ETP-KA/2024-11

Simultaneous cross section measurements of top quark-antiquark pair production with additional heavy flavor jets at the CMS experiment

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN (Dr. rer. nat.)

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

DISSERTATION

von

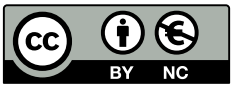
M.Sc. Emanuel Lorenz Pfeffer

aus Kassel

Tag der mündlichen Prüfung: 24.01.2025

Referent:	Prof. Dr. Ulrich Husemann	Institute of Experimental Particle Physics
Korreferent:	Prof. Dr. Thomas Müller	Institute of Experimental Particle Physics

CERN-THESIS-2025-002
24/01/2025



Erklärung:

Ich versichere wahrheitsgemäß, die Arbeit selbstständig verfasst, alle benutzten Hilfsmittel vollständig und genau angegeben und alles kenntlich gemacht zu haben, was aus Arbeiten anderer unverändert oder mit Abänderungen entnommen wurde sowie die Satzung des KIT zur Sicherung guter wissenschaftlicher Praxis in der Fassung vom 27.03.2017 beachtet zu haben.

Karlsruhe, den 18.12.2024

. . . . . .

Emanuel L. Pfeffer

Abstract

This thesis presents a simultaneous cross section measurement of top quark-antiquark pair ($t\bar{t}$) production in association with b jets, c jets, a Higgs boson, or a Z boson in the $H/Z \to b\bar{b}$ decay mode with exactly two charged leptons (e,μ), targeting the dilepton $t\bar{t}$ decay channel. The measurement is performed at the CMS experiment at the CERN LHC in proton-proton collisions at a center-of-mass energy of 13 TeV and the analyzed data corresponds to an integrated luminosity of approximately $60\,\mathrm{fb}^{-1}$. Collision events are modeled as a mathematical graph structure and processed using graph transformer neural network architectures based on multi-head attention mechanisms to perform multi-class classification. The measured cross sections are parameterized as signal strength parameters relative to the cross sections predicted by the Standard Model of particle physics (SM). Four signal strength parameters are simultaneously extracted in a maximum likelihood fit to data and result to $\mu^{\mathrm{obs}}_{t\bar{t}B} = 0.98^{+0.34}_{-0.25}$, $\mu^{\mathrm{obs}}_{t\bar{t}C} = 0.74^{+0.41}_{-0.41}$, $\mu^{\mathrm{obs}}_{t\bar{t}H} = 0.89^{+0.95}_{-0.93}$, and $\mu^{\mathrm{obs}}_{t\bar{t}Z} = 1.28^{+1.15}_{-1.06}$, corresponding to an observed (expected) significance of 15σ (16σ) for $t\bar{t}B$, 1.8σ (2.9σ) for $t\bar{t}C$, 1.0σ (1.1σ) for $t\bar{t}H$, and 1.2σ (1.0σ) for $t\bar{t}Z$ compared to the SM prediction without these processes. The results reveal a good agreement with the SM expectation for the signal strength of one within the 68% confidence interval.

Contents

Part I

Introduction

1 Overview

The quest to uncover the most fundamental constituents of nature has driven humanity throughout its entire history. The Standard Model of particle physics (SM) is a relativistic quantum field theory that describes the currently known elementary particles and the interactions between them. It describes three of the four known fundamental interactions, namely the electromagnetic, weak, and strong interactions, while gravity is not part of the theory. The SM has withstood numerous precision measurements and to date has not been falsified by any attempts to disprove it. One of the most significant accomplishments of the SM was the prediction of the Higgs boson [1–6], which was ultimately confirmed by the ATLAS and CMS Collaborations [7, 8]. However, the SM is limited due to its inability to explain certain phenomena such as Dark Matter, Dark Energy, the gravitational force, or size of the matter-antimatter asymmetry in the universe.

Two intriguing particles in the SM are the top quark and the Higgs boson. The top quark is the heaviest known particle in the SM and provides crucial insights into the interactions between quarks and the origin of mass. The Higgs boson, on the other hand, is pivotal for understanding the Higgs field, which imparts mass to fundamental particles and thereby determine the structure of the universe. Processes in which the top quark couples to the Higgs boson are of particular interest as they allow for the direct study of interactions between these two heaviest known particles. One such process is the top quark-antiquark pair production ($t\bar{t}$) in association with a Higgs boson ($t\bar{t}H$). These interactions offer to precisely determine the top quark's coupling to the Higgs boson, and may provide hints of physics beyond the Standard Model (BSM). The dominant decay mode of the Higgs boson is the decay into a pair of bottom quarks (b), denoted as $H(b\bar{b})$. However, investigating the $t\bar{t}H(b\bar{b})$ process rises significant challenges. Due to the relatively small predicted cross section of $\sigma_{t\bar{t}H(b\bar{b})} = 0.29\,\text{pb}$, this process is difficult to measure, as it is irreducibly overshadowed by other processes. Notably, the production of $t\bar{t}$ in association with an

additional b quark-antiquark pair ($t\bar{t}b\bar{b}$) has a predicted production cross section that is approximately ten times higher. Moreover, distinguishing $t\bar{t}H(b\bar{b})$ from the associated production of $t\bar{t}$ with a charm quark-antiquark pair ($t\bar{t}c\bar{c}$), where the flavor of the additional charm quark must be precisely identified, is an additional difficulty. Another irreducible background process is the production of $t\bar{t}$ in association with a Z boson and a subsequent decay of $Z \to b\bar{b}$. Despite being undesirable backgrounds, all processes are inherently intriguing. A precise investigation of these processes provides insights into the nature of quantum chromodynamics (QCD) and could shed light on BSM physics.

The objective of this thesis is a simultaneous measurement of all four processes. This allows for a homogeneous treatment of all processes as well as their uncertainties while all processes are treated with equal importance. The analysis strategy is built upon distinguishing the processes through a two-stage, multi-class classification approach, leveraging machine learning techniques as a branch of Artificial Intelligence (AI). Graph transformer architectures based on multi-head attention mechanisms are employed, as used in large-language models or foundation models across various domains. In the two-stage process, the jets of an event are first assigned to their respective origin, with this information then being used in a subsequent multi-class classification procedure for event classification. The signal strength parameters of the four processes are extracted through maximum likelihood fits to binned distributions of data and statistical inference methods.

The measurement is conducted at the Compact Muon Solenoid (CMS) experiment at the Large Hadron Collider (LHC) at Conseil Européen pour la Recherche Nucléaire (CERN). The analyzed data correspond to an integrated luminosity of $59.83\,\mathrm{fb}^{-1}$ in proton-proton collisions at a center-of-mass energy at $\sqrt{s} = 13\,\mathrm{TeV}$ during the data-taking in LHC Run-II. Collision events are included in the data analysis if they feature exactly two charged leptons, i.e. electrons or muons. This selection targets the decay channel of the $t\bar{t}$ system where both top quark decays include leptonic processes.

The thesis is structured in eleven chapters. This introduction forms the starting point. In Chapter 2, theoretical foundations are discussed, such as the SM based on relativistic quantum field theories, along with an experimental perspective introducing elementary particles. Chapter 3 focuses on the experimental setup, starting with an overview of the LHC and a discussion of the CMS experiment whose data are studied in this thesis. Chapter 4 explains the object reconstruction process, which translates detector data into physics objects and observables, enabling the inference of underlying physics processes during a collision. Additionally, the process of event simulation is discussed in this chapter, which is used to model and predict detector signals. In Chapter 5, the basic concepts of machine learning are introduced, playing a central role in differentiating between processes in this thesis. Chapter 6 presents key concepts of statistical data analysis, which form the basis for drawing conclusions about the analyzed processes in precision measurements at the end of this analysis. Chapter 7 offers an in-depth discussion of the motivation and

challenges of this thesis, positioning the thesis within the landscape of current research, and outlines the analysis strategy and goals of this thesis. In Chapter 8, the phase space and all physics objects of the analysis are defined, followed by a qualitative comparison between predictions based on simulated events and the recorded data. Chapter 9 provides a detailed explanation of the data set and the architecture of the AI model based on graph transformer neural networks. In Chapter 10, a statistical model is developed to infer the underlying processes from the data. The statistical model is thoroughly examined, with a comprehensive quantitative comparison between predictions based on simulated events and the recorded data. Finally, the key parameters are determined using maximum likelihood estimation and statistical inference techniques. The thesis concludes with a summary of the results in Chapter 11.

Part II

Foundations

2 Theoretical fundamentals

In this chapter, the key theoretical concepts essential for the subsequent thesis are discussed. The primary focus is on the Standard Model of particle physics (SM), as this thesis is a precision measurement of certain aspects of the SM. The chapter is structured as follows: In Section 2.1, the SM is introduced theoretically and formulated as a relativistic quantum field theory. Section 2.2 provides an experimental perspective, describing all known fundamental particles and their properties. Section 2.3 is dedicated specifically to the top quark, which plays a central role in this work. Finally, Section 2.4 addresses important aspects of physics at hadron colliders, including the factorization theorem, and concludes with a specific discussion for the $t\bar{t}b\bar{b}$ process, which is pivotal in this thesis.

2.1 Standard model of particle physics

The SM is a theoretical model describing the constituents of matter and their interactions. This fundamental theory describes all known elementary particles and their interactions through three of the four fundamental forces. The SM accounts for the strong interaction, the weak interaction, and the electromagnetic interaction, excluding gravity. The success of the SM lies in the fact that it underwent extensive testing and validation across numerous experimental aspects, e.g. the observation of the Higgs boson [7, 8]. However, some phenomena such as the aforementioned gravitation, Dark Matter, Dark Energy, or the mass of neutrinos are not accounted by the SM and require extended models, often summarized as physics beyond the Standard Model (BSM). The SM is designed as a relativistic, renormalizable quantum field theory based on symmetry principles of gauge theories. The underlying principles are outlined in the following sections, based on Ref. [9].

2.1.1 Quantum electrodynamics

In quantum field theories, particles are understood as manifestations of field excitations. In the context of a field theory, the Lagrangian density serves as the field-theoretic analog

of the Lagrangian function in classical mechanics, forming the basis of the Lagrangian formalism. The Lagrangian density $\mathcal{L}(\phi(x), \partial_\mu \phi(x))$ is a function of the field $\phi(x)$ and its derivative $\partial_\mu \phi(x)$, where x denotes space-time coordinates. For example, the Lagrangian density of quantum electrodynamics (QED) describes interactions with the photon field and allows for calculations of the interaction between fermions, e.g. electrons and photons [10–15]. Free fermions are spin[1]-1/2 particles and satisfy the Dirac equation, which is a first-order differential equation for relativistic particles that have spin-1/2 [16]. According to the Dirac equation, the Lagrangian density becomes

$$\mathcal{L}_{\text{fermion}} = \bar{\psi}(x)\left(i\gamma^\mu \partial_\mu - m\right)\psi(x) \quad , \tag{2.1}$$

where $\psi(x)$ denotes the Dirac spinor, $\bar{\psi}(x) = \psi^\dagger(x)\gamma^0$ refers to the Dirac adjoint spinor, γ^μ are the Dirac matrices, and m denotes the fermion mass. With respect to global and local symmetry principles, the equation is invariant under a global unitary transformation in a one-dimensional complex vector space $U(1)$. To impose the invariance of the system under a local $U(1)$ transformation, it is necessary to introduce the substitution

$$\partial_\mu \rightarrow D_\mu = \partial_\mu + iqA_\mu \quad , \tag{2.2}$$

where A_μ denotes a gauge field and D_μ is the covariant derivative. The constant q emerges from the continuous symmetry and, according to Noether's theorem, leads to a conserved quantity [17].

To obey local gauge symmetry, the gauge field A_μ is transformed according to

$$A_\mu(x) \rightarrow A_\mu(x) - \frac{1}{q}\partial_\mu \varphi(x) \quad , \tag{2.3}$$

where $\varphi(x)$ is a space-time dependent local phase and A_μ is the vector field associated with the gauge boson [18]. Bosons, in contrast to fermions, are particles characterized by having an integer spin value, such as 0, 1, or 2. In QED, A_μ is interpreted as the photon field with the Lagrangian density

$$\mathcal{L}_{\text{kin}} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad , \tag{2.4}$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ denotes the electromagnetic field strength tensor. Thus, the Lagrangian density describes the dynamics of the electromagnetic field. Together with a term $-q\bar{\psi}\gamma^\mu A_\mu \psi$ to describe the interaction of the fermion ψ with the photon field A_μ, the Lagrangian density of QED becomes

[1] The spin is a fundamental property of particles describing the intrinsic angular momentum in quantum mechanics.

$$\mathcal{L}_{\text{QED}} = i\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi - m\bar{\psi}\psi - qA_{\mu}\bar{\psi}\gamma^{\mu}\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad . \tag{2.5}$$

Hence, the coupling is proportional to the constant q, interpreted as the electric charge, m refers to the mass of the electron and the photon remains massless.

2.1.2 Quantum chromodynamics

To describe the strong interaction, the concept of gauge field theories used for QED is extended to non-Abelian Lie-groups [19–21]. Quantum chromodynamics (QCD) is based on the special unitary group $SU(3)$, which consists of eight generators and leads to eight different gauge bosons. In the $SU(3)_C$ group of QCD, the three charges (C) are associated with the fundamental particles, namely quarks and gluons, where quarks and gluons carry color charges and gluons mediate the strong force between them. The term ''chromo'' in QCD refers to the metaphorical use of color charges, which are denoted as red, green, and blue as well as their corresponding anti-colors.

As stated before, the Lagrangian density of the theory must remain invariant under both global and local $SU(3)$ transformations. Analogously to QED, gauge fields A^a_μ are introduced representing gluon fields. The covariant derivative is defined as $D_\mu = \partial_\mu - ig_s T^a A^a_\mu$, where g_s is a coupling constant, and T^a are the generators of the $SU(3)$ group. The dynamics of gluons are described by the field strength tensor

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g_s f^{abc} A^b_\mu A^c_\nu \quad , \tag{2.6}$$

where f^{abc} are the structure constants of the $SU(3)$ group. Finally, the Lagrangian density of QCD is given by

$$\mathcal{L}_{\text{QCD}} = \sum_{\text{q}} \bar{\psi}_{\text{q},i}(i\gamma^{\mu}\partial_{\mu} - m_{\text{q}})\psi_{\text{q},i} - g_{\text{s}}\left(\bar{\psi}_{\text{q},i}\gamma^{\mu}T^a_{ij}\psi_{\text{q},j}\right)A^a_\mu - \frac{1}{4}F^a_{\mu\nu}F^{\mu\nu}_a \quad , \tag{2.7}$$

where the generators T^a_{ij} are represented by the Gell-Mann matrices. The first two terms describe free quarks q with mass m_q, similar to electrons in QED, while the third term denotes the quark-gluon coupling, proportional to the coupling g_{s}. The last term represents the gluon propagation and self-interaction. Similar to the Lagrangian density of QED, Equation 2.7 does not include mass terms for the gauge bosons, implying that gluons, like photons, are massless. However, in contrast to QED, gluons carry the charge associated with their quantum field theory, i.e. the color charge. This leads to not only an interaction of gluons with quarks, but also the self-interaction as reflected in the Lagrangian density. This also leads to a phenomena known as confinement, where quarks and gluons are confined within hadrons and cannot exist as free particles at low energies or the equivalent

of large distances [22–24]. At high energies, or short distances, quarks behave effectively as free particles and the coupling becomes weaker. This characteristic is known as asymptotic freedom. Together with quantum fluctuations, this leads to an energy dependence of the coupling constants, which is referred to as running coupling. The coupling constant of the strong force α_S, which is frequently referenced throughout this thesis, is related to the coupling g_s in Equation 2.7 via the relation $\alpha_S = \frac{g_s^2}{4\pi}$.

2.1.3 Weak interaction and electroweak unification

When a neutron decays, a proton, an electron, and an anti-neutrino are produced. This process known as beta decay is mediated by the weak interaction [25]. In principle, the Lagrangian density for weak interactions can be derived in a similar way to those of QED and QCD, but the construction of a quantum field theory poses some challenges. It has been experimentally proven that only left-handed particles and right-handed anti-particles interact weakly, requiring a chiral theory [26, 27]. Furthermore, the electromagnetic and weak forces can be seen as different aspects of the same interaction at high energies and are therefore described using a combined gauge group. The concept of a unified discription of the electromagnetic and weak interactions is referred to as electroweak unification [28–30]. The symmetry group of the electroweak interaction is $SU(2)_L \times U(1)_Y$, where L represents the chiral characteristic of left-handed particles and Y is the weak hypercharge of the $U(1)$ gauge group. The conserved quantity of the $SU(2)_L$ group is the weak isospin T, i.e. a quantum number associated with the weak interaction [17]. The electric charge Q of a particle is related to the weak isospin and hypercharge via the Gell-Mann–Nishijima relation

$$Q = T_3 + \frac{Y}{2} \quad , \tag{2.8}$$

where T_3 is the third component of weak isospin for the particle, i.e. $\pm 1/2$ for weak isospin doublets [31–33].

The covariant derivative is given by

$$D_\mu \phi = (\partial_\mu - ig\frac{\sigma^i}{2}W_\mu^i - ig'\frac{Y}{2}B_\mu)\phi \quad , \tag{2.9}$$

where W_μ^a are the $SU(2)_L$ gauge fields, B_μ is the $U(1)_Y$ gauge field, g (g') is the weak (electromagnetic) coupling constant, and σ^i are the three Pauli matrices. With the covariant derivative, the Lagrangian density of the electroweak interaction results in

$$\mathcal{L}_{\mathrm{EW}} = \bar{\psi}_L(i\gamma^\mu D_\mu)\psi_L + \bar{\psi}_R(i\gamma^\mu D_\mu)\psi_R - \frac{1}{4}W_{\mu\nu}^i W^{i,\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} \quad , \tag{2.10}$$

where $W^{i,\mu\nu}$ with $i \in \{1,2,3\}$ are the field strength tensors of the weak isospin, $B^{\mu\nu}$ is the field strength tensor of the hypercharge, and ψ_L (ψ_R) are the left-handed (right-handed) components of the fermion. However, similar to QED and QCD, the gauge bosons remain massless, which contradicts experimental observations [34–36]. The introduction of mass terms into the Lagrangian density such as $\frac{1}{2}m_W^2 W_\mu W^\mu$ violate the underlying gauge invariance of the theory and leads to an inconsistent theory. A key to resolving this problem is offered by the principle of spontaneous symmetry breaking, as discussed in the following section. Together with the gauge group of QCD, the gauge group of the SM is based on the combination

$$SU(3)_C \times SU(2)_L \times U(1)_Y \quad . \tag{2.11}$$

2.1.4 Spontaneous symmetry breaking: the Higgs mechanism

To obtain masses for the gauge bosons of the weak force as experimentally proven, the idea of the Higgs mechanism is to introduce a complex scalar field ϕ constructed as a doublet under the $SU(2)_L$ gauge symmetry [1–6]. The doublet is constructed as

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \quad , \tag{2.12}$$

where ϕ^+ (ϕ^0) is the electrically charged (neutral) component of the field.

The Higgs potential is chosen such that the vacuum state is minimized at a nonzero value of the Higgs field. A two-dimensional illustration of this potential can be seen in Figure 2.1.

The potential is symmetrical in its equilibrium state, however, this state is unstable as a perturbation destroys the symmetry. The symmetry is broken once the Higgs field acquires a nonzero vacuum expectation value at any of the potential's minima in the figure. A potential reflecting this character is described by

$$V(\phi) = \mu^2|\phi|^2 + \lambda|\phi|^4 \quad , \tag{2.13}$$

where $\mu^2 < 0$ is required to allow for spontaneous symmetry breaking to occur and $\lambda > 0$ is a coupling constant.

The ground state is chosen as

$$\phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H \end{pmatrix} \quad , \tag{2.14}$$

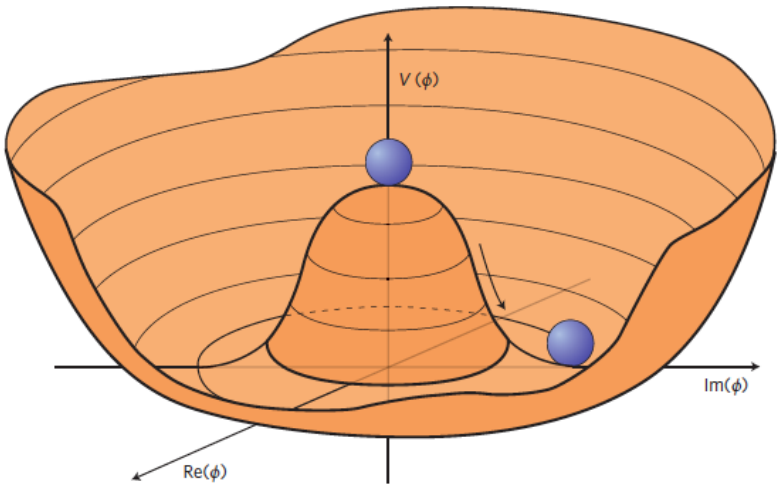


Figure 2.1: Illustration of the Higgs potential. Any point at the minimum breaks the rotational $U(1)$ symmetry, as illustrated by the blue pearl. Taken from Ref. [37].

where v denotes the vacuum expectation value $v = \sqrt{-\mu^2/\lambda}$ and H is an electrically neutral Higgs boson. The concept of spontaneous symmetry breaking requires the field to be isotropic. Hence, a scalar field is chosen, as any spin other than 0 would create a preferred direction or orientation in space, which would undermine the process of symmetry breaking. The Lagrangian density is formulated to satisfy the equation of motion defined by the Klein-Gordon equation and is given by

$$\mathcal{L}_{\text{Higgs}} = (D_\mu \phi)^\dagger \, (D^\mu \phi) - \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2 \quad , \tag{2.15}$$

where the first term describes the kinetic part and the second and third term reflect the Higgs potential of Equation 2.13.

The mass terms of the physical gauge bosons W and Z are generated when the gauge group symmetry is broken and the Higgs field develops the nonzero vacuum expectation value v. The covariant derivative acting on the Higgs field results to

$$\mathcal{L}_{\text{mass}}(\text{W}) = \frac{1}{2} g^2 v^2 W_\mu^+ W^{\mu -} \quad \text{and} \quad \mathcal{L}_{\text{mass}}(\text{Z}) = \frac{1}{2}(g^2 + g'^2) v^2 Z_\mu Z^\mu \quad , \tag{2.16}$$

with the W boson mass $m_W = \frac{1}{2} g v$ and the Z boson mass $m_\text{Z} = \frac{1}{2}(\sqrt{g^2 + g'^2} v)$. Consequently, the massless gauge bosons mix after spontaneous symmetry breaking to form the massive $\text{W}^\pm$ and Z bosons, whereas the photon remains massless. In the following section, the coupling and consequently the mass terms for the fermions of the SM are discussed.

2.1.5 Yukawa coupling

The previously introduced scalar Higgs field also interacts with spin-1/2 particles, i.e. fermions such as quarks and leptons of the SM (cf. Section 2.2) endowing them with mass via the mechanism of Yukawa coupling [38]. The Lagrangian density for the fermion-Higgs interaction term is

$$\mathcal{L}_{\text{Yukawa}} = -y_f \bar{\psi}_L \phi \psi_R + \text{h.c.} \quad , \tag{2.17}$$

where y_f is the fermion-Higgs Yukawa coupling, e.g. for an electron. In the case of quarks, the coupling is a matrix that allows flavor mixing.

After spontaneous symmetry breaking, the fermion-Higgs interaction gives mass to the fermion with

$$m_f = y_f \frac{v}{\sqrt{2}} \quad . \tag{2.18}$$

Thus, the fermion mass depends on the strength of the interaction, i.e. the Yukawa coupling, and the vacuum expectation value of the Higgs field.

2.2 Particles and interactions

In the previous section, the SM, based on the principles of quantum field theory and local gauge symmetries, described by the gauge group $SU(3)_C \times SU(2)_L \times U(1)_Y$, was introduced from a theoretical perspective. In this section, the SM is discussed from an experimental perspective, focusing on the introduction of all the known particles. The previously described coupling constants, along with the fermion and boson masses, are free parameters of the SM and require experimental assessment.

The fundamental particles are categorized into two groups: fermions and bosons. Fermions possess half-integer spin and follow Fermi-Dirac statistics [39, 40], whereas bosons have integer spin and follow Bose-Einstein statistics [41, 42]. The known fermions include quarks and leptons, whereby quarks form the building blocks of protons and neutrons. There are six flavors of quarks, namely up (u), down (d), charm (c), strange (s), top (t) and bottom (b). The u ,c, and t quark are also referred to as up-type quarks, whereas the d, c, and b quark are denoted as down-type quarks. Quarks carry an electric as well as a color charge, and their masses range over five orders of magnitude from 2 MeV for the up quark to 173 GeV for the top quark. The c, b, and top quarks are referred to as heavy flavor quarks due to their masses of 1 GeV or greater. These quarks and their corresponding anti-particles are the focus of this thesis and are examined in the analysis. Since the top quark plays a central role in this thesis, it is discussed separately in Section 2.3. In contrast,

Table 2.1: Fermions of the SM. The electric charge Q is given in units of the elementary charge e. The third component of the weak isospin T_3 corresponds to left-handed particles and right-handed anti-particles. Right-handed particles and left-handed anti-particles are isospin singlets with $T_3 = 0$. All anti-particles carry opposite electric charge. All values taken from Ref. [43].

	Gen.	Fermion	Interaction	Q	T_3	Mass
Quarks	1	up quark (u)	strong, electroweak	$+\frac{2}{3}$	$+\frac{1}{2}$	2.2 ± 0.1 MeV
		down quark (d)	strong, electroweak	$-\frac{1}{3}$	$-\frac{1}{2}$	4.7 ± 0.1 MeV
	2	charm quark (c)	strong, electroweak	$+\frac{2}{3}$	$+\frac{1}{2}$	1.27 ± 0.01 GeV
		strange quark (s)	strong, electroweak	$-\frac{1}{3}$	$-\frac{1}{2}$	93.5 ± 0.8 MeV
	3	top quark (t)	strong, electroweak	$+\frac{2}{3}$	$+\frac{1}{2}$	172.57 ± 0.29 GeV
		bottom quark (b)	strong, electroweak	$-\frac{1}{3}$	$-\frac{1}{2}$	4.18 ± 0.01 GeV
Leptons	1	electron neutrino (ν_e)	weak	0	$+\frac{1}{2}$	< 0.8 eV
		electron (e^-)	electroweak	-1	$-\frac{1}{2}$	0.51 MeV
	2	muon neutrino (ν_μ)	weak	0	$+\frac{1}{2}$	< 0.8 eV
		muon (μ^-)	electroweak	-1	$-\frac{1}{2}$	105.7 MeV
	3	tau neutrino (ν_τ)	weak	0	$+\frac{1}{2}$	< 0.8 eV
		tau lepton (τ^-)	electroweak	-1	$-\frac{1}{2}$	1776.9 ± 0.1 MeV

the u, d, and s quarks, which have masses up to 100 MeV, are referred to as light flavor quarks. Leptons are a family of particles that include electrons, muons, tau leptons, and their respective neutrinos. The electrons, muons, and tau leptons carry electric charge and range from 511 keV for the electron to 1.8 GeV for the tau lepton, whereas the neutrinos are electrically neutral particles. A summary of all fermions is given in Table 2.1

The SM consists of five bosons that mediate the fundamental forces, as discussed in the previous section and summarized in Table 2.2. The photon transmits the electromagnetic interaction and is massless. The W and Z bosons mediate the electroweak interaction, whereby the W bosons carry an electric charge and the Z boson is neutral. The W bosons possess a mass of around 80.4 GeV, while the Z boson has a mass of around 91.2 GeV. The gluons are responsible for the strong interaction and are massless. The Higgs boson is a consequence of the Higgs mechanism by which particles obtain their mass through interaction with the Higgs field, as discussed in Section 2.1.4.

2.3 Top quark physics

The top quark and its anti-particle play a central role in this analysis and are therefore also reflected in the title of the thesis. The top quark is the heaviest known elementary particle in the SM and possesses unique properties, which invite deeper exploration. The

Table 2.2: Bosons of the SM. All values taken from Ref. [43].

Boson	Spin	Interaction	Charges	Mass
Gluons (g)	1	strong	color	massless
Photon (γ)	1	electromagnetic		massless
$W^{\pm}$ bosons	1	electroweak	electric, weak	80.369 ± 0.013 GeV
Z boson	1	electroweak	weak	91.188 ± 0.002 GeV
Higgs boson (H)	0		weak	125.20 ± 0.11 GeV

particle was discovered in 1995 by the CDF [44] and D0 Collaborations [45] at the Fermilab Tevatron collider in searches for top quark-antiquark pair production ($t\bar{t}$) and is an active field of research today [46–49]. The top quark's Yukawa coupling to the Higgs boson, as discussed in Section 2.1.5, plays a special role in the SM as it is the strongest Yukawa coupling due to the exceptionally large top quark mass. According to Equation 2.18, with an vacuum expectation value of $v \approx 246$ GeV the top-Higgs Yukawa coupling is approximately $m_t = y_t \frac{v}{\sqrt{2}} \approx 1$ [43]. This property has profound physical implications, positioning the top quark as a pivotal figure in exploring the Higgs mechanism and uncovering phenomena in BSM physics. The coupling also affects the mass of the Higgs boson which obtains contributions from quantum fluctuations of all particles that couple to the Higgs field. Among these contributions, the top quark's is particularly dominant, leading to a divergence that grows quadratically with the energy scale Λ

$$\Delta m_H^2 \sim -\frac{3|y_t|^2}{4\pi^2}\Lambda^2 \quad , \tag{2.19}$$

where Λ is the Planck scale in the order $\mathcal{O}(10^{19}\,\text{GeV})$, at which quantum gravity effects become significant. Hence, an extremely large correction to a comparably small value of the Higgs mass requires a fine balance of the relevant parameters in Equation 2.19, which makes the value of y_t exceptionally intriguing.

At the Large Hadron Collider (LHC), the primary production mechanism for $t\bar{t}$ arises from gluon initial states ($gg \to t\bar{t}$). However, depending on the colliding particles and the center-of-mass energy, also quark-antiquark pairs can produce $t\bar{t}$ in an collison event. The dependency is depicted in Figure 2.2. The $t\bar{t}$ production process depends on the composition described by the proton's parton distribution function (PDF), as discussed in Section 2.4.

The top quark has a life time of 5×10^{-25} seconds and decays via the charged current of the weak interaction before hadronization. The top quark decays almost exclusively into a W boson with a branching fraction of

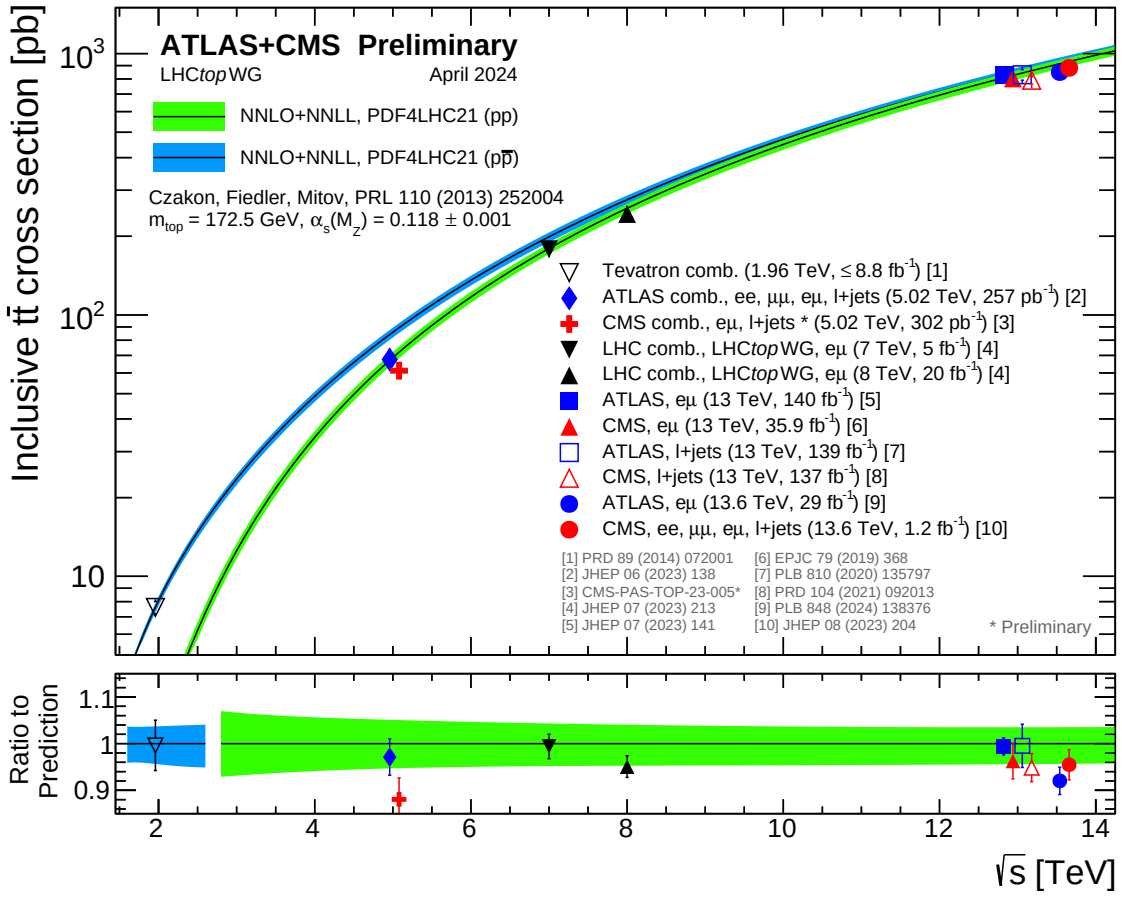


Figure 2.2: Cross section measurements at the LHC and Tevatron of $t\bar{t}$ production as a function of the center-of-mass energy. The uncertainties are based on Ref. [50], the figure is taken from Ref. [51].

$$\frac{\Gamma(t \to Wb)}{\Gamma(t \to Wq(q = b,s,d))} = 0.975 \pm 0.034 \quad . \tag{2.20}$$

This property is reflected in the Cabibbo-Kobayashi-Maskawa (CKM) matrix, which describes the transformation probabilities of quarks in weak decays [52, 53]. When considering the CKM matrix in the Wolfenstein parametrization, the corresponding entry for the transition from top to b quark is exactly one [54].

Since the top quark predominantly decays into a W boson and a b quark, the decays are usually characterized by the subsequent decay of the W boson. The branching fraction of the W boson to decay hadronically into a quark-antiquark pair ($W \to q\bar{q}'$) is 67%, while the branching fraction of the W boson to decay in a charged lepton and the corresponding neutrino is 33%. The decay combinations in a $t\bar{t}$ system are categorized into three channels: The first channel is the fully hadronic channel where both W bosons decay hadronically. In the second channel, one W boson decays hadronically, while the other one decays leptonically, which is referred to as the semileptonic channel. The third channel is the dilepton channel, where both W bosons decay leptonically. The resulting fractions of these channels are shown in Figure 2.3. In this thesis, in the following leptons refer exclusively to electrons and muons, as tau leptons are usually treated separately due to their additional decay processes. Each channel offers unique benefits and presents distinct challenges. The fully hadronic channel has the highest branching fraction with approximately 45%, but suffers from QCD background processes. In the semileptonic channel, the charged lepton in the signature plays a key role in enhancing the separation from background processes. With a branching fraction of about 30%, this channel has a reduced branching fraction compared to the fully hadronic channel, but features a signature that is easier to distinguish, as leptons are simpler to identify in collider detectors compared to hadronic decays. The dileptonic channel features the clearest signature due to the presence of two charged leptons. However, its low branching fraction of around 5% introduces significant statistical uncertainties, which can impact the measurements. The remaining 20% of the branching fraction are attributed to channels involving tau leptons. In this thesis, the dilepton channel is analyzed to take advantage of its clear signature. The dileptonic channel is further categorized by lepton combinations into three sub-channels as shown in Figure 2.3: the e^+e^- channel, the $e^\pm\mu^\mp$ channel, and the $\mu^+\mu^-$ channel.

2.4 Hadron collider physics

Protons used in collisions at the LHC consist of quarks and gluons, bound together by the strong interaction and described by the parton model. At an energy scale Q^2, a proton consists of a collection of partons, where the partons carrying the electric charge and the quantum numbers of the proton are referred to as valence quarks. In addition, quark-antiquark pairs from the vacuum can arise, which are known as sea quarks. Furthermore,

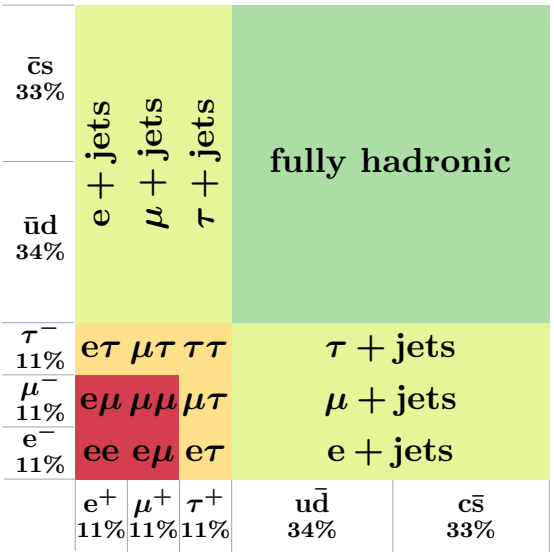


Figure 2.3: Decay channels of the $t\bar{t}$ system. Each axis represents the decay products of a W boson. The percentages denote the branching fractions of the W boson decays into electrons, muons, tau leptons, and the corresponding neutrinos as well as the quark-antiquark combinations $u\bar{d}$ and $c\bar{s}$. Adapted from [55].

all quarks can absorb and emit gluons according to the laws of QCD. The distribution of partons within a proton is described by PDFs, which represent the probability density of finding a parton with a given longitudinal momentum fraction x at a specific energy scale Q^2. Since in the case of proton-proton collisions, not necessarily the valence quarks, but in essence any existing parton may collide, an accurate understanding of the PDFs is crucial. The PDFs cannot yet be determined from first principles and need to be determined experimentally, e.g. in deep inelastic scattering processes. The evolution of PDFs as a function of the energy scale can be described using renormalization group equations (RGEs). The result from a measurement can be translated to a different energy scale using the RGEs, which are formulated in the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations [56–59]. Since PDFs are universal, they are determined through global fits across multiple experiments and can be applied to various experimental contexts [60]. Figure 2.4 illustrates the PDFs at two different energy scales and demonstrates how the quark and gluon PDFs are shifted at high energies, i.e. the probabilities for finding the valence quarks decrease, while the sea quarks and gluons gain relevance. Additionally, the b quark PDF becomes noticeable at higher energy scales, whereas it is not apparent at the lower energy scale.

To calculate the total cross section σ for a process in proton-proton collisions, the partonic cross section is integrated over all momentum fractions of the partons inside the protons. This is achieved by convoluting the partonic cross section with the previously described

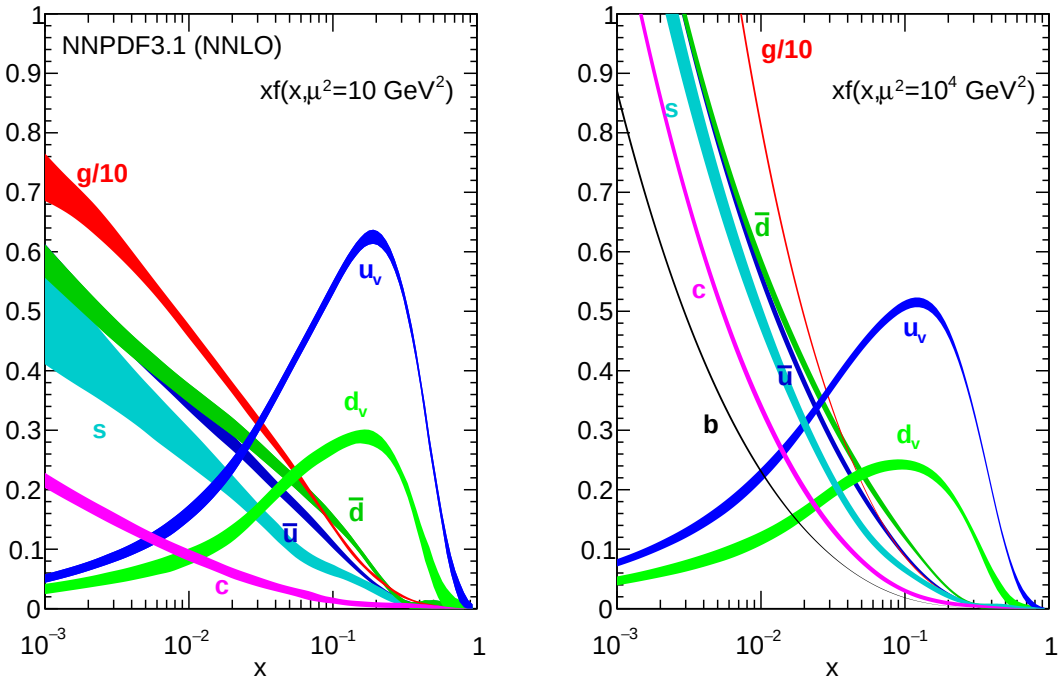


Figure 2.4: Example PDF sets as a function of the momentum fraction x at two different energy scales $Q^2 = 10\,\mathrm{GeV}^2$ (left) and $Q^2 = 10^4\,\mathrm{GeV}^2$ (right). The gluon distributions are scaled down by a factor of 10. At both energy scales, valence quarks dominate the high-momentum fractions, while gluons prevail at the lower momentum fractions. Taken from Ref. [60].

PDFs. This strategy is feasible due to the factorization theorem, which separates a process into contributions from different energy scales in QCD. At high energies and small values of the coupling constant α_S, sub-processes can be calculated using perturbation theory. For this, the quantum mechanical transition amplitudes, also known as the scattering matrix element (ME), are calculated for a given process. At low energies, corresponding to long-range effects, the coupling constant rises and leads to a non-perturbative QCD regime. Factorization separates these contributions, enabling the independent treatment of short-range and long-range parts. To separate these two cases, a factorization scale μ_F is introduced. This scale differentiates between the cross section for the hard partonic process $\hat{\sigma}$, calculated using perturbation theory, and further processes such as soft and collinear gluon radiation, while are included in the PDFs.

For a proton-proton collision process into a final state X, the QCD factorization results in a cross section of

$$\sigma_{pp\to X} = \sum_{jk} \int\int dx_j dx_k \; f_j\left(x_j, \mu_F^2\right) f_k\left(x_k, \mu_F^2\right) \cdot \hat{\sigma}_{\hat{p}_j \hat{p}_k \to X}\left(x_j p_1, x_k p_2, \mu_F^2, \alpha_S\left(\mu_R^2\right)\right) \quad , \tag{2.21}$$

where $f_{j,k}$ are the PDFs of the two partons $\hat{p}_{j,k}$ with a momentum fraction $x_{j,k}$ in the two protons with the momenta $p_{1,2}$. The factorization scale μ_F denotes the energy scale at which the PDFs are evaluated. The hard partonic process $\hat{\sigma}$ is a function of the coupling constant α_S, which, in turn, depends on the renormalization scale μ_R. The renormalization scale is the energy scale at which the strong coupling constant is evaluated, controlling how the interaction strength varies with energy. In Section 2.1.2, the running of the coupling constant in QCD is discussed. This running is described by RGEs which describe the dependence of α_S on the renormalization scale. It is a key aspect of the renormalization process, where infinities in quantum field theories are removed by redefining parameters to absorb these divergences. The connection between α_S and μ_R is discussed in detail in the context of the thesis results in Chapter 10.

In event simulation, the renormalization and factorization scales are typically defined dynamically, i.e. depending on the kinematics of each individual simulated process. This can be seen in the following example of an important process in this thesis. In the simulation of the production of a top quark-antiquark pair in association with additional b quark-antiquark pairs ($\mathrm{t\bar{t}b\bar{b}}$), where the $\mathrm{t\bar{t}b\bar{b}}$ process is calculated at ME level, the scales are defined as

$$\mu_R = \kappa_R \cdot \sqrt[4]{\prod_{\mathrm{t},\bar{\mathrm{t}},\mathrm{b},\bar{\mathrm{b}}} (m_{\mathrm{T},i})} = \kappa_R \cdot \sqrt[4]{m_{\mathrm{T,t}} \cdot m_{\mathrm{T},\bar{\mathrm{t}}} \cdot m_{\mathrm{T,b}} \cdot m_{\mathrm{T},\bar{\mathrm{b}}}} \quad , \tag{2.22}$$

and

$$\mu_F = \kappa_F \cdot \sum_{\mathrm{t},\bar{\mathrm{t}},\mathrm{b},\bar{\mathrm{b}},\mathrm{g}} (m_{\mathrm{T},i}) = \kappa_F \cdot \left(m_{\mathrm{T,t}} + m_{\mathrm{T},\bar{\mathrm{t}}} + m_{\mathrm{T,b}} + m_{\mathrm{T},\bar{\mathrm{b}}} + p_{\mathrm{T},g} \right) \quad , \tag{2.23}$$

where p_{T} is the transverse momentum as defined in Equation 3.2, m_{T} is the transverse mass with $m_{\mathrm{T},i} = \sqrt{m_i^2 + p_{\mathrm{T},i}^2}$, and $\kappa_{R,F}$ are constant pre-factors.

