **CMS Collaboration**, “The CMS experiment at the CERN LHC”, Journal of Instrumentation 3, S08004, no. 08, 2008.
- [68] **CMS Collaboration**, “The CMS tracker system project: Technical Design Report”. Technical Design Report CMS. CERN, Geneva, 1997.
- [69] **CMS Collaboration**, “The Phase-2 Upgrade of the CMS Tracker” Technical Report CERN-LHCC-2017-009. CMS-TDR-014, CERN, Geneva, Jun, 2017.
- [70] **CMS Collaboration**, “The CMS electromagnetic calorimeter project: Technical Design Report”. Technical Design Report CMS. CERN, Geneva, 1997.

- [71] **CMS Collaboration**, “Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at $\sqrt{s} = 7$ TeV”, *Journal of Instrumentation* 8, P09009, no. 09, 2013.
- [72] **CMS Collaboration**, “The CMS hadron calorimeter project: Technical Design Report”. Technical Design Report CMS. CERN, Geneva, 1997.
- [73] **CMS Collaboration**, “The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s} = 7$ TeV at the LHC”, *Journal of Instrumentation* 8, P11002, no. 11, 2013.
- [74] D. Barney, “CMS slice illustrator files.” <https://cms-docdb.cern.ch/cgi-bin/PublicDocDB/ShowDocument?docid=5581>. Date viewed: 29.12.2017.
- [75] R. E. Kalman, “A New Approach to Linear Filtering and Prediction Problems”, *Transactions of the ASME – Journal of Basic Engineering* 35–45, no. 82 (Series D), 1960.
- [76] R. Frühwirth, “Application of Kalman filtering to track and vertex fitting”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 262, 444 – 450, no. 2, 1987.
- [77] W. Waltenberger, R. Frühwirth, and P. Vanlaer, “Adaptive vertex fitting”, *Journal of Physics G: Nuclear and Particle Physics* 34, N343, no. 12, 2007.
- [78] **CMS Collaboration**, “Description and performance of track and primary-vertex reconstruction with the CMS tracker”, *Journal of Instrumentation* 9, P10009, no. 10, 2014.
- [79] **CMS Collaboration**, “Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET” Technical Report CMS-PAS-PFT-09-001, CERN, Geneva, 2009.
- [80] A. Sirunyan, A. Tumasyan, and W. Adam, “Particle-flow reconstruction and global event description with the CMS detector”, *Journal of Instrumentation* 12, P10003, no. 10, 2017.
- [81] A. Sirunyan, A. Tumasyan, and W. A. et al., “Particle-flow reconstruction and global event description with the CMS detector”, *Journal of Instrumentation* 12, P10003, no. 10, 2017.
- [82] H. Mildner, “Measurements of the Top-Higgs Coupling with the CMS Experiment”. PhD thesis, Karlsruher Institut für Technologie (KIT), 2016.
- [83] **CMS Collaboration**, “Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV”, *Journal of Instrumentation* 7, P10002, no. 10, 2012.
- [84] W. Adam et al., “Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC”, *Journal of Physics G: Nuclear and Particle Physics* 31, N9, no. 9, 2005.
- [85] S. Baffioni et al., “Electron reconstruction in CMS”, *The European Physical Journal C* 49, 1099–1116 Mar, 2007.
- [86] **CMS Collaboration**, “Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV”, *Journal of Instrumentation* 10, P06005, no. 06, 2015.
- [87] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- k_t jet clustering algorithm”, *JHEP* 04, 063 2008.