The choice of calculation at ME level also affects the PDF set. Since the PDFs are evaluated at an energy scale μ_F where the b quark mass is small compared to the energy scale ($m_{\mathrm{b}} \ll \mu_F$), the b quark mass can be dropped in the ME calculations. In this case, the PDF contains five active flavors, i.e. flavors that actively contribute to virtual and real emissions. This approach is referred to as five flavor scheme (5FS). Calculations for hard scattering with heavy quarks with mass m_{q} include terms $\log(\mu^2/m_{\mathrm{q}}^2)$ arising from collinear gluon splittings ($\mathrm{g} \to \mathrm{q}\bar{\mathrm{q}}$), and power-suppressed terms proportional to m_{q}^2/μ^2 at an energy scale μ. If m_{q} is set to zero in ME calculations, divergent logarithmic contributions are absorbed into the PDFs for initial states and into the parton shower (PS) for final states (cf. Section 4.6).

An alternative approach is the four flavor scheme (4FS), which treats b quarks as massive and assumes four active flavors in the PDF. In this scheme, the b quark PDFs are omitted, and initial state b quarks arise from $\mathrm{g} \to \mathrm{b}\bar{\mathrm{b}}$ splittings in the ME calculations. This approach is the preferred option for the description of the $\mathrm{t}\bar{\mathrm{t}}\mathrm{b}\bar{\mathrm{b}}$ process, as it represents a more precise characterization of this process.

3 Experimental environment

Fundamental theories, such as those discussed in Chapter 2, rely on experimental validation. The interplay between theory and experiment allows for theoretical models to be falsified and for experimental findings to enhance our understanding by determining parameters of a theory. The experimental environment at the Conseil Européen pour la Recherche Nucléaire (CERN) provides a unique opportunity to test and expand our knowledge of the Standard Model of particle physics (SM) and beyond. All data analyzed in this thesis were recorded at CERN. To achieve data gathering, Section 3.1 first describes the necessary particle accelerator concept for proton-proton collisions. Section 3.2 briefly describes the experimental setup of the detector as well as the main components employed for recording the data.

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC), the largest and most powerful particle accelerator in the world, is located at the CERN site. Positioned underground in the border area between Switzerland and France, the LHC has a circumference of approximately 27 km. As the LHC cannot be operated stand-alone, it relies on a series of pre-accelerators to develop its full operational potential. Either protons or lead nuclei can be accelerated in this chain of accelerators to a center-of-mass energy of up to 14 TeV and 5.6 TeV/nucleon, respectively. The data analyzed in this thesis were recorded during the operation of the accelerator with proton-proton collisions at a center-of-mass energy of 13 TeV during the period known as Run-II [61]. All subsequent explanations therefore refer to the accelerator design and detector complex during this phase of operation. The accelerator complex is shown schematically in Figure 3.1.

Initially protons are obtained from hydrogen atoms by ionization. Following this, the extracted protons are accelerated to an energy of 50 MeV using the linear accelerator

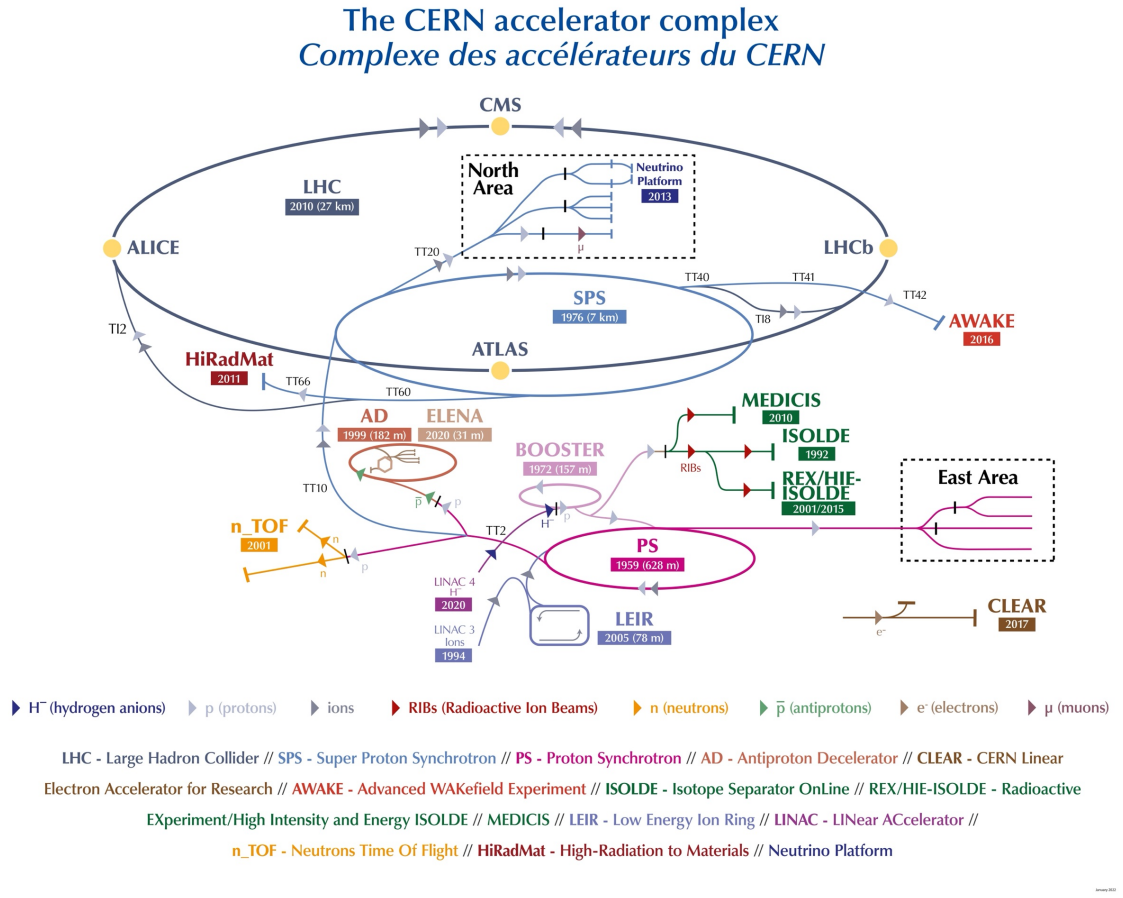


Figure 3.1: The accelerator complex at CERN, taken from Ref. [62]. A chain of coordinated accelerators ends with the LHC as last element (dark blue). Four experiments located at the LHC are indicated at their relative position as yellow dots.

LINAC2. Due to the nature of the acceleration process using radio-frequency cavities the protons are already broken down into bunches at this stage in the accelerator chain. This bunched structure runs through the entire subsequent accelerator chain and no continuous proton beam is produced. After the linear pre-accelerator, three circular accelerators follow in sequence, each with an increasingly larger circumference to achieve higher center-of-mass energies. The first circular machine is the BOOSTER with a targeted beam energy of 1.4 GeV, followed by the Proton Synchrotron with a beam energy of 26 GeV. The last circular pre-accelerator is the Super Proton Synchrotron with a beam energy of 450 GeV. The protons are then ejected from this synchrotron and injected into the final accelerator ring, the LHC [61]. Unlike in the preceding accelerators, the protons are now split into two beam pipes, one pipe running clockwise and the other one counter-clockwise in the accelerator.

The protons are accelerated by radio-frequency cavities that generate a strong oscillating electric field to boost the kinetic energy and continuously compensate for radiation losses due to synchrotron radiation. The radio-frequency cavities of the individual accelerators operate at different frequencies: 80 MHz for the Proton Synchrotron, approximately 200 MHz for the Super Proton Synchrotron, and 400 MHz for the LHC. At the collision rate of 40 MHz, two particle bunches, each consisting of approximately 10^{11} protons, collide every 25 ns. However, the limiting factor for achieving a maximum high center-of-mass energy is not the cavity design, but the bending magnets that keep the protons on an almost circular trajectory. For this purpose, superconducting dipole magnets operated at a temperature of 2 K generate a magnetic flux density of approximately 8 T. To focus the proton beams, quadrupole magnets are employed which compensate for the naturally induced divergence of protons within a bunch transversely to the beam axis [61].

In addition to the center-of-mass energy mentioned at the beginning, a further important key parameter for the physics program and important figure for an accelerator is the instantaneous luminosity. This parameter quantifies the particle flux and allows inferences about the number of successful particle collisions. The instantaneous luminosity for head-on collisions is given by

$$L_{\text{inst}} = f_{\text{rev}} \cdot n_b \cdot \frac{n_c n_{cc}}{4\pi\sigma_x\sigma_y} \quad , \tag{3.1}$$

where f_{rev} describes the revolution frequency, n_b the number of colliding bunches, and n_c (n_{cc}) the number of protons per bunch of the clockwise (counter-clockwise) beam. The spatial distribution of the protons in a bunch transversely to the beam axis is described by σ_x and σ_y, which denote the standard deviations of the Gaussian beam profile. The LHC is designed to reach an instantaneous luminosity of $10^{34}\,\text{cm}^{-2}\text{s}^{-1}$. The integral of the instantaneous luminosity with respect to time is a measure of the amount of data recorded in a given period of time.

The proton beams traveling in opposite directions collide at four distinct points at the LHC as shown in Figure 3.1. Protons within a bunch that have not collided remain in the beam, undergo the aforementioned acceleration and focusing procedures of the LHC again and can be induced to collide at one of their next beam crossings.

Data acquisition of collision events takes place with four detectors at the LHC. The ALICE detector focuses on heavy-ion collisions and examines the strong-interaction sector of the SM [63]. In contrast, the LHCb experiment is a general purpose detector in forward direction and follows a diverse physics program including measurements of B hadrons and charge conjugation parity symmetry (CP) violation [64]. The ATLAS and the Compact Muon Solenoid (CMS) experiments are also general purpose detectors and pursue a manifold physics program [65, 66]. Both experiments aim to test the SM through precision measurements and perform searches for new physics in the form of unknown particles and interactions. Despite sharing the same goals and being capable of independently verifying each other's results, the two experiments are conceptually similar, but differ in numerous detector structures, components and designs. The CMS detector used to record data analyzed in this thesis is described in Section 3.2.

3.2 The Compact Muon Solenoid experiment

As a versatile detector, the CMS experiment must meet a wide range of requirements. For precision measurements of the SM and searches for physics beyond the Standard Model (BSM), it is essential to reconstruct as many particles as possible with the utmost resolution and precision. In subsequent stages, different particles must be distinguished as accurately as possible. The CMS detector is able to reconstruct all known stable particles, but neutrinos remain undetected due to their vanishingly small interaction with the detector materials. The experiment is designed with a detector having 4π angular coverage for the most seamless reconstruction possible. The central part features a barrel design surrounding the beam pipe similar to an onion shell consisting of various components. End caps are installed at the detector's two ends in the forward and backward directions which also enable particle reconstruction. The key detector components are briefly outlined in the following section, a detailed description can be found in Ref. [66]. The experiment is also shown in Figure 3.2. The term ''compact'' in the detector's name needs to be understood in the light of comparable devices such as the ATLAS experiment, as the detector measures 15 m in diameter and a total length of 28.7 m.

3.2.1 Detector components

The CMS detector is designed to reconstruct a variety of particles produced through proton-proton collisions at high energies and high rates. In order to accomplish this mission, the CMS detector consists of various detector components, each targeting specific

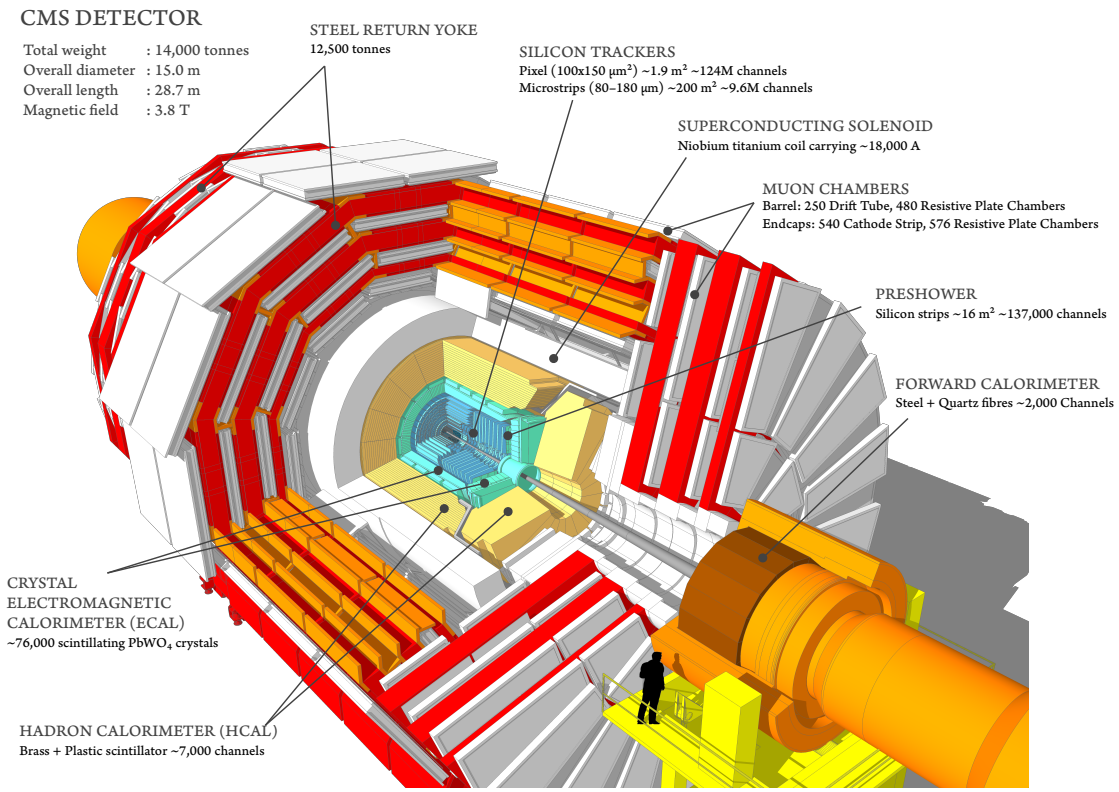


Figure 3.2: Graphical representation of the partially opened CMS detector, taken from Ref. [67]. Various detector elements form a dense cylindrical onion shape around the beam pipe (top left to bottom right).

objectives. Although each component is responsible for its respective role, it is the interplay and combination of all elements that provide a holistic picture and understanding of a collision event.

3.2.1.1 Tracking system

The tracking system is positioned at the innermost point of the detector as close as possible to the beam pipe and thus also to the collision point [66]. The purpose of the tracking system is to reconstruct the trajectory of particles produced during the collision event. Due to the presence of a magnetic field, which is described in more detail below, charged particles trace a curved trajectory from which a particle's momentum and the electric charge sign can be determined. Combining both of these elements with the energy deposit in the detector material, the mass and consequently the type of particle can be reconstructed using the energy-momentum relation. In addition, the tracking system can resolve vertices by using suitable reconstruction methods [68, 69]. For example, the reconstruction is important to identify vertices displaced from the collision point, which are a characteristic of the decay of B hadrons. The innermost unit of the tracking system consists of silicon semiconductor pixel detectors. The requirements for these detectors include a high granularity which is needed due to their position close to the collision point and the high particle flux. The material choice ensures minimal particle interactions to maintain the initial trajectory and a low material budget. Silicon also withstands radiation and fulfills the requirement of radiation hardness during operations of the LHC. The semiconductor detectors are reverse biased with a high voltage in order to create a maximized detector volume depleted of free electric charge carriers. If a charged particle passes through the semiconductor, an electric current is induced and recorded. The pixel size in the innermost layer is $100 \times 150\,\mu\text{m}^2$. A total of four layers of pixel detectors are arranged around the beam pipe with radii of $29, 68, 109,$ and $160\,\text{mm}$, enabling a reconstruction of three-dimensional space points. Microstrip detectors are mounted around these layers of pixel detectors and a reconstruction in two-dimensional space points is achieved. This results in a total of over $200\,\text{m}^2$ of active silicon [66, 70–72].

3.2.1.2 Electromagnetic calorimeter

The electromagnetic calorimeter (ECAL) surrounds the tracking system. Its purpose is to reconstruct the energy of particles with electromagnetic interactions, causing passing particles to deposit almost all of their energy [73, 74]. Among these particles are mainly photons as well as electrons and positrons. Other charged particles can also deposit energy to a small extent. The primary interaction mechanism is bremsstrahlung together with pair production of electrons and positrons, leading to electromagnetic showers. Characterized by the radiation length X_0, the electromagnetic showers can continue until the remaining energy drops below the threshold for electron-positron pair production. In this case, the

energy deposition process ends with ionization. The amount of energy deposited in the ECAL is determined by the number of photons produced via scintillation, which is recorded as an electrical signal by avalanche photodiodes. As the ECAL is located inside the solenoid there are limitations to the calorimeter's dimensions and the choice of material. The ECAL is designed as a homogeneous calorimeter where active material and absorber are identical. For this purpose lead tungstate crystals ($PbWO_4$) are chosen. To improve the separation of photons originating from bremsstrahlung processes ($e^{\pm} \rightarrow \gamma e^{\pm}$) from photon pair production through the decay of neutral pions ($\pi^0 \rightarrow \gamma\gamma$), a preshower detector is installed in the endcaps [75–77].

3.2.1.3 Hadronic calorimeter

The purpose of the hadronic calorimeter (HCAL) is the energy measurement of neutral and charged hadrons. It is the only detector element in which neutral particles such as hadrons can leave a signature. Therefore, the HCAL is a crucial component in the indirect reconstruction of neutrinos or exotic particles, as these particles can only be inferred from missing transverse momentum (see Section 3.2.2) in the entire experiment. The HCAL encases the ECAL and is also located within the magnet. Unlike the ECAL, the HCAL is a sampling calorimeter consisting of alternating slices of absorber material and scintillator. Incoming hadrons interact with the absorber material via scattering processes, creating new particles and hadronic showers. As many different production processes can occur and electromagnetic sub-showers are also possible, hadronic showers are more diverse compared to the effects in the ECAL. The hadronic interaction length λ_h, the equivalent to the radiation length X_0 in the ECAL, is considerably larger in comparison and hadronic showers are subject to higher fluctuations. As a result, the HCAL is noticeably larger and full absorption within the HCAL is not always guaranteed. The HCAL is composed of layers of brass for absorption and plastic scintillators [78, 79].

3.2.1.4 Superconducting solenoid

The superconducting magnet can reach field strengths of up to 4 T and is operated at 3.8 T. The solenoid encloses the tracking systems described before as well as the calorimeter units. A steel return yoke mounted around the magnet stabilizes the magnetic field and houses the muon system described in the following. Outside the magnet, the magnetic field strength is 2 T [66, 80]. Without the strong magnetic field that bends the trajectories of all electrically charged particles, measurements of momenta and the separation of particle types would not be feasible in the sub-detectors.

3.2.1.5 Muon system

The mission of the muon detector is the dedicated reconstruction of muons. As minimum ionizing particles, muons deposit very limited amounts of energy via ionization in the

detector layers, allowing them to pass through all preceding layers. Hence, dedicated muon detectors are installed in the return yoke of the solenoid which enable a reliable reconstruction of muons. Since large spatial areas have to be covered, gas detectors are chosen. Although they offer lower resolution compared to solid-state detectors like silicon, gas detectors provide a full solid-angle coverage over large volumes and are more cost-effective. During the fly-through of a muon, the gas mixture in the detector is ionized. By applying a high voltage, electrons are accelerated and can also generate additional free electrons. The arising cascade of electrons is read out and measured as an electric current. The gas mix used is argon and CO_2 [81, 82].

3.2.1.6 Trigger system

The collision rates of proton bunches at 40 MHz, as referred to in Section 3.1, with an average of 15 to 35 proton-proton collisions per bunch lead to rates of over 1 GHz. This results in raw data amounting to multiple terabytes per second in the experiment which precludes storing all data. In order to reduce the data rate and retain only relevant and interesting collisions a multi-level trigger system is implemented. This multi-stage procedure can trigger objects such as muons based on specific transverse momentum thresholds, among other kinematic quantities and combinations (cf. Section 3.2.2). First, a decision is made on a rough but fast basis using the Level-1 (L1) trigger as to whether, for example, measured ionizations in the muon chambers could in principle match a muon object. Within microseconds this hardware trigger decides whether formed clusters match a trigger pattern. For this, the data are massively parallelized and temporarily stored at the detector front end allowing for rapid processing. If no trigger pattern is matched, the event is rejected already at this stage. If the evaluation is successful, the entire event is fed into the succeeding trigger stage. With the help of the L1 trigger system the rate is reduced from a few GHz to the order of 100 kHz [83]. The subsequent trigger stage receives the pre-processed and filtered information and determines, within less than a second, whether more specific requirements for an interesting event candidate have been met. This so-called high-level trigger (HLT) operates in a highly parallelized manner on processor farms. Complex reconstruction methods and algorithms are avoided at this point in favor of high performance and a loose selection is required. A list of trigger paths is queried which have different requirements for objects or kinematic quantities and which may later be selected in dedicated analyses. Once an event meets one or more HLT requirements, the event is recorded. The HLT paths queried in this thesis are listed in Section 8.2.1. For some trigger paths an additional prescaling is implemented since the trigger rates are still too high to be recorded. In this case only every n-th event is recorded at a predefined level. This trigger system makes it possible to reduce the event rate down to 1 kHz and achieve reliable data storage [84].

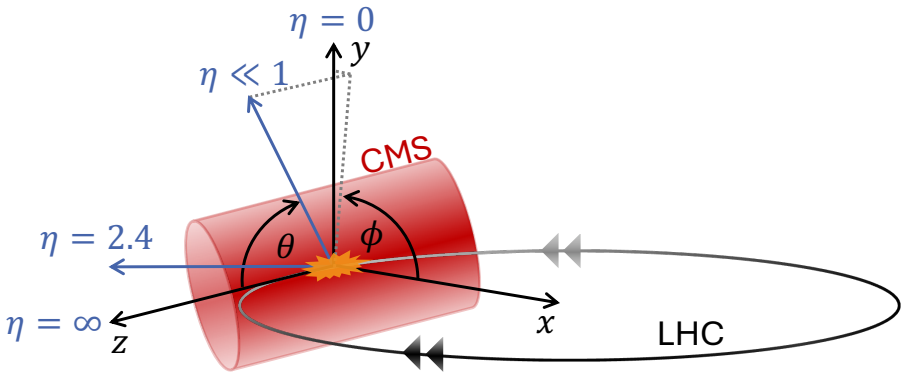


Figure 3.3: Schematic overview of the coordinate system at the CMS experiment. The pseudo rapidity of $\eta = 2.4$ is shown to illustrate the maximum polar angle considered in this analysis (cf. Chapter 8). The relative scales of the LHC and the CMS experiment do not reflect actual dimensions.

3.2.2 Coordinate system and kinematic quantities

The coordinate system at the CMS experiment is defined as a right-handed Cartesian coordinate system centered at the collision point. The x-axis points towards the center of the LHC, the y-axis points vertically upwards, and the resulting z-axis points in the beam direction towards the Jura mountains. Due to the cylindrical shape of the detector and the rotational invariance of a collision, cylindrical coordinates are a suitable choice. The polar angle θ is defined between the z-axis and a given vector while the azimuth angle ϕ is defined from the x-axis into the x, y-plane. The coordinate system is illustrated in Figure 3.3.

In a collision of the proton's constituents, the partons (cf. Chapter 2), the vectorial part of the momentum in the x, y-direction is insignificantly low. Furthermore, the part of the momentum of the proton in the z-direction carried by the parton is unknown. This leads to some useful considerations. Since the momentum component in the x, y-plane is negligibly small before the collision, the sum of the momentum vectors in this plane for all particles produced must be zero after the collision due to the conservation of momentum. The momentum vector in this plane is referred to as transverse momentum p_T and is defined by

$$p_\mathrm{T} \coloneqq \sqrt{p_x^2 + p_y^2} \quad . \tag{3.2}$$

Therefore, the transverse momentum is invariant under Lorentz transformations along the z-axis.

However, the sum of measured transverse momenta in an event is typically not zero. Because some electrically neutral particles interact only weakly or not at all with the detector material, their transverse momenta cannot be detected. Among these particles

are neutrinos and hypothetical dark matter candidates which could be produced in the collision. The negative sum of all measured transverse momenta then forms a quantity referred to as missing transverse energy (MET) [85], which is described in more detail in Section 4.5.

Another useful quantity is the rapidity. It is a measure of a particle's velocity and is defined as

$$y \coloneqq \tanh^{-1}\left(\frac{p_z}{E}\right) = \frac{1}{2}\ln\frac{E+p_z}{E-p_z} \quad , \tag{3.3}$$

where E denotes the particle's energy. The difference of this quantity between two particles is Lorentz invariant under boosts along the z-axis. For momenta significantly larger than the particle's mass ($p \gg m$), which is generally given at the detector as well in this thesis, the rapidity can be substituted by the pseudo rapidity η and simplified to

$$\lim_{p\gg m} y \coloneqq \eta = -\ln\tan\left(\frac{\theta}{2}\right) \quad . \tag{3.4}$$

Consequently, this quantity depends solely on the polar angle θ. As a result, values of $\eta \to 0$ point orthogonally to the z-axis while values of $\eta \to \infty$ lie in the direction of the z-axis, cf. Figure 3.3.

For a geometric distance between two objects i and j, the spatial distance ΔR is defined by

$$\Delta R_{ij} \coloneqq \sqrt{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2} \quad , \tag{3.5}$$

where $\eta_{i,j}$ and $\phi_{i,j}$ are the pseudo rapidity and azimuthal angle of the respective particles. The quantities listed here are frequently used in this thesis to describe properties of objects or relationships between them.

4 Object reconstruction and event definition

This chapter is dedicated to the reconstruction of physics objects in the Compact Muon Solenoid (CMS) detector, representing an elementary part of data analysis. The goal is to reconstruct particles and their properties based on the recorded signals in all sub-detectors, such as energy deposits in the calorimeters or tracks in the tracking system, as discussed in the previous chapter. Reconstruction requires precise processing and interpretation of the detector signals in order to identify physics objects such as leptons, jets, or missing transverse energy (MET) correctly. Thus, the object reconstruction forms the connection between the raw detector data and the physics analysis, which ultimately allows conclusions to be drawn about the underlying physics processes. The chapter is structured as follows: The track and vertex reconstruction is discussed in Section 4.1. In Section 4.2 the particle-flow algorithms are described and the overall reconstruction strategy of the most important objects in this analysis is outlined. The focus of Section 4.3 is on the reconstruction algorithm of jets. Based on this, in Section 4.4, the heavy flavor jet tagging algorithms are explored, which play an important role in this thesis. The reconstruction of MET is detailed in Section 4.5. At the end of this chapter, the event simulation is described in Section 4.6. This section also includes definitions of physics processes that are used throughout this thesis.

4.1 Track and vertex reconstruction

As outlined in Section 3.2.1.1, the tracking system constitutes the innermost layer of the CMS detector. The purpose of track reconstruction is to assemble hits into a consistent trajectory representing a particle's path through the tracker. At the CMS experiment, a combinatorial track finding algorithm is employed using Kálmán filters [86, 87].

Charged particles passing through the detector leave hits in the silicon pixel and strip detectors in a magnetic field generated by the solenoid. Each hit corresponds to a localized energy deposit caused by the interaction of the particle with the detector material. For the track reconstruction, potential seeds are created initially by combining neighboring hits to form track candidates, referred to as seed generation. These seeds are constructed using hits from adjacent layers of the pixel detector. Next, trajectories are developed by identifying hits across all detector layers that align with the track candidate's extrapolated path based on the initial seed. This step, referred to as track finding, is based on Kálmán filters [86, 87]. The magnetic field forces charged particles to follow a helical trajectory. In a subsequent track fitting step, the trajectories of the charged particle candidates are refined to closely match the observed hits. During this step, deviations and uncertainties of the reconstruction are reflected in the fitting procedure. Additionally, tracks are extrapolated to determine their origins, which is an important step to determine the primary collision vertex as described below. Finally, tracks are evaluated and discarded if they fail to meet quality criteria such as p_{T} thresholds. If a candidate satisfies all criteria, its associated hits are removed from the overall collection, and the iterative process continues with the remaining hit collection.

The algorithm for vertex finding clusters tracks into potential vertex candidates based on the z-position of their extrapolated origins. Each candidate's position is refined using an adaptive vertex fitting algorithm [88]. The primary vertex is determined as the one with the highest sum of the squared transverse momenta p_{T}^2 of the tracks assigned to it. An important observable is the impact parameter, defined as the distance to the primary vertex. In this thesis, the impact parameter is divided into two components, along the radial direction $|d_{xy}|$ and the longitudinal direction $|d_z|$. The vertex requirements for different objects, including leptons and jets, are detailed in Chapter 8.

4.2 Particle-flow algorithm

The particle-flow algorithm at the CMS experiment is a sophisticated method to reconstruct objects based on a comprehensive combination of all detector components [89]. Rather than reconstructing objects within isolated detector components, the particle-flow algorithm takes a unified approach by combining measurements from the tracking system, the electromagnetic calorimeter (ECAL), the hadronic calorimeter (HCAL), and the muon system. The reconstruction process proceeds sequentially, starting with particles that are easier to reconstruct based on energy resolution, trajectory, and momentum. The algorithm then proceeds to reconstruct particles that are more challenging to identify by utilizing the remaining signals. A summary of the individual particle reconstruction procedures is provided below. Further details on the particle-flow algorithm can be found in Ref. [89].

Electrons are reconstructed by correlating their trajectory in the tracking system (cf. Section 3.2.1.1) with energy deposits in the ECAL (cf. Section 3.2.1.2). This requires

identifying a track in the tracking system that matches the energy observed in the ECAL, enabling the measurement of their energy and momentum. Additionally, photons produced through bremsstrahlung in the tracking system are taken into account. Details on the electron selection criteria employed in this analysis are provided in Section 8.2.2.

Photons are also identified through their energy deposits in the ECAL. In contrast to electrons, photons do not produce a track in the tracking system. Additionally, the production of electron-positron pairs from photons in the tracking system must be considered. This is achieved by detecting secondary tracks that arise from photon interactions in the tracker material, displaced from the primary interaction vertex.

Muons are categorized into three different types based on their reconstruction. Standalone muons are identified solely in the muon system (cf. Section 3.2.1.5) of the CMS detector. Global muons additionally require the presence of a corresponding track in the tracking system. Tracker muons are reconstructed in the tracking system, extrapolated to the muon system, and require at least one segment or hit in the muon chambers that matches the extrapolated track. The object requirements for muons are detailed in Section 8.2.3.

Charged hadrons are reconstructed through clusters in the ECAL and HCAL (cf. Section 3.2.1.3) with matching tracks in the tracking system. For the reconstruction of charged hadrons, the momentum reconstruction in the tracking system is more precise than the energy measurement in the calorimeter.

Neutral hadrons do not leave tracks in the tracking system and solely deposit their energy in the HCAL. After reconstructing all other physics objects, the residual HCAL energy is assigned to neutral hadrons. Therefore, neutral hadrons rely exclusively on the HCAL, resulting in limited resolution.

4.3 Jet reconstruction

In high-energy proton-proton collisions, quarks and gluons are produced, each carrying a color charge (cf. Section 2.1.2). However, due to the strong force of quantum chromodynamics (QCD), quarks and gluons cannot exist freely and instead hadronize to form composite, color-neutral particles, i.e. hadrons. As a result, a stream of collimated particles is generated during hadronization, referred to as a jet. These jets are reconstructed with the anti-k_{T} algorithm at the CMS experiment [90]. This algorithm satisfies two key criteria: collinear and infrared safety. Collinear safety guarantees that a particle splitting with a narrow angular separation does not create a new jet, keeping the original jet candidate unaffected. By ensuring infrared safety, radiation at low momentum fractions is handled so that the jet candidate remains invariant without producing additional jets. The anti-k_{T} algorithm is based on a sequential recombination using a distance measure d_{ij} between all particle-flow candidates, which is defined as

$$d_{ij} = \min\left(p_{\mathrm{T},i}^{-2}, p_{\mathrm{T},j}^{-2}\right) \frac{\Delta R_{ij}^2}{R^2} \quad . \tag{4.1}$$

In this distance measure, ΔR_{ij} is the spatial distance of the two particle-flow candidates in the detector, as defined in Equation 3.5. The radius parameter R is set to $R = 0.4$ as a default choice for jets at the CMS experiment. Additionally, the distance value

$$d_i = p_{\mathrm{T},i}^{-2} \quad , \tag{4.2}$$

is defined for each particle and the values d_{ij} and d_i are calculated. If the distance measure d_{ij} is the smallest value, the two particles are merged into a pseudo-particle. Both particles are then removed from the collection and replaced with the pseudo-particle. If d_i is the smallest distance, the (pseudo-)particle i is removed from the collection and the collection is defined as a jet. The anti-k_{T} algorithm is applied iteratively until all particle-flow candidates are fully clustered into jets.

The resulting jets are cone-shaped with a radius of approximately 0.4 in the η-ϕ space. Particles with high p_{T} values play a significant role in the clustering algorithm, as the procedure incorporates squared momenta into the distance calculation, which effectively accounts for the primary energy flow. Consequently, particles with higher momentum are preferentially grouped when in close proximity. Lower-energy particles, with less influence on the distance calculation, are subsequently added to the jet candidate.

4.4 Heavy flavor jet tagging

The identification of the jet flavor plays a central role in this analysis. With the help of sophisticated algorithms, it is possible to identify the flavor of the initial particle. The flavor identification of jets harnesses fundamental properties of quarks. Heavy quarks, compared to light quarks or gluons, undergo fragmentation and hadronization over short distances. For example, B hadrons typically travel a few hundred micrometers before they decay in their rest frame. The reason is the suppressed transition in the Cabibbo-Kobayashi-Maskawa (CKM) matrix, as discussed in Section 2.1. This property results in a secondary decay vertex that is located at a distance from the original collision point, as illustrated in Figure 4.1. The secondary vertex information is an important input in flavor identification techniques.

Heavy flavor jet identification algorithms are an active field of research at the CMS experiment. The DeepCSV flavor tagger is an identification algorithm based on deep neural networks (DNNs) [91]. A more comprehensive approach, DeepJet, uses a convolutional neural network architecture and shows improved performance over DeepCSV [92]. Further developments such as ParticleNet are based on graph convolutional neural networks [93].

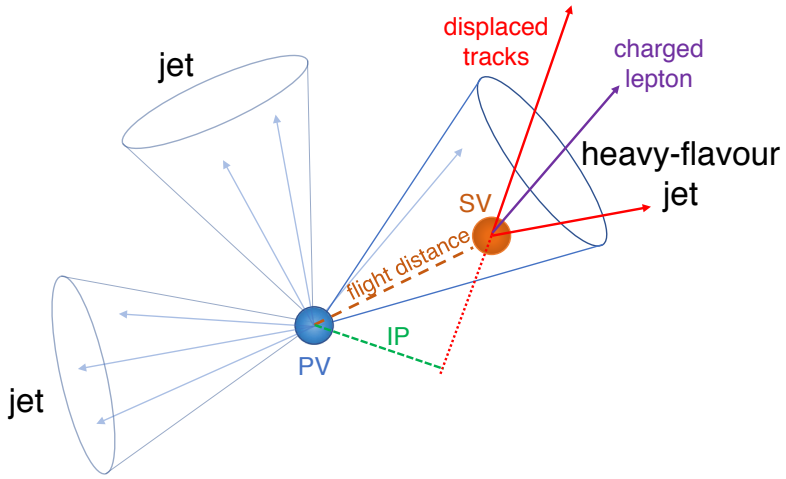


Figure 4.1: Illustration of a heavy flavor jet featuring a secondary vertex (SV) displaced from the primary vertex (PV), arising from the decay of a b or c hadron and producing charged-particle tracks. The decay results in a large impact parameter (IP) value. Taken from Ref. [91].

The Particle Transformer flavor tagger uses as a transformer architecture with attention mechanisms [94]. The continuous improvement of architectures and efficiencies enables the analysis of processes that are difficult to access, such as $t\bar{t}$ in association with additional c jets (cf. Section 7.2.4). Beyond this, the algorithms developed must also be capable of handling the future challenges of the experimental environment. In particular, this includes the challenges posed by the increased pileup in the high-luminosity Large Hadron Collider (LHC) era, in which tagging algorithms must operate reliably in hostile collision environments. The heavy flavor jet tagging algorithms have been examined and upgraded for such future scenarios [95] within this thesis.

The heavy flavor jet tagging algorithm employed in this thesis is the DeepJet tagger, which is the standard jet tagger in Run-II analyses and is also used in related measurements [96]. The DeepJet tagger is based on a convolutional DNN, replacing the track-based lifetime information used in DeepCSV with low-level observables of 25 charged and neutral particle-flow jet constituents with the highest p_T value. Additionally, the four leading secondary vertices associated with a jet are incorporated. For each input set, separate 1×1 convolutional layers are implemented with progressively smaller filter sizes, applied individually to each charged particle, neutral particle, or secondary vertex. Thus, the convolutional layers extract local features of a jet to identify key sub-structures. The output from the convolutional layers is fed into a set of multiple recurrent layers based on Long Short-Term Memory (LSTM) cells. The purpose of the LSTM is to capture additional dependencies between the jet constituents independently for charged particles, neutral particles, and secondary vertices. The LSTM outputs are then combined with

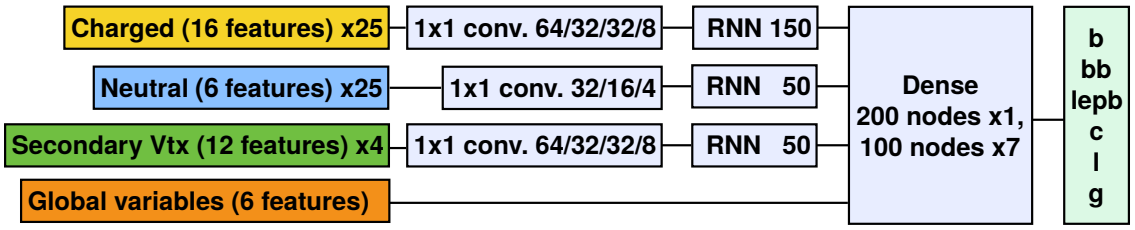


Figure 4.2: Architecture of the DeepJet heavy flavor jet tagging algorithm. Taken from Ref. [92].

information of global jet observables and propagated through a multilayer perceptron (MLP) (cf. Section 5.1) to concatenate all features. The output layer of the multi-class classification is composed of six nodes, each assigning a probability that a jet belongs to one of the categories. Three nodes correspond to different properties of b jets: with one B hadron, two B hadrons and B hadrons containing leptonic decays. The combination of these three output nodes defines the discriminant quantity used in this thesis, which is referred to as b tagging value. Jet and event classification performed with graph neural networks (GNNs) in this thesis relies significantly on this input feature, as described in Chapter 9. The remaining three output nodes represent the jet classes corresponding to c-, up/down/strange (uds)-, and gluon-initiated jets. The architecture of the DeepJet tagger is shown in Figure 4.2.

Two-dimensional discriminants can be constructed from the prediction scores of the DeepJet tagger. For this, ratio observables are defined based on the DeepJet output node probabilities, which reflect the predictions of the c vs. b (CvB) and c vs. light (CvL) jet discrimination. Light jets in this thesis refer to jets originating from up, down, and strange quarks, as well as gluons. The two observables are defined as

$$\text{CvB} = \frac{p(\text{c})}{p(\text{c}) + p(\text{b}) + p(\text{bb}) + p(\text{lepb})} \quad , \tag{4.3}$$

and

$$\text{CvL} = \frac{p(\text{c})}{p(\text{c}) + p(\text{uds}) + p(\text{g})} \quad , \tag{4.4}$$

where $p(i)$ refers to the probability in the output node i of the DeepJet tagging algorithm. The tagger performance is evaluated on simulated events, e.g. $\text{t}\bar{\text{t}}$+jets in the fully hadronic channel. For this purpose, the Receiver Operating Characteristic (ROC) is calculated, where the light jet misidentification rate is measured against the b jet efficiency. The ROC performance measure is depicted in Figure 4.3.

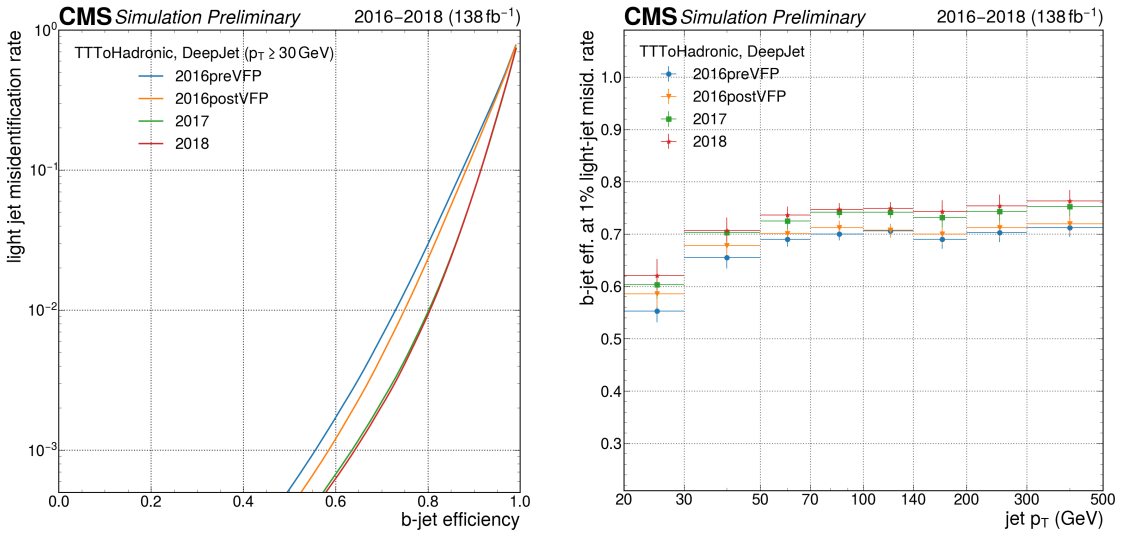


Figure 4.3: ROC of the DeepJet heavy flavor jet tagging algorithm on simulated $t\bar{t}$ events in the fully hadronic channel for all jets with transverse momentum of $p_T \geq 30\,\text{GeV}$ (left). The performance of the tagging algorithm is evaluated at a fixed 1% misidentification rate of light jets as a function of the jet p_T (right). The b jet efficiencies are separated into the four Run-II data-taking eras. Taken from Ref. [97].

Typically, three different working points (WPs) are defined at constant light-jet misidentification rates: 0.1% (tight WP), 1% (medium WP), and 10% (loose WP). Although the WPs are not utilized in this analysis as the phase space is chosen to be inclusive without selections on b-tagged jets, control distributions reflecting bins of the number of b-tagged jets defined with the medium WP can be found in Section 8.5. Figure 4.3 illustrates the dependence of the b jet efficiency on the jet's transverse momentum. For jets with $p_T \geq 30\,\text{GeV}$, the efficiency ranges from 70% to 80% for the medium WP in the 2018 data-taking era. For jets with a transverse momenta of 20 to 30 GeV, the efficiency is significantly lower. For this reason, only jets with $p_T \geq 30\,\text{GeV}$ are considered in this thesis, as outlined in Chapter 8.2.4.

4.5 Missing transverse energy

Particles produced during collisions that are only weakly interacting and electrically neutral cannot be reconstructed in the CMS detector. These undetectable particles include not only neutrinos but also hypothetical Dark Matter candidates. However, the presence of these particles can be inferred indirectly by calculating the sum of transverse momenta in an event. At the LHC, the transverse momenta of the colliding protons are negligibly small, as they collide head-on along the beam axis, which is depicted in Figure 3.3. Therefore,

the sum of the transverse momenta of all produced particles must also be zero after the collision. Deviations in the transverse momenta indicate that energy is missing and potentially carried away by undetected particles. Therefore, the MET is calculated with

$$\text{MET} = \left| - \sum_{\text{detected}} \vec{p}_{\text{T},i} \right| = \left| \sum_{\text{undetected}} \vec{p}_{\text{T},i} \right| \quad . \tag{4.5}$$

Two key aspects must be considered: First, the transverse energy of all particles must be fully reconstructed. Therefore, MET is sensitive to sub-detector resolutions and detector artifacts. Second, the number of undetected particles and their respective MET fractions cannot be determined.

4.6 Event simulation

Event simulation is an essential element in measurements at the CMS experiment to understand physics processes in proton-proton collisions and to interpret the measured data. As shown in Chapter 7, in a given recorded collision event the physics process that occurred is unknown. Particularly in events with identical reconstructed final states, it is not possible to unambiguously determine the process leading to the signature in the detector. Therefore, events from all processes expected to occur in the proton-proton collisions are simulated to estimate and predict the signatures of data observed in the measurements. The object reconstruction methods are identical for simulated events and recorded data. This approach enables the search for beyond the Standard Model (BSM) physics and the precise testing of existing theories, such as the Standard Model of particle physics (SM). In addition, simulated events are valuable as they enable the training of Artificial Intelligence (AI) models in supervised learning methods using labeled data, as discussed in Chapter 9.

Simulating an event is a comprehensive procedure involving a combination of theoretical and experimental knowledge at various stages. An overview of the main steps is given below. Additionally, an exemplary implementation is discussed for the most important simulation setup in this thesis.

The first step of every event simulation is the calculation of the matrix element (ME) for a given physics process. The ME refers to the quantum mechanical transition amplitude for a process resulting in a specific final state scenario. The next step is the simulation of the hard process, which is the primary collision in which high-energy particles are generated. The leading order (LO) calculation in QCD considers Feynman diagrams with the lowest order in the strong coupling constant α_S. Additional Next-to-leading order (NLO) corrections take into account contributions from radiation processes or loop diagrams. The POWHEG BOX (A Positive Weight Hardest Emission Generator) is an ME event generator for the implementation of these steps [98, 99]. After hard scattering, additional

radiation is simulated, referred to as parton shower (PS), taking into account higher-order contributions. The PSs consider additional radiation before the scattering, referred to as initial-state radiation (ISR), and after the scattering, called final-state radiation (FSR). As the name suggests, POWHEG BOX generates the hardest emission first and can be interfaced with a PS simulator such as PYTHIA for the remaining radiation [100]. For this, the MEs and PSs must be matched in a procedure that handles the transitions between the ME calculation for the hard process and the PS correctly to avoid double counting of radiation. The splitting into two partons in the PS is calculated with Sudakov form factors [101]. The factors describe the probability of a parton at an energy scale q_i^2 to evolve to an energy scale q_f^2 without emitting an additional parton. However, as the PS energy decreases and α_S consequently increases, simulating hadronization processes becomes necessary. When the energy scale drops below 1 GeV, the strong coupling constant assumes values of $\mathcal{O}(1)$, and perturbation theory can no longer be applied. After the PS, quarks and gluons transition into bound states through the process of hadronization. This process is described using phenomenological models such as the Lund string model implemented in PYTHIA [102]. Finally, the simulated particles are propagated to a detector simulation of the CMS experiment to simulate the interaction of all particles with the detector components. Thus, at this level of simulation, the event resembles the signal as it would appear in a recorded event from a real collision. For the detector simulation, the GEANT 4 software framework is employed [103]. Still, the simulation is inherently an approximation, constrained by the available knowledge of the detector, the underlying physics, and other relevant factors. Known discrepancies between simulated data and recorded data in the reconstruction process are corrected, as detailed in Section 8.4. Simplified detector simulations, for instance with DELPHES, can simulate the detector behavior up to two orders of magnitude faster [104]. However, a study on the event simulation with DELPHES for the production of training data for relevant processes in this thesis shows that the distributions of basic kinematic observables can deviate by up to 10% compared to the event simulation with an extensive detector description in GEANT 4 [105].

4.6.1 Jet flavor identification and jet origin

For the analysis strategy, as defined in Section 7.3, the flavor and origin of a clustered jet from stable final-state particles at the generator level, i.e. before the detector simulation, is crucial. However, this requires the information about the hadrons from hadronization at this level. The information is obtained in several steps. First, all B and C hadrons in the event are identified. For each hadron, the corresponding b or c quark is determined by tracing the particle chain back to the initial state particles. During this process, each B or C hadron is matched to a single quark. The parent of this quark is then considered the origin of the hadron, which can be a top quark, Higgs boson, Z boson, gluon, or any other particle. Finally, each hadron has to be associated with a jet. For this, a ghost clustering algorithm is employed [106, 107]. In this process, the momenta of the hadrons

are scaled to infinitely small values. These hadrons, each with negligible energy, are then clustered together with the stable final-state particles. Due to the chosen jet clustering algorithm, the original jet is not altered. The hadron among the jet constituents can then be identified, and the information about the flavor and the origin of this jet is revealed.

4.6.2 Event topologies

In case where a b jet originates from a top quark decay, it is referred to as a b jet from the $t\bar{t}$ system in this thesis. On the other hand, if a jet that does not stem from a top quark decay, e.g. from a gluon splitting or H/Z boson decay, it is referred to as an additional jet.

A distinction is made between different event topologies of simulated $t\bar{t}$ events based on the hadron flavor in the additional jets. Events in which two additional B hadrons enter two different jets in addition to the $t\bar{t}$ system they are referred to as $t\bar{t}b\bar{b}$. If both additional B hadrons appear in one additional jet, it is referred to as $t\bar{t}2b$. This occurs when both B hadrons are clustered in a jet due to a collinear $g \to b\bar{b}$ splitting, for example. In events in which one B hadron is matched to an additional jet, but the other B hadron is out of acceptance, this is referred to as $t\bar{t}b$. The group of $t\bar{t}b\bar{b}$, $t\bar{t}2b$, and $t\bar{t}b$ is collectively referred to as $t\bar{t}B$. Correspondingly, $t\bar{t}H(B)$ and $t\bar{t}Z(B)$ are defined with the equivalent sub-process definitions. If no additional B hadrons are present, but additional C hadrons exist, analogous logics are applied to $t\bar{t}C$ and its sub-processes. If neither additional B nor C hadrons are present, the process is referred to as $t\bar{t}LF$, where LF indicates the additional light flavor jets.

5 Graph neural networks

In proton-proton collisions, the reconstructed collision products result in unknown final state physics object multiplicity, e.g. the number of jets in an event. Furthermore, all decay products in an event form a permutation invariant set of final state objects. Mathematical graphs satisfy these two properties, making graph neural networks (GNNs) a natural choice for jet and event classification. In this chapter, the key properties of GNNs are explored, structured as follows: Section 5.1 provides an introduction to the fundamental principles of neural networks. Section 5.2 focuses on key properties of mathematical graphs, whose topology forms the basis of GNNs. Section 5.3 delves into the core principles of message passing and aggregation in GNN architectures. The application and the architecture employed for classification in this thesis are discussed in detail in Chapter 9.

5.1 Machine learning and neural networks

Modeled after their biological counterparts, neural networks emulate the core processes of natural neural systems. A single neuron can handle simple decision-making tasks, while an entire network can analyze complex relationships and recognize intricate patterns. The structure of a single perceptron is composed of n inputs, denoted as x_i for $i = 1, \ldots, n$. Each input x_i is assigned a weight w_i, which scales the input to reflect its relative importance. The sum of the weighted inputs serves as argument for a non-linear activation function ψ. The activation function processes the sum of the weighted inputs and determines a prediction p, calculated with

$$p = \psi \left(\sum_{i=1}^{n} x_i w_i \right). \tag{5.1}$$

The prediction p enables the perceptron to perform a decision. If p is less than or equal to a threshold b, the perceptron assigns the input to class A; otherwise, the input is

classified as class B. During the training process of the perceptron, both the weights w_i and the threshold b are adjusted to optimize the classification performance. Combining multiple perceptrons into a structured network enables the classification of more complex patterns. This is achieved by organizing multiple perceptrons in several layers, referred to as multilayer perceptron (MLP). In an MLP, each perceptron in a given layer is connected to each perceptron in the subsequent layer, known as fully connected feedforward network. The input layer is the first layer in an MLP, where the data is fed into the network, followed by a series of hidden layers. The final layer is called the output layer, which enables prediction or classification into categories, as discussed for a single perceptron. Hence, the output from a perceptron in the l-th layer serves as one of the inputs for the $(l+1)$-th layer in the MLP. An MLP with multiple hidden layers is considered a deep neural network (DNN).

In this thesis, the non-linear activation function employed is the LeakyReLU function, defined as

$$\psi(x') = \begin{cases} x' & \text{if } x' > 0 \\ \epsilon x' & \text{if } x' \leq 0 \end{cases} , \tag{5.2}$$

where ϵ is a tunable parameter and set to $\epsilon = 0.1$ in this thesis. The activation function is preferred for its simplicity. It is based on the ReLU function, but unlike ReLU, the LeakyReLU function has a non-zero slope for negative values. This prevents the vanishing gradient problem by ensuring small gradients for negative input values, which allows gradients to propagate through the network during backpropagation, as explained below.

The sigmoid activation function is used in the output layer for classification in this thesis, which is defined as

$$\sigma(x') = \frac{1}{1 + e^{-x'}} \quad . \tag{5.3}$$

Its ability to produce values between 0 and 1 is advantageous, as these values can be directly interpreted as probabilities, such as the likelihood of a jet belonging to a specific class.

Training a neural network involves adjusting its weights and biases, collectively referred to as trainable parameters. The training process seeks to adjust the trainable parameters to improve the network's ability to predict values that are close to the true value, e.g. in a classification task. Therefore, the loss function L is introduced as a measure of how good the predictions of the network are compared to the true class. The cross-entropy is a suitable choice, as it quantifies the difference between the predicted probability distribution and the actual target class distribution. It is defined as

$$L = -\sum_i y_i \log(\hat{y}_i) \quad , \tag{5.4}$$

where y_i is the one-hot encoded true label and $\hat{y}_i$ denotes the predicted probability for class i. Consequently, if the network assigns a high probability to the true class, the cross-entropy loss is minimized.

To minimize the loss function, a gradient descent approach is applied during training. In this procedure, the gradient of the loss function with respect to the weights is calculated and adjusted in the direction of the steepest descent.

The weight update in the gradient descent is calculated with

$$w_{ji}^{(l)} \leftarrow w_{ji}^{(l)} - \lambda \frac{\partial L}{\partial w_{ji}^{(l)}} \quad , \tag{5.5}$$

where $w_{ji}^{(l)}$ represents the assigned weight of the connection between the j-th perceptron in the l-th hidden layer and the i-th perceptron in the preceding layer. The hyperparameter λ known as the learning rate determines the step size for updating the network's weights and is a tunable parameter of the model. Hence, the gradient descent compels the network to progressively minimize the loss function. This process is operated iteratively over multiple epochs, during which the loss function is evaluated in each epoch and the weights updated incrementally. An epoch refers to a full cycle through the entire training data set during the network's training procedure.

The gradient descent procedure is executed using the Adam optimizer in this thesis [108]. Adam uses an adaptive learning rate λ for each parameter, adjusting it individually based on the size of the previous gradient, which improves optimization efficiency. The optimizer employs exponential moving averages of the gradients combined with squared gradients, which speeds up and stabilizes the convergence by considering past gradients.

A potential risk in the neural network minimization process is overfitting the trainable parameters to the training data, which can compromise the network's generalization ability. To avoid such overfitting during the training procedure, regularization techniques are employed. One regularization technique, called dropout [109], randomly deactivates perceptrons during training to prevent overfitting. At each epoch, each perceptron is dropped out with a predefined probability, effectively excluding it from all computations. This enforces the network to learn robustly and not rely on individual perceptrons. During testing or validation, dropout is disabled, and all perceptrons remain active.

5.2 Graph properties

A mathematical graph is a structure defined by a set of nodes (V) and a set of edges (E). A graph is then an ordered structure defined as the tuple $G = (V, E)$, where $V = \{v_1, v_2, \ldots, v_N\}$ represents the set of nodes and $E \subseteq V \times V$ is the set of edges connecting pairs of nodes [110]. Each edge $(u, v) \in E$ connects two nodes u and v. Furthermore, graph structures carry additional information stored in nodes and edges, referred to as features in this thesis. A weight w_{uv} is associated with each edge (u, v) that represents the strength or importance of the respective connection between two nodes. Similarly, each node v of a graph is assigned features, i.e. multiple properties representing the characteristics of the node object. However, properties can also be assigned to the entire graph, for example a label for the overall characteristic of the graph. A practical application of a structure with the aforementioned properties is, for example, a molecule. In an exemplary graph, nodes represent atoms (e.g. carbon or hydrogen atoms), edges represent the chemical bonds (e.g. single bonds between carbon and hydrogen) and the chemical properties of the molecule can be attributed to the entire graph (e.g. the type of molecule).

A key element in the representation of the graph structure is the adjacency matrix A. For a graph with N nodes, A is an $N \times N$ matrix in which the matrix elements A_{ij} are

$$A_{ij} = \begin{cases} 1 & \text{if an edge from node } i \text{ to } j \text{ exists} \\ 0 & \text{otherwise} \end{cases} . \tag{5.6}$$

As mentioned before, in case of weighted graphs, A_{ij} represents the weight w_{ij} of the edge between the node i and j. Generally, graphs can be categorized into two groups: directed and undirected graphs. For undirected graphs, the adjacency matrix is symmetric, i.e. $A_{ij} = A_{ji}$, whereas this does not apply to directed graphs. All graphs in this thesis are directed graphs.

As the name indicates, GNNs are a type of neural networks specifically designed to process graph-structured data. Unlike feedforward DNNs, which are typically applied to grid-like data such as images or sequential data, GNNs are specifically designed to capture the relationships between nodes in irregular, non-Euclidean structures. While both GNNs and feedforward DNNs share the fundamental principle of learning representations through layers of transformation as discussed in Section 5.1, GNNs differ in their ability to process and update information based on the graph's topology. GNNs leverage the structure of the graph to learn meaningful representations for nodes, edges, or the entire graph, by iteratively aggregating and transforming information from neighboring nodes. The paper in Ref. [111] systematically analyzes the differences between DNNs and GNNs for typical data structures of proton-proton collision decay products and constitutes results of the research associated with this thesis.

GNNs are capable to perform predictions on three different levels. In the following, exemplary applications across various domains are demonstrated. The first level is the edge level prediction. At this level, predictions are made for individual edges, focusing on their properties or existence, for instance, predicting whether two users in a social network might become friends. The second level is the node level prediction (NLP). In this case, the task is a prediction for individual nodes based on their own information and neighboring nodes, e.g. if a product is relevant for a customer under scrutiny. The third level is the graph level prediction (GLP). At this level, predictions are made about the entire graph as a unit. A typical application is classifying a molecule based on its structure, such as predicting the toxicity. This thesis employs the latter two levels: NLP and GLP. In Section 9.1, the transformation of the event topology into a graph structure is presented. The classification of jets using NLPs is described in Section 9.2, whereas the classification of events based on GLPs is discussed in Section 9.3.

5.3 Message passing and aggregation

The core mechanism in the training procedure of a GNN is called message passing. The goal is to collect information from the neighborhood of a node and iteratively incorporate it into new node representations, referred to as embedding. Consequently, a node can update its properties not only based on its own information, but also on the information of its neighbors, with connections represented by the adjacency matrix. Message passing, aggregation, and update in GNNs involve three major steps.

In the first step, the message $m_{u\to v}^{(i)}$ sent from node u to node v in the i-th iteration[1] is calculated with

$$m_{u\to v}^{(i)} = \phi(h_u^{(i)}, h_v^{(i)}, e_{uv}) \quad , \tag{5.7}$$

where the message depends on the embedding $h_u^{(i)}$ and $h_v^{(i)}$ of the nodes involved and connected with the edge e_{uv}. The function ϕ can be chosen as an MLP or any weighted linear function.

The second step is an aggregation of all messages sent by all connected nodes to the node under scrutiny. All incoming messages $m_{u\to v}^{(i)}$ for the node v are collected and aggregated with

$$\bar{m}_v^{(i)} = \mathrm{AGG}(\{m_{u\to v}^{(i)} \mid u \in \mathcal{N}(v)\}) \quad , \tag{5.8}$$

where AGG is the aggregation function. The aggregation function can be chosen arbitrarily; however, the function must be independent of the order of the neighbors. This choice

[1] The term iteration instead of layer reflects the fact that certain GNN architectures involve multiple iterations per layer.

enables permutation invariance. Functions that fulfill this requirement are the sum, average, or maximum. With this, the aggregation process reduces any number of incoming messages to a consolidated message. Therefore, the calculation is also invariant to the size of the graph, enabling different node multiplicities to be processed identically. This property is exploited in this thesis, as it allows all possible jet multiplicities in events to be considered by the GNN (cf. Chapter 9). In this thesis, a weighted aggregation is implemented based on an attention mechanism as introduced in transformer models [112]. As a result, the aggregation function is given an additional attention weighting α_{uv} and is calculated with

$$\text{AGG} = \sum_{u \in \mathcal{N}(v)} \alpha_{uv} m_{u \to v}^{(i)} \quad . \tag{5.9}$$

The third step is updating the current embedding, in which the aggregated messages are processed to update the representation of each node. The new embedding of the node v is determined with

$$h_v^{(i+1)} = \psi(h_v^{(i)}, \bar{m}_v^{(i)}) \quad , \tag{5.10}$$

where ψ is any non-linear transformation implemented as an activation function such as LeakyReLU (cf. Equation 5.2). Hence, the updated embedding can, for instance, be calculated with

$$h_v^{(i+1)} = \sigma\left(W[h_v^{(i)} \| \bar{m}_v^{(i)}]\right) \quad , \tag{5.11}$$

where $\|$ is the concatenation function and W a weight matrix containing trainable parameters. The dimensions of the vector can be changed $h_v^{(i+1)}$ compared to $h_v^{(i)}$ with the choice of the weight matrix W. The technical implementation, along with the detailed selection of GNN models, is thoroughly examined in Chapter 9.

6 Statistical data analysis

In particle physics and this thesis, statistical data analysis methods are a fundamental part when extracting small signals from vast data sets in precision measurements. Processes in particle physics are inherently probabilistic, since the underlying quantum mechanical wave functions are interpreted as probability density functions (PDFs). Yet, stochastic phenomena not only play a central role from a theoretical perspective, but are also crucial in experiments. Since the determination of cross sections is essentially a counting experiment that is affected by statistical fluctuations, statistical tools are required for evaluation. Eventually, hypothesis tests are used to analyze the large amount of data in order to determine the parameter of interests (POIs) of the measurement and compare the results with theoretical predictions. This chapter summarizes the most important statistical methods used in this thesis. In Section 6.1, the concept of a cross section measurement is introduced and the POI based on it is defined. The principle of a maximum likelihood fit and a parameter determination thereof are described in Section 6.2. In Section 6.3, hypothesis tests are discussed to evaluate a result and assess the significance of the measurements performed.

6.1 Cross sections and signal strengths

The cross section is a measure of the probability that a particular process, e.g. $\mathrm{t\bar{t}b\bar{b}}$ production, will occur in an event. It is denoted as σ in this thesis and is expressed in units of area denoted as barn (b), where $1\,\mathrm{b} = 10^{-28}\,\mathrm{m}^2$. All cross sections in this thesis are stated in pb. The predicted cross sections of all processes considered in this thesis are discussed in Chapter 8 and are documented in Table 8.1. The expected number of events N within a point in time t then results from

$$N = \sigma \cdot \int_0^t L_{\mathrm{inst}}\, \mathrm{d}t' \quad , \tag{6.1}$$

where L_{inst} is the instantaneous luminosity defined in Equation 3.1. To determine the cross section of a specific process in the Compact Muon Solenoid (CMS) experiment, this relation is used by subtracting the number of expected background events (N_{bkg}) from the number of observed events (N_{obs}). Combined with the product of the acceptance and reconstruction efficiency denoted as ϵ, which reflects the fraction of actually reconstructed events determined by auxiliary measurements or simulation, the cross section of interest is defined as

$$\sigma = \frac{N_{\text{obs}} - N_{\text{bkg}}}{\epsilon \cdot \int_0^t L_{\text{inst}} \, \mathrm{d}t'} \quad . \tag{6.2}$$

The simplicity of Equation 6.2 conceals the complexity of the actual application in a precision measurement. Since precision measurements aim to detect small signals and, as mentioned at the beginning of this chapter, involve stochastic processes alongside numerous systematic uncertainties (cf. Section 10.1), the actual determination of a cross section is far more complex. Therefore, the remainder of this chapter is devoted to this concept and based on Ref. [113–115].

Typically, the cross section in measurements is parameterized and expressed as a signal strength parameter, which is defined as

$$\mu = \frac{\sigma^{\text{obs}}}{\sigma^{\text{SM}}} \quad , \tag{6.3}$$

where σ^{obs} denotes the observed cross section and σ^{SM} refers to the cross section predicted by the SM. Accordingly, the signal strength parameter μ is defined as the cross section relative to the expectation from a theory prediction. If the signal strength parameter equals one, the measured cross section corresponds precisely to its SM prediction.

6.2 Maximum likelihood fits and parameter estimation

Generally, the observed data and predictions are represented as histograms for data reduction in this analysis. The parameter estimation, which determines the parameters that best fit the data given a particular model, is performed using a maximum likelihood fit. In this method, the content of each bin i of the histogram is expressed by a probability $\mathcal{P}$, while the products over all bins form the likelihood with

$$\mathcal{L}\left(\vec{\mu}, \vec{\theta}\right) = \prod_i \mathcal{P}\left(n_i | \lambda_i(\vec{\mu}, \vec{\theta})\right) \quad . \tag{6.4}$$

In this equation, $\mathcal{P}$ describes the probability of observing n_i events, whereas a value of λ_i is expected. This expected yield λ_i depends on two sets of parameters, which are summarized

to $\vec{\mu}$ and $\vec{\theta}$. The former describes the POIs, such as signal strength parameters from Equation 6.3, while the latter describes the nuisance parameters (NPs), which are also part of the model, but are of indirect interest. In contrast to other analyses that search for only one signal, in this thesis $\vec{\mu}$ is a set of four parameters, which are defined in Chapter 10. With the help of the NPs, it is possible to incorporate systematic uncertainties into the statistical model, which is discussed in Section 10.1. The expected yield λ_i is derived from the number of signal events (s_i) and background events (b_i) with

$$\lambda_i(\vec{\mu}, \vec{\theta}) = \vec{\mu} \cdot \vec{s}_i(\vec{\theta}) + b_i(\vec{\theta}) \quad . \tag{6.5}$$

Thus, the signal strength parameters scale the signal linearly. Since four signal strength parameters are measured in this thesis, the product $\vec{\mu} \cdot \vec{s}_i(\vec{\theta})$ is given by

$$\vec{\mu} \cdot \vec{s}_i(\vec{\theta}) = \mu_{\mathrm{t\bar{t}b\bar{b}},i}\ s_{\mathrm{t\bar{t}b\bar{b}},i}(\vec{\theta}) + \mu_{\mathrm{t\bar{t}c\bar{c}},i}\ s_{\mathrm{t\bar{t}c\bar{c}},i}(\vec{\theta}) + \mu_{\mathrm{t\bar{t}H},i}\ s_{\mathrm{t\bar{t}H},i}(\vec{\theta}) + \mu_{\mathrm{t\bar{t}Z},i}\ s_{\mathrm{t\bar{t}Z},i}(\vec{\theta}) \quad . \tag{6.6}$$

The probability $\mathcal{P}$ in bin i follows a Poisson distribution, as it is a natural choice describing independent events with a constant mean rate within a given interval. The Poisson distribution with respect to the parameters introduced is described by

$$\mathcal{P}\left(n_i|\lambda_i(\vec{\mu}, \vec{\theta})\right) = \frac{\lambda_i^{n_i}(\vec{\mu}, \vec{\theta})}{n_i!} e^{-\lambda_i(\vec{\mu}, \vec{\theta})} \quad . \tag{6.7}$$

To incorporate systematic uncertainties into the model via the NPs, the likelihood from Equation 6.4 combined with Equation 6.5 is modified with an additional constraint term, which results in

$$\mathcal{L}\left(\vec{\mu}, \vec{\theta}\right) = \prod_i \mathcal{P}\left(n_i|\vec{\mu} \cdot s_i(\vec{\theta}) + b_i(\vec{\theta})\right) \cdot P(\vec{\tilde{\theta}}|\vec{\theta}) \quad . \tag{6.8}$$

The constraint term is associated with the degree of belief $P(\vec{\theta}|\vec{\tilde{\theta}})$ for values of $\vec{\theta}$ given $\vec{\tilde{\theta}}$ and indicates what the true values of the NPs $\vec{\theta}$ could be via Bayes' theorem

$$P(\vec{\theta}|\vec{\tilde{\theta}}) \propto P(\vec{\tilde{\theta}}|\vec{\theta}) \cdot \pi_{\vec{\theta}}(\vec{\theta}) \quad , \tag{6.9}$$

where $P(\vec{\tilde{\theta}}|\vec{\theta})$ is a conditional probability obtained from auxiliary measurements or theoretical calculations, $\vec{\tilde{\theta}}$ are the default values of the NPs, and $\pi_{\vec{\theta}}(\vec{\theta})$ is the prior probability of $\vec{\theta}$, often assumed to be uniformly distributed [113].

As previously mentioned, systematic uncertainties incorporated into the likelihood via NPs can be considered correlated or uncorrelated (cf. Section 10.1). In the case of a correlation,

systematic uncertainties are described by the same NP, while in the uncorrelated case, they are described by different NPs. To determine the optimal set of parameters, the NPs $\vec{\theta}$ are profiled as functions of the POIs $\vec{\mu}$, which reduces the total number of parameters to be estimated. The optimal test statistic results from the likelihood ratio according to the Neyman-Pearson lemma [116]. This ratio is defined as the likelihood, which is maximized by a set of parameters $\{\vec{\mu}, \hat{\vec{\theta}}(\vec{\mu})\}$ for fixed values of $\vec{\mu}$ relative to the likelihood whose set of parameters $\{\hat{\vec{\mu}}, \hat{\vec{\theta}}(\hat{\vec{\mu}})\}$ maximizes the likelihood function globally. For example, $\vec{\mu} = 0$ can be chosen for the fixed value of the parameter set in the numerator, whose choice is motivated in the following section. By constructing a profiled negative log-likelihood ratio, it is possible to perform a numerically preferable minimization procedure instead of a maximization method. As a result, the test statistic $q_{\vec{\mu}}$ to be minimized is

$$\tilde{q}_{\vec{\mu}} = -2 \log \left(\frac{\mathcal{L}\left(\vec{\mu}, \hat{\vec{\theta}}(\vec{\mu})\right)}{\mathcal{L}\left(\hat{\vec{\mu}}, \hat{\vec{\theta}}(\hat{\vec{\mu}})\right)} \right) \quad . \tag{6.10}$$

It should be noted that this method determines the parameter values that are most compatible with the data for a given model rather than finding the correct model. The construction of an accurate model therefore remains crucial. The correlation of parameters $\alpha \in \{\vec{\mu}, \vec{\theta}\}$ can be calculated from the inverse of the Hessian matrix of the likelihood function, which is defined as matrix of second derivatives at the maximum of the likelihood. The components of this covariance matrix C are then calculated from

$$C_{ij}^{-1} = \left. \frac{\partial \mathcal{L}}{\partial \alpha_i \partial \alpha_j} \right|_{\hat{\vec{\alpha}}} \quad . \tag{6.11}$$

In addition, confidence intervals can be obtained from the covariance matrix with

$$\alpha_i^{\uparrow\downarrow} = \hat{\alpha}_i \pm \sqrt{C_{ii}} \quad . \tag{6.12}$$

Alternatively, the test statistic $q_{\vec{\mu}}$ can be used in the limit of large samples to form confidence intervals according to the Wilks' theorem [117]. The confidence intervals then correspond to Gaussian standard deviations, since $q_{\vec{\mu}}$ asymptotically follows a χ^2 distribution. The interval boundaries can then be read from the values of the test statistic, as all parameter values that lie within s Gaussian standard deviations belong to the considered interval

$$\tilde{q}_{\vec{\mu}}^{\uparrow\downarrow}(s) = \tilde{q}_{\vec{\mu}}^{\min} \pm s^2 \quad . \tag{6.13}$$

6.3 Hypothesis tests and significance

In particle physics, the null hypothesis H_0 is typically chosen as ‘‘background-only’’, i.e. the absence of the signal processes with a prediction of the data by a model with all known processes. The alternative hypothesis H_1 is the combination of both, which is referred to as ‘‘signal + background’’. Therefore, hypothesis tests are performed to determine whether the data are compatible with the null hypothesis or whether they deviate significantly. The probability distribution under the H_0 hypothesis, representing the background-only model, is used to evaluate the significance of a result quantified by a p-value. Due to the conservative approach of defining the signal + background hypothesis as H_1, the null hypothesis is only rejected if H_0 can be excluded with a high level of certainty. This approach minimizes the probability of false-positive results, also known as Type I error. Hence, this procedure ensures an unbiased approach for testing the alternative hypothesis H_1.

The test statistic $\tilde{q}_{\vec{\mu}}$ from Equation 6.10 is the optimal test statistic for performing the hypothesis test [116]. To represent the null hypothesis in the test statistic, $\vec{\mu} = 0$ is set in the numerator. The corresponding test statistic is referred to as $\tilde{q}_0$ and the optimal values from the minimization in data (‘‘observed’’) are designated $\vec{\hat{\theta}}_0^{\text{obs}}$. This p-value is then defined as

$$p = \mathcal{P}\left(\tilde{q}_0 \geq \tilde{q}_0^{\text{obs}}\right) \quad . \tag{6.14}$$

Pseudodata can be generated to evaluate $\mathcal{P}$ based on the model with $\vec{\hat{\theta}}_0^{\text{obs}}$ as NP allowing for Poisson-distributed fluctuations of the event yields in each bin. The resulting PDFs can then be used to determine how likely it is to observe a value equal to or greater than $\tilde{q}_0^{\text{obs}}$ if the null hypothesis is true:

$$p = \int_{\tilde{q}_0^{\text{obs}}}^{\infty} f\left(\tilde{q}_0 | 0, \vec{\hat{\theta}}_0^{\text{obs}}\right) \mathrm{d}\tilde{q}_0 \quad . \tag{6.15}$$

The smaller the p-value, the stronger the observed result contradicts the null hypothesis, which can be rejected in favor of the H_1 hypothesis.

The decision to reject the null hypothesis can be based on a previously specified threshold. However, the p-value is usually converted into a significance, which is also referred to as Z-score. The Z-score represents how many standard deviations a result differs from the mean of a Gaussian distribution. This significance can be calculated with

$$Z = \Phi^{-1}(1 - p) \quad , \tag{6.16}$$

where Φ is the cumulative distribution function of the standard normal distribution with

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \mathrm{d}x \quad . \tag{6.17}$$

By convention, passing a 3σ threshold ($Z = 3$) is considered an ‘‘evidence’’, while exceeding 5σ ($Z = 5$) is considered a ‘‘discovery’’, leading to the rejection of the null hypothesis, as in the case of the Higgs boson discovery by the ATLAS and CMS Collaborations [7, 8]. A significance of 5σ corresponds to a p-value of $3 \cdot 10^{-5}\%$.

In this thesis, results that are calculated on pseudodata are referred to as ‘‘expected’’, e.g. ‘‘expected significance’’. Results that are determined on data are referred to as ‘‘observed’’.

Part III

$\mathbf{t\bar{t}X}$ analysis

7 Introduction

This chapter serves as an overview and introduction to the experimental part of this thesis. First, the motivation for performing the analysis and its relevance are outlined in Section 7.1. Additionally, the methodological difficulties are highlighted. This is followed by a review of related measurements, if any, and their varying approaches and findings in Section 7.2. Finally, the objective of the analysis is defined and the overall strategic concept is summarized in Section 7.3.

7.1 Motivation and challenges

Top quark physics plays an important role at the Large Hadron Collider (LHC), as elaborated in Section 2.3. Also, the key fundamentals of Higgs physics have been discussed in Section 2.1.4. This thesis addresses both topics, for which there are numerous reasons to pursue open questions. However, this combination inherently includes many associated challenges, which are outlined in the following.

The interaction of the Higgs boson with another particle is proportional to the particle's mass. Hence, the coupling to the particle with the largest known mass, the top quark, is particularly interesting. Since the direct decay of a Higgs boson into a top quark-antiquark pair ($H \to t\bar{t}$) is kinematically forbidden, the associated production of a Higgs boson and a top quark-antiquark pair ($t\bar{t}H$) can be used for probing the top-Higgs Yukawa coupling. Precision measurements of the top-Higgs Yukawa coupling are thus an important test of the Standard Model of particle physics (SM). In addition, these measurements may answer questions about physics beyond the Standard Model (BSM) that predict different coupling strengths. In this thesis, the $t\bar{t}H$ production with a subsequent decay of the Higgs boson into a b quark-antiquark pair ($H \to b\bar{b}$) is investigated. With a branching fraction of $(53 \pm 8)\%$, the $H \to b\bar{b}$ channel represents the highest branching fraction of all channels and offers the advantage of a pure coupling of the Higgs boson to third generation fermions,

which allows direct access to the Higgs Yukawa sector [43]. However, the signals of $t\bar{t}H(b\bar{b})$ productions in the data are overshadowed by tremendous backgrounds that are quantified later.

An irreducible process to $t\bar{t}H(b\bar{b})$ is the associated production of a top quark-antiquark pair with a Z boson ($t\bar{t}Z$). If the Z boson, analogously to the Higgs boson, decays into b quark-antiquark pair ($Z \to b\bar{b}$), both processes constitute identical final states. Therefore, the $t\bar{t}Z(b\bar{b})$ process must be thoroughly understood in order to enable a precise measurement of $t\bar{t}H(b\bar{b})$. However, the $t\bar{t}Z(b\bar{b})$ process is not only interesting as a background to $t\bar{t}H(b\bar{b})$, but also in its own terms. The weak coupling of the Z boson to the top quark is a test of the SM and contributions from BSM physics can in principle be explored in this process. Although these tests are more efficient when considering leptonic Z boson decay channels due to their cleaner signatures and lower background levels, the $Z \to b\bar{b}$ final state provides a complementary probe offering valuable insights.

Another process in the phase space of $t\bar{t}H(b\bar{b})$ and $t\bar{t}Z(b\bar{b})$ is the associated production of a top quark-antiquark pair with additional jet radiation ($t\bar{t}$+jets). In particular, if the additional radiation results in a b quark-antiquark pair (e.g. $g \to b\bar{b}$), the so-called $t\bar{t}b\bar{b}$ process also forms an irreducible background for the two processes discussed so far. However, the $t\bar{t}b\bar{b}$ process is not just an undesired background for $t\bar{t}H(b\bar{b})$ and $t\bar{t}Z(b\bar{b})$. Due to the two different energy scales of the top and the b quark masses, the modeling of the process is challenging. An accurate measurement of the $t\bar{t}b\bar{b}$ process is therefore key to a profound understanding of its multiscale quantum chromodynamics (QCD) nature. While $t\bar{t}b\bar{b}$ processes have a comparably distinct signature with respect to $t\bar{t}$ events with additional light flavor radiation ($t\bar{t}$LF) from u, d, s, and g, this characteristic is less pronounced for $t\bar{t}$ processes with additional c jets ($t\bar{t}c\bar{c}$). However, dedicated heavy flavor tagging algorithms, which also enable c jet identification, allow for a distinction between $t\bar{t}b\bar{b}$, $t\bar{t}c\bar{c}$, and $t\bar{t}$LF signatures. Therefore, $t\bar{t}c\bar{c}$ measurements are a crucial element to thoroughly understand the underlying processes in the $t\bar{t}$ + heavy flavor phase space.

In this thesis, the dileptonic $t\bar{t}$ channel is chosen due to its particular clear signature at the sacrifice of lower rates. Therefore, reduced data statistics compared to other channels is an additional challenge in this analysis. Detailed advantages and disadvantages of the top decay channels are described in Section 2.3. The leading order (LO) Feynman diagrams of the processes $t\bar{t}b\bar{b}$, $t\bar{t}H(b\bar{b})$, and $t\bar{t}Z(b\bar{b})$ are depicted in Figure 7.1. In all processes, the dominant production mode at the LHC is the interaction of two gluons. The top quark (antiquark) decays into a W boson and a b quark (antiquark), as detailed in Section 2.3. In both cases, the W boson decays leptonically to a charged lepton (e, μ) and the corresponding (anti-)neutrino ($W \to \ell\nu_\ell$). The only difference in the three diagrams is the center part where a different boson is created. In the $t\bar{t}b\bar{b}$ process, an additional $b\bar{b}$ pair arises from gluon splitting ($g \to b\bar{b}$). In the other two cases, the quark-antiquark pair originates from the decay of the H/Z boson.

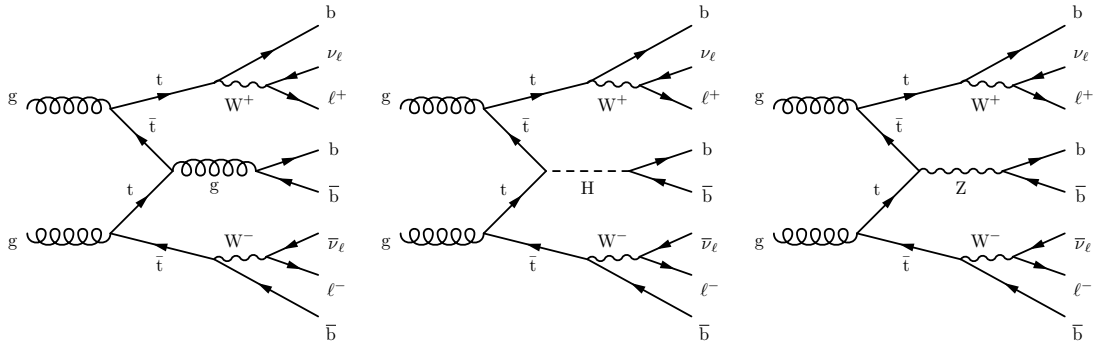


Figure 7.1: Examples of LO Feynman diagrams for the $t\bar{t}b\bar{b}$ (left), $t\bar{t}H(b\bar{b})$ (middle), and $t\bar{t}Z(b\bar{b})$ (right) processes in the dilepton channel.

In Figures 7.2 and 7.3, an example of a simulated $t\bar{t}H(b\bar{b})$ event is shown in three dimensions from different angles. In these figures, tracks and energy depositions can be observed in the sub-detectors. In addition to an electron and a muon, the event contains other low p_T leptons that are removed after an event cleaning procedure, as they do not pass the requirements defined in Section 8.2. Six jets can be identified in these figures, represented as cones of clustered particles. However, it is not possible to infer the underlying production process causing a given final state in an individual data event based on the measured signatures. Such determinations cannot be achieved on an event-by-event basis. Instead, statistical analysis must be employed, relying on expected rates, kinematic distributions, and other properties of the final state objects, as introduced in Chapter 6.

To summarize, the precision measurements of these four processes are essential for understanding the SM and exploring BSM theories in multiple aspects. Furthermore, all these processes are also crucial in other analyses where they form important backgrounds. The $t\bar{t}$+jets processes can be found in numerous phase spaces of analyses at the LHC. Together with the $t\bar{t}H(b\bar{b})$ and $t\bar{t}Z(b\bar{b})$ processes, these processes form important backgrounds in analyses such as four top quark production in channels with up to two leptons [119]. These analyses are of interest as they can complementarily measure the top quark Yukawa coupling. Four top quark measurements also serve as a probe for BSM physics, since some models predict an extended Higgs sector leading to additional scalar particles with Yukawa-like couplings to top quarks [120].

In previous analyses each of the four processes have been only treated independently as a signal process while others have been treated as background. Therefore, as a new methodology, instead of defining one process as a signal and performing a separate measurement, all processes are treated as signals simultaneously. This allows for a homogeneous treatment of all processes as well as a correct treatment of correlations. Additionally, all processes are treated with equal importance. For example, the event classifier in Chapter 9 is not tuned for any specific process in a way that would compromise the sensitivity to other

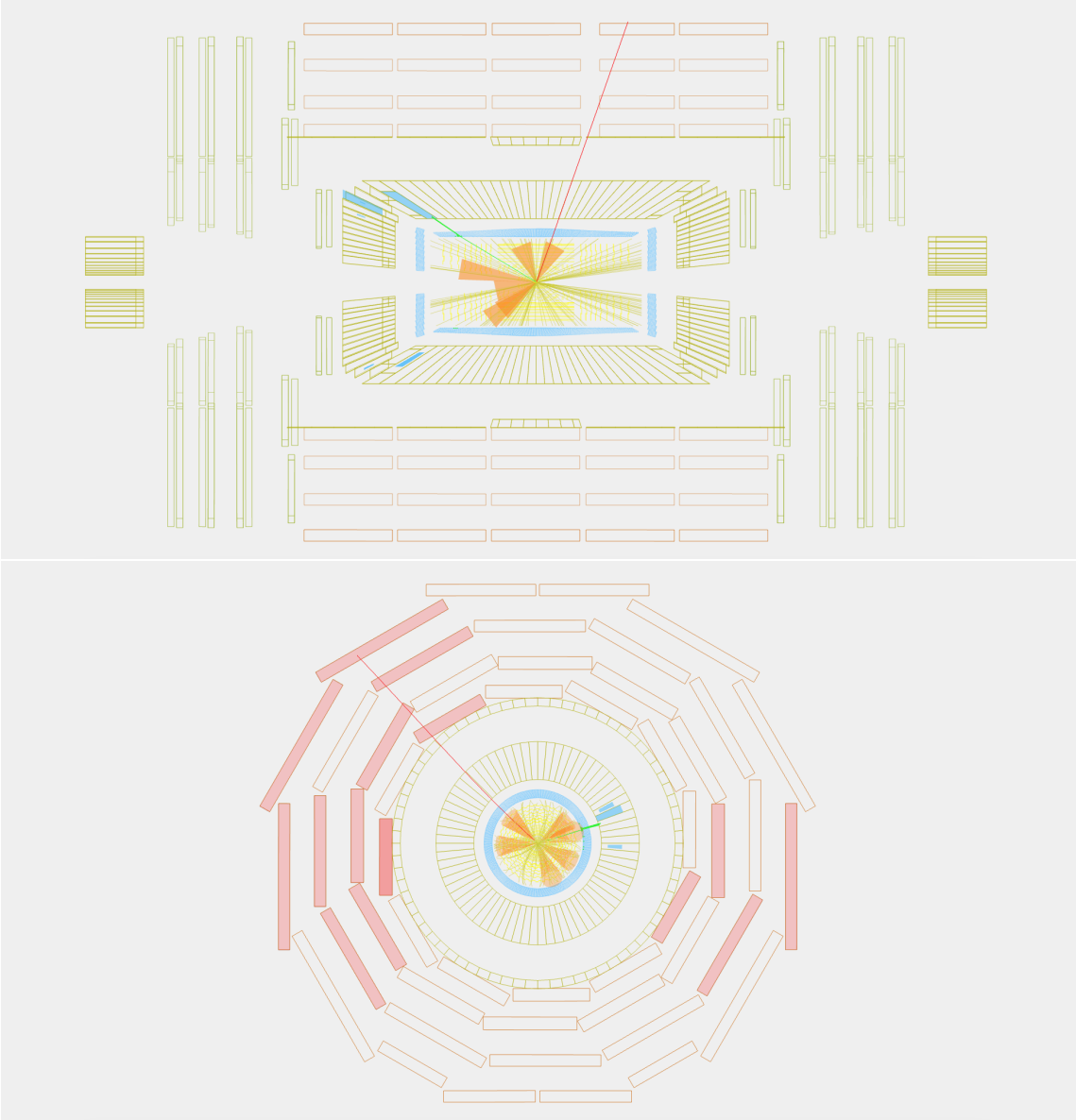


Figure 7.2: Display of a simulated $t\bar{t}H(b\bar{b})$ event in the dilepton channel projected onto the xz (top) and yx (bottom) planes. Reconstructed tracks are visualized in yellow. Energy depositions in the electromagnetic calorimeter (ECAL) are indicated in green and in the subsequent hadronic calorimeter (HCAL) in blue. A muon is represented as a red line. Clustered jets are illustrated in orange. Simulated event visualized with [118].