- [88] M. Cacciari, G. P. Salam, and G. Soyez, “FastJet user manual”, The European Physical Journal C 72, 1896 Mar, 2012.
- [89] **CMS Collaboration**, “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV”, [arXiv:1712.07158](#) (2017).
- [90] **CMS Collaboration**, “Heavy flavor identification at CMS with deep neural networks” Technical Report CMS-DP-2017-005, CERN, Geneva, 2017.
- [91] **DØ Collaboration**, “Observation of Single Top Quark Production - March 2009.” https://www-d0.fnal.gov/Run2Physics/top/singletop_observation/singletop_observation_updated.html. Date viewed: 17.01.2018.
- [92] E. E. Boos et al., “The top quark (20 years after its discovery)”, Physics-Uspekhi 58, 1133, no. 12, 2015.
- [93] R. Hawkings, “Top quark physics at the LHC”, Comptes Rendus Physique 16, 424 – 434, no. 4, 2015.
- [94] U. Husemann, “Top-quark physics: Status and prospects”, Progress in Particle and Nuclear Physics 95, 48 – 97 2017.
- [95] H. Harari, “A new quark model for hadrons”, Physics Letters B 57, 265 – 269, no. 3, 1975.
- [96] S. W. Herb et al., “Observation of a Dimuon Resonance at 9.5 GeV in 400-GeV Proton-Nucleus Collisions”, Phys. Rev. Lett. 39, 252–255 Aug, 1977.
- [97] L. Lederman, C. Brown, and M. Pearson, “Discoveries at Fermilab - Discovery of the Bottom Quark.” http://www.fnal.gov/pub/inquiring/physics/discoveries/bottom_quark_pr.html. Date viewed: 17.01.2018.
- [98] **DØ Collaboration**, “Search for High Mass Top Quark Production in $p\bar{p}$ Collisions at $\sqrt{s} = 1.8$ TeV”, Phys. Rev. Lett. 74, 2422–2426 Mar, 1995.
- [99] **CDF Collaboration**, “Observation of Top Quark Production in $\bar{p}p$ Collisions with the Collider Detector at Fermilab”, Phys. Rev. Lett. 74, 2626–2631 Apr, 1995.
- [100] **ATLAS Collaboration**, “Measurement of the top quark mass in the $t\bar{t} \rightarrow$ dilepton channel from $\sqrt{s} = 8$ TeV ATLAS data”, Phys. Lett. B761, 350–371 2016.
- [101] **CMS Collaboration**, “Measurement of the top quark mass using proton-proton data at $\sqrt{s} = 7$ and 8 TeV”, Phys. Rev. D93, 072004, no. 7, 2016.
- [102] **CDF and DØ Collaboration**, “Combination of CDF and DØ results on the mass of the top quark using up 9.7 fb⁻¹ at the Tevatron” 2016.
- [103] M. Brucherseifer, F. Caola, and K. Melnikov, “On the NNLO QCD corrections to single-top production at the LHC”, Phys. Lett. B736, 58–63 2014.
- [104] E. L. Berger, J. Gao, and H. X. Zhu, “Differential Distributions for t-channel Single Top-Quark Production and Decay at Next-to-Next-to-Leading Order in QCD”, JHEP 11, 158 2017.
- [105] N. Kidonakis, “Top Quark Production” in *Proceedings, Helmholtz International Summer School on Physics of Heavy Quarks and Hadrons (HQ 2013): JINR, Dubna, Russia, July 15-28, 2013*, pp. 139–168. 2014. [arXiv:1311.0283](#).
- [106] T. M. P. Tait and C.-P. Yuan, “Single top quark production as a window to physics beyond the standard model”, Phys. Rev. D 63, 014018 Dec, 2000.

- [107] Q.-H. Cao, J. Wudka, and C.-P. Yuan, “Search for new physics via single-top production at the LHC”, *Physics Letters B* 658, 50 – 56, no. 1, 2007.
- [108] V. Barger, M. McCaskey, and G. Shaughnessy, “Single top and Higgs associated production at the LHC”, *Phys. Rev. D* 81, 034020 Feb, 2010.
- [109] **CMS Collaboration**, “Evidence for Associated Production of a Single Top Quark and W Boson in pp Collisions at $\sqrt{s} = 7$ TeV”, *Phys. Rev. Lett.* 110, 022003 Jan, 2013.
- [110] **ATLAS and CMS Collaboration**, “Top Quark Decay Properties” in *5th Large Hadron Collider Physics Conference (LHCP 2017) Shanghai, China, May 15-20, 2017*. 2017. [arXiv:1710.04277](https://arxiv.org/abs/1710.04277).
- [111] **CMS Collaboration**, “Observation of $t\bar{t}H$ production” 2018.
- [112] **ATLAS Collaboration**, “ATLAS finds evidence of the Higgs boson produced in association with a pair of top quarks.” <https://atlas.cern/updates/physics-briefing/atlas-finds-evidence-higgs-boson-produced-association-pair-top-quarks>. Date viewed: 31.01.2018.
- [113] **ATLAS Collaboration**, “Evidence for the associated production of the Higgs boson and a top quark pair with the ATLAS detector”, *Phys. Rev. D* 97, 072003, no. 7, 2018.
- [114] **ATLAS Collaboration**, “Search for the standard model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”, *Phys. Rev. D* 97, 072016 Apr, 2018.
- [115] **CMS Collaboration**, “Search for $t\bar{t}H$ production in the $H \rightarrow b\bar{b}$ decay channel with leptonic $t\bar{t}$ decays in proton-proton collisions at $\sqrt{s} = 13$ TeV” 2018.
- [116] G. Cowan et al., “Asymptotic formulae for likelihood-based tests of new physics”, *The European Physical Journal C* 71, 1554 Feb, 2011.
- [117] T. W. Anderson et al., “An introduction to multivariate statistical analysis”, volume 2. Wiley New York, 1958.
- [118] R. A. FISHER, “The use of multiple measurements in taxonomic problems”, *Annals of Eugenics* 7, 179–188 9, 1936.
- [119] A. Dresden, “The fourteenth western meeting of the American Mathematical Society”, *Bull. Amer. Math. Soc.* 26, 385–396 06, 1920.
- [120] A. Höcker et al., “TMVA - Toolkit for Multivariate Data Analysis”, *PoS ACAT*, 040 2007.
- [121] W. G. Cochran, “On the Performance of the Linear Discriminant Function”, *Technometrics* 6, 179–190, no. 2, 1964.
- [122] D. E. Rumelhart and J. L. McClelland, eds., “Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations”. MIT Press, Cambridge, MA, USA, 1986.
- [123] R. Hecht-Nielsen, “Theory of the backpropagation neural network” in *International 1989 Joint Conference on Neural Networks*, pp. 593–605 vol.1. 1989.
- [124] D. E. Rumelhart et al., “Backpropagation” in *Backpropagation*, Y. Chauvin and D. E. Rumelhart, eds., ch. Backpropagation: The Basic Theory, pp. 1–34. L. Erlbaum Associates Inc., Hillsdale, NJ, USA, 1995.

- [125] D. Clevert, T. Unterthiner, and S. Hochreiter, “Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)”, CoRR abs/1511.07289 2015.
- [126] M. Buckland and M. Collins, “AI techniques for game programming”. Premier press, 2002.
- [127] D. M. Kline and V. L. Berardi, “Revisiting squared-error and cross-entropy functions for training neural network classifiers”, *Neural Computing & Applications* 14, 310–318 Dec, 2005.
- [128] S. Ruder, “An overview of gradient descent optimization algorithms”, ArXiv e-prints September, 2016.
- [129] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization”, ArXiv e-prints December, 2014.
- [130] S. Han et al., “Learning both Weights and Connections for Efficient Neural Network” in *Advances in Neural Information Processing Systems 28*, C. Cortes et al., eds., pp. 1135–1143. Curran Associates, Inc., 2015.
- [131] T. Ježo et al., “New NLOPS predictions for $t\bar{t} + b$ -jet production at the LHC” technical report, University of Zürich, 2018.
- [132] A. Bredenstein et al., “NLO QCD corrections to $pp \rightarrow t\bar{t} + b\bar{b} + X$ at the LHC”, *Phys. Rev. Lett.* 103, 012002 2009.
- [133] G. Bevilacqua et al., “Assault on the NLO Wishlist: $pp \rightarrow t\bar{t} + b\bar{b}$ ”, *JHEP* 09, 109 2009.
- [134] A. Bredenstein et al., “NLO QCD corrections to $t\bar{t} + b\bar{b}$ production at the LHC: 2. Full hadronic results”, *Journal of High Energy Physics* 2010, 21 Mar, 2010.
- [135] F. Cascioli et al., “NLO matching for $t\bar{t} + b\bar{b}$ production with massive b-quarks”, *Physics Letters B* 734, 210 – 214 2014.
- [136] M. V. Garzelli, A. Kardos, and Z. Trócsányi, “ $t\bar{t} b\bar{b}$ hadroproduction at NLO accuracy matched with parton shower”, *PoS EPS-HEP2013*, 253 2013.
- [137] M. V. Garzelli, A. Kardos, and Z. Trócsányi, “Hadroproduction of $t\bar{t}b\bar{b}$ final states at LHC: predictions at NLO accuracy matched with Parton Shower”, *JHEP* 03, 083 2015.
- [138] S. Höche et al., “A critical appraisal of NLO+PS matching methods”, *Journal of High Energy Physics* 2012, 49 Sep, 2012.
- [139] P. Skands, S. Carrazza, and J. Rojo, “Tuning PYTHIA 8.1: the Monash 2013 Tune”, *Eur. Phys. J. C* 74, 3024, no. 8, 2014.
- [140] V. Hirschi et al., “Automation of one-loop QCD computations”, *Journal of High Energy Physics* 2011, 44 May, 2011.
- [141] G. Bevilacqua et al., “HELAC-NLO”, *Computer Physics Communications* 184, 986 – 997, no. 3, 2013.
- [142] A. Kardos and Z. Trócsányi, “Hadroproduction of t anti- t pair with a b anti- b pair using PowHel”, *Journal of Physics G: Nuclear and Particle Physics* 41, 075005, no. 7, 2014.
- [143] T. Gleisberg et al., “Event generation with SHERPA 1.1”, *Journal of High Energy Physics* 2009, 007, no. 02, 2009.

- [144] F. Cascioli, P. Maierhöfer, and S. Pozzorini, “Scattering Amplitudes with Open Loops”, *Phys. Rev. Lett.* 108, 111601 Mar, 2012.
- [145] F. Cascioli et al., “Precise Higgs-background predictions: merging NLO QCD and squared quark-loop corrections to four-lepton + 0,1 jet production”, *Journal of High Energy Physics* 2014, 46 Jan, 2014.
- [146] N. Bartosik, “Associated top-quark-pair and b-jet production in the dilepton channel at $\sqrt{s} = 8$ TeV as test of QCD and background $t\bar{t}$ +Higgs production”. PhD thesis, University of Hamburg, Hamburg, 2015.
- [147] W. Hollik and M. Kollár, “NLO QED contributions to top-pair production at hadron colliders”, *Phys. Rev. D* 77, 014008 Jan, 2008.
- [148] J. H. Kühn, A. Scharf, and P. Uwer, “Electroweak corrections to top-quark pair production in quark-antiquark annihilation”, *The European Physical Journal C - Particles and Fields* 45, 139–150 Jan, 2006.
- [149] **CMS Collaboration**, “GenHFWadronMatcher - revision 30 from 30.01.2017.” <https://twiki.cern.ch/twiki/bin/view/CMSPublic/GenHFWadronMatcher>. Date viewed: 13.02.2018, only internally accessible.
- [150] **ATLAS Collaboration**, “Search for the Standard Model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”, Submitted to: *Phys. Rev. D* 2017.
- [151] R. Odorico, “Telling top jets from QCD jets using energy flow”, *Physics Letters B* 120, 219 – 223, no. 1, 1983.
- [152] R. Marshall, “The separation of quark flavours in e^+e^- annihilation and its applications”, *Zeitschrift für Physik C Particles and Fields* 26, 291–299 Nov, 1984.
- [153] B. Denby, “Neural networks and cellular automata in experimental high energy physics”, *Computer Physics Communications* 49, 429 – 448, no. 3, 1988.
- [154] C. Peterson, “Track finding with neural networks”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 279, 537 – 545, no. 3, 1989.
- [155] C. Peterson and T. Rognvaldsson, “An Introduction to artificial neural networks” in *1991 CERN School of Computing Ystad, Sweden, August 23-September 2, 1991*, pp. 0113–170. 1991.
- [156] L. Lönnblad, C. Peterson, and T. Rognvaldsson, “Using neural networks to identify jets”, *Nuclear Physics B* 349, 675 – 702, no. 3, 1991.
- [157] L. Lönnblad, C. Peterson, and T. Rognvaldsson, “Pattern recognition in high energy physics with artificial neural networks — JETNET 2.0”, *Computer Physics Communications* 70, 167 – 182, no. 1, 1992.
- [158] C. Peterson, T. Rognvaldsson, and L. Lönnblad, “JETNET 3.0—A versatile artificial neural network package”, *Computer Physics Communications* 81, 185 – 220, no. 1, 1994.
- [159] B. Brandl et al., “Multivariate analysis methods to tag b-quark events at LEP/SLC”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 324, 307 – 316, no. 1, 1993.