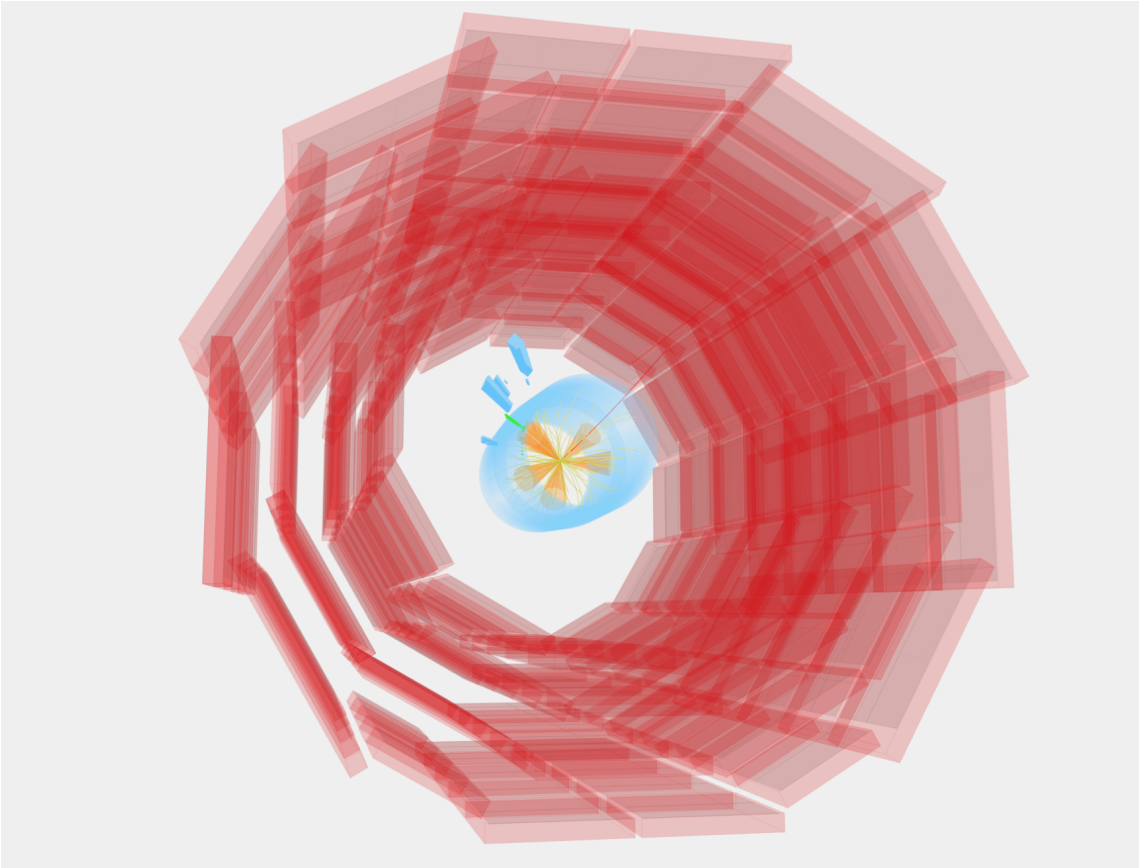


Figure 7.3: Display of a simulated $t\bar{t}H(b\bar{b})$ event in the dilepton channel. Reconstructed tracks are visualized in yellow. Energy depositions in the ECAL are indicated in green and in the subsequent HCAL in blue. A muon is represented as a red line. Clustered jets are illustrated in orange. Simulated event visualized with [118].

processes. However, the ambition to measure all these processes at the same time leads to numerous challenges.

The biggest challenge is the irreducible final state and the highly imbalanced production cross sections of the processes in $t\bar{t}$ + heavy flavor phase space. Figure 7.4 shows the cross section of selected processes and indicates how the $t\bar{t}$ cross sections exceeds the $t\bar{t}Z$ and $t\bar{t}H$ cross sections by three orders of magnitude.

Considering the branching fractions of Z and H bosons decaying into $b\bar{b}$, the cross section for the $t\bar{t}b\bar{b}$ process is an order of magnitude higher than the cross sections times branching fraction for $t\bar{t}H(b\bar{b})$ and $t\bar{t}Z(b\bar{b})$, which are approximately $\sigma_{t\bar{t}H(b\bar{b})} = 0.29\,\text{pb}$ and $\sigma_{t\bar{t}Z(b\bar{b})} = 0.13\,\text{pb}$ [122, 123]. Additionally, the $t\bar{t}b\bar{b}$ process is difficult to model due to the different scales mentioned above. The relevant energy scale of $pp \to t\bar{t}$ is in the order of several hundred GeV, whereas the energy scale of the additional b quark-antiquark pair is of the order of 10 GeV. This results in two widely separated energy scales, without the scale of the $b\bar{b}$ being negligible. The $t\bar{t}b\bar{b}$ process, which is already substantially more prevalent compared to $t\bar{t}H(b\bar{b})$ and $t\bar{t}Z(b\bar{b})$, additionally suffers from large uncertainties depending on the modeling approach [96, 124]. The choice of the renormalization and factorization scales leads to an uncertainty of up to 50 % on the cross section [125]. This makes $t\bar{t}b\bar{b}$ the leading factor of sensitivity in previous measurements of $t\bar{t}H(b\bar{b})$ at the CMS experiment [126]. As of the time of this thesis, there is no clear answer regarding the best possible modeling approach of $t\bar{t}b\bar{b}$. The current state of the art is summarized in Section 7.2.

Another challenge in distinguishing events in the $t\bar{t}$ + heavy flavor phase space for a simultaneous measurement is the difficulty in the identification of the b jets. In LO calculations four jets are expected (cf. Figure 7.1), but the reconstruction of b jets is only possible at certain levels of precision (cf. Section 4.4). This leads to the misidentification of light (u, d, s, g) and c jets as b jets, as well as b jets that are not identified. Consequently, it is possible, among other scenarios, for events with a $t\bar{t}$ pair and an additional c quark-antiquark pair to resemble the $t\bar{t}b\bar{b}$ process. Similarly, a $t\bar{t}Z(b\bar{b})$ event might not be recognized as having a $Z \to b\bar{b}$ decay because the b jets were not identified as such. Further difficulties for simultaneous measurement are numerous uncertainties, which are discussed in Section 10.1.

7.2 Landscape of related measurements

In this section, the current theoretical and experimental understanding of the four processes examined in this thesis is reviewed. The section addresses the state-of-the-art modeling of the processes, the current experimental results, which are sometimes in contradiction with each other, as well as unexplored domains. General introductions to top quark physics and hadron collider physics are provided in Sections 2.3 and 2.4. The following sections build upon these foundations.

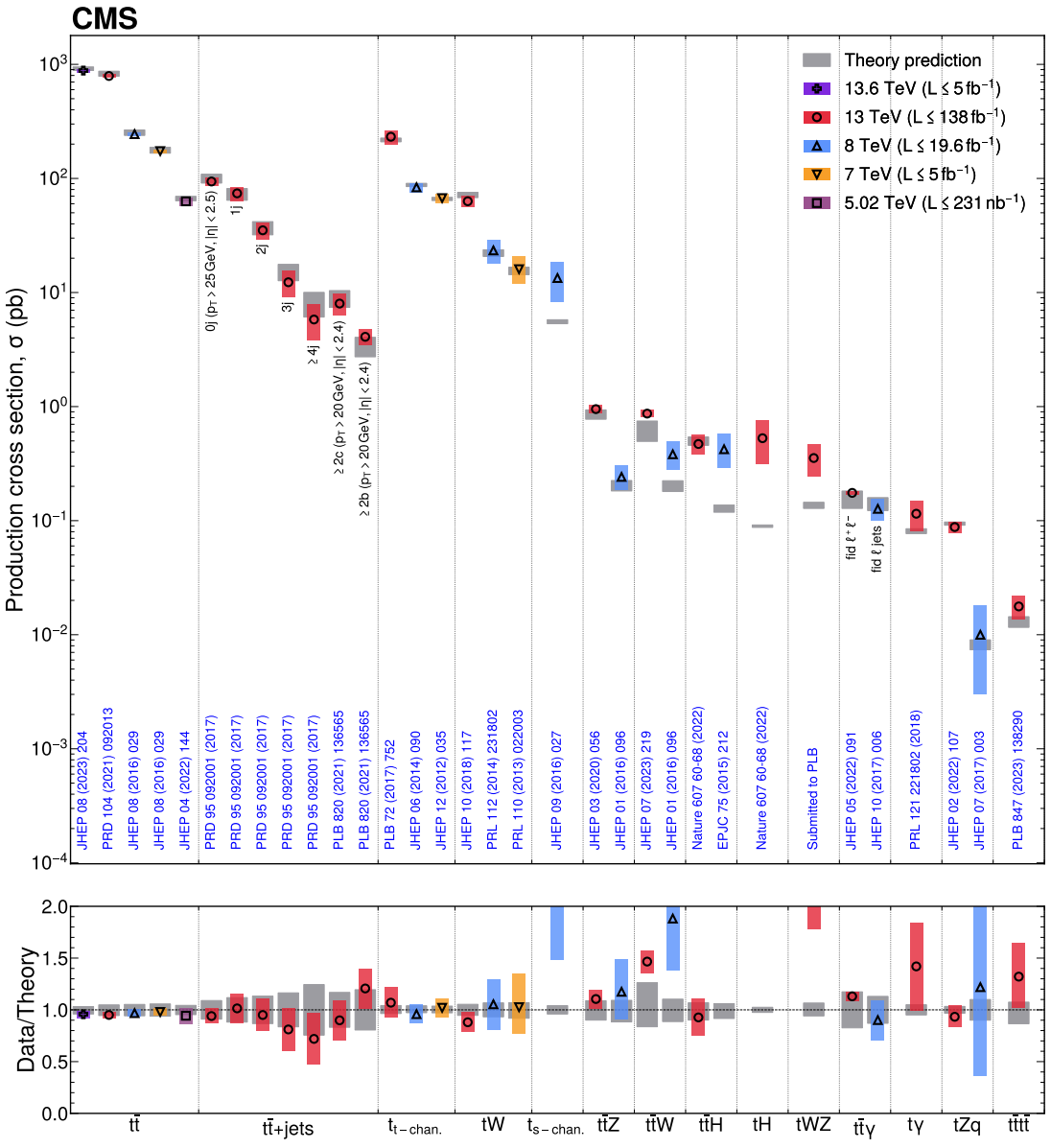


Figure 7.4: Cross sections measured by Compact Muon Solenoid (CMS) at various energies for selected processes. Gray bands refer to the SM theory uncertainty prediction while colored bands indicate the total uncertainty of the measurement. The logarithmic scale reveals that the $t\bar{t}$ cross sections is around a factor of 1000 higher than the $t\bar{t}Z$ and $t\bar{t}H$ cross section. Figure taken from Ref. [121].

7.2.1 $t\bar{t}b\bar{b}$ measurements

The ATLAS and CMS Collaborations have both published several measurements on $t\bar{t}b\bar{b}$ in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$. Three of these measurements cover data recorded in 2016 at the LHC. The ATLAS experiment has published a measurement in the semileptonic and dilepton channels with an integrated luminosity of $36.1\,\text{fb}^{-1}$ [127]. An analog measurement by the CMS Collaboration is based on an integrated luminosity of $35.9\,\text{fb}^{-1}$ with one or two leptons in the final state [128]. Additionally, an analysis in the fully hadronic channel has been published using the same data set [129]. In these measurements, the $t\bar{t}b\bar{b}$ process is described by the inclusive simulation of $t\bar{t}$ at Next-to-leading order (NLO) accuracy in QCD + parton shower (PS) in Powheg Box and Pythia in the five flavor scheme (5FS) (cf. Section 2.4). In contrast, the following analysis employ a dedicated modeling of $t\bar{t}b\bar{b}$ simulated with Powheg Box and Pythia at NLO accuracy in QCD in the four flavor scheme (4FS). Both collaborations have performed two measurements covering the full Run-II data set. The ATLAS analysis includes the semileptonic and dilepton channel; however, only the combination of exactly one electron and one muon [130]. The $t\bar{t}b\bar{b}$ model uses $\kappa_R = 0.5$ and $\kappa_F = 0.5$ as pre-factors for the renormalization and factorization scale in Equations 2.22 and 2.23. In this way, the model applies identical scales to those used in a combined study for modeling of $t\bar{t}b\bar{b}$ as an important background for $t\bar{t}H(b\bar{b})$ measurements [124]. This study, as well as the inclusive and differential measurement of $t\bar{t}b\bar{b}$ in the semileptonic channel, shows that the CMS Collaboration uses a different setting for the factorization scale [96, 131]. While the pre-factor of the renormalization scale is also set to $\kappa_R = 0.5$, a lower pre-factor of $\kappa_F = 0.25$ is chosen for the factorization term. A dedicated analysis of scale choice in a $t\bar{t}b\bar{b}$ measurement indicates that a suitable choice to describe the fiducial cross section is a combination of $\kappa_R = 0.5$ and $\kappa_F = 0.5$ [132]. This is consistent with findings from studies in the following section, in which the $t\bar{t}b\bar{b}$ process forms an important background. In addition to the simulation of $t\bar{t}b\bar{b}$ with Powheg Box, Pythia and the renormalization and factorization scales mentioned before, alternative simulations for $t\bar{t}b\bar{b}$ were also tested in Refs. [96, 132]. All simulations are tested in different phase spaces, which partially overlap, but investigate different aspects of $t\bar{t}b\bar{b}$ with increasing requirements for the number of jets, number of b jets, and light flavor jets if applicable. However, none of the $t\bar{t}b\bar{b}$ simulations simultaneously describe all measured distributions in the various phase space regions equally well. In the most inclusive phase space region, the fiducial cross section is best described by the previously mentioned Powheg Box and Pythia $t\bar{t}b\bar{b}$ model, which is also used in this thesis to describe the $t\bar{t}b\bar{b}$ process. However, this model predicts cross sections larger than the measured cross sections in the more exclusive fiducial phase space regions. Additionally, the modeling demonstrates limited accuracy in describing the distributions of jet multiplicity and b jet multiplicity, as illustrated in Figure 7.5. In Chapter 10, the strategic approach for addressing and incorporating

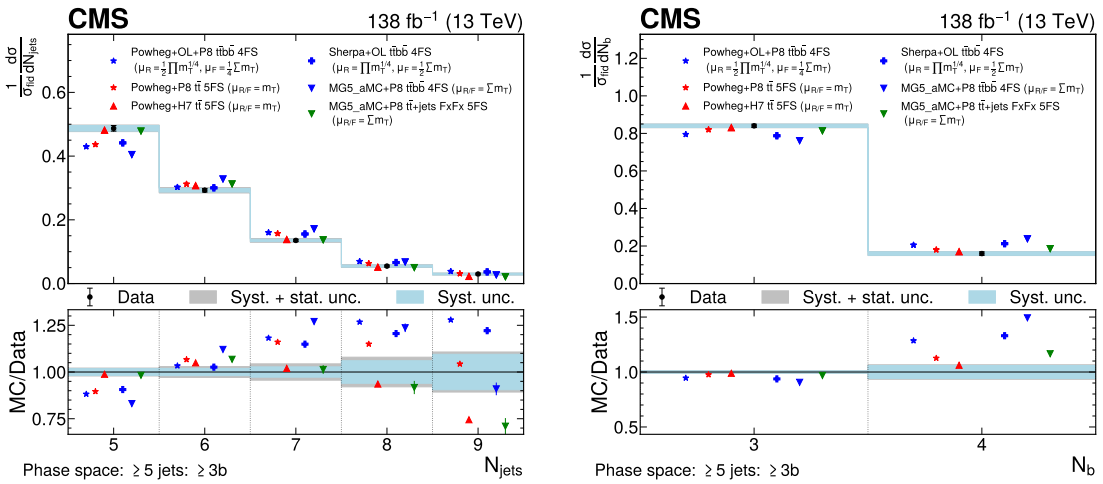


Figure 7.5: Predicted and observed normalized differential cross sections in the ≥ 5 jets, ≥ 3b jets fiducial phase space for the inclusive jet multiplicity (left) and the b jet multiplicity (right). The cross section predictions are obtained at the particle level from various simulation approaches. The "Powheg+OL+P8 $\mathrm{t\bar{t}b\bar{b}}$ 4FS" simulation (blue star) corresponds to the $\mathrm{t\bar{t}b\bar{b}}$ model employed in this thesis. Taken from Ref. [96].

this mismodeling into the analysis is detailed. The "Robust Independent Validation of Experiment and Theory" (Rivet) framework allows modeling of physics processes in Monte Carlo event simulation (MC) to be tested independently of the experiment. A Rivet routine is available in Ref. [133] for the $\mathrm{t\bar{t}b\bar{b}}$ measurement by the CMS Collaboration Ref. [96] and allows future modeling approaches to be compared with the measured data along all analyzed models.

7.2.2 $\mathrm{t\bar{t}H(b\bar{b})}$ measurements

The ATLAS and CMS Collaborations have each published several analyses for $\mathrm{t\bar{t}H}$ production in the $\mathrm{H} \to \mathrm{b\bar{b}}$ decay channel in proton-proton collisions at $\sqrt{s} = 13\,\mathrm{TeV}$. The CMS experiment published a signal strength (cf. Section 6.1) of 0.72 ± 0.45 in an analysis with at least one lepton and a signal strength of 0.9 ± 1.5 in a measurement in the fully hadronic channel on data recorded in 2016 with an integrated luminosity of $35.9\,\mathrm{fb}^{-1}$ [134, 135]. The former measurement has an observed (expected) significance of 1.6 (2.2) standard deviations. Due to the limited statistics compared to an analysis on full Run-II data, the relative uncertainty is dominated by the statistical uncertainty. Beyond these measurements, both experiments have published analyses of $\mathrm{t\bar{t}H(b\bar{b})}$ on the full Run-II data set and a combination of all channels. The ATLAS Collaboration has reported a search for $\mathrm{t\bar{t}H(b\bar{b})}$ with data corresponding to an integrated luminosity of $139\,\mathrm{fb}^{-1}$ and published a measured signal strength of $0.35^{+0.36}_{-0.34}$. This measurement corresponds to an observed (expected) significance of 1.0 (2.7) standard deviations [136]; however, the analysis is

superseded by Ref. [137] as discussed below. The corresponding measurement of the CMS experiment, performed on Run-II data with an integrated luminosity of $138\,\mathrm{fb}^{-1}$ and a combination of all channels, revealed a signal strength of 0.33 ± 0.26 with an observed (expected) significance of 1.3 (4.1) standard deviations [126]. With these results, the two independent measurements match well in the ratio of the measured cross section to the SM prediction and feature a compatible level of sensitivity. Although the two analyses differ in details such as the parameter choice in event simulation for $\mathrm{t\bar{t}b\bar{b}}$, analysis methods like artificial neural networks, and the characterization of theoretical uncertainties, their findings remain consistent and compatible. This also applies to other findings in the analyses. The nuisance parameter pulls of both analyses suggest that the cross section of $\mathrm{t\bar{t}b\bar{b}}$ is higher than assumed in the modeling. Likewise, in both cases it is found that the description of $\mathrm{t\bar{t}b\bar{b}}$ is a decisive factor in the sensitivity of a $\mathrm{t\bar{t}H(b\bar{b})}$ measurement. For this reason, a joint effort is conducted in the LHC Higgs Working Group to synthesize findings and derive a refined strategy for describing the theoretical uncertainties in $\mathrm{t\bar{t}H(b\bar{b})}$ measurements across the two experiments [124]. A Rivet routine is also available for this study [138].

However, in a re-analysis of the same data set published two years later, the ATLAS Collaboration reported a signal strength of $0.81^{+0.22}_{-0.19}$ corresponding to an observed (expected) significance of 4.6 (5.4) standard deviations [137]. The measured $\mathrm{t\bar{t}H}$ cross section is considerably closer to the expected cross section in the SM and larger compared to the previous full Run-II data analyses. Also, this measurement demonstrates an increased observed significance of the $\mathrm{t\bar{t}H}$ signal hypothesis compared to a SM background-only scenario. All measured signal strengths, including the signal strengths in the dilepton channels for a better comparison with the results from this thesis, are summarized in Table 7.1. The result can be attributed to a number of factors, albeit the individual impact of each change compared to the initial analysis on the overall result is not evident and the main driver cannot be explicitly discerned. The changes include improved b jet identification and looser selection criteria, which increase the signal acceptance by a factor of three. Furthermore, the classification in control regions is improved with an advanced neural network using transformer architectures. A new feature of the analysis are data-driven modeling corrections for $\mathrm{t\bar{t}}{+}{\geq}1\mathrm{c}$ and $\mathrm{t\bar{t}LF}$. Additionally, there is a new dedicated event simulation for $\mathrm{t\bar{t}}{+}{\geq}1\mathrm{b}$ processes using updated scale settings and an extended uncertainty scheme. As a result, the modeling of $\mathrm{t\bar{t}}$ in association with additional b jets is no longer the leading driver of the total systematic uncertainty, which is a change compared to analyses before. Instead, the modeling of the signal now dominates the uncertainty of the measurement. The estimated probability that the result of the re-analysis is compatible with the previous result is 21% [137]. This is mainly based on a correlation study of both analyses, which determined a statistical correlation of 19% using a bootstrapping method [137].

Table 7.1: Results of $t\bar{t}H(b\bar{b})$ measurements by ATLAS and CMS in various publications. The signal strength of all channels is reported as $\mu_{t\bar{t}H}$, as well as the signal strength for the dilepton channel only $\mu_{t\bar{t}H,\,DL}$, if applicable. The last two columns quote the expected and observed significance in standard deviations, if reported in the paper. The abbreviations SL, DL, and FH refer to the semileptonic, dileptonic, and fully hadronic channel, respectively.

Analysis	$\mu_{t\bar{t}H}$	$\mu_{t\bar{t}H,\,DL}$	$\sigma_{exp.}$	$\sigma_{obs.}$
CMS $36\,fb^{-1}$, SL/DL [134]	0.72 ± 0.45	$-0.24^{+1.21}_{-1.12}$	2.2	1.6
CMS $36\,fb^{-1}$, FH [135]	0.9 ± 1.5	N/A	N/A	N/A
ATLAS $139\,fb^{-1}$ [136]	$0.35^{+0.36}_{-0.34}$	$0.60^{+0.69}_{-0.65}$	2.7	1.0
ATLAS $139\,fb^{-1}$, re-analysis [137]	$0.81^{+0.22}_{-0.19}$	$1.03^{+0.38}_{-0.34}$	5.4	4.6
CMS $138\,fb^{-1}$ [126]	0.33 ± 0.26	$-0.23^{+0.42}_{-0.41}$	4.1	1.3

The three full Run-II data measurements of $t\bar{t}H(b\bar{b})$ by ATLAS and CMS apply different pre-factors $\kappa_{R,F}$ for the renormalization and factorization scales of their $t\bar{t}b\bar{b}$ model, which is simulated with POWHEG BOX and PYTHIA with $t\bar{t}b\bar{b}$ at matrix element (ME) level in both cases. While the first ATLAS measurement applied $\kappa_R = 1$ and $\kappa_F = 0.5$ [136], in the re-analysis the $t\bar{t}b\bar{b}$ process was modeled with a reduced renormalization scale pre-factor of $\kappa_R = 0.5$ while fixing κ_F to $\kappa_F = 0.5$ [137]. The corresponding measurement by the CMS Collaboration used $\kappa_R = 0.5$ and $\kappa_F = 0.25$ [126]. In this analysis, the observed $t\bar{t}B$ normalization is scaled by a factor of 1.19 and thus favors higher cross sections for these processes. Yet, the $t\bar{t}b\bar{b}$ cross section is scaled to the cross section prediction of the $t\bar{t}b\bar{b}$ process described with $t\bar{t}$ at ME level, which is shown to predict a cross section of 10% to 24% lower than the dedicated measurements in Refs. [96, 132]. Consequently, although the same model is used for $t\bar{t}b\bar{b}$ in the $t\bar{t}H(b\bar{b})$ analysis by the CMS Collaboration in [126] as in this thesis, it is expected that the $t\bar{t}B$ normalization differs from the measured signal strength parameter for these processes in this analysis.

An examination of the scale choice variation of μ_R and μ_F in a dedicated $t\bar{t}b\bar{b}$ analysis demonstrates the behavior of $t\bar{t}b\bar{b}$ models in the $t\bar{t}H$ signal phase space and provides additional insights on the impact of $t\bar{t}b\bar{b}$ in $t\bar{t}H(b\bar{b})$ measurements [132]. Given the behavior of varied $t\bar{t}b\bar{b}$ models based on distinct scale combinations in Ref. [132], it cannot be concluded that the increased $t\bar{t}H$ signal strength in the re-analysis by ATLAS (Ref. [137]) is solely based on the improved model and therewith the updated choice of the two scales. However, the choice of the revised scale of $\kappa_R = 0.5$ and $\kappa_F = 0.5$ in Ref. [137], which is suggested in Section 7.2.1 based on the findings in Ref. [132], is consistent with the determined normalization factor for $t\bar{t}b\bar{b}$. In contrast to the previous studies, the $t\bar{t}b\bar{b}$ normalization factor with this scale choice is now closer to one and thus the predicted cross section closer to the measured cross section.

The sensitivity in a $t\bar{t}H(b\bar{b})$ analysis is usually driven by the semileptonic channel, where the systematic uncertainties dominate the measurement. In contrast, the sensitivity in the dilepton channel is limited due to the reduced event yield and therefore the size of the data set. The measurements of the $t\bar{t}H$ signal strength parameters exclusively in the dilepton channel show larger variance. While the ATLAS experiment determines a signal strength of $0.60^{+0.69}_{-0.65}$ in the initial analysis, the re-analysis revealed a measured signal strength of $1.03^{+0.38}_{-0.34}$ agreeing well with the SM expectation [136, 137]. Given the uncertainties, both signal strength measurements are mutually compatible. At the CMS experiment, a significantly lower value was found in the dilepton channel. A signal strength estimate of $-0.23^{+0.42}_{-0.41}$ is technically possible since the fit is intentionally designed to allow negative values based on the structure of the model's likelihood function. However, statistical methods can be applied to transform these estimates into physically meaningful ranges.

7.2.3 $t\bar{t}Z(b\bar{b})$ measurements

A measurement of the associated production of a Z boson with a top quark-antiquark pair where the Z boson subsequently decays to a b quark-antiquark pair ($t\bar{t}Z(b\bar{b})$) has not been performed in the dilepton channel yet. In a thesis analyzing data recorded in 2018, a signal strength of $-0.77^{+1.19}_{-1.26}$ was measured in the semileptonic $t\bar{t}$ channel for $t\bar{t}Z(b\bar{b})$ [139]. Therefore, this is the first cross section measurement of $t\bar{t}Z$ in the dilepton phase space with resolved hadronic Z boson decays. Analyses with the closest thematic proximity usually deal with leptonic decays of the Z boson, which allow for very clean reconstruction of the Z boson [140]. However, these processes are outside the targeted phase space of this thesis and are explicitly removed, as discussed in Chapter 8. In addition, some studies are carried out in the boosted regime of the Z boson [141]. In this case, the event signature is different, as the $Z \to q\bar{q}$ decay products are merged together forming a single but larger jet. The resulting jet in this case is clustered with a distance parameter of 0.8 (cf. Section 4.3). This applies to cases where the Z boson has transverse momenta of more than 200 GeV, which is explicitly required in these analyses [141]. The choice of this phase space is beneficial for the examination of effective field theories because deviations from the SM are expected at large p_T values of the Z boson.

7.2.4 $t\bar{t}c\bar{c}$ measurements

Compared to the $t\bar{t}b\bar{b}$ process, the production of a top quark-antiquark pair in association with additional charm jets ($t\bar{t}c\bar{c}$) has been explored to a lesser extent both experimentally and theoretically. The first measurement of the inclusive $t\bar{t}c\bar{c}$ cross section was performed by the CMS Collaboration on recorded data in 2017 in proton-proton collisions at a center-of-mass energy at $\sqrt{s} = 13\,\text{TeV}$, corresponding to an integrated luminosity of $41.5\,\text{fb}^{-1}$ [142]. The analysis focuses exclusively on the dilepton $t\bar{t}$ channel and the $t\bar{t}$+jets phase space is modeled with an inclusive description of $t\bar{t}$ with up to two additional jets at NLO in

QCD, merged using the FxFx scheme [143]. Thus, no dedicated simulation is employed for $t\bar{t}b\bar{b}$ and $t\bar{t}c\bar{c}$. One event simulation of $t\bar{t}$+jets is conducted with POWHEG BOX and PYTHIA, the other one with MADGRAPH5_aMC@NLO. The heavy flavor jet tagging algorithm employed is DEEPCSV, which is the predecessor of the DEEPJET tagger used in this thesis and features a lower b jet identification efficiency [91]. Both of the tested simulation approaches in Ref. [142] are consistent with observed data at the level of one to two standard deviations, but both under-predict the measured cross section for $t\bar{t}c\bar{c}$.

The ATLAS Collaboration has performed a measurement of the $t\bar{t}c\bar{c}$ process on recorded data between 2015 and 2018 at $\sqrt{s} = 13\,\text{TeV}$, corresponding to an integrated luminosity of $140\,\text{fb}^{-1}$ [144]. The measurement considers events in the semileptonic and dilepton channel and applies a dedicated flavor tagging algorithm to distinguish c and b jets. As in the previously described analyses by ATLAS, a $t\bar{t}$+jets simulation with POWHEG BOX and PYTHIA as well as a dedicated $t\bar{t}b\bar{b}$ simulation with POWHEG BOX and PYTHIA are also used in this $t\bar{t}c\bar{c}$ analysis. The scales employed for the renormalization and factorization scale are $\kappa_R = 0.5$ and $\kappa_F = 0.5$ and therefore identical to the scales used in the $t\bar{t}H(b\bar{b})$ re-analysis by the ATLAS Collaboration discussed before [137]. Further alternative models are tested additionally. In this measurement, all $t\bar{t}c\bar{c}$ predictions at NLO accuracy in QCD are consistent with the observed values and the POWHEG BOX and PYTHIA models agree with the measurement within 0.5 to 1.1 standard deviations. However, all models under-predict the measured cross sections. The alternative models tested under-predict the measured values by up to 40%. Limiting factors driving the uncertainties are the modeling of all $t\bar{t}$ processes as well as the heavy flavor jet tagging algorithm calibration and data statistics [144].

7.3 Analysis strategy and goal

The objective of this analysis is the simultaneous measurement of processes in the $t\bar{t}X$ phase space in the dilepton channel where X stands for $b\bar{b}$, $c\bar{c}$, $H \to b\bar{b}$, and $Z \to b\bar{b}$. Due to the equivalent final state of these processes a separation based on cuts is not feasible. Even though measurements focusing on single processes, as described above, already exist for some cases, this is the first time that a multi-dimensional measurement of all processes is being attempted simultaneously. Also, previous analyses which target the measurement of a single process use refined machine learning methods in order to cope efficiently with the event classification. In this thesis, state-of-the-art Artificial Intelligence (AI) methods are used to achieve powerful jet and event classifications. For this purpose, graph transformer neural networks are employed. This allows the conversion of simulated events and data with an intrinsic graph structure into the equivalent of a mathematical graph, but at the same time the use of attention mechanisms that are successfully applied in natural language processing. In contrast to other analyses such as [126], in which the phase space was subdivided and multiple neural networks were trained, in this analysis strategy the graph

transformer models cover the entire phase space. This means no cuts are applied, e.g. on the number of b-tagged jets to create signal and background enriched phase space regions in which different graph transformer models are trained. Moreover, compared to using deep neural networks, the implementation of graph neural network architectures provides advantages in performance, convergence speed, and generalizability, while needing fewer trainable parameters [111]. Beyond this, properties such as permutation invariance and the independence of the model architecture in terms of the number of jets in an event constitute an ideal setting. Details of graph neural network (GNN) architectures are discussed with a theoretical focus in Chapter 5, while the technical implementation, the network modeling, as well as the resulting performance are described in Chapter 9. Using maximum likelihood fits, the signal strength parameters, their uncertainties, correlations, etc. are determined relying on the preceding GNN-based event classification of the processes. This allows conclusions to be drawn about the agreement between expectations from SM theory given the data from proton-proton collisions. For the large statistics $t\bar{t}b\bar{b}$ and $t\bar{t}c\bar{c}$ production channels, this is an important test of the current predictions and the SM. In the case of $t\bar{t}H(b\bar{b})$ production, there may be indications as to why signal strengths deviating from one were found in the past. For $t\bar{t}Z(b\bar{b})$, new insights are gained for the first time.

8 Object and event definition

In this chapter, the object and event selection is described that forms the basis of the analysis. The simulated events, which are considered in the phase space of the analysis, are discussed in Section 8.1. Section 8.2 covers the definition of all analysis objects in an event. Based on this, the event selection is defined in Section 8.3. In Section 8.4, corrections are discussed which adjust emerging discrepancies between data and simulation. Finally, Section 8.5 presents the resulting distributions, validating the agreement between data and simulation.

8.1 Event simulation

The event simulation introduced in Section 4.6 focusing on technical aspects, serves three central purposes in this thesis. The first purpose is to estimate signal and background contributions across all measurement regions using simulated events for each process. Second, it ensures that a solid physics model is built before a measurement is performed. This allows for a neutral strategy and unbiased measurements after successfully completing a series of statistical tests. As a result, this procedure prevents wrong hypotheses from being favored due to mismodeling. The third purpose is to provide labeled inputs for training classifiers, which would not be feasible without simulated events in this thesis. All samples of simulated events considered in the analysis phase space are summarized in the following. A breakdown is given in Table 8.1 along with cross sections σ for the normalizations of the samples. In this thesis, the term "sample" always refers to a simulated data set of a specific process.

$t\bar{t}$+jets: The production of a top quark-antiquark pair in proton-proton collisions is simulated with the POWHEG BOX version 2 matrix element (ME) generator at Next-to-leading order (NLO) precision in quantum chromodynamics (QCD) [98, 99]. In contrast to a simulation at leading order (LO), this implies up to one real emission of a gluon or

quark that can occur in addition to the $t\bar{t}$ pair. For the simulation of the parton shower (PS) as well as the hadronization, the ME generator is interfaced with PYTHIA version 8.240 [100]. An additional b quark-antiquark pair in this simulation therefore always arises from the splitting of a gluon ($g \to b\bar{b}$) simulated in the PS. The b quarks are assumed to be massless in this event simulation approach. Hence, for the measurement of the $t\bar{t}b\bar{b}$ process as shown in Figure 7.1, a dedicated simulation is used which already defines $t\bar{t}b\bar{b}$ at ME level as explained below. The protons' inner structure is described with five active flavors, which is referred to as five flavor scheme (5FS). The parton distribution function (PDF) set chosen is 5FS NNPDF3.1 at Next-to-next-to-leading order (NNLO) in QCD [60]. A dynamic renormalization and factorization scale is applied similar to Equations 2.22 and 2.23. However, since only the top quark-antiquark scale is relevant in this simulation the scales are reduced to

$$\mu_{R,F} = \sqrt{m_{\mathrm{T,t}} + m_{\mathrm{T},\bar{\mathrm{t}}}} \quad . \tag{8.1}$$

In POWHEG BOX, a damping function is used to regulate the contribution of high-p_{T} emissions in the ME calculation, preventing excessive weights and ensuring a consistent description with the PS across energy scales. The corresponding damping parameter h_{damp} is set to 1.379 times the top quark mass. The damping parameter value has been determined in a tuning procedure by the Compact Muon Solenoid (CMS) Collaboration [145]. In comparison, the ATLAS Collaboration chooses an h_{damp} parameter value of 1.5 times the top quark mass for their $t\bar{t}$ simulations [124]. The top quark mass is set to $m_{\mathrm{t}} = 172.5\,\mathrm{GeV}$ for all simulated samples. In PYTHIA, a series of parameters must be specified to describe the underlying event structure. The chosen set of parameters is known as CP5 tune [145]. Each $t\bar{t}$ decay channel is simulated separately, resulting in three different samples for the dilepton, semileptonic, and fully hadronic $t\bar{t}$+jets simulation.

$t\bar{t}b\bar{b}$: The $t\bar{t}b\bar{b}$ signal process, which is fundamental in this thesis, is not sufficiently described in the previous inclusive $t\bar{t}$+jets simulation. Therefore, a dedicated $t\bar{t}b\bar{b}$ calculation is performed at ME level. This ME calculation is implemented in POWHEG BOX RES and OPENLOOPS 2 [146, 147]. As in the $t\bar{t}$+jets simulation described above, the PS simulation is accomplished with PYTHIA8. However, in contrast to the $t\bar{t}$+jets simulation, the b quarks are no longer massless. Instead, the b quark mass is set to $m_{\mathrm{b}} = 4.75\,\mathrm{GeV}$ and a different flavor scheme is used. Since the initial state of the b quark is now no longer obtained from the PDF but the calculation is performed at ME level, only four active flavors are required in the proton PDF. The flavor scheme is referred to as four flavor scheme (4FS) and the PDF set chosen is 4FS NNPDF3.1 NNLO [60]. The h_{damp} parameter and the tune are fixed to the same values as in the $t\bar{t}$+jets simulation. The renormalization and factorization scales are chosen as dynamical scales as defined in Equations 2.22 and 2.23. Hence, the renormalization and factorization scales used reflect the relevant mass and p_{T}

scales of all expected final state quarks calculated at ME. In comparison to the previous $t\bar{t}$+jets simulation, the relevant scales of the b quarks also appear in the $t\bar{t}b\bar{b}$ simulation for $\mu_{R,F}$. As the b quarks are not taken into account in the ME calculation for the $t\bar{t}$+jets simulation, their scale was not considered in the renormalization and factorization scales. A consequence of adding the b quark scale is that the associated scale uncertainty is expected to increase [125]. This is also shown in a modeling study of $t\bar{t}b\bar{b}$ processes at ATLAS and CMS in Ref. [124]. The resulting uncertainties are dominated by the μ_R scale compared to the μ_F scale, as the former has a strong $\alpha_S^4(\mu_R)$ dependency in the ME calculations. The pre-factors of the renormalization and factorization scales in Equations 2.22 and 2.23 are set to $\kappa_R = 0.5$ and $\kappa_F = 0.25$, respectively.

Combination of **$t\bar{t}$+jets** and **$t\bar{t}b\bar{b}$**: Since the $t\bar{t}$+jets simulation does not adequately describe the $t\bar{t}b\bar{b}$ sub-processes, whereas the dedicated $t\bar{t}b\bar{b}$ simulation does not cover the entire phase space of $t\bar{t}$+jets events, a combination is employed to unify the strengths of the two simulation approaches while preserving a complete description of the entire phase space. A ghost-hadron matching procedure is used to determine the relevant simulated events that need to be extracted from the samples, which is explained in Section 4.6.1. Here, jets play a central role in the selection process which are present in addition to the $t\bar{t}$ system and referred to as *additional jets* in this thesis. If an event contains at least one additional particle-level jet with a ghost-matched B hadron, this event is taken from the dedicated $t\bar{t}b\bar{b}$ simulation in the 4FS scheme. Analogously, events with this properties are cut out from the $t\bar{t}$+jets 5FS event simulation. Events bearing this signature are referred to as $t\bar{t}$B. Thus, the majority of the inclusive analysis phase space defined in Section 8.3 is described by the $t\bar{t}$+jets simulation. Additionally, two further processes are defined with a similar logic. If there is at least one additional particle-level jet with a ghost-matched C hadron but no B hadron, it is referred to as $t\bar{t}$C. All remaining events that are neither $t\bar{t}$B or $t\bar{t}$C are referred to as $t\bar{t}$LF. An exemplary LO Feynman diagram of the $t\bar{t}$B process is shown in Figure 7.1.

$t\bar{t}$H, $t\bar{t}$Z, and $t\bar{t}$W: The associated production of a top quark-antiquark pair with a Higgs boson is simulated with the POWHEG BOX version 2 ME generator at NLO precision in QCD. The Higgs boson mass is set to $m_{\mathrm{H}} = 125\,\mathrm{GeV}$. One dedicated sample considers only the $\mathrm{H} \to \mathrm{b}\bar{\mathrm{b}}$ decays, while the other event simulation targets the remaining decay channels ($\mathrm{H} \to \mathrm{nonb}\bar{\mathrm{b}}$). Differently from the $t\bar{t}$H and $t\bar{t}$+jets simulations, the $t\bar{t}$Z process is simulated with MADGRAPH5_AMC@NLO version 2.6.5 at NLO in QCD [148]. The decay of the Z boson is modeled using the MADSPIN package to preserve the spin correlation of the decay products [149]. In the $t\bar{t}$Z event simulation, a distinction is made between Z boson decay channels. One sample solely addresses the leptonic Z boson decays ($Z \to \ell\ell, \nu\nu$), while the other covers the hadronic part of the decay ($\mathrm{Z} \to \mathrm{q}\bar{\mathrm{q}}$). Two exemplary LO Feynman diagrams of the $t\bar{t}$H and $t\bar{t}$Z processes are shown in Figure 7.1. Similar to the signal process simulations of $t\bar{t}$H and $t\bar{t}$Z, the background contributions of the $t\bar{t}$ production in

association with a W boson ($t\bar{t}W$) are simulated with MadGraph5_aMC@NLO version 2.6.5 at NLO in QCD and MadSpin. While one sample is dedicated to the $W \to \ell\nu$ decay, the other handles the hadronic W boson decay. In Figure 8.1, an exemplary Feynman diagram at LO accuracy in QCD is shown.

single top: The single top quark production forms a background in this analysis and is simulated in varying ways depending on the top quark production channel. In the s-channel it is simulated with the MadGraph5_aMC@NLO ME generator at NLO accuracy in QCD. In the t-channel as well as in the associated production with a W boson, it is simulated with the Powheg Box toolbox at NLO accuracy in QCD. An exemplary Feynman diagram in the associated production with a W boson is shown in Figure 8.1.

VV and V+jets (Vx): Additional backgrounds stem from the product of one or two vector bosons (V = W,Z) with leptonic decays in association with jets. Diboson production includes WW, WZ, and ZZ production, abbreviated as VV, and is simulated with Pythia at LO accuracy. A separate simulation is performed for each of the three boson combinations. The WW sample is normalized to the NNLO QCD prediction, while WZ and ZZ are normalized to the NLO QCD prediction. Processes of vector boson production in association with jets, i.e. W+jets and Z+jets, are simulated with MadGraph5_aMC@NLO at LO accuracy in QCD and are abbreviated as V+jets. The simulation of V+jets production includes processes with up to four jets at ME level and the samples are binned in H_T and $m_{\ell\ell}$ where applicable. The inclusive cross sections are normalized to NNLO. As long as no distinction is needed, VV and V+jets are summarized as Vx in this thesis. Exemplary LO Feynman diagrams are shown together with the other backgrounds in Figure 8.1.

In addition to the samples discussed, the simulation of multi-jet events from QCD processes is examined. However, tests on the contribution of these QCD events in the considered phase space revealed a fraction of QCD events compared to all other simulated events of less than 0.5% . Therefore, QCD multi-jet production is negligible in this analysis with respect to all other processes and their systematic uncertainties discussed in Section 10.1. The vanishing contribution is particularly driven by the selection of exactly two leptons discussed in Section 8.3. Due to the negligible contribution, QCD events are not considered further in this analysis and the event simulation is not discussed in similar detail compared to simulations actually used in this analysis.

All simulated processes are summarized in Table 8.1 associated with the cross sections for normalization.

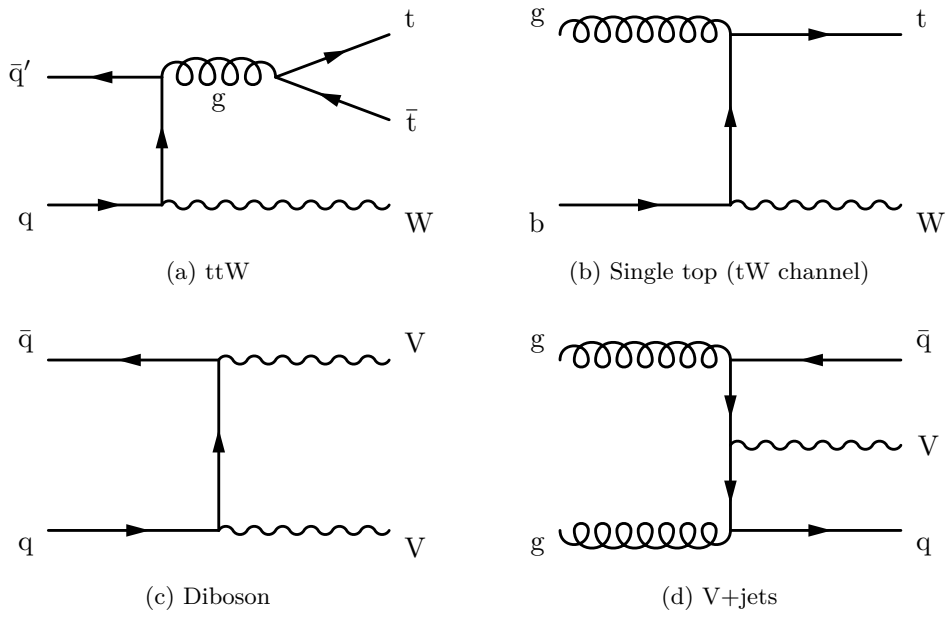
8.2 Object cleaning, definition and selection

8.2.1 Pre-selection: Trigger requirements

In order to narrow down the total amount of recorded data relevant for the analysis phase space, dedicated high-level trigger (HLT) paths are queried as described in Section 3.2.1.6.

Table 8.1: Simulated processes considered in the analysis phase space. Sub-categories are defined based on the simulation approach for each process. Cross sections are chosen based on theory calculations or directly from the simulation.

Process	Sub-category	σ [pb]	Source
$t\bar{t}$+jets	dilepton channel	88.5	[50, 150]
	semileptonic channel	366.3	[50, 150]
	fully hadronic channel	378.9	[50, 150]
$t\bar{t}b\bar{b}$	dilepton channel	4.6	simulation
	semileptonic channel	19.2	simulation
	fully hadronic channel	19.9	simulation
$t\bar{t}H$	$H \to b\bar{b}$	0.2953	[122]
	$H \to nonb\bar{b}$	0.2118	[122]
$t\bar{t}Z$	$Z \to \ell\ell$ and $Z \to \nu\nu$	0.2529	[123]
	$Z \to q\bar{q}$	0.5297	[123]
$t\bar{t}W$	$W \to \ell\nu$	0.2043	[123]
	$W \to q\overline{q}'$	0.4062	[123]
single top	s-channel, $W \to \ell\nu$	3.30	[151]
	t-channel, $\bar{t}$	80.95	[152]
	t-channel, t	136.02	[152]
	tW-channel, $\bar{t}$	19.56	[151]
	tW-channel, t	19.56	[151]
VV	WW	118.7	[153]
	WZ	65.5443	[154]
	ZZ	15.8274	[155]
W+jets with $W \to \ell\nu$	$70\,\text{GeV} \leq H_T < 100\,\text{GeV}$	1443.7	[156]
	$100\,\text{GeV} \leq H_T < 200\,\text{GeV}$	1434.5	[156]
	$200\,\text{GeV} \leq H_T < 400\,\text{GeV}$	383.19	[156]
	$400\,\text{GeV} \leq H_T < 600\,\text{GeV}$	51.68	[156]
	$600\,\text{GeV} \leq H_T < 800\,\text{GeV}$	12.53	[156]
	$800\,\text{GeV} \leq H_T < 1200\,\text{GeV}$	5.62	[156]
	$1200\,\text{GeV} \leq H_T < 2500\,\text{GeV}$	1.32	[156]
	$2500\,\text{GeV} \leq H_T$	0.009	[156]
Z+jets with $Z \to \ell\ell$	$10\,\text{GeV} \leq m_{\ell\ell} < 50\,\text{GeV}$	18 610.0	[156]
	$50\,\text{GeV} \leq m_{\ell\ell}$ and $70\,\text{GeV} \leq H_T < 100\,\text{GeV}$	211.21	[156]
	$50\,\text{GeV} \leq m_{\ell\ell}$ and $100\,\text{GeV} \leq H_T < 200\,\text{GeV}$	183.2	[156]
	$50\,\text{GeV} \leq m_{\ell\ell}$ and $200\,\text{GeV} \leq H_T < 400\,\text{GeV}$	55.29	[156]
	$50\,\text{GeV} \leq m_{\ell\ell}$ and $400\,\text{GeV} \leq H_T < 600\,\text{GeV}$	7.846	[156]
	$50\,\text{GeV} \leq m_{\ell\ell}$ and $600\,\text{GeV} \leq H_T < 800\,\text{GeV}$	1.933	[156]
	$50\,\text{GeV} \leq m_{\ell\ell}$ and $800\,\text{GeV} \leq H_T < 1200\,\text{GeV}$	0.831	[156]
	$50\,\text{GeV} \leq m_{\ell\ell}$ and $1200\,\text{GeV} \leq H_T < 2500\,\text{GeV}$	0.183	[156]
	$50\,\text{GeV} \leq m_{\ell\ell}$ and $2500\,\text{GeV} \leq H_T$	0.004	[156]



(a) ttW (b) Single top (tW channel)

(c) Diboson (d) V+jets

Figure 8.1: Examples of LO Feynman diagrams for the background processes in the dilepton channel.

All HLT paths chosen in this thesis are not pre-scaled, which means that all events passing the trigger requirement are recorded, in contrast to only selecting a subset of events. In addition, HLT paths are chosen to have the loosest object selection requirements. This approach allows to achieve a maximum data utilization. Additionally, not only dedicated dilepton HLT paths are queried, but single lepton HLT paths are considered as well. With this procedure a fraction of events can additionally be considered in this analysis even though the dilepton HLT paths have not accepted those events. This can be attributed to events that meet the subsequent event selection, however one of the two leptons has not been captured by the dilepton HLT paths. The amount of events that can be recovered through this procedure and used additionally for the analysis is approximately 7.8%. The list of all triggers used in the analysis is summarized in Table 8.2. The trigger paths are linked with logical disjunctions (ORs). Hence, an event in a given lepton channel satisfies at least one of the triggers. The lepton channel refers to the possible combinations of electrons and muons as discussed in Section 2.3. The HLT path names indicate the conditions placed on the event. Among these requirements are the calorimeter, track, and isolation conditions of the lepton according to the Particle Flow algorithm, as discussed in Section 4.2. The number following the lepton name tag (`Ele` and `Mu` for electrons and muons respectively) indicates the p_{T} threshold in GeV. This is followed by further encoded conditions, which are explained in Table 8.2. Leptons in this analysis refer to electrons or muons, as defined in the following.

Table 8.2: HLT paths for the e^+e^-, $e^\pm\mu^\mp$, and $\mu^+\mu^-$ dilepton decay channels used in the analysis. The value following the lepton name tag (`Ele` and `Mu` for electrons and muons, respectively) indicates the p_T threshold in GeV. `CaloIdL` and `TrackIdL` refer to the loose working point (WP) of the lepton identification in the calorimeter and tracker, respectively. `IsoVL` denotes the very loose WP of the lepton isolation. The abbreviation GSF refers to the Gaussian sum filter, a dedicated track fitting algorithm for electrons [157]. `TrkIsoVVL` signifies a very-very loose isolation of the muon in the tracker. `IsoMu` implies an isolated muon. The condition `Mass3p8` pertains to a condition to reject muon pairs from J/Ψ decays. `DZ` indicates the application of a longitudinal impact parameter d_z filter to ensure that the leptons originate from the primary vertex. The trigger paths are linked with logical disjunctions (ORs).

Channel	HLT path
e^+e^-	`HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ` OR `HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL` OR `HLT_Ele32_WPTight_Gsf`
$e^\pm\mu^\mp$	`HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL` OR `HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ` OR `HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ` OR `HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ` OR `HLT_Ele32_WPTight_Gsf` OR `HLT_IsoMu24`
$\mu^+\mu^-$	`HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8` OR `HLT_IsoMu24`

8.2.2 Electron selection

Electrons are required to have transverse momenta of $\geq 25\,\text{GeV}$ ($\geq 20\,\text{GeV}$) for the leading (subleading) electron and a pseudorapidity range of $|\eta| < 2.4$. The pseudorapidity range is indicated in Figure 3.3. This decision is driven by the limited coverage of the silicon tracker. Although the coverage of the inner tracker and calorimeter reaches up to $|\eta| < 2.5$ with a suitable efficiency, it is set to a slightly lower threshold in order to keep the angular definition identical to muons that are limited to $|\eta| < 2.4$, as explained in the following section. The effects of bremsstrahlung are taken into account by clustering all photons within a cone of $\Delta R \leq 0.1$ to the electron. This method is also known as ''dressing''. The calorimeter supercluster (SC) region of $1.4442 < |\eta_{\text{SC}}| < 1.566$ is excluded since this region corresponds to the transition section between the barrel and endcap regions of the electromagnetic calorimeter (ECAL). Additional cuts are performed on the impact parameters (cf. Section 4.1), dependent on the calorimeter SC region

$$\begin{aligned} |d_{xy}| < 0.05 \text{ AND } |d_z| < 0.1 \text{ , if } |\eta_{\text{SC}}| \leq 1.479 \\ |d_{xy}| < 0.1 \text{ AND } |d_z| < 0.2 \text{ , if } |\eta_{\text{SC}}| > 1.479 \quad , \end{aligned} \tag{8.2}$$

where $|d_{xy}|$ is the impact parameter in the transverse direction and $|d_z|$ in the longitudinal direction. This ensures an upper distance boundary from the interaction vertex. Additionally, observables are defined that require particular conditions in the detector elements to ensure desired properties of the electrons, as described in the following. As recommended by the CMS Collaboration, the additional requirements are set for genuine electrons which correspond to the ''tight'' WP in Ref. [158]. This setting corresponds to an efficiency of approximately 70% for genuine electrons [158]. Isolation observables can be constructed by summing the transverse momenta of charged hadrons, photons, and neutral hadrons inside an isolation cone with a distance of $\Delta R < 0.3$ with respect to the electron direction. This is particularly helpful since electrons with higher momenta correspond to higher energy spreads along their direction in the sub-detectors. Therefore, thresholds for the required electron isolation are parameterized as a function of the electron p_{T}. To mitigate effects from pileup, the pileup contributions are subtracted from the isolation requirements through an effective term ρA_{eff}, where ρ is the median of the E_{T} density per unit area in the event and A_{eff} denotes the effective area of the electron's isolation region. The full relative isolation variable I_{e} reads as

$$I_{\text{e}} = p_{\text{T,e}}^{-1} \cdot \left(\sum_{\Delta R<0.3} p_{\text{T,ch}} + \max \left\{ 0, \sum_{\Delta R<0.3} p_{\text{T,n}} + \sum_{\Delta R<0.3} p_{\text{T},\gamma} - \rho A_{\text{eff}} \right\} \right) \quad , \tag{8.3}$$

Table 8.3: Tracker and SC related electron identification requirements used in this analysis. The observable $\sigma_{i\eta i\eta}$ refers to the log-weighted distribution of energies in the supercluster as defined in Ref. [158]. The variables η_{tr} and ϕ_{tr} refer to the extrapolated innermost track position and p_{tr} refers to the track momentum at the closest point to the vertex. Prompt electrons arising from the beamline are separated from secondary background electrons through the number of expected hits in the innermost pixel detector layers. A conversion veto ensures that no tracks originating from charged particles with a hit in the innermost layer are unmatched to a reconstructed conversion vertex [158].

Observable	Barrel ($\lvert\eta_{\mathrm{SC}}\rvert \leq 1.479$)	Endcaps ($\lvert\eta_{\mathrm{SC}}\rvert > 1.479$)
Relative isolation I_{e} (Eq. 8.3)	$< 0.0287 + 0.506 p_{\mathrm{T}}$	$< 0.0445 + 0.963/p_{\mathrm{T}}$
$\sigma_{i\eta i\eta}$	< 0.010	< 0.035
$\lvert\eta_{\mathrm{SC}} - \eta_{\mathrm{tr}}\rvert$	< 0.0026	< 0.0050
$\lvert\phi_{\mathrm{SC}} - \phi_{\mathrm{tr}}\rvert$	< 0.022	< 0.024
$E_{\mathrm{Hadr}}/E_{\mathrm{EM}}$	$< 0.026 + \frac{1.15}{E_{\mathrm{SC}}} + \frac{0.032\rho}{E_{\mathrm{SC}}}$	$< 0.019 + \frac{2.06}{E_{\mathrm{SC}}} + \frac{0.183\rho}{E_{\mathrm{SC}}}$
$\lvert E_{\mathrm{SC}}^{-1} - p_{\mathrm{tr}}^{-1}\rvert$	< 0.16	< 0.0197
No. of missing inner hits	≤ 1	≤ 2
Pass conversion-safe veto	True	True

where the indices of p_{T} denote the transverse momenta of charged hadrons (ch), neutral hadrons (n), and photons (γ) in a clustered cone around the electron with a distance parameter of $\Delta R = 0.3$. Additional requirements are summarized in Table 8.3.

8.2.3 Muon selection

Identical to electrons, muons are required to have transverse momenta of $\geq 25\,\mathrm{GeV}$ ($\geq 20\,\mathrm{GeV}$) for the leading (subleading) muon and a pseudorapidity range of $|\eta| < 2.4$. The dressing method is equivalently used for muons as for electrons, even though it plays a minor role for muons. Although the inner tracker is capable of reconstructing momenta of charged particles up to $|\eta| < 2.5$, the muon system outside the solenoid is limited to $|\eta| < 2.4$. Additional cuts are applied on the impact parameters with a requirement of $|d_{xy}| < 0.2$ and $|d_z| < 0.5$. The "tight" WP criteria are applied to the muon candidates in order to separate prompt muons from background muons as recommended by the CMS Collaboration [159]. Such background muons can arise from meson decays or even from cosmic rays produced outside the detector. With the help of additional criteria summarized in Table 8.4, these unwanted background muons are suppressed as they usually originate from decays in flight and from hadronic punch-through. Similar to the procedure with electrons, for muons these criteria are combinations of characteristics in the tracker as well as the dedicated muon chambers. The relative isolation variable for muons is defined as

Table 8.4: Tracker and muon chambers related muon identification requirements used in this analysis. The chosen cuts correspond to the "tight" WP of the CMS recommendation [159].

Requirement	Cut
Relative isolation I_μ (Eq. 8.4)	< 0.15
Tracker and global muon	True
No. of matching segments in muon stations	> 1
χ^2/ndf of global muon track fit	< 10
No. of muon chamber hits	≥ 1
No. of pixel hits	≥ 1
No. of tracker layer hits	≥ 6

$$I_\mu = p_{\mathrm{T},\mu}^{-1} \cdot \left(\sum_{\Delta R<0.4} p_{\mathrm{T,ch}} + \max \left\{ 0, \sum_{\Delta R<0.4} p_{\mathrm{T,n}} + \sum_{\Delta R<0.4} p_{\mathrm{T},\gamma} - \frac{1}{2} \sum_{\Delta R<0.4} p_{\mathrm{T,PU}} \right\} \right) \quad (8.4)$$

and structured in an analogous way to I_e for electrons (cf. Equation 8.3). In contrast to I_e, the distance parameter for the anti-k_T clustering is increased to $\Delta R = 0.4$ for I_μ. Pileup (PU) contributions are subtracted in a similar way by summing the transverse momenta $p_{\mathrm{T,PU}}$ of all pileup interactions. With the application of these methods for muons, three of the four muons in the exemplary $\mathrm{t\bar{t}H(b\bar{b})}$ event simulation in Figures 7.2 and 7.3 are rejected as these most likely do not stem from the beamline.

8.2.4 Jet selection

Jets, as introduced in Section 4.3, must pass a series of quality requirements in a similar way to the leptons. All jets are required to have transverse momenta of $p_\mathrm{T} > 30\,\mathrm{GeV}$ and a pseudorapidity of $|\eta| < 2.4$. Consequently, all considered objects are in the same pseudorapidity range. Jets are reconstructed from Particle Flow (PF) candidates using the anti-k_T clustering algorithm discussed in Section 4.2 with a distance parameter of $R = 0.4$. However, all jet candidates within a distance of $R \leq 0.4$ of any electron or muon object according to the previous definitions are excluded from the jet collection. Jets with $p_\mathrm{T} \leq 50\,\mathrm{GeV}$ need to fulfill an additional loose pileup rejection criterion constructed with a boosted decision tree to reject jets originating from pileup interactions. Additional criteria for jets are summarized in Table 8.5 which comply with the "tight" WP recommendations by the CMS Collaboration [160, 161].

In principle, a lower p_T threshold value would also be possible and thereby increase the data yield, however, this is waived in this thesis. As the relevant and interesting jets in this

Table 8.5: Jet identification requirements used in this analysis. The chosen cuts correspond to the "tight ID" for jets of the CMS recommendation [160, 161].

Requirement	Cut
Charged hadron energy fraction	> 0
No. of charged hadrons	≥ 1
Charged EM energy fraction	< 0.99
Neutral hadron energy fraction	< 0.9
Neutral EM energy fraction	< 0.9
No. of constituents	≥ 2
Muon energy fraction	< 0.8

analysis are primarily b jets, the required threshold is raised to 30 GeV since the efficiency of the jet flavor tagging algorithm used is significantly reduced for jets with p_T between 20 and 30 GeV. This behavior can be observed in Figure 4.3. In order to determine the jet flavor, the DeepJet heavy flavor jet tagging algorithm is chosen [97]. A jet needs to pass the "medium" WP of the DeepJet classifier to be considered a b jet in this analysis. Details on the heavy flavor tagging procedure, the algorithm, and the WP are described in Section 4.4. However, no selection is made on the basis of the number of b-tagged jets. A cut to a small number of b-tagged jets would indeed increase the signal to background ratio in the analysis phase space, but this separation is carried out by the event classifier later, which is given the DeepJet heavy flavor jet tagging information for each jet in an event. As a result, no assumption is made about the number of b-tagged jets, which depends on the b tagging modeling and may increase the model dependency. Moreover, a more inclusive phase space is enabled and analyzed.

8.2.5 MET selection

A cut on missing transverse energy (MET) of ≥ 40 GeV is applied in the e^+e^- and $\mu^+\mu^-$ channels to reduce contributions from Drell-Yan events. This is only necessary in two dilepton channels with the same lepton flavor, as the Drell-Yan contribution in the $e^\pm\mu^\mp$ channel is negligible.

8.3 Event selection

All objects in both simulated events and data must fulfill the requirements described above. Moreover, each event must fulfill a set of defined requirements. At least three jets must be present in an event in order to pass the selection, which is lower than the expected number of jets in LO calculations. This allows for some background enriched regions and limited reconstruction efficiencies to be taken into account. This also accounts for events in which, for example, a jet is out of acceptance. To select preferably dileptonic $t\bar{t}$ events,

Table 8.6: Data yields and expected event yields from simulation of the signal and background processes for a combination of all lepton channels as well as a breakdown to the e^+e^-, $e^\pm\mu^\mp$ and $\mu^+\mu^-$ channels. Due to rounding to integers, the sum of all individual channels for a given processes may not be equal to the corresponding value in the all channels column.

Process	all channels	e^+e^- channel	$e^\pm\mu^\mp$ channel	$\mu^+\mu^-$ channel
Data	266138	38289	142686	85163
$t\bar{t}B$	12638	1588	7475	3574
$t\bar{t}C$	26698	3392	15764	7542
$t\bar{t}H$	461	60	267	135
$t\bar{t}Z$	599	95	321	183
$t\bar{t}LF$	193178	24566	114939	53674
$t\bar{t}W$	594	83	342	169
single top	17316	2384	10091	4841
VV	1279	246	582	451
W+jets	485	81	299	104
Z+jets	17838	4903	1644	11290
Data/MC	0.982	1.024	0.940	1.039

dedicated selection criteria are constructed which require exactly two isolated leptons to prevent contributions from unwanted backgrounds. The two leptons are required to carry an opposite electric charge. In addition, the invariant mass of the dilepton system $m_{\ell\ell}$ needs to be outside a 15 GeV mass window from the Z boson mass of $m_Z = 91$ GeV. This requirement reduces contributions from Drell-Yan events as well as $t\bar{t}Z$ events with a leptonically decaying Z bosons ($Z \to \ell\ell$) in the fully hadronic $t\bar{t}$ channel. This rule only applies to events in which the two leptons have the same lepton flavor, i.e. the e^+e^- and $\mu^+\mu^-$ channels. Furthermore, the invariant mass of the dilepton system must exceed 20 GeV to suppress contributions from low mass resonances in the phase space. Additionally, events are vetoed if further leptons are present with $p_T > 15$ GeV. The event yields resulting from object and event definitions, broken down for the respective processes in Monte Carlo event simulation (MC) and for data recorded in 2018 at the CMS experiment are summarized in Table 8.6. The simulated events deviate by up to 6% relative to the recorded data, which is justified, as it is a pre-fit comparison and the likelihood function will adjust the yields further (cf. Chapter 10).

8.4 Event corrections

Although simulated events are based on deep theoretical understanding and sophisticated simulation techniques, differences remain in certain aspects when compared predictions

with data. Such effects can be corrected with the help of auxiliary measurements. The correction determined in such a procedure can then be applied to other phase spaces or measurements. For this recalibration, scale factor (SF) corrections are derived which adjust observed discrepancies in auxiliary measurements. An SF is the ratio of the efficiency ϵ in data to the efficiency in simulation and is calculated with

$$\text{SF} = \frac{\epsilon_{\text{data}}}{\epsilon_{\text{simulation}}} \quad . \tag{8.5}$$

The efficiencies for data and simulation are defined as the fraction of events that fulfill certain criteria, e.g. a dedicated selection, relative to the total number of events examined. Usually, the SFs are a function of certain parameters, typically a number of kinematic observables and properties, e.g. the p_{T}, η, and flavor of a jet. In this case, each jet in an event is evaluated and the whole event is given a weight. Through this method, each event receives a dedicated weight for each correction, resulting in a whole series of weights that take all corrections into account. This results in reweighting the kinematic distributions of simulated events to better match the distributions observed in data. This form of calibration is explained below for a series of corrections applied in this thesis. Each correction possesses its own uncertainty, which is explained in the context of systematic uncertainties in Section 10.1.

8.4.1 Pileup reweighting

Pileup, i.e. the occurrence of multiple interactions within the same proton bunch crossing, differs in data and simulated events. In simulation, the pileup in an event is described by a mean expected number of pileup interactions. The behavior is then characterized by means of a Poisson distribution. In addition, a collision before and after the bunch crossing under investigation can be mistakenly considered as part of it. A pileup reweighting is determined by a dedicated group within the CMS Collaboration. An inclusive data set is taken for this calculation, the details can be found in Ref. [162]. The corresponding observable is the number of primary vertices in an event, which is shown in Figure 8.2 and already includes the determined corrections. It can be seen that after the application of the reweighting a substantial mis-agreement between data and simulated events still exists. This is a well-known problem and also occurs in other analyses, e.g. in the semileptonic $\text{t}\bar{\text{t}}\text{b}\bar{\text{b}}$ phase space [132]. However, since the reweighting leaves all other relevant observables invariant and the number of primary vertices is irrelevant for this analysis, no further correction is performed.

8.4.2 L1 pre-fire reweighting

Muon candidates identified by the trigger system (cf. Section 3.2.1.6) can be associated to the wrong proton bunch due to a limited time resolution of the muon detector elements in

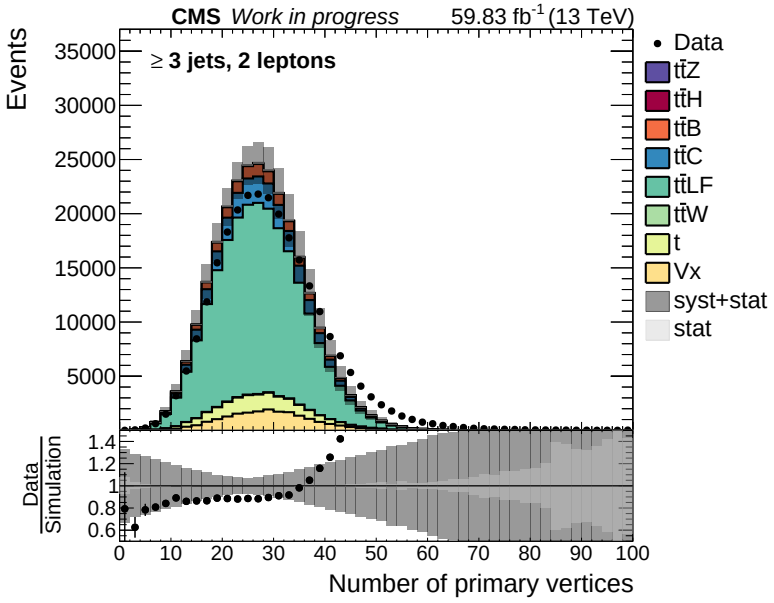


Figure 8.2: Number of primary vertices in an event after pileup reweighting. The systematic uncertainties include a 4.6% variation of the estimated inelastic proton-proton cross section.

the 2018 data-taking era. However, this effect is not incorporated in the event simulation and therefore leads to a discrepancy between data and simulated events. To accommodate the effect post simulation, the probability p for an event not to pre-fire is calculated. The calculation considers all offline photons, jets, and muons present in an event and takes individual efficiencies for each object into account. The Level-1 (L1) pre-fire event weight is obtained by

$$w_{\text{pre-fire}} = 1 - p\,(\text{pre-fire}) = \prod_{i=\gamma,\text{jets},\mu} \left(1 - \epsilon_i^{\text{pre-fire}}\,(p_\text{T}, \eta)\right) \quad . \tag{8.6}$$

While the effect is constant for objects with $p_\text{T} > 25\,\text{GeV}$ and thus stable for almost all objects in this thesis, the core dependency lies with η. The scale of the effect is between 0% and 3%.

8.4.3 Trigger efficiencies

In this analysis, numerous HLT trigger paths are queried as discussed in Section 8.2.1. The efficiencies of these trigger combinations differ only slightly between data and simulated events, nevertheless SFs are calculated resulting in values close to one. For this purpose, individual efficiencies must be calculated for each HLT path combination in Table 8.2, from which the SFs can then be determined according to Equation 8.5. For the calculation of the SFs, a set of p_T^miss baseline triggers (cf. Section 4.5) is used whose efficiency is

independent of the HLT paths used in this thesis. Still, a possible correlation is included as systematic uncertainty on the calculated SFs. Simulated events of $t\bar{t}$ processes are used for the efficiency calculations on simulation, while a data set primarily containing MET enriched events are used for the efficiency determination in data. For each channel, SFs are obtained that depend on the transverse momenta of the two leptons. Since the chosen trigger combinations were already deployed in analyses at the CMS experiment, the trigger SF are available. Details of the trigger SF calculations can be found in Refs. [163, 164]. The SFs used for all three channels are shown in Figures 8.3 and 8.4.

8.4.4 Electron reconstruction efficiencies

The reconstruction efficiency of electrons in data and simulated events are determined in auxiliary measurements in a phase space where $Z \to e^+e^-$ events are enriched using a tag-and-probe method. This process provides a clean, well-defined signal with high purity and therefore a high significance. The total efficiency for electrons is composed of

$$\epsilon_{e,\,\text{total}} = \epsilon_{e,\,\text{ID}} \cdot \epsilon_{e,\,\text{reco}} \quad , \tag{8.7}$$

with efficiencies for the identification (ID) and reconstruction (reco) as discussed in Section 8.2.2. The electron efficiencies are determined by specialized groups at the CMS experiment [158]. In the e^+e^- channel of this thesis, the efficiencies for both electrons are determined and multiplied.

8.4.5 Muon reconstruction efficiencies

The efficiency calculations for muons are similar to the procedure for electrons. The overall efficiency results from

$$\epsilon_{\mu,\,\text{total}} = \epsilon_{\mu,\,\text{ID}} \cdot \epsilon_{\mu,\,\text{reco}} \cdot \epsilon_{\mu,\,\text{iso}} \quad , \tag{8.8}$$

whereby an additional efficiency is added here for the muon isolation (iso), which does not exist for the electrons. Details are discussed in Section 8.2.3. The calculation of the efficiencies is performed in events with $Z \to \mu^+\mu^-$ processes and is determined by dedicated groups within the CMS Collaboration [159]. In the $\mu^+\mu^-$ channel of this thesis, the efficiencies for both muons are determined and multiplied.

8.4.6 Jet energy corrections

Similar to leptons, reconstructed jets need to be calibrated to match the true jet energy. Due to differences between reconstructed jets in data and simulated events, a jet energy correction (JEC) is performed [160]. In essence, the JEC consists of two components: the

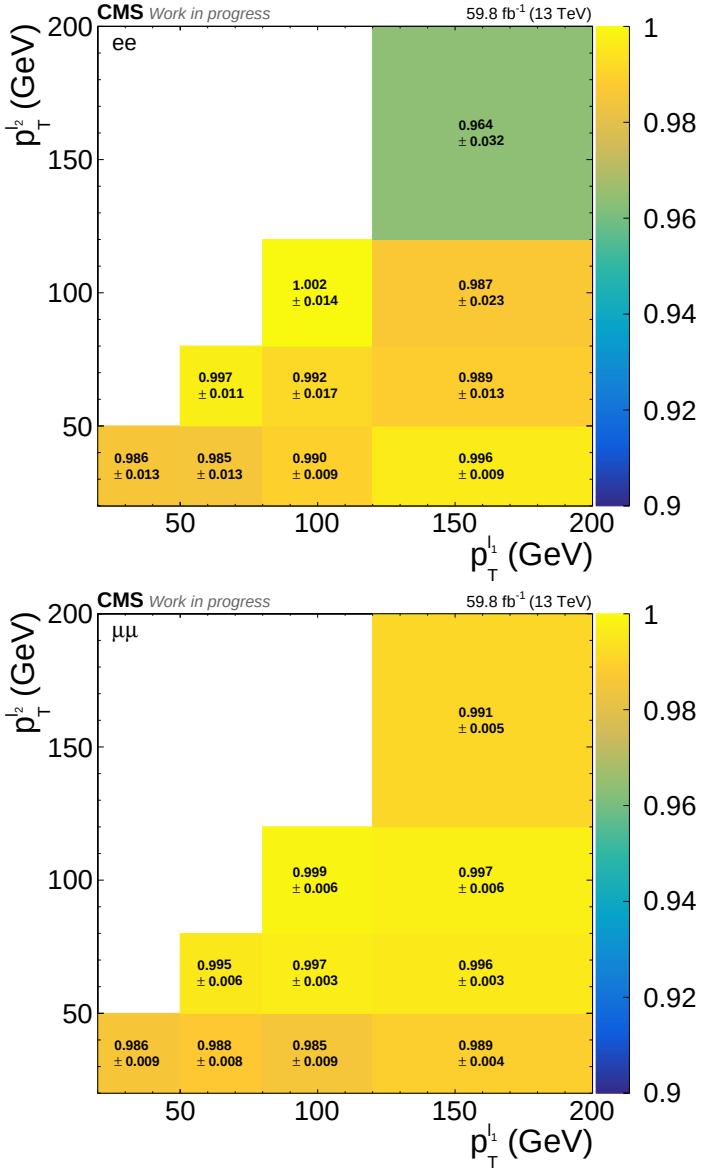


Figure 8.3: Trigger SFs for the e^+e^- channel (top) and the $\mu^+\mu^-$ channel (bottom). The SFs are functions of the leading and subleading lepton p_T. The SFs are shown for each bin including the total uncertainties. Taken from Refs. [163, 164].

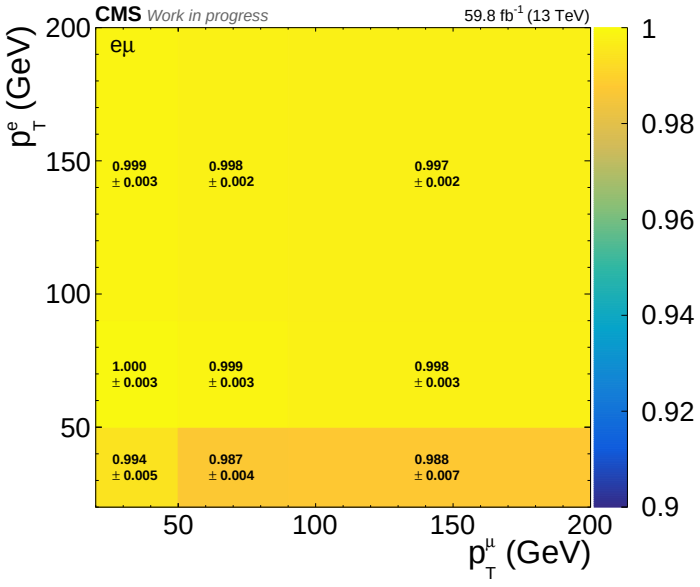


Figure 8.4: Trigger SFs for the $e^{\pm}\mu^{\mp}$ channel. The SFs are functions of the muon and electron p_T. The SFs are shown for each bin including the total uncertainties. Taken from Refs. [163, 164].

jet energy scale (JES) and the jet energy resolution (JER) correction. In general, the JECs are determined using simulated events of various processes. The calculated corrections are subsequently applied to data and simulated events. The multi-stage process of the calculation is sketched in Figure 8.5. The initial step is the pileup correction, which removes energy deposits in the detector from pileup interactions from the jets. Differences between data and simulated events are determined using the random cone method [160]. This is followed by corrections to the jet response based on a comparison of reconstructed jets in simulated events relative to the true information of jets at generator level. Remaining differences are applied using so-called residual corrections which are determined in $\gamma/Z+$jet and multi-jet events as a function of η and p_T. The differences in the JER between data and simulated events are corrected in simulation. For this purpose, a jet matching from detector level to particle level is performed and a SF is calculated for the p_T value of the corresponding jet. If the matching fails, the resolution is corrected with a SF that is randomly smeared based on a Gaussian distribution.

8.4.7 Top p_T reweighting

According to earlier measurements by the CMS Collaboration, the top quark p_T in simulated $t\bar{t}$ events at NLO precision in the ME calculation shows discrepancies when compared to data [165]. The deviation of simulated events and data in the spectrum of the transverse momenta of the $t\bar{t}$ system is reduced if NNLO predictions are used instead. Hence, a reweighting procedure is used which maps the NLO spectrum of the top quark p_T observable

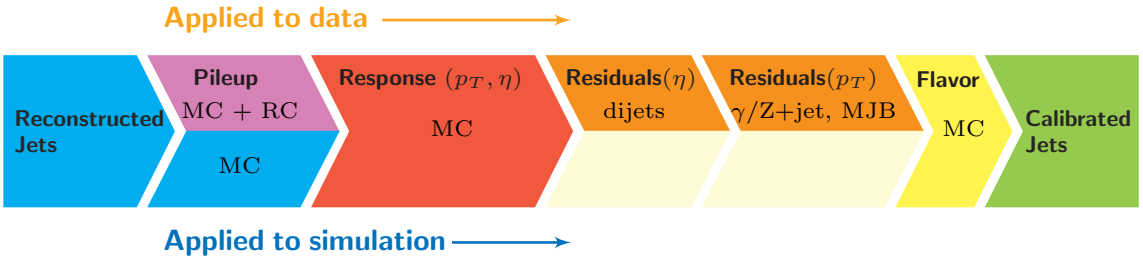


Figure 8.5: Steps of the JEC procedure for data and simulated events. The source of the calculation is indicated, whereby RC refers to random cone and MJB refers to multi-jet events. Taken from Ref. [160].

to the corresponding spectrum at NNLO. With this technique, an event weight $w_{\text{NNLO/NLO}}$ is calculated, which consists of a combination of the two SFs for the top quark and the top antiquark as a function of their transverse momenta, which is calculated with

$$w_{\text{NNLO/NLO}} = \sqrt{\text{SF}(p_{\text{T}}^{\text{t}})\text{SF}(p_{\text{T}}^{\bar{\text{t}}})} \quad , \tag{8.9}$$

with $\text{SF}(p_{\text{T}}) = \alpha e^{-\beta p_{\text{T}}/\text{GeV}} + \gamma p_{\text{T}}/\text{GeV} + \delta$, where $\alpha = 0.103$, $\beta = 0.0118$, $\gamma = -0.000134$, and $\delta = 0.973$. This parameterized weight function is based on Ref. [166].

8.4.8 Heavy flavor jet tagging efficiency

Generally, there are two ways to calibrate the efficiency of the heavy flavor jet tagging algorithm. In the first approach, the efficiencies for the three predefined WPs (cf. Section 4.4) are corrected. These corrections adjusts the yields of the b jet multiplicities. In the second approach, the full b jet tagging discriminant shape is corrected. This is particularly important if the complete information from the heavy flavor tagger is used. In this thesis, this procedure is necessary since the entire b jet tagger information is used as an input feature for the graph transformer networks for jet and event classification in Chapter 9. To achieve a better agreement of the b jet tagging discriminant in data and simulated events, the iterative Fit (itFit) method is used based on Ref. [97]. To determine the weights that compensate for the discrepancies, data are compared to simulated events in different event topologies such as $\text{t}\bar{\text{t}}$ or QCD events, specifically chosen for their varying jet flavor compositions. The iterative procedure is based on a tag-and-probe method requiring exactly two jets and simultaneously determines the SFs for both b and light flavor jets. The SFs are derived as a function of the DeepJet discriminant value in bins of p_{T} and η of the jet and are provided by the CMS Collaboration [91, 97].

However, the application of the centrally provided SFs, which are intended to correct only the shape of the b tagging discriminant, also affects other observables and their distributions. This is due to the fact that the phase space for determining the corrections can deviate

significantly from the analysis phase space. Observables that should be invariant under the application of SFs are, for example, the number of jets (N_{jets}) and the sum of all transverse momenta (H_T) in an event. For this reason, a correction is applied that reverses the effects of the applied itFit b tagging weight w_b, which represent the SF in these observables. This is achieved by calculating the ratio R of the sum of the event weights before the application ($\sum w_{before}$) relative to the sum of the event weights after the application ($\sum w_{after}$) in bins of N_{jets} and H_T:

$$R\left(N_{jets}, H_T\right) = \frac{\sum w_{before}\left(N_{jets}, H_T\right)}{\sum w_{after}\left(N_{jets}, H_T\right)} \quad . \tag{8.10}$$

The new, corrected event weight is obtained by multiplying the ratio R with the b tagging event weight, which is the product of the individual jet weights

$$w_{b,\,corr} = R\left(N_{jets}, H_T\right) \cdot \prod_{i}^{N_{jets}} w_b\left(D_i, p_{T\,i}, \eta_i\right) \quad , \tag{8.11}$$

where D_i is the b tagging value of the i-th jet in an event. The distribution of the b tagging value can be seen in Figure 8.6. Three scenarios are shown: no application of the itFit b tagging weights, application of the itFit b tagging weights w_b, and the application of the reweighted itFit b tagging weights $w_{b,\,corr}$ according to Equation 8.11. It can be seen that the application of the itFit b tagging weights changes the rate and the shape of the unweighted discriminant. The application of $w_{b,\,corr}$ compared to the itFit b tagging weights causes a small rate change for the majority of the discriminant. This behavior is different for the observables N_{jets} and H_T that are used to calculate the reweighting depicted in Figure 8.7. It can be seen how the number of events increases due to the application of the itFit b tagging weights for increasing jet multiplicities compared to the distribution without the application of these corrections. This unwanted effect is compensated if the corrected weights $w_{b,\,corr}$ are applied instead. With the application of the corrected weights $w_{b,\,corr}$ the distribution of N_{jets} does not change, which reflects the desired behavior. Analogous behavior can be seen for the H_T observable in this figure.

8.5 Control distributions

Control distributions verify the agreement between data and simulated events, ensuring that the predictions of all simulated processes reflect real conditions at the CMS experiment. Likewise, control distributions can be utilized to verify the accuracy and conformity of object reconstruction in both data and simulation. For example, possible systematic effects in certain observables can be identified. Known and relevant corrections for these effects are discussed in the previous section and are applied to all events.

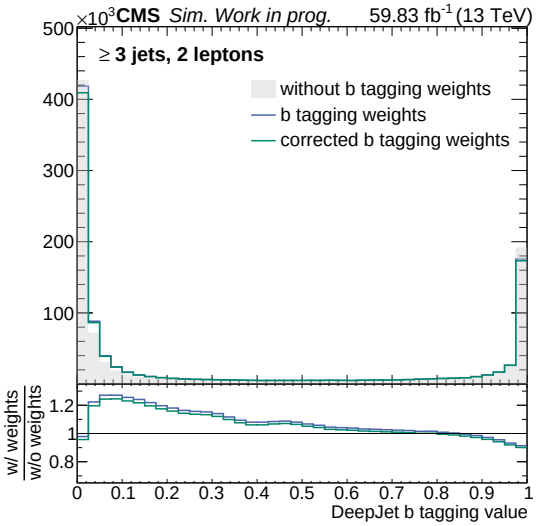


Figure 8.6: Distributions of the b tagging discriminant of the DeepJet heavy flavor tagging algorithm without itFit b tagging weights, with itFit b tagging weights and with corrected itFit b tagging weights. The application of the weights results in a rate and shape changing effect compared to the distribution without the weights, as can be seen in the ratio plot in the lower panel.

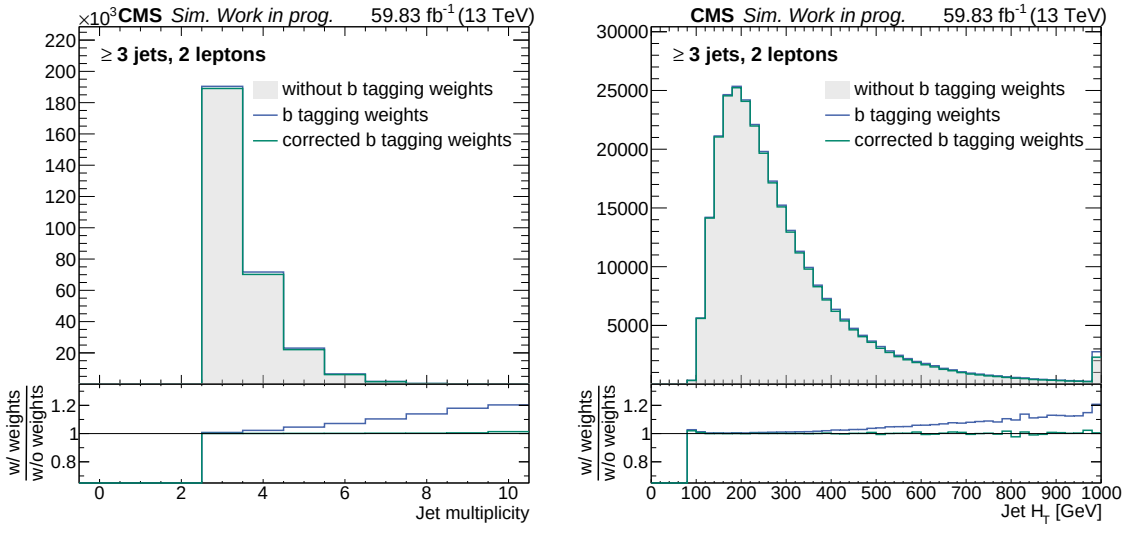


Figure 8.7: Distributions of the N_{jets} (top) and the H_T (bottom) observable without itFit b tagging weights, with itFit b tagging weights and with corrected itFit b tagging weights. The application of the itFit b tagging weights results in a rate and shape changing effect compared to the distribution without the weights, as can be seen in the ratio plot in the lower panel. The application preserves the original distribution of both observables.

Jet, lepton, and MET observables are depicted in Figure 8.8 and 8.9. It can be seen that the uncertainty bands, which include all a-priori uncertainties described in Section 10.1, enclose the majority of data points. For events with three jets in the analysis phase space, the ratio of the data yield to simulated events is approximately one. For event with increased jet multiplicities, larger event yields are predicted. Comparing this behavior with the event yields as a function of the number of b-tagged jets in the events that fulfill the medium WP, it can be recognized that events without b-tagged jets tend to be under-predicted. Events without b-tagged jets form a region containing predominantly background processes. In this region, the uncertainty of the prediction does not cover the data point. Events with one or two b-tagged jets, on the other hand, are over-predicted in the event simulation. In this region, $t\bar{t}C$ events in particular are enriched, but also $t\bar{t}B$ and $t\bar{t}H(B)$ as well as $t\bar{t}Z(B)$ events can occur, in cases where some b jets do not fulfill the medium WP or jets are out of acceptance. In the phase space region with three b-tagged jets, an over-prediction prevails, while the prediction matches the data in the region with four b-tagged jets. In this region, fully resolved $t\bar{t}B$, $t\bar{t}H(B)$ as well as $t\bar{t}Z(B)$ are enriched. However, it should be noted that the statistical uncertainty increases with rising b-tagged jet multiplicities. In Section 10.2.4, a quantitative evaluation is performed using a goodness-of-fit test for the statistical model of the predictions. This test demonstrates that the predictions meet the required criteria. The only exceptions are the ratio observables c vs. b (CvB) and c vs. light (CvL) (cf. Equations 4.3 and 4.3), which, as discussed in Section 9.1.1, do not fulfill these requirements and are therefore excluded.