- [160] M. Feindt and U. Kerzel, “The NeuroBayes neural network package”, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 559, 190 – 194, no. 1, 2006.
- [161] W. A. Belson, “Matching and Prediction on the Principle of Biological Classification”, Journal of the Royal Statistical Society. Series C (Applied Statistics) 8, 65–75, no. 2, 1959.
- [162] C. E. Brodley and P. E. Utgoff, “Multivariate decision trees”, Machine Learning 19, 45–77 Apr, 1995.
- [163] **CDF Collaboration**, “Measurement of the cross section for $t\bar{t}$ production in $p\bar{p}$ collisions using the kinematics of lepton + jets events”, Phys. Rev. D72, 052003 2005.
- [164] M. Abadi et al., “TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems” 2015. Software available from tensorflow.org.
- [165] “Die Geschichte von NVIDIA - Von der Grafikkarte zum Mobilprozessor.” <http://www.nvidia.de/object/corporate-timeline-de.html>. Date viewed: 05.02.2018.
- [166] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning”, Nature 521, 436, no. 7553, 2015.
- [167] “The Intelligent Industrial Revolution.” <https://blogs.nvidia.com/blog/2016/10/24/intelligent-industrial-revolution/>. Date viewed: 20.03.2018.
- [168] **CMS Collaboration**, “Identification of b-quark jets with the CMS experiment”, Journal of Instrumentation 8, P04013, no. 04, 2013.
- [169] **CMS Collaboration**, “Performance of heavy flavour identification algorithms in proton-proton collisions at 13 TeV at the CMS experiment” Technical Report CMS-DP-2017-012, CERN, Geneva, 2017.
- [170] **ATLAS Collaboration**, “Measurements of fiducial cross-sections for $t\bar{t}$ production with one or two additional b-jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector”, The European Physical Journal C 76, 11 Jan, 2016.
- [171] S. Alioli et al., “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX”, JHEP 06, 043 2010.
- [172] M. Cacciari et al., “Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation”, Physics Letters B 710, 612 – 622, no. 4, 2012.
- [173] P. Bärnreuther, M. Czakon, and A. Mitov, “Percent-Level-Precision Physics at the Tevatron: Next-to-Next-to-Leading Order QCD Corrections to $q\bar{q} \rightarrow t\bar{t}+X$ ”, Phys. Rev. Lett. 109, 132001 Sep, 2012.
- [174] M. Czakon and A. Mitov, “NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels”, Journal of High Energy Physics 2012, 54 Dec, 2012.
- [175] M. Czakon and A. Mitov, “NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction”, Journal of High Energy Physics 2013, 80 Jan, 2013.
- [176] M. Czakon, P. Fiedler, and A. Mitov, “Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $\mathcal{O}(\alpha_s^4)$ ”, Phys. Rev. Lett. 110, 252004 Jun, 2013.