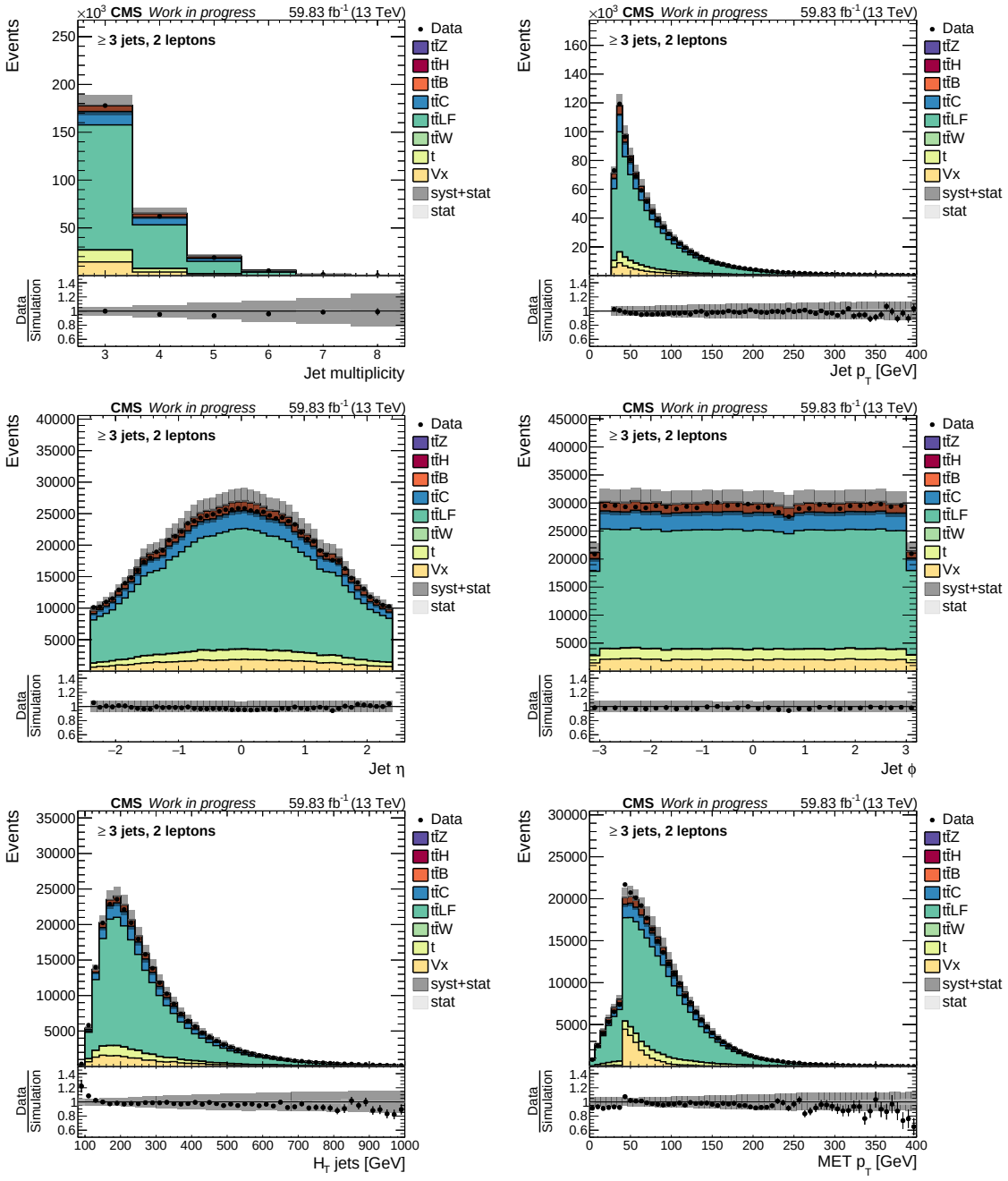


Figure 8.8: Jet multiplicity (top left), jet p_{T} (top right), jet η (middle left), jet ϕ (middle right), $\mathrm{H_T}$ of all jets (bottom left), and MET ϕ (bottom right) in the analysis phase space. The contributions of all processes are displayed as stacked histograms, with data represented as black dots. The bottom pad depicts the ratio of data to the expected yields from event simulation. The gray bands include the a-priori uncertainties described in Section 10.1.

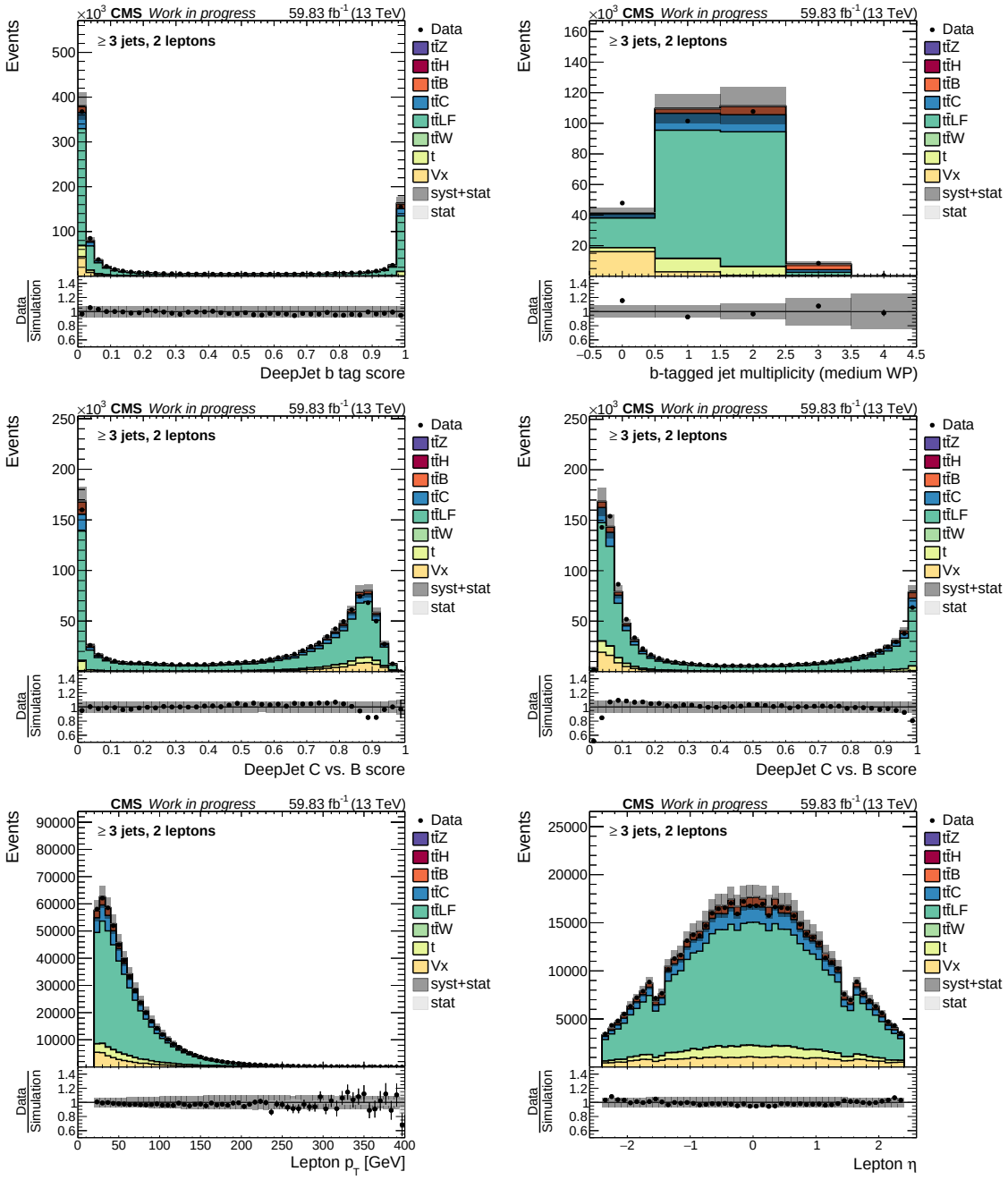


Figure 8.9: DeepJet b tagging discriminant (top left), number of b-tagged jets at the medium WP p_{T} (top right), DeepJet CvB tagging discriminant (middle left), and DeepJet CvL tagging discriminant (middle right), lepton p_{T} (bottom left), and lepton η (bottom right), in the analysis phase space. The contributions of all processes are displayed as stacked histograms, with data represented as black dots. The bottom pad depicts the ratio of data to the expected yields from event simulation. The gray bands include the a-priori uncertainties described in Section 10.1.

9 Data set and AI model design

In this chapter, the classification of events using Artificial Intelligence (AI) techniques, particularly leveraging its machine learning branch to apply multivariate analysis methods, is examined in the analysis phase space defined in Chapter 8. First, the transformation of simulated events into a fully connected graph is explained and the overall design of the graph training structure is discussed. The two stages of classification, the particular designs and the results of trainings with graph neural network (GNN) are described in Sections 9.2 and 9.3.

9.1 Data transformation and network architecture

Simulated events and recorded data at the Compact Muon Solenoid (CMS) experiment are stored in various data formats. Only a few of these formats are suitable for extensive data analyses, as the rapidly increasing storage requirements of more expensive formats that store details of particle candidates make effective data handling impossible. Data storage in the raw format can be in the order of megabytes per event while the smallest data format takes only a few kilobytes of memory per event. Therefore, when processing several billions of data points[1] in this analysis, a technically feasible data format must be chosen. In the smallest format, known as nano Analysis Object Data (nanoAOD), the total storage requirement amounts to approximately 6.3 terabytes. A skimming procedure reduces the amount of data and simulation to approximately 110 gigabytes. Skimming refers to the process of extracting only relevant information and removing event candidates that do not pass the selection criteria stated in Chapter 8. An additional 80 gigabytes are generated after skimming due to additional simulations of systematic uncertainty variations. In

[1] A total of 4 611 694 287 events in data and simulation are generally candidates in this analysis phase space and require processing. This does not include any additional simulations for the consideration of specific systematic uncertainties.

the nanoAOD data format, only high-level information on physics objects such as jets and leptons are available. For example, individual components for clustered jets such as information on particle-flow candidates (cf. Section 4.2) are truncated. The jet origin information crucial for the training in this chapter does not exist in the standard nanoAOD format. This information needs to be taken from the next higher compression format, the mini Analysis Object Data (miniAOD) format, in a dedicated procedure for all $t\bar{t}$+jets, $t\bar{t}b\bar{b}$, $t\bar{t}H$, and $t\bar{t}Z$ event simulations. In total, about 10 million unweighted simulated events remain in the nanoAOD format after the skimming procedure.

To enable supervised training on simulated events, all required information must be present at the detector simulation level. At this level the degree of information is equivalent to the degree of information in recorded data. However, information about whether a jet originated from a top quark, for instance, only exists at generator simulation level at most. In the common event simulation procedure at the CMS experiment as described in Section 4.6.1, there is no information at detector level about the origin of a jet. This lack of information is remedied by a matching procedure transferring the jet origin information between the two simulation levels. Jets at the generator simulation level are compared to jets at the detector simulation level and all distances are calculated. If the distance between a generator level jet and a detector level jet is less than $\Delta R = 0.4$, the jet origin information is transferred. The matching efficiency of jets at the two different levels is 99.5% to 99.7% depending on the event simulation in this thesis.

In this thesis, GNNs are used due to their ability to represent relationships and patterns between objects in an event. Unlike a feedforward deep neural network (DNN), GNNs offer structural advantages such as permutation invariance and flexible input dimensions. However, GNNs do not outperform DNNs by default. In fact, GNNs only unfold their full potential when domain knowledge of the physics processes is incorporated in various aspects into the graph structure. In a preceding comparison study between GNNs and DNNs conducted for a generic Large Hadron Collider (LHC) experiment with a similar event topology and a closely related classification task associated to this thesis, the advantages of GNNs over DNNs are elaborated in detail [111]. This study demonstrates how carefully designed GNNs are able to outperform comparable DNN structures, converge faster, generalize more strongly, and require fewer free parameters for the classification task. The initial design of the graph structure in this thesis based on domain knowledge is conceptually equivalent with the construction of the graphs in this study and is explained in the following sections.

Several Bachelor and Master theses thoroughly explored various aspects of jet and event classification in the $t\bar{t}$ + heavy flavor phase space in preparation of this thesis. The jet classification using DNNs in Ref. [167] represents the first step and cornerstone of subsequent analyses. In Ref. [168], jet classification using GNNs is examined. Ref. [169] focuses on the Bayesian optimization of GNNs as well as hyper-graph structures, which

reflect properties of groups in a GNN such as the top quark decay products with hadronically decaying W bosons. The studies in Ref. [170] provide dedicated performance comparisons of GNNs with equivalent DNN structures. In addition, this study includes investigations on properties of GNNs for a deeper understanding of important factors impacting the stability and reliability of GNNs. The thesis in Ref. [105] is dedicated to simulations of $t\bar{t}B$, $t\bar{t}H$, and $t\bar{t}Z$ processes with a simplified detector simulation for fast event simulation with reduced computational complexity. The studies in Ref. [171] analyze implementation strategies for a measurement in the $t\bar{t}$ + heavy flavor phase space. Finally, the study in Ref. [172] examines the GNNs used in this thesis to determine the key factors driving their performance, focusing on the importance of the input features.

9.1.1 Input features

The graph structures are composed of the kinematic observables that are accessible from the physics objects at nanoAOD level, except for the jet origin. Each final state object is thus translated into a node of the graph structure. The transformation of an exemplary event into a graph structure is illustrated in Figure 9.1. Generally, the maximum number of nodes is arbitrary and can be freely configured for each event. In contrast to DNNs, it is not necessary to set an upper boundary on the number of jets for the graph or to apply a padding procedure to fill empty nodes due to a fixed input layer size [111]. The initial size of the graph representation of an event in this thesis is a function of the jet multiplicity. The total number of nodes n in a graph results from the jet multiplicity plus three, as there are always exactly two leptons and missing transverse energy (MET) in the event after the selection, cf. Section 8.3. Consequently, a graph comprises a minimum of six nodes. Initially, the features of each node in the graph follow a fixed pattern represented by a vector referred to as node feature vector. The kinematic information of an object is initially embedded in a predefined order in this node feature vector: p_T, η, ϕ, M, E, electric charge, and the b tagging value. The b tagging value refers to the DeepJet heavy flavor jet tagging algorithm score as described in Section 4.4. If a feature does not exist, this value is artificially set to zero, e.g. the b tagging value for a lepton. In addition, the type of object is encoded in the node feature vector with a flag in the form of a one-hot encoding. The first flag denotes a jet, the second refers to a lepton, and the third corresponds to MET. This offers the GNN a simple but effective way of distinguishing the type of object.

The relation between two physics objects in the detector is also embedded in the graph. In the comparison study between GNNs and DNNs, it was revealed that GNNs are able to process this information superior to DNNs [111]. Furthermore, the GNN benefits significantly from the embedding of this domain-knowledge and can therefore also increase its performance compared to an equivalent DNN model and a GNN model without this information. In the benchmark study, relational information is added individually to the graph structure, demonstrating the strong performance gain from embedding a single

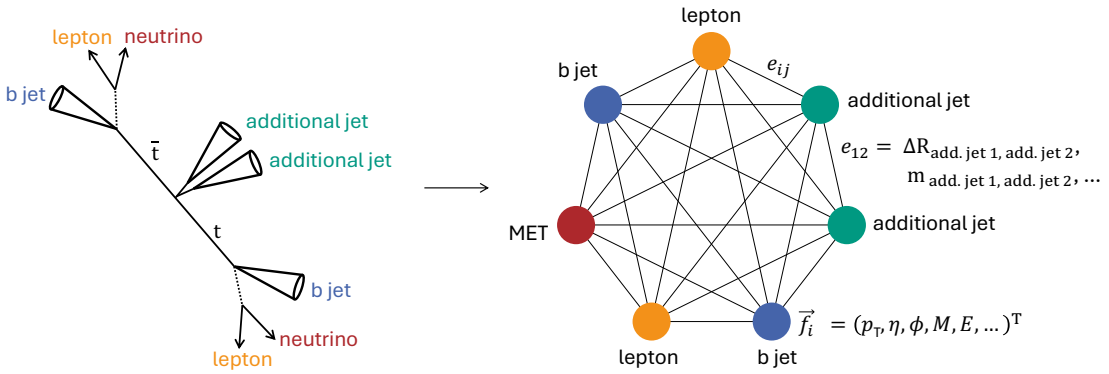


Figure 9.1: Each object in an event (jet, lepton, MET) (left) corresponds to a node in the graph (right). Each node possesses kinematic properties of p_T, η, ϕ, M, E, charge sign, and the b tagging value if applicable as its node feature vector. The edges of the fully connected graph structure are characterized by relational information of the node pairs, for example the mass m_{ij} or distance ΔR_{ij} between the two objects i and j.

quantity such as the distance ΔR_{ij} or mass m_{ij} between two objects. In contrast, multiple quantities based on domain-knowledge are embedded simultaneously in the graph structure in this thesis in order to benefit more effectively from the combination of all observables.

To prevent certain features from dominating the training, e.g. p_T, all features are normalized to a range of $[0, 1]$ for continuous variables. Such normalization avoids an imbalance in the importance of a sub-set of features and prevents convergence issues in the optimization algorithm during training. All node features are summarized in Table 9.1. Relational information, of which a total of six features are constructed, is implemented as edges in the graph. These are calculated for every combination of nodes i and j, and include m_{ij}, $p_{T,ij}$, $m_{T,ij}$, ΔR_{ij}, $\Delta\eta_{ij}$, and $\Delta\phi_{ij}$.

In Figure 9.2 the jet distributions of the node features p_T and η as well as the edge features m_{ij} and ΔR_{ij} are shown as selected examples for $t\bar{t}$+jets, $t\bar{t}H$, and $t\bar{t}Z$ events. All remaining node and edge feature distributions can be found in Appendix A. A distinction is made between b jets from the $t\bar{t}$ system and additional jets. The term additional jets denotes jets that do not stem from the $t\bar{t}$ system, but from gluon splitting such as $g \to b\bar{b}$ or the boson decays $H, Z \to b\bar{b}$, as defined in Section 8.1. Generally, the four observables shown in this figure reveal particularly strong separation power between the jets from the $t\bar{t}$ system and the additional jets (cf. Section 4.6.2) for all three processes. A multivariate method can therefore be used to identify the origin of a jet, which is referred to as *jet classification* in this thesis. Beyond that, the figure also show that not only a jet separation is possible, but also an event separation between $t\bar{t}$+jets, $t\bar{t}H$, and $t\bar{t}Z$ simulated events. It is expected that the $t\bar{t}$ system looks rather similar in $t\bar{t}$+jets, $t\bar{t}H$, and $t\bar{t}Z$ events, however,

Table 9.1: Input features used for jet and event classification.

Input feature	Jets	Leptons	MET
p_{T}	$x \in [0, 1]$	$x \in [0, 1]$	$x \in [0, 1]$
η	$x \in [0, 1]$	$x \in [0, 1]$	0
ϕ	$x \in [0, 1]$	$x \in [0, 1]$	0
m	$x \in [0, 1]$	$x \in [0, 1]$	0
E	$x \in [0, 1]$	$x \in [0, 1]$	0
Electric charge sign	0	$x \in \{-1, 1\}$	0
b tagging value	$x \in [0, 1]$	0	0
Jet flag (one hot)	1	0	0
Lepton flag (one hot)	0	1	0
MET flag (one hot)	0	0	1

the behavior of the additional jets in these features differs across the three processes. Hence, an event classification using multivariate methods is feasible especially when the attention centers on the additional jets. This procedure is referred to as *event classification* in this thesis.

The p_{T} distribution in Figure 9.2 shows that the additional jets carry smaller transverse momenta compared to the jets from the $\mathrm{t\bar{t}}$ system. Also, it is apparent that while the additional jets in $\mathrm{t\bar{t}H}$ and $\mathrm{t\bar{t}Z}$ events are similar in p_{T}, both reveal larger values compared to the additional jets in $\mathrm{t\bar{t}}$+jets events. The observable η indicates that the jets from the $\mathrm{t\bar{t}}$ system tend to be located in the more central region of the detector compared to the additional jets. Particularly small pseudo rapidities occur in the case of additional jets in $\mathrm{t\bar{t}}$+jets events compared to additional jets from $\mathrm{t\bar{t}H}$ and $\mathrm{t\bar{t}Z}$ simulated events. The quantity ϕ does not contain any direct separation power by definition due to the given azimuthal symmetry in the detector. However, as a spatial component in combination with other quantities, ϕ can form a strong observable, for example when geometric distances such as ΔR to other objects in the event are calculated. Even though the distances ΔR_{ij} are integrated as edges separately into the network, the quantity ϕ is added to the network in order to determine potential relations to other observables. Furthermore, the observables m and E are added to the network as node features. Despite the fact that m and E are linked to the momenta via the energy-momentum relation and thus contain correlated information, all three quantities are added to the network. In this way, it is the task of the network to decide which information is most suitable for classification and it is not required to learn the energy-momentum relation, for example. The b tagging value information also contains distinguishing characteristics. While the b jets from the $\mathrm{t\bar{t}}$ system tend to have large b tagging values, the additional b jets tend to have smaller values relative to these b tagging values. This is a result of the fact that, as shown above, the b jets from

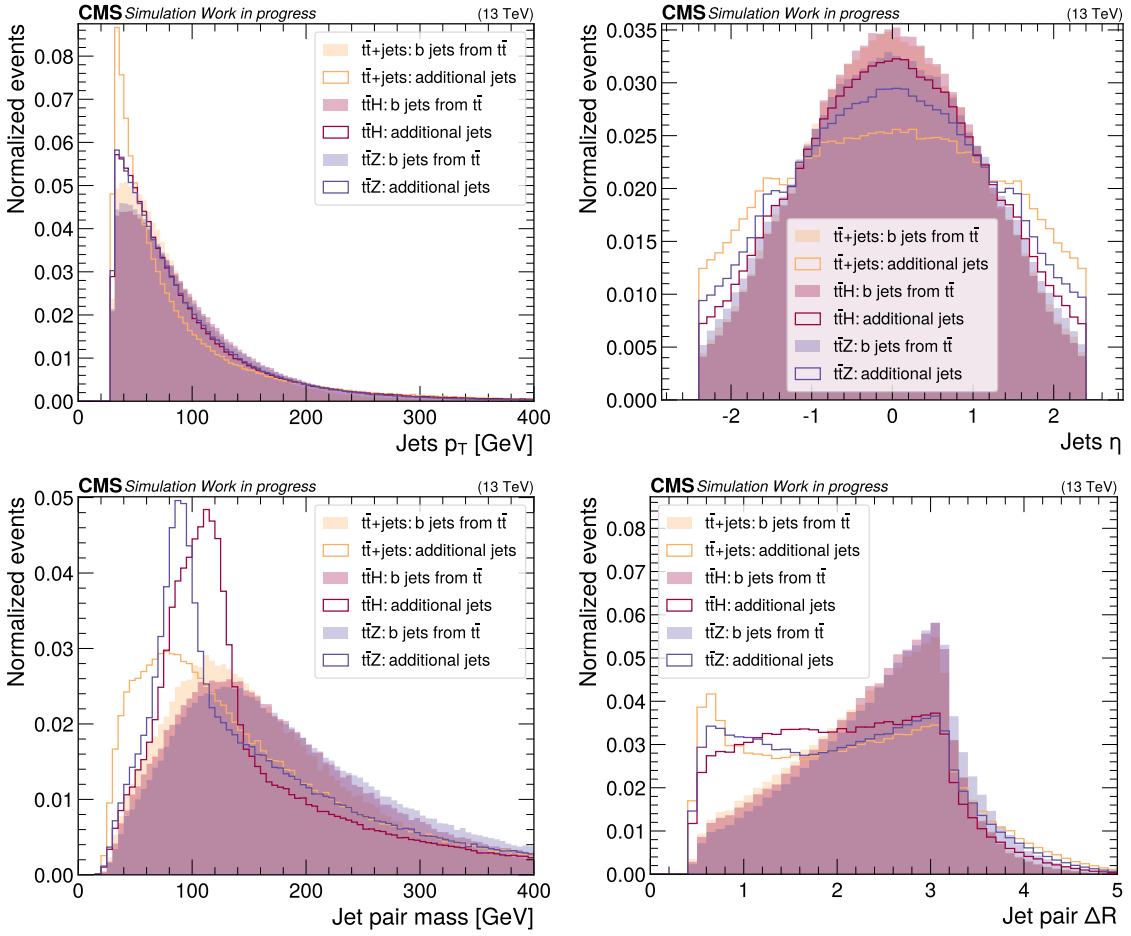


Figure 9.2: Input feature distributions of jet p_T (top left) and jet η (top right) included as node features for simulated $t\bar{t}$+jets, $t\bar{t}H$, and $t\bar{t}Z$ events. The distributions are separated into b jets from the $t\bar{t}$ system and additional jets. The bottom row shows the jet pair mass (left) and the distance ΔR (right) of jet pairs for b jets from the $t\bar{t}$ system and additional jets. Jet pair information is used as edge features in the graph structure.

the $t\bar{t}$ system have larger transverse momenta. Consequently, secondary vertices can be more pronounced, which in turn is an important parameter for heavy flavor tagging. Also, Figure 4.3 reveals that the tagging efficiency increases with higher transverse momenta.

The reconstruction of the mass from jet pairs through the combination of the two four-vectors is one of the most distinct observables. While the mass distributions of jet pairs from the $t\bar{t}$ system show no remarkable characteristics, the additional jets in the case of $t\bar{t}H$ and $t\bar{t}Z$ events reveal the mass peaks of the Z and H bosons prominently shown Figure 9.2. In contrast, in the case of $t\bar{t}$+jets events the additional jets show no resonances, as expected, and the distribution of the jet pair masses is shifted to smaller values relative to the masses from the b jet pairs of the $t\bar{t}$ system. Another important observable is the spatial distance ΔR_{ij} in the detector between two jets. With this distance measure it can be noticed that the additional jets generally show smaller angular separation compared to the b jets from the $t\bar{t}$ system in an event. Since there is an angular correlation in additional jets resulting from the splitting of a gluon, for example, this pattern is expected. The ΔR_{ij} distribution of the b jet pairs from the $t\bar{t}$ system show relatively larger distances and reaches a maximum at π before dropping rapidly. This behavior results from the definition of ΔR_{ij} in Equation 3.5, as $\Delta\phi_{ij}$ can only attain a maximum value of π. In differential measurements of the $t\bar{t}b\bar{b}$ process the ΔR observable is intentionally chosen in order to achieve a good accuracy in finding the additional b jet radiation in the event without employing a multivariate method in favor of a straightforward kinematic definition [96]. In the analysis in Ref. [96], the additional jets are defined as the jets with the smallest distance among all b jets in the event.

The training of models in this thesis is performed exclusively on simulated events, as this allows for supervised learning with labeled training samples. Subsequent inference using the trained models to evaluate data implies the assumption of identical behavior between data and simulation in all input features. To verify the validity of this assumption, each feature used is subjected to a goodness of fit test. This statistical test examines whether the distribution of simulated events deviates from the respective observable in data for each feature. The required threshold at which it is assumed that the distributions do not match is a p-value below 5%. Further details about the method can be found in Section 10.2.4, where it is discussed in detail within the context of the statistical model. The summary of the goodness of fit tests for all input features can be seen in Figure 9.3. Each feature embedded in trainings satisfies the required p-value threshold.

Two variables that are not utilized in this analysis as they do not pass the goodness of fit test are the DeepJet tagger scores CvB and CvL, as defined in Equations 4.3 and 4.4. These two-dimensional jet flavor discriminants allow a distinction between b jets and c jets as well as c jets and light flavor jets, as described in Section 4.4. In the appendix, Figure A.1 demonstrates that these two observables provide a strong distinction between additional jets and b jets from the $t\bar{t}$ system. However, a difference in the behavior of

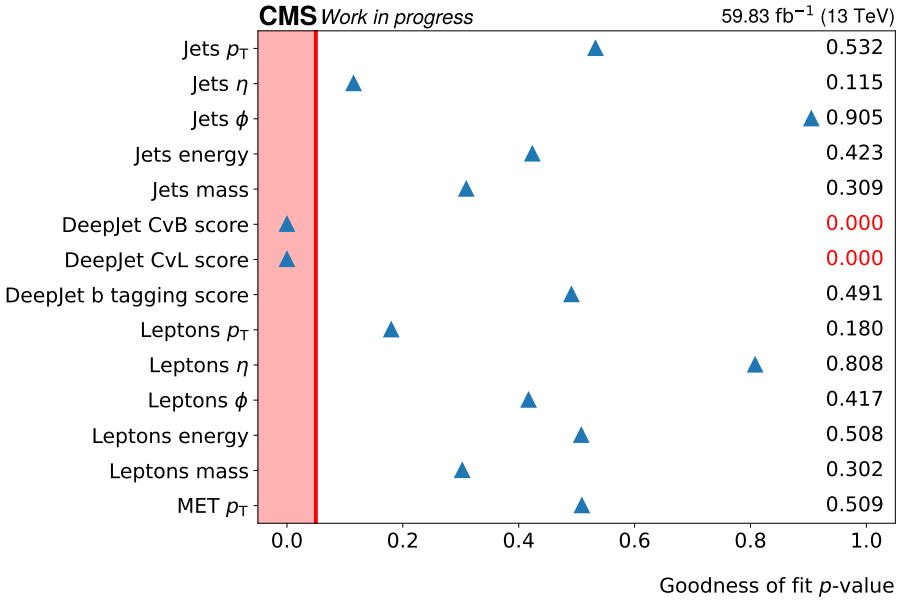


Figure 9.3: Resulting p-values of the goodness of fit tests performed for each observable between data and simulation. Observables with p-values greater than 5% are used as input features in the training. The two DeepJet jet flavor tagger discriminants c vs. b (CvB) and c vs. light (CvL) do not pass the test, all other kinematic features are considered in the training.

the two scores between simulation and data can be observed in areas with particularly high and particularly low scores. In addition, these two value ranges account for the majority of jets. Performing a calibration or applying a scale factor (SF) as discussed in Section 8.4.8 does not resolve the divergence. Instead of the two-dimensional tagger information the one-dimensional b tag score is used. By definition, this feature is correlated with the two two-dimensional observables, but the differing behavior of the b tagging value in simulation and data can be corrected with a SF provided by the CMS Collaboration. Although incorporating CvB and CvL as node features improves the performance of all analyzed models, both are removed from the training due to the mismatch between event simulation and data.

9.1.2 Training strategy

Since the additional jets play a key role in event separation, a two-stage training procedure is designed. First, a classification algorithm is implemented which is tasked with identifying the additional jets in an event. This information, i.e. a probability estimation from a jet classifier that an object in an event is an additional jet, is used in a subsequent event classification algorithm. The event classification constitutes the second stage. The overall transformation of the event into a graph structure is conceptually identical in both stages. The node and edge features are constructed as previously described. However, the jet

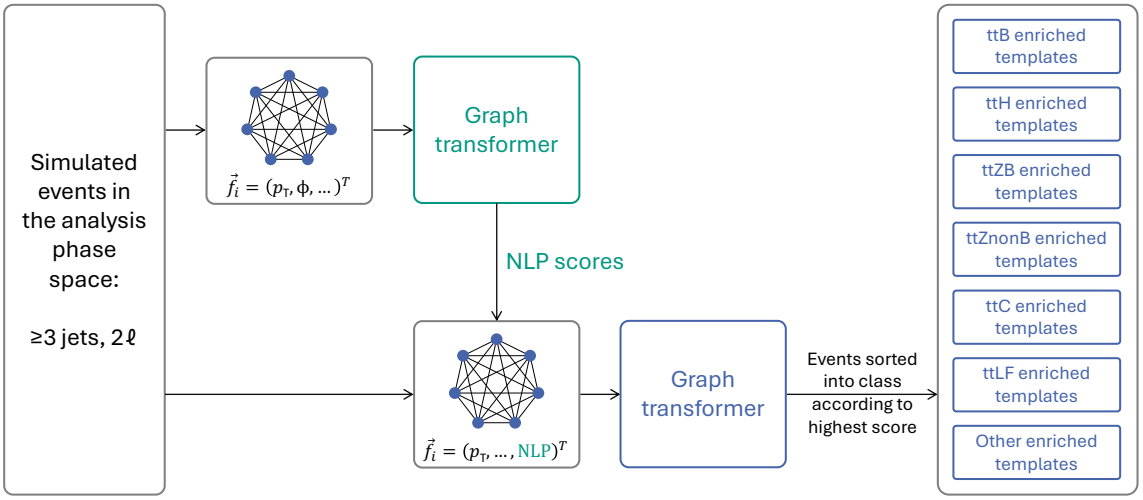


Figure 9.4: Conceptual representation of the steps from event simulation to event classification. Graphs are constructed and trained based on simulated events. First, a binary node level prediction is performed which predicts for each node of a graph whether it is an additional jet or not. This information is then incorporated into an equivalent graph structure, which now also contains this information from the pre-classifier. In the second training stage, a graph level prediction is performed that predicts the type of event for each graph. The events are then categorized into the class for which the highest score is achieved.

prediction score from the first classification stage is concatenated to each node feature vector in the graph structure of the second stage. Based on an overall similar input structure, the main technical difference in the two stages is the type of classification of the GNN. Given the different classification objective, a node level prediction (NLP) is performed in the first classification stage to identify the jet-type. This classification is a binary decision as to whether a node in the graph is an additional jet or not. The evaluation is performed for each node in the graph structure. In the second stage, a graph level prediction (GLP) is performed to distinguish signal from background events. Here, a multi-class classification is performed in which the classifier assigns a score to each graph structure for belonging to a predefined event class. This score can be interpreted as a probability for an event to be a $t\bar{t}B$ event, for example. This two-stage procedure is depicted in Figure 9.4. Since the graphs are constructed from identical domain knowledge, the output of the NLP as additional input to the GLP model is redundant information from a physics perspective. However, it is demonstrated in Section 9.3.3 how this pre-classifier procedure boosts the event classification performance with explicit, refined information about the jet origin.

While the initial graph structure of the two-stage training procedure remains almost identical, numerous differences exist in the training architecture. For the jet classification with NLP these can be found in Section 9.2, while those for the event classification with

GLP can be found in Section 9.3. With the help of the event classification the relevant parameters of the constructed physics model can then be extracted in a fit model. The application of the event classifier is described in Chapter 10.

Two additional conceptual principles are implemented in both training stages. The first principle relates to running and testing each network architecture multiple times. In this thesis, the term ‘‘architecture’’ refers to the overall structure of a network with various layers, while the term ‘‘model’’ refers to a specific configuration of the architecture with a fixed choice of hyperparameters, random weight initialization, and the subsequent execution of the optimization on the training sample. In general, each configuration is trained ten times in this thesis. This enhances the robustness and reliability of the model and prevents a particularly good or relatively poor optimum from being found by coincidence during the training process. Performing only one training could result in the false selection or rejection of a model, even though the setting may (not) constitute the optimal configuration. Instead, training identical network configurations ten times with different random seeds ensures that the best architecture is found objectively.

The second principle is to perform the training procedure with a two-fold training sample. The reason for splitting the simulated events into two orthogonal subsets arises mainly from the use of the trained models in the subsequent fit model rather than the performance evaluation during the training procedure. For the fit model of this analysis, both simulated events and data must be evaluated with the trained model. To prevent trained models from being evaluated on simulated events that were already seen in the training procedure the simulated events are split into two folds. While one model is trained on all events with an even event ID, the trained model for the subsequent statistical model (cf. Chapter 10) is only evaluated on simulated events with an odd event ID. Similarly, a second model with identical settings and thus the same architecture is trained on all events with odd event IDs and selected for the evaluation of all simulated events with even IDs. In this way, a possible bias in the behavior of the AI model on simulated events relative to data, for example when performing statistical tests on simulated events, can be prevented.

9.2 Jet classification: Node level prediction

As described in Section 9.1, the additional jets are particularly crucial for event separation, which is why special emphasis is put on their identification. The two-staged process illustrated in Figure 9.4 can in principle be standardized into a single training step. However, as the jet identification step is extremely important, it is designed in two separate training stages. The main advantage of this procedure is that the results of the jet identification can be investigated independently from the event classification. Likewise, the quality of the classification performance can be benchmarked against other methods of jet identification and, if necessary, replaced by a different approach.

9.2.1 NLP architecture and training

For the training of jet identification, state of the art machine learning methods are employed. To exploit the advantages of both GNNs and transformer architectures discussed in this section, a combined graph transformer network architecture is employed. Two key reasons advocate for the use of graph transformer network architectures. First, the graph structure described above is exploited, into which events are translated in a natural manner. Second, the advantages of transformer architectures with self-attention mechanisms are utilized, which learn and reflect relationships in sequences [112]. The latter is used successfully in particular for natural language processing in large language and foundation models. The TransformerConv operator in Pytorch Geometric is used for this purpose [173].

After the normalization and translation of the observables into node and edge features of the graph structure as described in the previous section, the graph is first translated into the embedding space. This is done with an encoding layer that changes the dimension of the initial node feature vector and transforms all features into a representation in the embedding space. The dimension E of the vector in the embedding space is a hyperparameter of the model and remains unchanged in the transformer architecture. At this point, the straightforward interpretation and readability of the graph structure built from components of physics quantities is no longer given. Following the encoding step, a further normalization is performed and passed on to the activation function. This is followed by a series of graph transformer blocks in which the key operations of the training are performed. These graph transformer blocks are repetitive units of identical operations, analogous to multiple hidden layers of a feedforward neural network. Similar to the embedding dimension, the number of blocks is also a hyperparameter of the architecture. In these blocks, the general idea is identical to other GNNs, in which the nodes of the graph are updated using message passing. The first step in the message passing procedure is to compute the message for each node. Each node calculates a message for each of its connected neighbors. In the model of this thesis the calculation is done for each encoded representation that was initially created from physics objects into the embedding space since it is implemented as a fully connected graph. The calculated message for a node is a function of its own feature vector, the edge to the neighbor, and the feature vector of the neighbor. After the message calculations, the messages are sent and each node in the graph aggregates the received messages. The sum or mean can be used for message aggregation, in this thesis, the former is used. In the final step, the feature vector is updated by combining the aggregation with a node's own feature vector to create an updated node feature vector. This entire process runs in parallel and synchronously for all nodes in the graph. After the graph transformer blocks, a multilayer perceptron (MLP) is placed to generate the prediction. The architecture is shown in Figure 9.5 and the details about the mathematical operations in the graph transformer blocks are explained in the following.

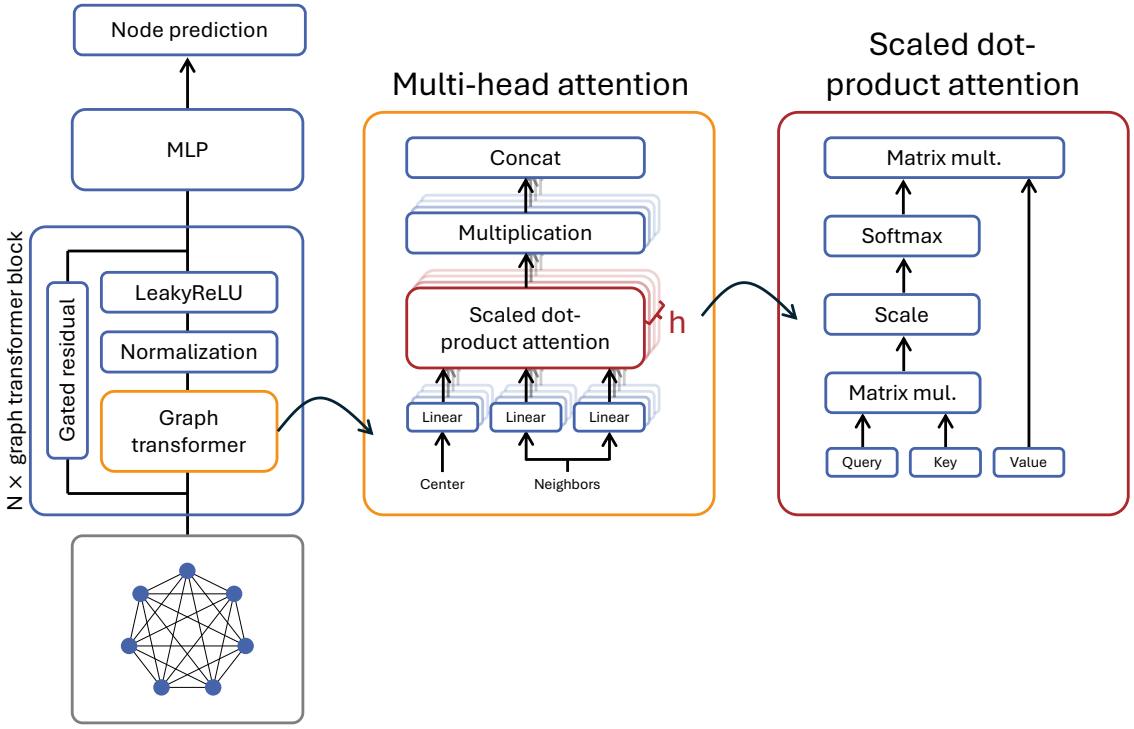


Figure 9.5: Jet identification architecture designed with graph structures and transformer architectures using multi-head attention mechanisms. After a total of N transformer blocks the embedding dimension of the feature vectors is decreased with an MLP to a scalar predicting the type of node. Adapted from [173].

Instead of the typical adjacency matrix in GNNs, which reflects the weighted connections for message passing, in this architecture it is replaced by the multi-head attention matrix to enable self-attention mechanisms. The attention mechanism of the transformer models in its originally proposed form [112] is adapted for GNNs [173]. The self-attention is calculated in a sub-block referred to as head and based on Ref. [173]. As a first component for the calculation of the attention in a head of the GNN, query (q) and key (k) vectors are required. For the embedded node features $F^{(l)} = \{f_1^{(l)}, f_2^{(l)}, ..., f_N^{(l)}\}$ of the graph with a total of N nodes, these are calculated using matrix multiplications and bias vectors b for a connection from node j to i through

$$\begin{aligned} q_{c,i}^{(l)} &= W_{c,q}^{(l)} f_i^{(l)} + b_{c,q}^{(l)} \\ k_{c,j}^{(l)} &= W_{c,k}^{(l)} f_j^{(l)} + b_{c,k}^{(l)} \quad , \end{aligned} \tag{9.1}$$

where c denotes the current head and l the current layer of the calculation. With this, the feature vector $f_i^{(l)}$ and connected feature vector $f_j^{(l)}$ are transformed into a query vector $q_{c,i}^{(l)} \in \mathbb{R}^d$ and key vector $k_{c,j}^{(l)} \in \mathbb{R}^d$ with dimension d identical in all heads. The transformation is performed through a multiplication of the feature vectors with the matrices $W_{c,q}^{(l)}$, $W_{c,k}^{(l)} \in \mathbb{R}^{d \times E}$ and biases $b_{c,q}^{(l)}$, $b_{c,k}^{(l)} \in \mathbb{R}^d$ which contain trainable parameters of the model. The trainable parameters are optimized during the training procedure. Since each head operates independently, it is possible to learn unique patterns. Multiple heads therefore allow the model to learn various relationships in parallel, enhancing the overall expressiveness and improve the generalization of the model.

The grade of attention, which can be calculated from a scaled dot-product of two vectors in the query-key space, is given by

$$\langle q_{c,i}^{(l)}, k_{c,j}^{(l)} \rangle = \exp\left(\frac{(q_{c,i}^{(l)})^{\mathrm{T}} (k_{c,j}^{(l)})}{\sqrt{d}} \right) \quad . \tag{9.2}$$

This involves an exponential scalar product of the two vectors with a normalization relative to the dimension d of the two vectors. The dimension d of the query and key vectors is smaller than the embedding dimension E of the node features $f_{i,j}^{(l)}$ and is given by $d = E/C$, where C is the number of multi-heads. The property of the scalar product is used to determine the attention. In this way the relevance of the key vector with respect to the query vector is determined based on linear algebra principles. If the two vectors point in the same direction in query-key space, the value of the dot-product is large, whereas orthogonal vectors result in a dot-product of zero. Vectors pointing in opposite directions cause the scalar product to be negative which is interpreted as a low attention, i.e. importance of the key vector for the query vector. The exponential function transforms the scale of this

dot-product. Hence, the more similar the query and key vectors are, the greater the scalar product and the greater the attention of node j to i.