- [177] M. Czakon and A. Mitov, “Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders”, *Comput. Phys. Commun.* 185, 2930 2014.
- [178] **CMS Collaboration**, “Measurement of $t\bar{t}$ production with additional jet activity, including b quark jets, in the dilepton decay channel using pp collisions at $\sqrt{s} = 8$ TeV”, *The European Physical Journal C* 76, 379 Jul, 2016.
- [179] J. Alwall et al., “MadGraph 5: going beyond”, *Journal of High Energy Physics* 2011, 128 Jun, 2011.
- [180] P. Artoisenet et al., “Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations”, *JHEP* 03, 015 2013.
- [181] G. Corcella et al., “HERWIG 6: an event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)”, *Journal of High Energy Physics* 2001, 010, no. 01, 2001.
- [182] **CMS Collaboration**, “Measurements of $t\bar{t}$ cross sections in association with b jets and inclusive jets and their ratio using dilepton final states in pp collisions at $s=13\text{TeV}$ ”, *Physics Letters B* 776, 355 – 378 2018.
- [183] T. Sjöstrand, S. Mrenna, and P. Skands, “A brief introduction to PYTHIA 8.1”, *Computer Physics Communications* 178, 852 – 867, no. 11, 2008.
- [184] G. Bevilacqua and M. Worek, “On the ratio of $t\bar{t}b\bar{b}$ and $t\bar{t}jj$ cross sections at the CERN Large Hadron Collider”, *JHEP* 07, 135 2014.
- [185] **CMS Collaboration**, “Event generator tunes obtained from underlying event and multiparton scattering measurements”, *The European Physical Journal C* 76, 155 Mar, 2016.
- [186] **NNPDF Collaboration**, “Parton distributions for the LHC Run II”, *JHEP* 04, 040 2015.
- [187] **CMS Collaboration**, “CMS Luminosity Measurements for the 2016 Data Taking Period” Technical Report CMS-PAS-LUM-17-001, CERN, Geneva, 2017.
- [188] M. Aliev et al., “HATHOR – HAdronic Top and Heavy quarks crOss section calculator”, *Computer Physics Communications* 182, 1034 – 1046, no. 4, 2011.
- [189] P. Kant et al., “HatHor for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions”, *Computer Physics Communications* 191, 74 – 89 2015.
- [190] K. Melnikov and F. Petriello, “Electroweak gauge boson production at hadron colliders through $\mathcal{O}(\alpha_s^2)$ ”, *Phys. Rev. D* 74, 114017 Dec, 2006.
- [191] R. Gavin et al., “FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order”, *Computer Physics Communications* 182, 2388 – 2403, no. 11, 2011.
- [192] Y. Li and F. Petriello, “Combining QCD and electroweak corrections to dilepton production in the framework of the FEWZ simulation code”, *Phys. Rev. D* 86, 094034 Nov, 2012.
- [193] R. Gavin et al., “W Physics at the LHC with FEWZ 2.1”, *Comput. Phys. Commun.* 184, 208–214 2013.
- [194] T. Gehrmann et al., “ W^+W^- Production at Hadron Colliders in Next to Next to Leading Order QCD”, *Phys. Rev. Lett.* 113, 212001, no. 21, 2014.

- [195] J. M. Campbell and R. K. Ellis, “Update on vector boson pair production at hadron colliders”, *Phys. Rev. D* 60, 113006 Nov, 1999.
- [196] J. M. Campbell, R. K. Ellis, and C. Williams, “Vector boson pair production at the LHC”, *Journal of High Energy Physics* 2011, 18 Jul, 2011.
- [197] **CMS Collaboration**, “CMS Offline Software.” <https://cms-sw.github.io/index.html>. Date viewed: 20.02.2018.
- [198] **CMS Collaboration**, “SWGuideBTagMCTools - revision 41 from 16.05.2016.” <https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideBTagMCTools>. Date viewed: 13.02.2018, only internally accessible.
- [199] **CMS Collaboration**, “pp Collision16 runs, BCDEFG era (23 Sep ReReco) + H era Prompt deco - FINAL json file.” <https://hypernews.cern.ch/HyperNews/CMS/get/physics-validation/2766.html>. Date viewed: 19.02.2018, only internally accessible.
- [200] **CMS Collaboration**, “Luminosity and normtags 2016 data.” <https://hypernews.cern.ch/HyperNews/CMS/get/physics-announcements/4495.html>. Date viewed: 19.02.2018, only internally accessible.
- [201] S. van der Meer, “Calibration of the effective beam height in the ISR” Technical Report CERN-ISR-PO-68-31. ISR-PO-68-31, CERN, Geneva, 1968.
- [202] M. Lamont et al., “Luminosity Optimization and Calibration in the LHC” May, 2009.
- [203] **CMS Collaboration**, “ttH(bb) Analysis Definitions for Moriond17.” <https://gitlab.cern.ch/ttH/reference/blob/master/definitions/Moriond17.md>. Date viewed: 19.02.2018, only internally accessible.
- [204] **CMS Collaboration**, “Search for ttH production in the $H \rightarrow b\bar{b}$ decay channel with $\sqrt{s} = 13$ TeV pp collisions at the CMS experiment” Technical Report CMS-PAS-HIG-16-004, CERN, Geneva, 2016.
- [205] **CMS Collaboration**, “Search for ttH production in the $H \rightarrow b\bar{b}$ decay channel with 2016 pp collision data at $\sqrt{s} = 13$ TeV” Technical Report CMS-PAS-HIG-16-038, CERN, Geneva, 2016.
- [206] M. Ramos, “Continuous Integration, Delivery, and Deployment with GitLab - Aug 5, 2016.” <https://about.gitlab.com/2016/08/05/continuous-integration-delivery-and-deployment-with-gitlab/>. Date viewed: 20.02.2018.
- [207] M. Weiss, “The Benefits of Continuous Integration - Last updated: 2017-06-19.” <https://blog.codeship.com/benefits-of-continuous-integration/>. Date viewed: 20.02.2018.
- [208] A. Taylor, “5 Advantages of Continuous Integration - May 22, 2017.” <https://pantheon.io/blog/5-advantages-continuous-integration>. Date viewed: 20.02.2018.
- [209] **KIT Campus Nord analysis group**, “NNFlow - a framework to train binary and multiclass neural networks for classification of events in particle physics.” <https://github.com/kit-cn-cms/NNFlow>. Date viewed: 22.02.2018.
- [210] **CMS Collaboration**, “Search for ttH, $H \rightarrow b\bar{b}$ decays using the full 2016 data sample” CMS Note 2017/063, CERN, 2017.

- [211] F. James and M. Roos, “Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations”, *Comput. Phys. Commun.* 10, 343–367 1975.
- [212] **CMS Collaboration**, “CMS Higgs Combination toolkit.” <https://cms-hcomb.gitbooks.io/combine/>. Date viewed: 24.02.2018.
- [213] **CMS Collaboration**, “CMS CombineHarvester framework.” <https://cms-analysis.github.io/CombineHarvester/>. Date viewed: 24.02.2018.
- [214] R. Brun and F. Rademakers, “ROOT – An object oriented data analysis framework”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 389, 81 – 86, no. 1, 1997.
- [215] E. Bagnaschi and A. Vicini, “The Higgs transverse momentum distribution in gluon fusion as a multiscale problem”, *Journal of High Energy Physics* 2016, 56 Jan, 2016.
- [216] **CMS Collaboration**, “Utilities for Accessing Pileup Information for Data - revision 28.” <https://twiki.cern.ch/twiki/bin/view/CMS/PileupJSONFileforData?rev=28>. Date viewed: 25.02.2018, only internally accessible.
- [217] **CMS Collaboration**, “Instructions for applying electron and photon ID - revision 35.” https://twiki.cern.ch/twiki/bin/view/CMS/EgammaIDRecipesRun2#Electron_efficiencies_and_scale. Date viewed: 24.02.2018, only internally accessible.
- [218] **CMS Collaboration**, “Work in progress (evolving results) and dedicated studies from PAGs - revision 16.” https://twiki.cern.ch/twiki/bin/viewauth/CMS/MuonWorkInProgressAndPagResults#Work_in_progress_on_central_resu. Date viewed: 24.02.2018, only internally accessible.
- [219] **CMS Collaboration**, “JECUncertaintySources TWiki.” <https://twiki.cern.ch/twiki/bin/view/CMS/JECUncertaintySources>. Date viewed: 25.02.2018, only internally accessible.
- [220] **CMS Collaboration**, “Jet Energy Resolution - revision 54.” https://twiki.cern.ch/twiki/bin/view/CMS/JetResolution#JER_Scaling_factors_and_Uncertai. Date viewed: 24.02.2018, only internally accessible.
- [221] **CMS Collaboration**, “Performance of b tagging algorithms in 2016 data at 13 TeV” CMS Note 2017/018, CERN, 2017.
- [222] **CMS Collaboration**, “Calibration of the Combined Secondary Vertex b-Tagging discriminant using dileptonic $t\bar{t}$ and Drell-Yan events” CMS Note 2013/130, CERN, 2013.
- [223] N. Moretti et al., “Measuring the signal strength in $t\bar{t}H$ with $H \rightarrow b\bar{b}$ ”, *Phys. Rev. D* 93, 014019 Jan, 2016.
- [224] **CMS Collaboration**, “Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of $t\bar{t}$ at $\sqrt{s} = 8$ and 13 TeV” Technical Report CMS-PAS-TOP-16-021, CERN, Geneva, 2016.
- [225] **CMS Collaboration**, “CMS Luminosity – Public Results.” <https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults>. Date viewed: 25.05.2018.

Appendix

A Pull and impact distributions of the three reference analyses

In the following the combined pull and impact distributions containing the 30 systematic uncertainties which have the largest impact on the $t\bar{t} + \geq 1\text{b-jet}$ signal strength result are presented for the three reference analyses.