The edge $e_{c,ij}$ between the two nodes under consideration is calculated in the analogous way to the node features in Equation 9.1, but with independent trainable weight matrices $W_{c,e}$ and biases $b_{c,e}$:

$$e_{c,ij} = W_{c,e} e_{ij} + b_{c,e} \quad . \tag{9.3}$$

The embedded edge vector is added to the previously determined key vector $k_{c,j}$ and the result of the calculated attention matrix is

$$\alpha_{c,ij}^{(l)} = \frac{\langle q_{c,i}^{(l)}, k_{c,j}^{(l)} + e_{c,ij} \rangle}{\sum_{u \in \mathcal{N}(i)} \langle q_{c,i}^{(l)}, k_{c,u}^{(l)} + e_{c,iu} \rangle} \quad . \tag{9.4}$$

The applied sum runs over all nodes $\mathcal{N}(i)$ connected to node i to ensure a normalized attention. The calculation of the value vector is done using the weight matrices $W_{c,v}^{(l)}$ with the biases $b_{c,v}^{(l)}$. Both contain further trainable parameters of the model. The value vector $v_{c,j}^{(l)}$ results from

$$v_{c,j}^{(l)} = W_{c,v}^{(l)} f_j^{(l)} + b_{c,v}^{(l)} \quad . \tag{9.5}$$

This value vector together with the edge is then weighted based on the previously determined attention

$$\hat{f}_i^{(l+1)} = \Big\|_{c=1}^{C} \Big[\sum_{j \in \mathcal{N}(i)} \alpha_{c,ij}^{(l)} (v_{c,j}^{(l)} + e_{c,ij}) \Big] \quad , \tag{9.6}$$

where $\|$ is the concatenation function for all heads C that are calculated in parallel. With this step a new direction is calculated that shifts the original vector to a new, optimized position based on the direction of the value vector $v_{c,j}^{(l)}$ and the edge $e_{c,ij}$. This directional translation is weighted with the previously determined attention $\alpha_{c,ij}^{(l)}$.

To prevent the model from oversmoothing, a gated residual connection is implemented which bypasses the self-attention mechanism. For this, an additional residual vector $r_i^{(l)}$ is calculated with

$$r_i^{(l)} = W_r^{(l)} f_i^{(l)} + b_r^{(l)} \tag{9.7}$$

in an analogous way to Equation 9.1, but with independent matrices $W_r^{(l)}$ and biases $b_r^{(l)}$.

The new final node representation results from the gated residuals and the message passing through attention mechanisms as discussed using

$$\begin{aligned} \beta_i^{(l)} &= \text{sigmoid}(W_g^{(l)}[\hat{f}_i^{(l+1)}; r_i^{(l)}; \hat{f}_i^{(l+1)} - r_i^{(l)}]) \\ f_i^{(l+1)} &= \textsc{LeakyReLU}(\text{LayerNorm}((1-\beta_i^{(l)})\hat{f}_i^{(l+1)} + \beta_i^{(l)} r_i^{(l)})) \quad , \end{aligned} \tag{9.8}$$

where $\beta_i^{(l)}$ is determined from the concatenation of the vectors $\hat{f}_i^{(l+1)}$, $r_i^{(l)}$, and $\hat{f}_i^{(l+1)} - r_i^{(l)}$, multiplied with the weight matrix $W_g^{(l)}$, and transformed into a range of $(0, 1)$ with the sigmoid function, as defined in Equation 5.3. The $\beta_i^{(l)}$ vector can be applied to add the gated residual information in a weighted manner to the output value of the multi-head attention mechanism. The application of $\beta_i^{(l)}$ is considered a hyperparameter of the architecture. This involves the application of the sigmoid function to stabilize the numerical calculations. With the LeakyReLU function, a further non-linear function is employed, as defined in Equation 5.2. An important aspect in the design of the architecture is the implementation of a normalization referred to as LayerNorm at the end of each graph transformer block. The first problem that can occur when no normalization is applied is known as internal covariate shift. This occurs especially in deep neural networks where the output from the first layer enters the second, the second enters the third, etc., so that changes in leading layers affect layers further back in the network. From this, two consequences arise. Firstly, without suitable normalization, the training is slowed down since the trailing layers have to adapt continuously. Secondly, the network stability can suffer if gradients become too large or too small and thus almost vanish, which in turn leads to worse generalization. By renormalizing each batch, the internal covariate shift can be prevented. This is called batch normalization (BatchNorm) [174, 175]. The second problem that can occur without performing a normalization is the effect of unwanted oversmoothing. In GNNs, the node embeddings become too similar after several layers, which reduces the differentiation capability of the nodes and renders them almost indistinguishable. As a result, the performance of the model can gradually decrease with the number of layers. In GNNs, this is remedied by normalizing the node embedding in each layer so that differences are retained. This normalization is referred to as pair normalization (PairNorm) [176]. The choice of layer normalization, whether it is BatchNorm or PairNorm, is considered a hyperparameter of the architecture and is part of the hyperparameter optimization process to identify the optimal model for the classification task.

After the completion of multiple graph transformer blocks in the training, the dimension E of the embedded node feature is changed. The dimension of the node vector was previously always kept constant, but is now reduced to one to enable the NLP indicating the additional jet-ness of a node. To decrease the node feature dimension, an MLP is constructed for the embedded node, consisting of multiple layers. In each consecutive hidden layer of the MLP, the dimension of the nodes is gradually reduced by a fixed factor. The output layer of the

Table 9.2: Event yields per process used for training of the NLP model.

Process	No. of events
$t\bar{t}b\bar{b}$	34682
$t\bar{t}b$	50000
$t\bar{t}H(b\bar{b})$	69881
$t\bar{t}H(b)$	43036
$t\bar{t}Z(b\bar{b})$	27998
$t\bar{t}Z(b)$	23290
$t\bar{t}Z(nonB)$	50000
$t\bar{t}c\bar{c}$	57618
$t\bar{t}c$	50000
$t\bar{t}LF$	100000
$\sum_{events}$	485505

MLP comprises one node with a sigmoid activation function. In this way, the resulting scalar per node can be interpreted as a probability prediction generated by the model that a given node mapped to a physics object is an additional jet.

A total of approximately half a million events are used to train the NLP model. Details on the composition of the training sample are given in Table 9.2. Differences in the number of simulated events for each process are compensated via reweightings in the loss function. In all networks, the cross-entropy serves as loss function. The training sample is based on the event simulations in Table 8.1, taking into account different scenarios that can occur at the particle level. For instance, the training sample is composed of $t\bar{t}b\bar{b}$ and $t\bar{t}b$ events, with only one additional b jet in the latter. This is the case if, for example, a jet is out of acceptance. The terminology for $t\bar{t}H(b\bar{b})$ and $t\bar{t}H(b)$, $t\bar{t}Z(b\bar{b})$, and $t\bar{t}Z(b)$, as well as $t\bar{t}c\bar{c}$ and $t\bar{t}c$ is chosen analogously. No additional heavy flavor jets are present in the $t\bar{t}LF$ event case. Details on the differences are described in Section 4.6. The validation and test samples are composed in a similar way, each with a size of 5% of the training sample. Training, validation, and test samples are randomly composed from the event simulation and are orthogonal to each other. The event simulation for the 2018 data-taking era is additionally enriched with the simulation from 2017 for the $t\bar{t}b\bar{b}$, $t\bar{t}H$, and $t\bar{t}Z$ processes in order to increase the training statistics.

The optimal architecture is determined by performing a grid search hyperparameter optimization. As stated before, to avoid choosing a stochastically arbitrary optimum, each architecture is trained for ten randomized weight initializations in parallel. The Receiver Operating Characteristic - Area Under the Curve (ROC AUC) is used as a metric for evaluating a model's capability to identify additional jets. With this, the ROC AUC serves

Table 9.3: Hyperparameter settings for NLP trainings.

Hyperparameter	Setting
Batch size	2048
Loss function (Eq. 5.4)	Binary cross-entropy
Optimizer (cf. Section 5.1)	Adam
Learning rate (cf. Section 5.1)	0.01
Number of graph transformer blocks	8
Normalization	PairNorm
Activation function (Eq. 5.2)	LeakyReLU
Node aggregation	Sum
Node embedding dimension	32
Number of multi-head attentions	5
Multi-head consolidation (concat / average) (Eq. 9.6)	Concat
Gated residual information weighting $\beta_i^{(l)}$ (Eq. 9.8)	False (unweighted)
Dropout probability of normalized attention coefficients	0.2
Additive bias b	True
Root weight (add transformed root node to aggregation)	True
MLP layers	3
MLP embedding dimension	32
MLP embedding dimension reduction factor per layer	0.4
MLP output layer dimension	1

as a single measure of the overall performance, with values closer to one indicating a better performance. The optimal network achieves a ROC AUC of 0.957. The standard deviation of the mean value of all training repetitions with identical architectures is below 0.1%. The hyperparameters of this network are summarized in Table 9.3.

The loss values and the ROC AUC for each epoch during training and validation are shown in Figure 9.6. A total of 43 epochs are required to complete the training. An epoch is defined as a full iteration through the entire training sample. The loss function values decrease rapidly within the first training epochs and the ROC AUC increases steeply as the initialized weights are quickly adjusted. After the first saturation in the loss curve is reached in the 19th epoch, the learning rate (cf. Section 5.1) is scaled down by a factor of 0.01. This adjustment ensures that a local minimum can be better exploited, as evidenced by the subsequent sharp drop in the loss function. Once a new saturation point is reached, the training is stopped early. Overfitting of the model, which would manifest in an increase of the calculated loss value on the validation sample, does not occur. A similar pattern is observed in the ROC AUC performance.

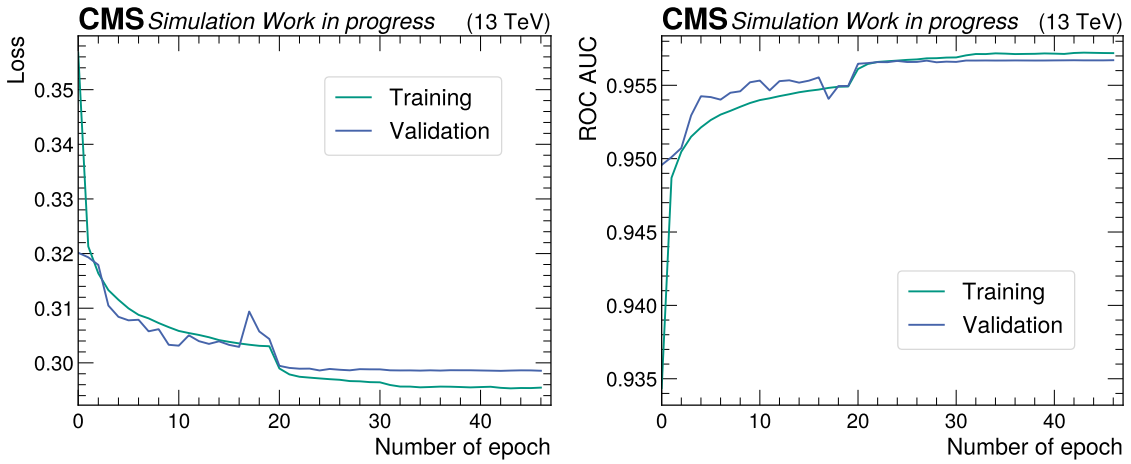


Figure 9.6: Evolution of the graph transformer network for NLP on simulated events: loss function values (left) and ROC AUC performance measure (right) per epoch for training (green) and validation (blue).

The performance of the trained model on the test sample is shown in Figure 9.7, with performance broken down into the event classes of the training. It can be seen that especially the performance in events with additional light and c jets is better compared to events with additional b jets. For example, the performance is best in $t\bar{t}$ events with one or more light jets, followed by $t\bar{t}Z$(nonB) events. The model performance in $t\bar{t}c$ and $t\bar{t}c\bar{c}$ events is worse than the performance in $t\bar{t}$ events with additional light jets, but better compared to $t\bar{t}$ events with additional b jets. This is expected because it is difficult for the network to separate additional jets from b jets originating from the $t\bar{t}$ system, especially if the observables show strong similarities. This is the case if the additional jets are also b jets, resulting in similar b tagging feature values. Additionally, other kinematic variables such as jet p_T are similar, making separation more challenging. In the case of events with additional c or light jets, the distinction between a b jet from the $t\bar{t}$ system and an additional jet can easily be made based on the b tagging information as a powerful feature. The most challenging determination of the additional jets is in $t\bar{t}H(b\bar{b})$ events.

ROC AUC values of over 0.9 are achieved for the events of all processes on the test sample. This is also due to the fact that theoretically possible misclassifications of the model never occur. For example, this includes avoiding the simple confusion of an additional jet with a node representing a lepton. The ability to make this rather simple distinction already leads to high ROC AUC values. Therefore, in the following section the quality of the model is examined under various aspects that go beyond a direct ROC AUC determination after training on the test sample.

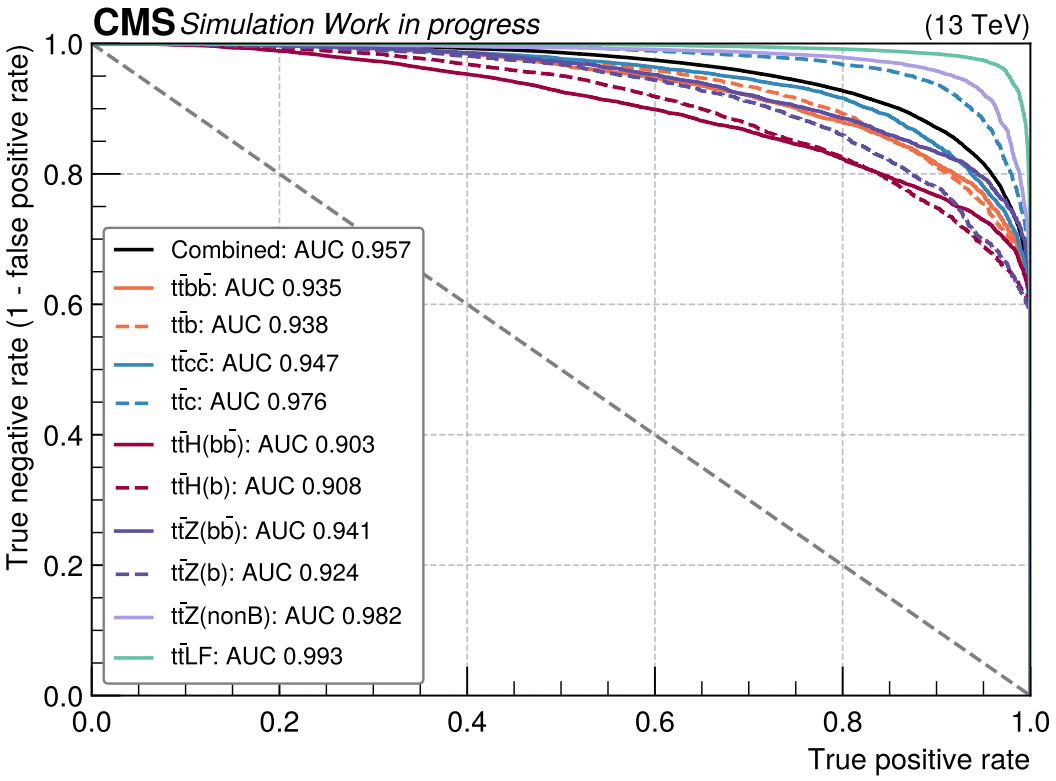


Figure 9.7: Receiver Operating Characteristic (ROC) curves show specificity (true negative rate) over sensitivity (true positive rate). The Area Under the Curve (AUC) serves as a single measure of the overall performance, with values closer to one indicating a better performance. The total performance on the full test sample is indicated by a black line. In addition, the performance is broken down for particular event classes. The dotted gray line represents the lower boundary for a random classifier.

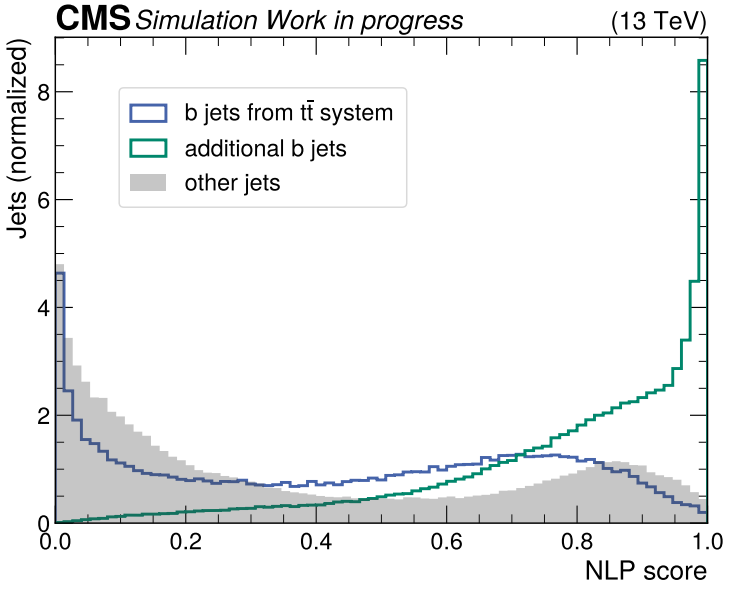


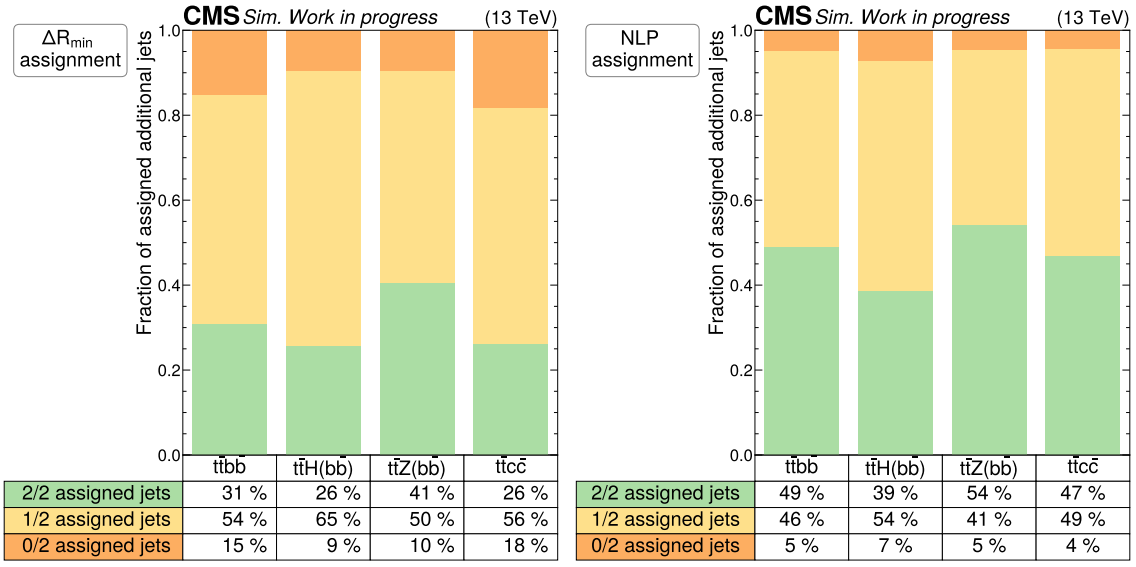
Figure 9.8: Evaluation of the NLP model for jets of $t\bar{t}b\bar{b}$, $t\bar{t}H(b\bar{b})$, and $t\bar{t}Z(b\bar{b})$ events. The NLP score of the binary classifier shows separation between additional jets and b jets from the $t\bar{t}$ system as well as other jets.

9.2.2 Node level prediction evaluation

In this section, the trained model is examined in detail and benchmarked to provide a better understanding of the model performance. With this, three important aspects are assessed. First, the discrimination power of the NLP model to distinguish additional jets from other jets in an event is investigated. Second, it is verified that the model learns a variety of features and does not fail to predict additional jets with particular properties. Third, the method's performance is compared to other established methods to rank its effectiveness.

To assess the separation quality of the NLP model, the output score of the prediction is examined. By design through the sigmoid function, the score ranges from zero to one. The distribution of the NLP scores for the additional jets is shown in Figure 9.8. Additionally, the distribution of the score for all remaining jets is shown, split into b jets from the $t\bar{t}$ system and other jets, which do not belong to either additional jets or jets from the top quark decays. The distribution of the NLP scores shows high purity for scores greater than 0.9. Overall, there is a clear separation between additional jets and b jets from the $t\bar{t}$ system as well as other jets. Only a relatively low number of additional jets result in small NLP scores. At scores between 0.6 and 0.8, the relative number of b jets from the $t\bar{t}$ system increases slightly before dropping again at large scores.

An alternative method for determining the additional jets using a combinatorial approach is to calculate the distances of all b-tagged jets and select those with the minimum distance



	tt̄bb̄	tt̄H(bb̄)	tt̄Z(bb̄)	tt̄cc̄
2/2 assigned jets	31 %	26 %	41 %	26 %
1/2 assigned jets	54 %	65 %	50 %	56 %
0/2 assigned jets	15 %	9 %	10 %	18 %

	tt̄bb̄	tt̄H(bb̄)	tt̄Z(bb̄)	tt̄cc̄
2/2 assigned jets	49 %	39 %	54 %	47 %
1/2 assigned jets	46 %	54 %	41 %	49 %
0/2 assigned jets	5 %	7 %	5 %	4 %

Figure 9.9: Jet identification accuracy for the ΔR_{min} method (left) and the NLP model (right). It is determined how many events have either both additional jets (2/2), one out of two additional jets (1/2), or no additional jets (0/2) identified.

as the additional jets. This procedure is used, for example, in the differential measurement of the $\text{t}\bar{\text{t}}\text{b}\bar{\text{b}}$ process at the CMS experiment [96], and is referred to as the ΔR_{min} method. To compare the accuracy of the ΔR_{min} method with the NLP model, it is determined how many events have either both additional jets, one out of two additional jets, or no additional jets identified. For $\text{t}\bar{\text{t}}\text{b}\bar{\text{b}}$ events using the ΔR_{min} method, both additional jets can be correctly assigned in 31% of all events, one additional jets in 54% of all events, and in the remaining 14% of events the method fails completely with no additional jet being found via the distance determination. The evaluation of the NLP model shows significantly better results: here, two out of two additional jets are found in 49% of the events, one out of two additional jets in 46% of the events and no additional jet is found in only 5% of all events. With this, the NLP model demonstrates a relative improvement in precision of more than 58% in finding both additional jets in $\text{t}\bar{\text{t}}\text{b}\bar{\text{b}}$ events. The detailed performance breakdown is shown in Figure 9.9. The assignment of additional jets works best in $\text{t}\bar{\text{t}}\text{Z}(\text{b}\bar{\text{b}})$ simulated events. In 41% of all events both additional jets can be found with the ΔR_{min} method, whereas 54% of all events can be found with the NLP method. The most difficult distinction between additional jets and other jets is in simulated $\text{t}\bar{\text{t}}\text{H}(\text{b}\bar{\text{b}})$ events compared to simulations of the $\text{t}\bar{\text{t}}\text{b}\bar{\text{b}}$ and $\text{t}\bar{\text{t}}\text{Z}(\text{b}\bar{\text{b}})$ processes. Here, both additional jets are found in 26% of all events with the ΔR_{min} method, while both additional jets are found in 39% of all events with the NLP method. The ΔR_{min} method definition for evaluating the assignment accuracy differs slightly for $\text{t}\bar{\text{t}}\text{c}\bar{\text{c}}$ events. While the distances between all b-tagged jets are calculated in $\text{t}\bar{\text{t}}\text{b}\bar{\text{b}}$, $\text{t}\bar{\text{t}}\text{H}(\text{b}\bar{\text{b}})$, and $\text{t}\bar{\text{t}}\text{Z}(\text{b}\bar{\text{b}})$ events, the distance between all existing jets in the event is calculated in the case of $\text{t}\bar{\text{t}}\text{c}\bar{\text{c}}$ events. Whichever jet combination show the

smallest distance, the two jets are then assigned as additional c jets. This results in the largest difference in the assignment accuracy of two out of two additional jets between both methods. The assignment accuracy of the ΔR_{min} method is 26% while it is 47% with the NLP method. As a result, both true additional c jets are correctly assigned in over 80% more events with the NLP method.

As discussed, the ΔR_{min} method offers an effective way of reconstructing the additional jets, albeit being inferior to the NLP method in terms of accuracy. However, the reduced accuracy is not the sole caveat. Jets determined as additional jets by the ΔR_{min} method possess properties that are not typical of the true additional jets. The distribution of the ΔR_{min} observable for true additional jets exhibits larger values compared to the distances of the additional jets identified by the ΔR_{min} method, i.e. the closest two b jets in an event. This is the case by definition as the two closest jets do not necessarily form the true additional jets. The differences in these distance distributions are shown in Figure 9.10 for $t\bar{t}b\bar{b}$, $t\bar{t}H(b\bar{b})$, and $t\bar{t}Z(b\bar{b})$ events, without breaking down the processes in favor of clarity. Depicted in this figure is the distribution of the true additional jets as well as the distribution of the distances between the two closest b jets in an event and thereby those jets that would be determined by the ΔR_{min} method. In addition, the distribution of the distances between the two additional jets determined by the NLP method is shown. It is evident from the distributions that the ΔR_{min} method does not reflect the true inherent distribution, whereas the NLP method is considerably closer to the true ΔR pattern. This also indicates that the NLP method does not rely excessively on the ΔR observable which is included as an edge weight into the graph structure. Figure 9.10 also depicts the jet pair mass of the additional jet distributions. In this case, the jet pair mass is determined for the true additional jet pairs as well as the additional jet pairs predicted by the NLP model. For the true additional jet pairs, the expected distributions can be identified in the breakdowns for $t\bar{t}b\bar{b}$, $t\bar{t}H(b\bar{b})$, and $t\bar{t}Z(b\bar{b})$ events: a continuous spectrum for $t\bar{t}b\bar{b}$ events, a peak at the Z boson mass for $t\bar{t}Z(b\bar{b})$ events, and a peak at the Higgs boson mass for $t\bar{t}H(b\bar{b})$ events. The equivalent distributions of the additional jets assigned by the NLP model also show this characteristic. However, distinctions are less pronounced and the distributions are skewed towards the average of the three true distributions.

A further important examination is the behavior of the NLP model on data. Since no labels are available for data, the trained NLP model is evaluated on both data and simulation, and the distributions of the predictions are compared. The agreement of the predictions on data and simulated events is shown in Figure 9.11. Generally, it can be seen that all data points are within the statistical and systematic uncertainty of the prediction based on simulated events. Details on the uncertainties are discussed in Chapter 10.

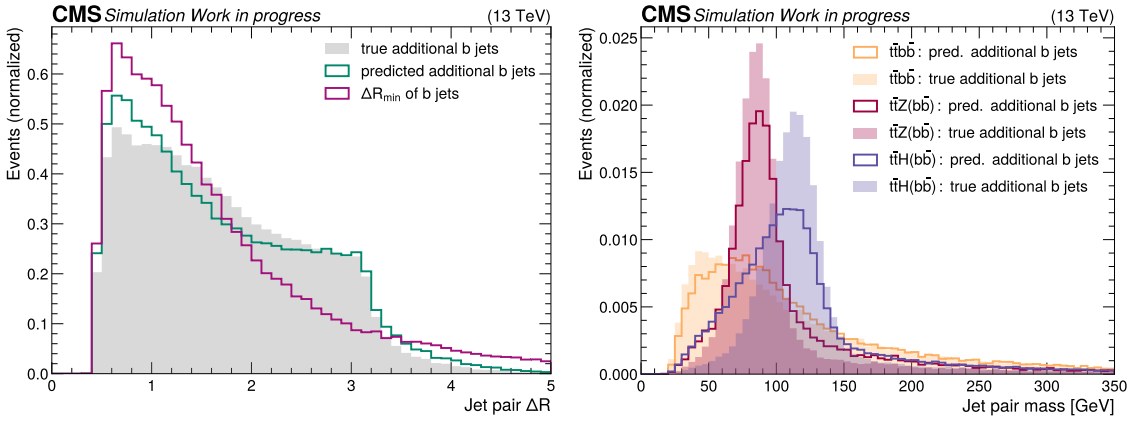


Figure 9.10: Distributions of the distance ΔR between pairs of true additional b jets, predicted additional b jets by the NLP model, and additional b jets determined with the ΔR_{min} method on combined $t\bar{t}b\bar{b}$, $t\bar{t}H(b\bar{b})$, and $t\bar{t}Z(b\bar{b})$ events (left). Distributions of masses of the jet pair combinations for true additional b jets and the additional b jets predicted by the NLP model sub-divided for $t\bar{t}b\bar{b}$, $t\bar{t}H(b\bar{b})$, and $t\bar{t}Z(b\bar{b})$ events (right).

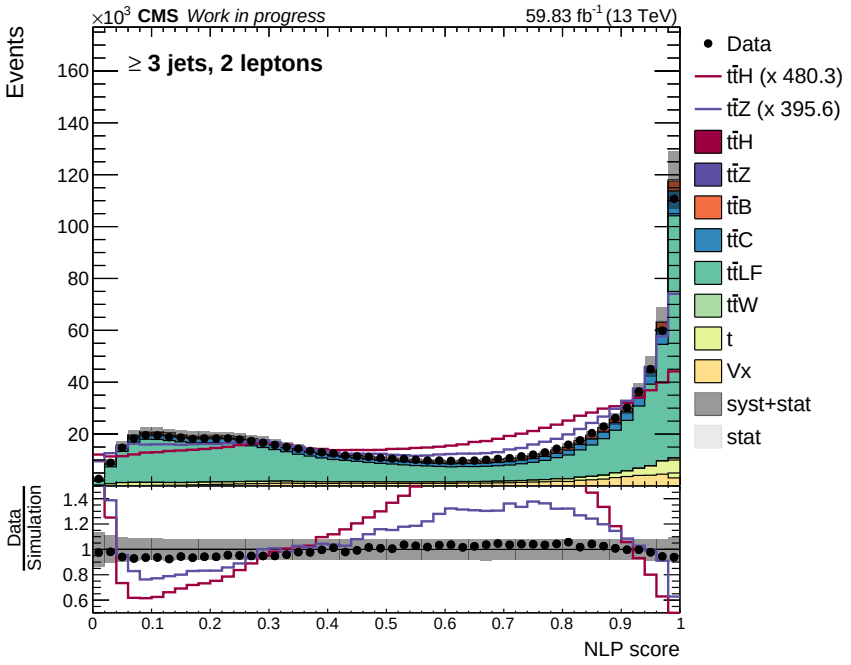


Figure 9.11: Evaluation of the NLP model on the entire phase space. The NLP score distribution for data and event simulation as well as their ratio is displayed. The simulated events of $t\bar{t}H$ and $t\bar{t}Z$ processes are shown both stacked and as a line, scaled to the integral of all events.

9.3 Event classification: Graph level prediction

In this section, the second stage of the classification strategy shown in Figure 9.4 is discussed. This stage involves the classification of events in the analysis phase space based on an additional AI model. The transformation of events into a graph structure as discussed in Section 9.1 is identical to the procedure for jet identification with the NLP model discussed in Section 9.2. However, while the identical event information in the form of observables is translated into features of the graphs for event classification, in this stage the node feature vector is additionally enriched with binary information about each node from the evaluated NLP model. The biggest difference, however, is the type of prediction of the graph. In contrast to the NLP model for jet identification, which predicts individual nodes of the graph, this approach predicts the class of the entire graph for event classification. Compared to NLP, the GLP task is not a binary but a multi-class classification. Also, the training data set must be refined to ensure that the properties of the entire phase space are accurately represented, e.g. including all backgrounds. The details of the GLP architecture and training are explained in the following section.

9.3.1 GLP model and training

The goal of event classification is to build a robust model trained on simulated events that can distinguish events both in simulation and in data. The resulting distributions of evaluated events are then analyzed in the fit model in Chapter 10 to determine the signal strengths of the processes of interest. Similar to the NLPs, the architecture for event classification is generally based on graph transformer blocks, which employ transformers and self-attention mechanisms to perform message passing within the GNN model. However, since neither an NLP is performed nor the classification is binary, the architecture is modified accordingly. Whereas previously the representation of the embedded node features with dimension E was reduced to a scalar with a different MLP architecture, now the entire graph is reduced to a linear, fixed-size representation with an MLP. This flattened representation can then be used to predict the class of the graph and eventually the event.

To flatten the graph representation, a pooling method is employed for downsampling, similar to techniques used in image recognition with convolutional neural networks. For this purpose, the so-called top-k pooling is added to the graph transformer architecture, based on [177–179]. The top-k pooling procedure selects a number of kN nodes from a total of N nodes in the graph that are retained for subsequent processing. All other $N - kN$ nodes and their connections are removed from the graph. Essentially, this procedure consists of two hyperparameters, the fraction $k \in [0, 1)$, and the decision after how many graph transformer blocks the pooling should be executed. In general, it is also possible to drop a fixed number of nodes, but in this thesis a ratio is chosen in order to respond more

flexible to varying node multiplicities. Since a ratio is chosen, the resulting product kN is always rounded to an integer. The decision on which nodes to keep is based on a projection of the node embeddings. Initially, the N nodes of the graph, each with E features, are combined to form a matrix $F^l \in \mathbb{R}^{N \times E}$ in the l-th graph transformer block. This matrix F^l is multiplied by a vector[2] p^l in the l-th layer, which contains trainable parameters and is learned by the model

$$y^l = \frac{F^l p^l}{\|p^l\|} \quad , \tag{9.9}$$

where $\|p^l\|$ denotes the 2-norm of the vector p^l. The projection y^l resulting from the multiplication is used to select the highest k entries from the vector y^l

$$i^l = \text{top-}k\left(y^l, k\right) \quad . \tag{9.10}$$

While y^l is a vector with dimension N, the vector i^l therefore has the dimension kN. To obtain the top-k matching embedded nodes

$$F^{l+1} = \left(F^l \odot \tanh\left(y^l\right)\right)_{i^l} \tag{9.11}$$

is used to calculate a new representation of the remaining nodes passed to the subsequent layer $l+1$. Here, $\odot$ is the element-wise product and with the tanh function, a non-linearity is incorporated into the calculation. The operation $(\,)_{i^l}$ selects slices according to the given indices by the vector i^l. The dimension of node embedding is now reduced to $F^{l+1} \in \mathbb{R}^{kN \times E}$ compared to $F^l \in \mathbb{R}^{N \times E}$ before.

After executing a total of L_p pooling layers, a readout layer is designed. This readout layer is intended as a flattening layer of a fixed size to enable a later processing of the encoded graph representation by an MLP. For this purpose, the node representations arising in each pooling operation are summarized, which contain refined information of the graph based on the optimization steps during training. First, the sum of all node representations after the top-k selection is added up in each pooling operation performed, concatenated with the maximum

$$s^{(l)} = \left(\frac{1}{N^{(l)}} \sum_{i=1}^{N^{(l)}} f_i^{(l)}\right) \Big\| \max_{i=1}^{N^{(l)}} f_i^{(l)} \quad , \tag{9.12}$$

where $f_i^{(l)}$ is the i-th node's feature vector with a total of $N^{(l)}$ remaining nodes of the graph in layer l. The concatenation is denoted by the $\|$ symbol. According to Ref. [178],

[2] For improved readability, explicit vector notation is omitted.

Table 9.4: Hyperparameter settings for GLP trainings.

Hyperparameter	Setting
Batch size	2048
Optimizer (cf. Section 5.1)	Adam
Learning rate (cf. Section 5.1)	0.01
Loss function (Eq. 5.4)	Cross-entropy
Number of graph transformer blocks	4
Normalization	BatchNorm
Activation function (Eq. 5.2)	LeakyReLU
Node aggregation	Sum
Node embedding dimension	32
Number of multi-head attentions	5
Multi-head consolidation (concat / average) (Eq. 9.6)	Average
top-k	0.5
Number of poolings	2
Gated residual information weighting $\beta_i^{(l)}$ (Eq. 9.8)	False (unweighted)
Dropout probability of normalized attention coefficients	0.5
Additive bias b	True
Root weight (weight of node relative to aggregation)	True
MLP layers	3
MLP embedding dimension	64
MLP embedding dimension reduction factor	0.4
MLP output layer dimension	7

the concatenation of the mean and maximum of $f_i^{(l)}$ improves the resulting representations. Finally, the sum of all sums is calculated

$$s = \sum_{l=1}^{L_p} s^{(l)} \quad , \tag{9.13}$$

which represents an average of all learned node embeddings and therefore a representation of the initial graph with a fixed size E. Either directly or via linear transformation, the resulting vector s serves as an input layer for an MLP. The choice of the embedding dimension for the MLP as well as the number of layers and thus the size of the MLP are further hyperparameters of the architecture. All hyperparameters of the GLP architecture are summarized in Table 9.4.

The pooling operations and the associated MLP are shown in Figure 9.12. The output layer of the MLP consists of a total of seven nodes with sigmoid as the activation function

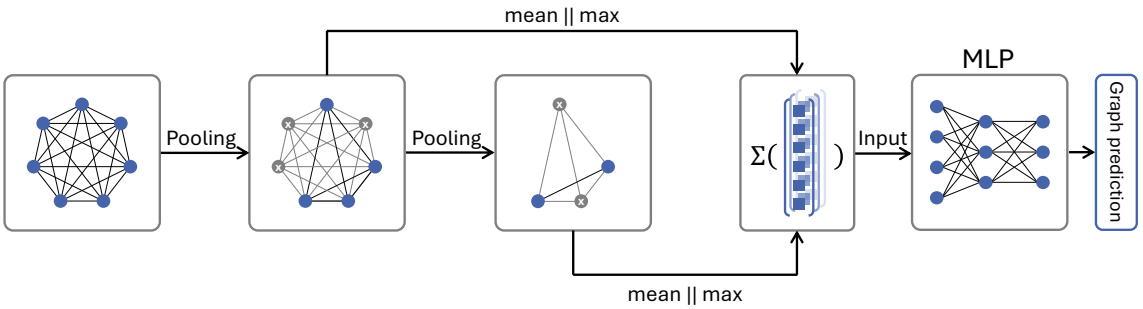


Figure 9.12: Additional pooling operations in the graph transformer network for GLP. In each pooling step, embedded node representations and their connections are removed from the graph, creating new graph representations through downsampling. For the flattening operation, averages are calculated and concatenated with the maximum of all learned node embeddings. All resulting representations are summed up to form an input vector, which is transferred to an MLP with a linear transformation. The MLP performs a multi-class classification of the initial graph. Adapted from [178].

in order to obtain a probability estimate from the score of a node that the event in question belongs to the respective class.

A total of seven different classes are defined for the training. The event class definition is based on the simulated process and the sub-category as defined in Table 8.1, along with the presence of additional jets and their flavors. The seven classes consist of $t\bar{t}B$, $t\bar{t}H(B)$, $t\bar{t}Z(B)$, $t\bar{t}Z(nonB)$, $t\bar{t}C$, $t\bar{t}LF$, and Other. The $t\bar{t}B$ class consists of simulated dileptonic $t\bar{t}B$ events with one or two additional b jets. Accordingly, the $t\bar{t}H(B)$ ($t\bar{t}Z(B)$) class is defined as all simulated $t\bar{t}H$ ($t\bar{t}Z$) events with a subsequent decay of $H \to b\bar{b}$ ($Z \to b\bar{b}$), which can lead to one or two additional b jets. All events of simulated $t\bar{t}Z$ processes that do not end in $Z \to b\bar{b}$ final states are defined as a separate class, which is referred to as $t\bar{t}Z(nonB)$. The $t\bar{t}C$ class includes all events that do not match the $t\bar{t}B$ class but have one or two additional c jets. If no additional b or c jets are present in $t\bar{t}$+jets simulated events, they belong to the $t\bar{t}LF$ class. All events that do not fit into the preceding classes belong to the Other class. These include, for example, other $t\bar{t}$+jets decays such as the semileptonic decay channel, but most importantly all backgrounds in the phase space.

A total of approximately 642 000 simulated events are used for the training of the GLP model. As with the NLP training sample, the $t\bar{t}H$ and $t\bar{t}Z$ simulated events are enriched from event simulations of the 2017 era. The individual classes' shares can be seen in Table 9.6. Since not all classes have an equal number of events, the loss function is weighted according to the proportions so that all classes contribute equally to the loss without inducing a bias. Reweighting generally allows for greater differences in the number of events in individual classes. Particularly for $t\bar{t}B$, $t\bar{t}C$, and especially $t\bar{t}LF$ events, the

Table 9.5: Training class definitions based on the simulated processes and sub-categories in Table 8.1. The presence of additional jets based on hadron matching is the decisive criterion. Cases that cannot occur are indicated with "−". Cases that are not considered for training are marked with "∘".

Process [Sub-category]	Additional jet criteria			
	≥ 1 add. b jets	no add. b jets ≥ 1 add. c jets	no add. b jets no add. c jets	None
$t\bar{t}$+jets [dilepton]	∘	$t\bar{t}C$	$t\bar{t}LF$	−
$t\bar{t}b\bar{b}$ [dilepton]	$t\bar{t}B$	∘	∘	−
$t\bar{t}H$ [$H \to b\bar{b}$]	$t\bar{t}H$	other	other	−
$t\bar{t}Z$ [$Z \to q\bar{q}$]	$t\bar{t}Z(B)$	$t\bar{t}Z(nonB)$	$t\bar{t}Z(nonB)$	−
all other	−	−	−	Other

Table 9.6: Event yields per class used for training of the GLP model.

Class	No. of events
$t\bar{t}B$	100000
$t\bar{t}H$	83428
$t\bar{t}Z(B)$	75326
$t\bar{t}Z(nonB)$	100313
$t\bar{t}C$	100000
$t\bar{t}LF$	100000
other	88278
$\sum_{events}$	642045

event simulations yield far higher numbers of simulated events and thus candidates for training. This can provide a more comprehensive description of these processes for the trained model and thus increases the variability of the training sample. However, trainings on samples with a much higher number of simulated $t\bar{t}$+jets events, resulting in a greater inequality of the class ratios in the overall training sample, do not show any performance improvements. In most cases, the performance results are even worse, despite the fact that the enlarged imbalance in the training sample was compensated by reweighting the loss function. With the training sample composition in Table 9.6, an optimum in the performance is achieved.

The loss and ROC AUC trends on the training and validation sample for the individual training epochs are shown in Figure 9.13. After epoch 19, the learning rate is adjusted and saturation sets in, where the loss values barely change and the ROC AUC values do not

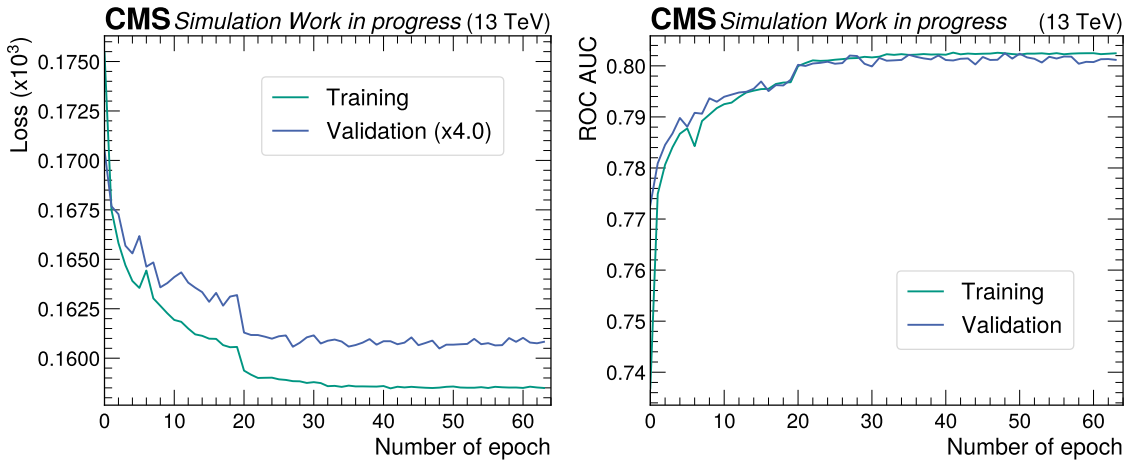


Figure 9.13: Evolution of the graph transformer network for GLP on simulated events: loss function values (left) and ROC AUC performance measure (right) per epoch for training (green) and validation (blue). The loss values of the validation are scaled by a factor of four due to dropout and to improve the comparability.

increase any further. The evaluation on the validation set shows slightly more fluctuations than the evaluation of the training, but no overfitting can be recognized. The average ROC AUC score is 0.805 ± 0.001.