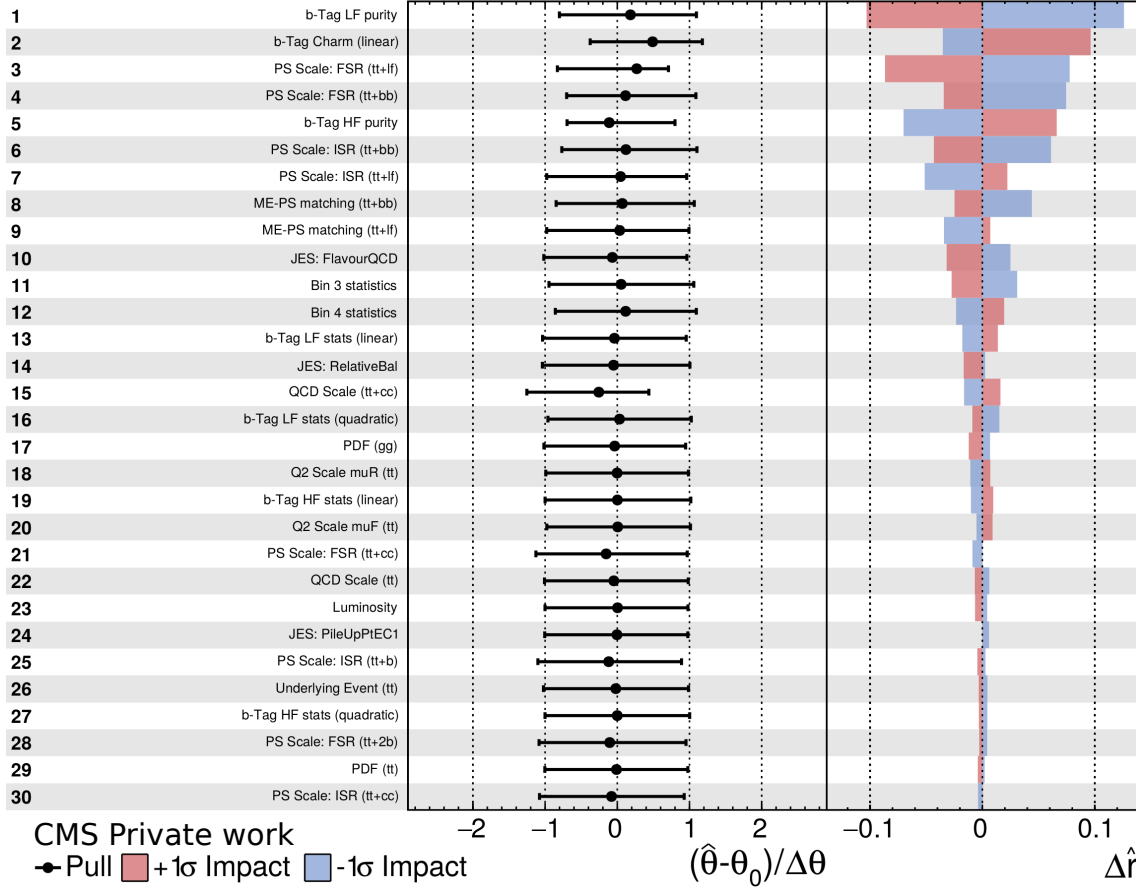


Figure A.1: Pull and impact distribution of the B-jet multiplicity based analysis: The pull of a nuisance parameter θ (black point) corresponds to the difference between its initial value and its best-fit value after the fit divided by the uncertainty on it before the fit $\frac{\hat{\theta} - \theta_0}{\Delta\theta}$. The 1-sigma confidence interval according to its prefit value is shown as a black line. The impact of a nuisance parameter on the $t\bar{t} + \geq 1\text{b-jet}$ signal strength result $\Delta\hat{r}$ is obtained by repeating the fit while fixing this parameter to the upper or lower bound of its nominal postfit 1-sigma confidence interval. Positive (blue/red) or negative correlations (red/blue) between the nuisance parameter and the signal strength result can be concluded from the coloured bars. In the plot the 30 systematic uncertainties having the largest impact are presented.

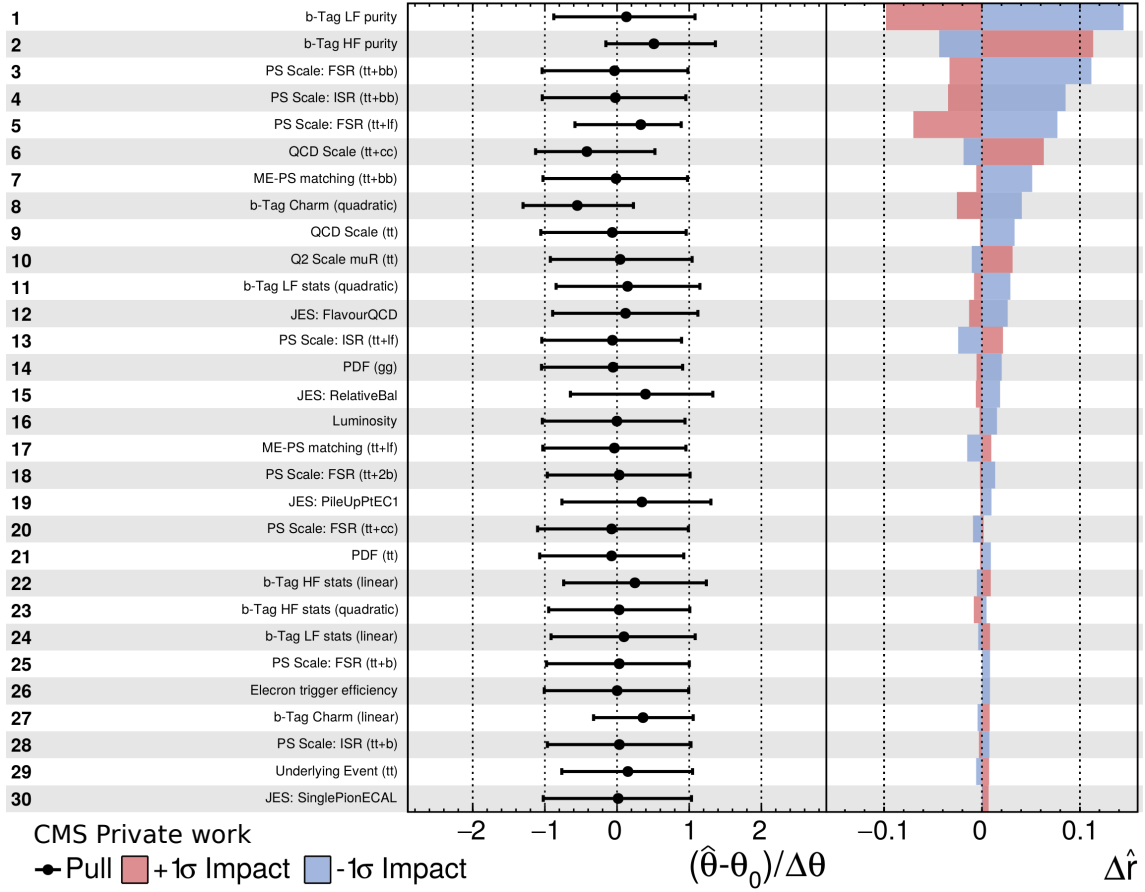


Figure A.2: Pull and impact distribution of the linear discriminant based analysis: The pull of a nuisance parameter θ (black point) corresponds to the difference between its initial value and its best-fit value after the fit divided by the uncertainty on it before the fit $\frac{\hat{\theta} - \theta_0}{\Delta\theta}$. The 1σ -confidence interval according to its prefit value is shown as a black line. The impact of a nuisance parameter on the $t\bar{t} + \geq 1b$ -jet signal strength result Δr is obtained by repeating the fit while fixing this parameter to the upper or lower bound of its nominal postfit 1σ -confidence interval. Positive (blue/red) or negative correlations (red/blue) between the nuisance parameter and the signal strength result can be concluded from the coloured bars. In the plot the 30 systematic uncertainties having the largest impact are presented.

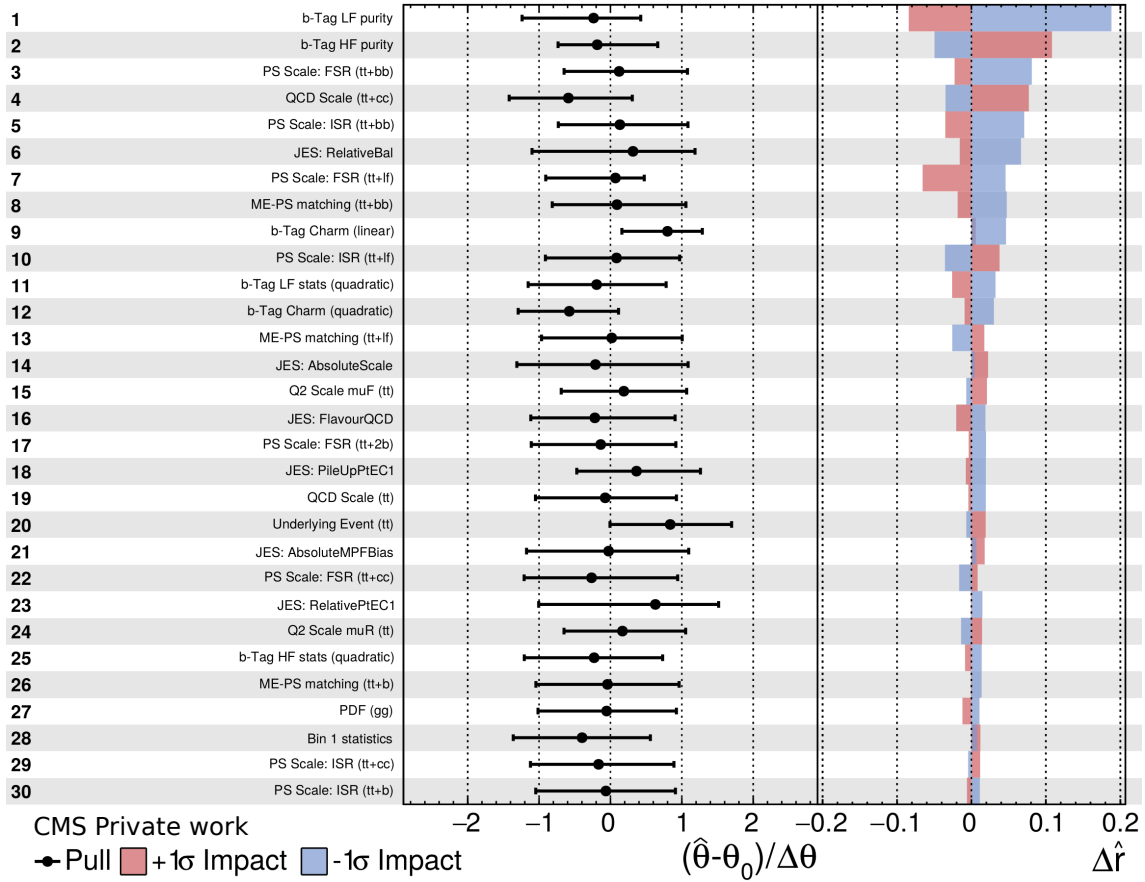


Figure A.3: Pull and impact distribution of the neural network based analysis: The pull of a nuisance parameter θ (black point) corresponds to the difference between its initial value and its best-fit value after the fit divided by the uncertainty on it before the fit $\frac{\hat{\theta} - \theta_0}{\Delta\theta}$. The 1-sigma confidence interval according to its prefit value is shown as a black line. The impact of a nuisance parameter on the $t\bar{t} + \geq 1b$ -jet signal strength result Δr is obtained by repeating the fit while fixing this parameter to the upper or lower bound of its nominal postfit 1-sigma confidence interval. Positive (blue/red) or negative correlations (red/blue) between the nuisance parameter and the signal strength result can be concluded from the coloured bars. In the plot the 30 systematic uncertainties having the largest impact are presented.