After training, the trained GLP model is evaluated on the test sample. For this purpose, each event is assigned to the class whose associated MLP node predicts the highest GLP score. With the help of a confusion matrix, it is then possible to identify which confusions occur more frequently, or which predictions work particularly well. For better interpretability, the confusion matrix is normalized to its respective class shares and visualized in two versions. One version refers to the efficiency, indicating how a true class is distributed among the different predicted classes. The other version illustrates the purity, which shows the proportion of true classes that actually constitute a predicted class. Both versions can be seen in Figure 9.14. It should be noted that these confusion matrices reflect equal class sizes, which means that migrations by the GLP model in a general fashion can be observed. This allows a global evaluation of the performance of the GLP. However, the confusion matrices do not consider event weights, which are used to describe the expected event yields and contributions in the phase space. The physically expected event yields, for example in the $t\bar{t}H$ node of the GLP, can therefore not be directly deduced and are examined in more detail in Section 9.3.2 and Chapter 10. The expected yields are then crucial for the physics model. Overall, the diagonality in the confusion matrix is clearly pronounced. In addition, the separation of the phase space is recognizable through block-diagonal elements. A b flavor enriched phase space block consisting of events in the $t\bar{t}B$, $t\bar{t}H$, and $t\bar{t}Z(B)$ classes, as well as a separated phase space for the remaining classes, is visible in both versions of the confusion matrix. Although migrations between these phase

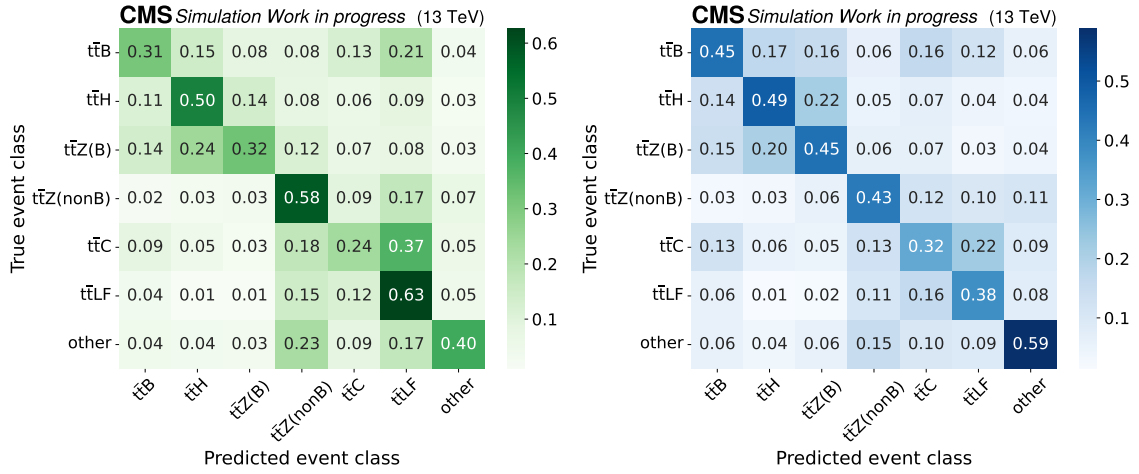


Figure 9.14: Confusion matrices for the GLP. The true class is placed on the vertical axis and the predicted class on the horizontal axis. Both confusion matrices contain the same underlying pattern, but with a different normalization. The green confusion matrix shows the prediction *efficiency* of each event class; the sum of each row is equal to one. The blue confusion matrix shows the *purity* of each event class; the sum of each column is equal to one.

space regions generally occur, the migrations are suppressed. The b jet tagging information is available to the model via the DeepJet b tagging score observable implemented as one of the node features. Furthermore, some confusions are pronounced, e.g. $\mathrm{t\bar{t}Z(B)}$ events ending up in the $\mathrm{t\bar{t}H}$ class. Also, the confusion between $\mathrm{t\bar{t}C}$ and $\mathrm{t\bar{t}LF}$ events is more prominent than with other predictions. A possible method of increasing the focus on events that are more difficult to classify is to reweight the loss function. For this purpose, the cross-entropy loss function is modified by an additional term $(1 - p_t)^\gamma$, so that the new loss function results in

$$\mathrm{FL}\left(p_t\right) = -\left(1 - p_t\right)^\gamma \log\left(p_t\right) \quad , \tag{9.14}$$

where p_t is the predicted probability of the true class and γ is a tunable focusing parameter with $\gamma \geq 0$. This modified loss function is referred to as focal loss [180]. However, applying a focal loss with various focusing parameters does not increase the performance of the GLP model. The confusion can also be reduced with the help of the two-dimensional heavy flavor tagger information CvB and CvL; however, as discussed in Section 9.1, these observables are not taken into account in the final model due to the failure of the goodness of fit test of the input features. Trainings conducted with this information included in the node feature vector show reduced confusion and improved performance.

9.3.2 GLP evaluation

The evaluation of the GLP model on data and simulated events across the entire phase space including the consideration of the event weights is shown in Table 9.7. Here, the expected event yields of each simulated process from Table 8.1 are categorized into the defined classes of the GLP as defined in Table 9.5. An event is assigned to the class whose corresponding node has the highest GLP score. Each class is predominantly composed of $t\bar{t}$LF events which contribute the largest relative shares. Given that $t\bar{t}$LF events are the most significant contribution in the phase space of this analysis, this behavior is anticipated. However, it should be noted that Table 9.7 does not provide any differential information. For instance, even if a comparatively large fraction of $t\bar{t}$LF events dominates the $t\bar{t}$H class, it is expected that the distribution of the GLP score differs for $t\bar{t}$LF and $t\bar{t}$H events within this class.

In the $t\bar{t}$B class it can be seen that the $t\bar{t}$B event yield is enriched. The signal (S) over background (B) ratio is 0.33, where $t\bar{t}$B is defined as the signal in this class and B as all remaining events. The $S/\sqrt{B}$ ratio, which is an estimate of the significance, is 30. This estimate is chosen since a Poisson distribution is assumed. This implies that the number of background events for a background-only hypothesis B corresponds to the expectation value B in a Poisson distribution with a standard deviation $\sqrt{B}$. The signal excess normalized to the uncertainty of the background expectation is then an approximate indicator of the significance. In the $t\bar{t}$H class, $t\bar{t}$H events are enriched but not dominant due to the higher cross sections of other processes. This class, where the $t\bar{t}$H events are defined as a signal, exhibits an S/B ratio of 0.2 and the estimated significance is approximately 1.3. The proportion of $t\bar{t}$Z events defined as signal in the $t\bar{t}$Z(B) class normalized to the background is approximately 0.01 and is therefore relatively small. Expressed in absolute figures, the number of $t\bar{t}$Z events in this class is even smaller compared to the migrated $t\bar{t}$Z events in the $t\bar{t}$H class. The estimated significance of $t\bar{t}$Z in the $t\bar{t}$Z(B) class is 0.57. In the $t\bar{t}$C class, the $t\bar{t}$C events are defined as signal and the $t\bar{t}$C events are significantly enriched. The $t\bar{t}$C events in this class reveal a S/B ratio of 0.18 and a significance estimation of 35. Furthermore, Table 9.7 shows the ratio of data to Monte Carlo event simulation (MC). It can be seen that the data/MC ratio for the heavy flavor classes $t\bar{t}$B, $t\bar{t}$H, $t\bar{t}$Z(B), and $t\bar{t}$C is close to one, while the light flavor classes $t\bar{t}$Z(nonB) and $t\bar{t}$LF as well as the Other class containing remaining backgrounds deviate further from the expectation.

Another test of the GLP model is the behavior of the model on events used for optimization in the training process (seen events) compared to events that were withheld from the training procedure (unseen events). For this test, despite even/odd event ID splitting, the model trained on events with even event IDs is evaluated on events with even and odd event IDs. The resulting distributions of all output node scores of the multi-class classifier are then analyzed. The expected result is a model prediction that does not differ between events that were seen during training and events that were not seen. Figure 9.15

Table 9.7: Event yield breakdown of each process for each class of the GLP model. Due to the multi-signal definition, the S/B and $S/\sqrt{B}$ ratios refer to the respective signal in the dedicated column. For example, the signal in the $t\bar{t}B$ column is defined as $t\bar{t}B$, while the signal in the $t\bar{t}H$ column is defined as $t\bar{t}H$. The backgrounds are defined as all remaining processes in the respective column.

	$t\bar{t}B$	$t\bar{t}H$	$t\bar{t}Z$(B)	$t\bar{t}Z$(nonB)	$t\bar{t}C$	$t\bar{t}LF$	Other	Total
Data	13999	7472	4208	38799	46826	114675	40159	266138
$t\bar{t}B$	3385	1944	1040	1365	1925	2362	619	12638
$t\bar{t}H$	33	115	32	126	49	54	54	461
$t\bar{t}Z$	33	47	37	235	66	98	84	599
$t\bar{t}C$	1914	1284	631	5129	6942	8987	1811	26698
$t\bar{t}LF$	6046	3461	1782	32828	33129	104112	11821	193178
$t\bar{t}W$	27	38	25	225	82	111	87	594
single top	2066	848	610	2448	2800	5814	2729	17316
VV	13	5	10	195	100	115	841	1279
W+jets	4	1	6	100	37	40	296	485
Z+jets	193	39	30	804	743	1192	14837	17838
Data/MC	1.02	0.96	1.00	0.89	1.02	0.93	1.21	0.98
S/B	0.33	0.02	0.01	-	0.18	-	-	-
$S/\sqrt{B}$	30	1.3	0.57	-	35	-	-	-

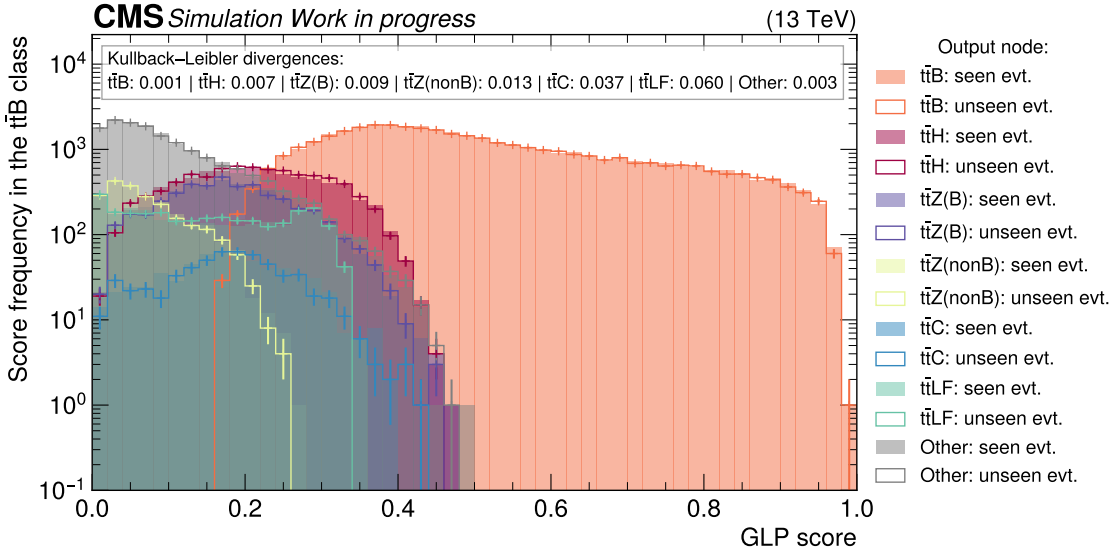


Figure 9.15: Evaluation of the GLP model on events seen during the training versus events not seen during the training (evt.). The distributions of GLP scores on events assigned to the tt̄B class are shown for each individual node of the multi-class classifier. The KL divergence for the distributions of the GLP node scores for seen (filled histograms) and unseen events (enveloping distribution) is calculated and given at the top.

shows this behavior. All events classified as tt̄B, i.e. those whose highest score is the tt̄B node score of the GLP, are shown in this plot. However, not only the score of the tt̄B node is indicated, but all node scores of the multi-class classifier are shown. It can be noted how the distribution of node scores of seen events (filled histograms) corresponds to the distribution of scores of unseen events (enveloping distribution). Only a few bins are outside the statistical uncertainty. In addition, the Kullback-Leibler (KL) divergences are calculated and stated for each node. The KL divergence measures the disparity between two probability distributions, where smaller values signify a stronger similarity between them. The results show that the GLP model behaves, within the statistical uncertainty, for events seen in training in the same way as for unseen events. The distributions for simulated events that fall into the remaining classes are shown in Appendix B.

9.3.3 GLP performance boost through NLP

In order to quantify the extent to which the GLP architecture benefits from the NLP pre-classifier, different scenarios are examined. A fundamental benchmark of the GLP model's performance discussed in the previous section is a comparison with an equivalent model that does not share the NLP information. For this purpose, the additional jet prediction score is removed from the node feature vectors, while the embedding dimension E is retained in order to maintain the same number of trainable parameters in both models.

In addition, all other hyperparameters also remain fixed so that both GLP models possess an identical architecture. The mean ROC AUC value achieved for the model that does not use the NLP pre-classifier information is 0.793 ± 0.002. In contrast, the model that utilizes the NLP score and is described in the previous section achieves a score of 0.805 ± 0.001. To identify a theoretical maximum performance, another GLP model of the same architecture is constructed, which directly obtains the truth information about the additional jets instead of the NLP scores. The true information in this procedure is transferred to the node feature vector as 0 or 1, depending on whether a node is an additional jet or not. With this, the GLP model is equivalent to a model that retrieves information from a perfect pre-classifier. The GLP model achieves a mean ROC AUC performance of 0.861 ± 0.001. Beyond these, multiple GLP models are constructed that reflect the behavior of GLP models between the two described extrema of no information versus full truth information about the additional jets as input features. For this purpose, the true information is injected into the GLP models at different levels and the performance is measured. The level of true information in all events is artificially increased from zero in 10% increments up to a maximum of 100%. For example, the label "GLP training with 20% truth information" denotes that the true additional jet information is correctly encoded in a random 20% of the events in the training sample, while in the other 80% of the events the true flags are randomly permuted among all jets nodes. In the latter, this information is therefore pure noise. Due to the randomized construction of events containing true information and events with noise in this node feature, there are no distinguishing characteristics between events with truth information and noise. The performance of all models is shown in Figure 9.16. It can be seen how the performance increases more strongly with the level of true information and the performance gaps between two models increase. The GLP model featuring the NLP pre-classifier information is ranked between the models with 30% and 40% truth information. It can also be seen that the GLP model with the NLP pre-classifier demonstrates the smallest standard deviation of the mean ROC AUC of all models.

In summary, this chapter provides a detailed examination of a two-stage training procedure. The first stage is to search for the additional jets that do not originate from the $t\bar{t}$ system, as their identification significantly improves the accuracy of the subsequent event classification. In the second stage, a multi-class classification is performed categorizing events into signal and background classes. The trained GLP model, based on transformer architectures with attention mechanisms, is utilized to evaluate the entire phase space of the analysis. The classified events are then used to measure the signal strength parameters and their significances of the $t\bar{t}B$, $t\bar{t}H$, $t\bar{t}Z$, and $t\bar{t}C$ processes are discussed in the following chapter.

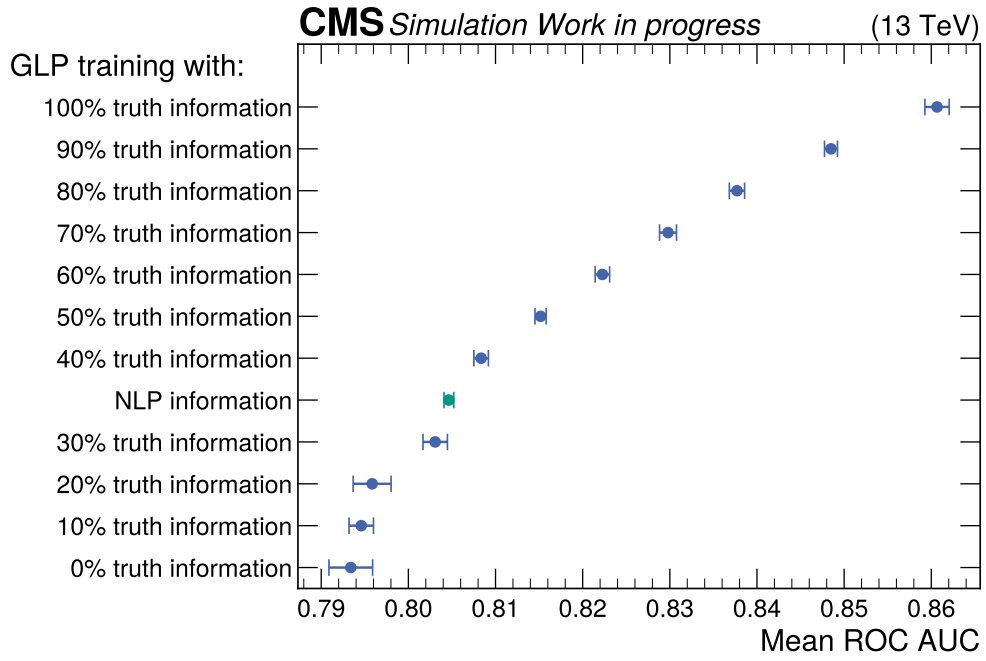


Figure 9.16: Mean performance scores of identical GLP architectures with different levels of information about the additional jets (blue dots). The extent of true information about the additional jets is increased from 0 in 10% increments up to a maximum of 100%. This means that in x% of all training events the true information about the additional is passed to the model, while in $1 - x$% this information is smeared. The GLP model which employs the information from the NLP pre-classifier is also displayed (green dot).

10 Statistical model and inference

In this chapter, the construction of the statistical model is tested and discussed in detail. The statistical model is a representation of the underlying theory for statistical inference of physics parameters. It should not be confused with the Artificial Intelligence (AI) models presented in Chapter 9, which perform jet and event classification using graph neural networks (GNNs). In this chapter, the theoretical methods of statistical data analysis introduced in Chapter 6 are applied to the analysis objectives outlined in Chapter 7. In Section 10.1 the systematic uncertainties that enter the fit model as nuisance parameters (NPs) in the maximum likelihood fit are described. Section 10.2 covers the fit model and its detailed tests on pseudodata. Finally, in Section 10.3 the fit model is applied to data and the signal strength parameters, uncertainties, correlations, and dependencies are presented.

10.1 Systematic uncertainties

Systematic uncertainties of different sources are considered in this thesis as NP in the maximum likelihood fit to binned distributions of observables. The concept is introduced in Section 6.2. Most systematic uncertainties affect the distribution of the individual processes and are therefore referred to as ‘‘shape’’ uncertainties. However, some sources of systematic uncertainties are employed as ‘‘rate’’ uncertainties, as they only alter the yield of the processes. The sources of systematic uncertainties are divided into two categories: experimental uncertainties and theoretical uncertainties. Experimental uncertainties result from imperfect experiments and are estimated from dedicated auxiliary measurements, some of which are discussed in the context of event corrections in Section 8.4. The theoretical uncertainties, on the other hand, stem from known inaccuracies in the calculation of the individual processes, e.g. due to the choice of scales and settings in event simulation, as discussed in the following.

10.1.1 Sources of systematic uncertainties from experiments

Luminosity – The integrated luminosity of the Large Hadron Collider (LHC) in the 2018 data-taking era is determined in auxiliary measurements at the Compact Muon Solenoid (CMS) experiment. Based on van der Meer scans, the uncertainty of the luminosity is determined to be 2.5% [181]. This systematic uncertainty is applied to all processes considered and is implemented as a correlated, rate-changing effect.

Pileup reweighting – As discussed in Section 8.4.1, the number of pileup events reveals a discrepancy between recorded data and simulated events and is therefore corrected. The systematic uncertainty is determined by a 4.6% variation of the inelastic proton-proton cross section in simulation. The resulting changes are propagated to all fit distributions and treated as shape variation in the fit. The uncertainty is considered correlated across all simulated processes.

Level-1 (L1) trigger pre-fire correction – As described in Section 8.4.2, muon candidates can be associated to the wrong proton bunch in the 2018 data-taking era. This results in incorrect firing of the L1 trigger. The total uncertainty on this correction is determined by summing the statistical uncertainty of the pre-fire efficiency and a flat 20% relative uncertainty in quadrature. The uncertainty is considered correlated between all simulated processes.

High-level trigger (HLT) efficiencies – As the HLT efficiencies in data and simulated events vary, the deviations are corrected using scale factors (SFs) as described in Section 8.4.3. The uncertainties on the SFs, which are composed of several sources, are also determined during the efficiency calculations in data and simulation. These include trigger correlations, phase space dependencies, run era dependencies, and statistical uncertainties, with the latter accounting for the largest contribution [163, 164]. The SFs including the total uncertainties are shown in Figures 8.3 and 8.4. The uncertainties are treated as correlated across all simulated processes.

Lepton efficiencies – The measurement of the lepton efficiencies in terms of tracking, reconstruction, identification, and isolation of leptons is described in Section 8.4.4 for electrons and in Section 8.4.5 for muons. Both the corrections and uncertainties are relatively small compared to other uncertainties discussed in this thesis. Nonetheless, all systematic uncertainties are included and propagated to the final discriminant. Even though these systematic uncertainties are almost negligible, each uncertainty is added individually as an NP instead of summarizing all lepton uncertainties. The uncertainties are treated as correlated across all simulated processes.

Jet energy corrections (JECs) – The JECs, consisting of the jet energy scale (JES) and jet energy resolution (JER), are calibrated as described in Section 8.4.6. Several systematic uncertainties for the jet energy are considered in the statistical model, which

originate from different sources and take different aspects of the JES into account. As there are numerous JES uncertainties, but not all are relevant for this analysis, only applicable groups of JES uncertainties are considered, following the recommendations of the CMS Collaboration [182]. The uncertainties in the JECs are evaluated by shifting the reconstructed jet energies by the amount defined as one standard deviation of the JES uncertainty. The event is then re-analyzed to assess the change of these shifts on the jet observables. This process can cause some events to no longer satisfy the selection criteria, while others, previously excluded, may enter the analysis phase space. The latter occurs when the required p_{T} threshold of a jet is exceeded and the jet multiplicity selection criteria are fulfilled.

Two sources of systematic uncertainties arise from the calibration of the absolute JES determination, referred to as ''absolute'' and ''absolute 2018'', with the latter being a specific uncertainty for the 2018 data set studied in this thesis. An additional source of uncertainty describes the difference in calibration fits between the p_{T}-balance and log-linear fits of the missing transverse energy (MET) projection fraction methods, designated as ''relative balance''. An additional systematic uncertainty is added to account for a calibration of discrepancies in jets with varying flavor, referred to as ''flavor QCD''. An η-dependent source of uncertainty describes the relative JES calibration between relative residuals observed in di-jet, Z+jets, and γ+jets systems between various simulations. This uncertainty is referred to as ''relative sample''. Two systematic uncertainty sources describe the residual uncertainties of the calibration, designated as ''BBEC1'' and ''BBEC1 2018'' [182]. Due to a malfunctioning hadronic calorimeter (HCAL) module in the $-2.5 < \eta < -1.3$ and $-1.57 < \phi < -0.87$ detector region (called ''HEM issue''), an additional systematic uncertainty with a jet energy variation of 20% is applied to all jets in this region. All uncertainties are propagated to the final distributions and considered as shape-changing effects. The JER is corrected as described in Section 8.4.6 and an uncertainty is calculated. This uncertainty is obtained by increasing and decreasing the difference between reconstructed and particle-level jets to reflect variations of resolution effects measured in data. The resulting uncertainty is treated identically to the JES described before. All uncertainties are treated as correlated across all simulated processes.

Heavy flavor jet tagging efficiencies – In this analysis, the full b tagging discriminant is corrected using the iterative Fit (itFit) method, as described in Section 8.4.8. The itFit method also determines a set of uncertainties accounting for various sources of systematic uncertainties. The uncertainties are applied to jets in simulated events based on their flavor. One source of uncertainty characterizes the light flavor contamination in the sample to derive the itFit SFs for the heavy flavor jets, while another one describes the contamination from b and c jets in the light flavor region. Two NPs for systematic uncertainties for heavy flavor jets and another two NPs for uncertainties of light flavor jets are employed to reflect the statistical uncertainty associated with the sample size used to derive the itFit SFs.

The statistical uncertainties are determined by multiplying linear and quadratic functions to the statistical uncertainty for each bin in the b tag discriminant. These uncertainties are applied to both light flavor and b jets. For c jets, two additional, separate sources of systematic uncertainties apply. All uncertainties are treated as correlated across all simulated processes.

10.1.2 Sources of systematic uncertainties from theory

Renormalization and factorization scales – The choice of the dynamic μ_R and μ_F scales in the event generator is based on theoretical considerations and experimental observations. As discussed in Section 7.2, different scales are chosen for simulated events in the various analyses. As the optimal choice of the renormalization and factorization scales and their pre-factors is not unambiguous, both scale choices are halved and doubled independently of each other. All scale variations are included in the matrix element (ME) calculations and stored in the form of a weight in each event, which is propagated to the final maximum likelihood fit distributions. Each variation is assigned a different NP for each of the processes $t\bar{t}B$, $t\bar{t}H$, $t\bar{t}Z$, $t\bar{t}C$, and $t\bar{t}LF$. The rate changing effects of all processes are considered separately, as explained in the following. The uncertainties are considered uncorrelated.

Quantum chromodynamics (QCD) rate effects – The inclusive cross sections are derived from theory calculations at Next-to-leading order (NLO) or higher accuracy in QCD. The calculations in QCD perturbation theory depend on the scale of μ_R and μ_F. As mentioned before, the rate effect has been removed for all processes. A rate uncertainty of $\pm 34.1\%$ ($\pm 11.9\%$) for the $t\bar{t}B$ ($t\bar{t}C$) processes is determined in simulation and included as log-normal rate-changing effect. The uncertainties from theory calculations at (N)NLO accuracy for all other processes are: $+5.8\%/-9.2\%$ for $t\bar{t}H$ processes, $+8.1\%/-9.3\%$ for $t\bar{t}Z$ processes, $+2.5\%/-3.6\%$ for $t\bar{t}LF$ processes, $+3.1\%/-2.1\%$ for single top processes, $+25.5\%/-16.4$ for $t\bar{t}W$ processes, $\pm 3.1\%$ for VV processes, $\pm 3.8\%$ for W+jets processes, and $\pm 2\%$ for Z+jets processes. The values are based on calculations by specialized groups within LHC working groups and Refs. [123, 183]. The systematic uncertainties are considered correlated among the Vx processes (VV, W+jets, and Z+jets), as well as between $t\bar{t}LF$ and $t\bar{t}W$, while being uncorrelated for all other processes.

Parton distribution function (PDF) shape – The inner structure of the colliding protons is described by the NNPDF3.1 PDF set at Next-to-next-to-leading order (NNLO) accuracy. The PDF set is determined through a global fit of multiple measurements, including results from D0, LHCb, ATLAS, and CMS [60]. Numerous variations in the fits of the PDF set are provided as residuals by the CMS Collaboration. Since the individual variations are small, they are summarized into an envelope. The rate effect is determined separately and is discussed in the subsequent uncertainty. As the type of residuals differs in the four flavor scheme (4FS) compared to the five flavor scheme (5FS) PDF sets, they

are treated differently. In the case of the 4FS PDF set, the final shape is determined by the root-mean-square of all individual variations, while in the 5FS PDF set the quadratic sum is calculated from all residuals corresponding to the eigenvectors of the hessian matrices. Additionally, a variation of the strong coupling constant α_S is incorporated into the resulting envelope. This procedure follows the recommendations in Ref. [184]. The uncertainties are considered correlated between the $\mathrm{t\bar{t}H}$, $\mathrm{t\bar{t}Z}$, $\mathrm{t\bar{t}C}$, $\mathrm{t\bar{t}LF}$, and $\mathrm{t\bar{t}W}$ processes and uncorrelated from the $\mathrm{t\bar{t}B}$ processes.

PDF rate – As described above, the shape and the rate effects of PDF uncertainties are separated, as the latter do not originate from the simulations but from dedicated calculations. The only exceptions are the $\mathrm{t\bar{t}B}$ and $\mathrm{t\bar{t}C}$ processes, as no precise calculations exist for these cases. In these two cases, the rate change is characterized by the variation in simulation. The theory calculations at (N)NLO accuracy for all other processes are as follows: $\pm 3.6\%$ for $\mathrm{t\bar{t}H}$ processes, $\pm 3.5\%$ for $\mathrm{t\bar{t}Z}$ processes, $\pm 4.2\%$ for $\mathrm{t\bar{t}LF}$ processes, $\pm 2.8\%$ for single top processes, $\pm 3.6\%$ for $\mathrm{t\bar{t}W}$ processes, $\pm 5\%$ for VV processes, $+0.8/-0.4\%$ for W+jets processes, and $\pm 0.2\%$ for Z+jets processes. The results are based on calculations by specialized groups within LHC working groups and Refs. [123, 183]. A distinction is made for the correlations of these rate uncertainties based on the initial state from which the processes originate. All signal processes are uncorrelated. The $\mathrm{t\bar{t}LF}$ processes are also uncorrelated, as they emerge from a gg initial state. Analogously, the single top processes are uncorrelated, which arises from qg initial states in the dominating t-channel. The processes $\mathrm{t\bar{t}W}$, VV, W+jets, and Z+jets processes are treated correlated, as they all arise from $\mathrm{q\bar{q}}$ initial states.

Parton shower (PS) scales – The PS is evaluated at a scale of the strong coupling constant α_S which affects the amount of additional gluon radiation. The scale is independently varied up and down by a factor of two for both initial-state radiation (ISR) and final-state radiation (FSR). The ISR uncertainties are considered as uncorrelated for all processes, while the FSR uncertainties are considered correlated. This procedure follows the latest recommendation of the CMS Collaboration.

PS matching to ME – In order to match the PS to the ME calculations at NLO, the damping of high p_T radiation is controlled via the h_damp parameter in the Powheg Box event generator. An increased h_damp value leads to greater suppression of real emissions and therefore softer radiation (cf. Section 8.1). The nominal setting of $h_\mathrm{damp} = 1.379 m_\mathrm{t}$ is varied to $h_\mathrm{damp} = 2.305 m_\mathrm{t}$ and $h_\mathrm{damp} = 0.8739 m_\mathrm{t}$. The expected effect of the varied h_damp parameter is applied by reweighting the simulated events, based on a machine learning technique [185]. Instead of an additional simulation with varied h_damp parameter, this approach enables the use of existing simulations with comparably high statistics. The uncertainties are treated as correlated across all $\mathrm{t\bar{t}X}$ processes.

Collinear gluon splitting – The collinear splitting of a gluon into a b quark-antiquark pair ($g \to 2b$), as in the $t\bar{t}2b$ process, leads to a divergence in QCD calculations for massless b quarks, as discussed in Section 2.4. Due to the resulting difficulty in the simulation of the process, a specific uncertainty is attributed to the $t\bar{t}2b$ process. This uncertainty is a 50% log-normal constrained rate relative to the normalization of all $t\bar{t}B$ processes. The uncertainties are treated as correlated across all $t\bar{t}B$ processes. In Section 7.2.1, it is discussed that the $t\bar{t}b\bar{b}$ model utilized in this thesis shows deviations compared to the measured data when examined in bins of the jet multiplicity and in bins of the b jet multiplicity. The rate and shape of the QCD scales addresses potential jet multiplicity mismodeling, however, this is not the case for the b jet multiplicity, i.e. a separation of events with three b jets versus four b jets. This uncertainty therefore differentiates within this separation of the b jet multiplicity, as the $t\bar{t}2b$ processes primarily occur in the phase space of three b jets, whereas the $t\bar{t}b\bar{b}$ process occurs in the phase space with four b jets. Since no dedicated separation into the number of b jets is performed in this thesis, the mismodeling does not impose any risk, but it is considered within the associated uncertainty. It is a strategic choice to avoid sensitivity to these issues by assigning a single signal strength parameter $\mu_{t\bar{t}B}$ to all $t\bar{t}B$ processes.

Underlying event – The modeling of the underlying event, which is the soft background radiation of processes not related to the hard scattering process, is tuned by the CMS Collaboration in Pythia8 with a set of parameters referred to as CP5 Tune [145]. Event simulations with variations of the CP5 Tune are also provided by the CMS Collaboration, and are taken into account as one sigma standard deviation uncertainties [186]. The uncertainties are treated as correlated across all $t\bar{t}X$ processes.

Color reconnection – PS algorithms are limited in the description of color lines in multi-parton interactions. However, color reconnections in multi-parton interactions impact the kinematics of generated particles and the resulting jet multiplicity [187, 188]. Two phenomenological models are considered in addition to the default color reconnection model in this thesis, allowing for a more comprehensive description of color connection topologies. These phenomenological models are parameterized and tuned to match observations in data. The first color reconnection model is inspired by QCD, where each parton is indexed, indicating possible combinations. In the second color reconnection model, gluons can function as mediators moving partons to another pair of partons, which is referred to as ‘‘gluon move’’. Additionally, an event simulation with the nominal color reconnection settings is considered as an uncertainty. In this simulation, instead of a top decay after the color reconnection, the top decays first. This is referred to as early resonance decays (ERD). Each of these three simulations is considered as a symmetric one sigma standard deviation variation of the nominal event simulation. The uncertainties are treated as correlated across all $t\bar{t}X$ processes.

top p_{T} modeling – As described in Section 8.4.7, the disagreement between data and simulation in the $\mathrm{t\bar{t}}$ NLO modeling is corrected. For this purpose, the nominal distribution is corrected according to Equation 8.9, which reflects the NNLO expectation. A systematic uncertainty is constructed by applying the correction, equivalent to the size of the effect itself, i.e. the application and non-application of the correction are constructed as a one sigma standard deviation symmetrical around the reweighted nominal distribution. The effect is applied to all $\mathrm{t\bar{t}X}$ processes and is considered correlated.

10.1.3 Statistical uncertainties from event simulation

Due to the finite number of simulated events, statistical fluctuations occur in the nominal prediction for each bin of all fit distributions. To describe the uncertainty of these fluctuations, a Gaussian-constrained NP is added to the fit for each bin of the distributions varying the total bin yield [189]. This method is known as Barlow-Beeston-lite approach [190].

10.1.4 Smoothing of systematic uncertainties

Certain previously described variations for the characterization of systematic uncertainties exhibit fluctuations, both in control and fit distributions. However, these fluctuations are not based on real physics effects in the estimation of the systematic uncertainty, but are caused by reduced modeling statistics. For example, this can result in the prediction of the ISR scales' up and down variation both being above the nominal prediction in a bin of the fit distribution. Without a treatment of these fluctuations, non-physical constraints and correlations of the NPs can emerge, as well as biases in the fit model. To compensate for the fluctuations, the LOWESS smoothing algorithm is applied [191, 192].

First, the ratios of the up and down variations to the nominal yield prediction from the simulated events in bin i are calculated, which are obtained with

$$\overline{r}_i = \frac{1}{2} \cdot \left(\frac{\lambda_i^{\uparrow}}{\lambda^0} - \frac{\lambda_i^{\downarrow}}{\lambda^0} \right) \quad . \tag{10.1}$$

In this equation, $\lambda_i^{\uparrow/0/\downarrow}$ is the yield in bin i for the up ($\uparrow$), nominal (0), or down ($\downarrow$) variation. The smoothed event yields $\hat{\lambda}_i^{\uparrow\downarrow}$ of the up and down variation are calculated with

$$\hat{\lambda}_i^{\uparrow\downarrow} = (1 \pm s_{\uparrow\downarrow} \hat{r}_i) \cdot \lambda_i^0 \quad , \tag{10.2}$$

where $\hat{r}_i$ is the smoothed ratio in bin i obtained with the LOWESS smoothing algorithm. The scales $s_{\uparrow\downarrow}$ are chosen to minimize

$$\chi^2_{\uparrow\downarrow} = \sum_i \left(\frac{\hat{\lambda}_i^{\uparrow\downarrow} - \lambda_i^{\uparrow\downarrow}}{\sigma_i} \right)^2 \quad , \tag{10.3}$$

reflecting the difference between smoothed and un-smoothed systematic distributions. The variations derived from Equation 10.2 not only result in smoothed distributions but also lead to symmetrical uncertainties around the nominal prediction. In this analysis, the μ_R, μ_F, ISR, FSR, JEC, b tagging, color reconnection, and underlying event uncertainties are smoothed. An analogous procedure is employed in the $\mathrm{t\bar{t}b\bar{b}}$ analysis by the CMS Collaboration [96].

Unless stated otherwise, the systematic uncertainties in all plots in this thesis include the quadratic sum of all variations described in this section.

10.2 Fit model

The concept of maximum likelihood fits and signal extraction is discussed conceptually in Chapter 6. The output score of the multi-class graph level prediction (GLP) model, as discussed in Section 9.3, forms the basis for the statistical model to build a likelihood with both histogrammed predictions as well as histogrammed data. The distributions derived from the output score of the evaluated GLP model are used to extract the parameter of interests (POIs) through a multi-dimensional maximum likelihood fit. A total of four POIs are defined according to Equation 6.3 for the multi-dimensional fit with

$$\mu_{\mathrm{t\bar{t}B}} = \frac{\sigma^{\mathrm{obs}}_{\mathrm{t\bar{t}B}}}{\sigma^{\mathrm{SM}}_{\mathrm{t\bar{t}B}}} \quad , \quad \mu_{\mathrm{t\bar{t}C}} = \frac{\sigma^{\mathrm{obs}}_{\mathrm{t\bar{t}C}}}{\sigma^{\mathrm{SM}}_{\mathrm{t\bar{t}C}}} \quad , \quad \mu_{\mathrm{t\bar{t}H}} = \frac{\sigma^{\mathrm{obs}}_{\mathrm{t\bar{t}H}}}{\sigma^{\mathrm{SM}}_{\mathrm{t\bar{t}H}}} \quad , \quad \mu_{\mathrm{t\bar{t}Z}} = \frac{\sigma^{\mathrm{obs}}_{\mathrm{t\bar{t}Z}}}{\sigma^{\mathrm{SM}}_{\mathrm{t\bar{t}Z}}} \quad , \tag{10.4}$$

where σ_i^{obs} refers to the observed cross section of process i, and σ_i^{SM} denotes the cross section predicted by the SM. Consequently, the signal strength parameters μ_i scale the number of the respective events for each process in each bin of the fit distributions, which are obtained from the GLP model. The uncertainties, discussed previously in Section 10.1, are considered as NPs θ in the fit, totaling 376 NPs.

To obtain the discriminant distributions for the maximum likelihood fit, the simulated events of all processes are evaluated with the node level prediction (NLP) and the GLP model according to the classification strategy described in Chapter 9. An event is then assigned to the pre-defined class whose GLP output score shows the highest probability. This ensures that each event only enters one class of the discriminant and guarantees the orthogonality of all classes. A total of seven classes are defined in Section 9.3 as output nodes of the GLP model: $\mathrm{t\bar{t}B}$, $\mathrm{t\bar{t}H(B)}$, $\mathrm{t\bar{t}Z(B)}$, $\mathrm{t\bar{t}Z(nonB)}$, $\mathrm{t\bar{t}C}$, $\mathrm{t\bar{t}LF}$, and Other. As described in the previous chapter, the event classes are constructed to enrich the events in which the respective processes took place. The GLP output score distributions are shown for the four signal classes in Figure 10.1 and for the background classes in Figure 10.2. The corresponding event yields are discussed in Section 9.3.2 and recorded in Table 9.7. The statistical and systematic uncertainties are also shown, as discussed in Section 10.1. The

$t\bar{t}H$ and $t\bar{t}Z$ predictions are additionally displayed as a line scaled to match the integral of the stack of background histograms, demonstrating their behavior with respect to the score in each class, which otherwise would not be discernible due to the relatively small predicted cross sections. The binning of the GLP output score distributions is designed to optimize the sensitivity in fits to pseudodata, while avoiding excessively fine binning, which can result in artifacts of artificial constraints leading to exceptionally high sensitivities. Almost all data points in the $t\bar{t}B$, $t\bar{t}H$, and $t\bar{t}Z$ classes lie within the uncertainty bands of the predictions derived from the statistical and systematic uncertainties. Furthermore, it is evident that the distribution of predictions varies across different processes within the same class. This suggests that the differential information about a process within a class is an important piece of information for the maximum likelihood fit. The discriminant of the $t\bar{t}C$ class shows that the $t\bar{t}C$ classification tends to be challenging for the GLP model, with only a few events showing large scores relative to the remaining events in this class. As discussed in Chapter 9, the core challenge in differentiating $t\bar{t}C$ events from other simulated events is identified as the lack of c tagging information embedded in the classifier, which is also evident in the discriminant distribution. Additionally, the evaluated data events exhibit a slight tendency towards smaller scores in the classifier compared to the prediction. However, all data points are within the uncertainties.

The background class $t\bar{t}Z$(nonB) indicates a discrepancy of around 11% between data and simulated events, with a tendency towards smaller scores for data compared to the model predictions. However, this minor trend is driven by the contribution of Vx processes. The majority of the data points are within the overall uncertainties of the statistical model. In the $t\bar{t}LF$ class, an over-prediction of simulated events relative to data by 7% can be seen, which is within the overall uncertainty. Conversely, in the Other class, the data are under-predicted by 21%, a discrepancy that is not accounted for by the uncertainties. This class clearly demonstrates that the divergence in data compared to simulation stems from the description of the background processes Vx. The signal regions remain unaffected by these processes, as the Vx contributions are negligible in the signal classes. The $t\bar{t}LF$ and Other classes are excluded from the fit, as they are not well modeled and do not improve the sensitivity in any test performed on pseudodata.

10.2.1 Pseudodata: fit results

Prior to extracting the signal strength parameters from a fit to data, tests are performed on a pseudodata set. The pseudodata corresponds to the sum of the event weights in each bin. In this way, the pseudodata set represents the alternative hypothesis to the null hypothesis, i.e. that the signals are present as predicted by the SM, corresponding to signal strength parameters of $\mu = 1$ for all processes [114]. This procedure allows a prediction of the sensitivity and significance of the measurement without relying on measured data, referred to as expected results (cf. Section 6.3). The approach also offers the advantage of

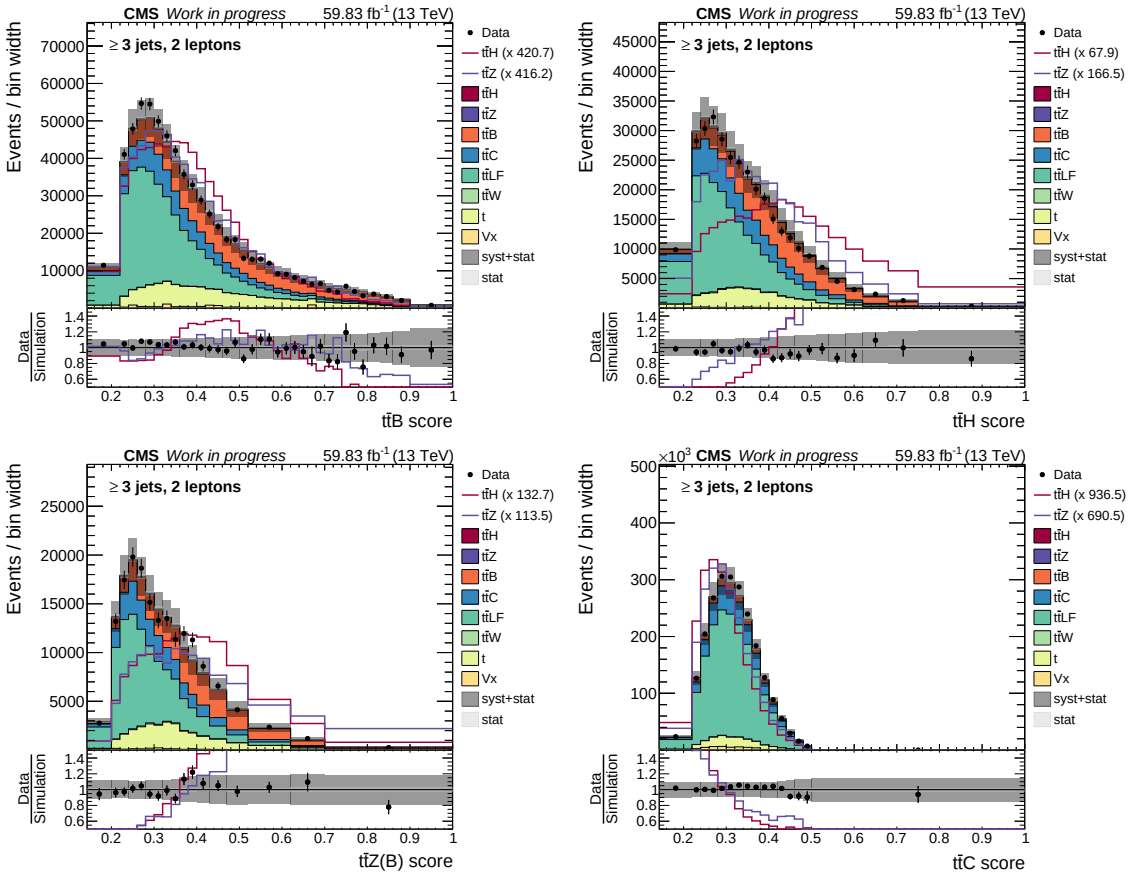


Figure 10.1: Pre-fit distributions of the GLP output score of all signal classes. The contributions of all processes are displayed as stacked histograms, with data represented as black dots. The $t\bar{t}H$ and $t\bar{t}Z$ predictions are additionally displayed as a line scaled to the integral of the stacked histogram. The bottom panel depicts the ratio of data to the expected yields from event simulation. The gray bands represent the a-priori uncertainties described in Section 10.1.

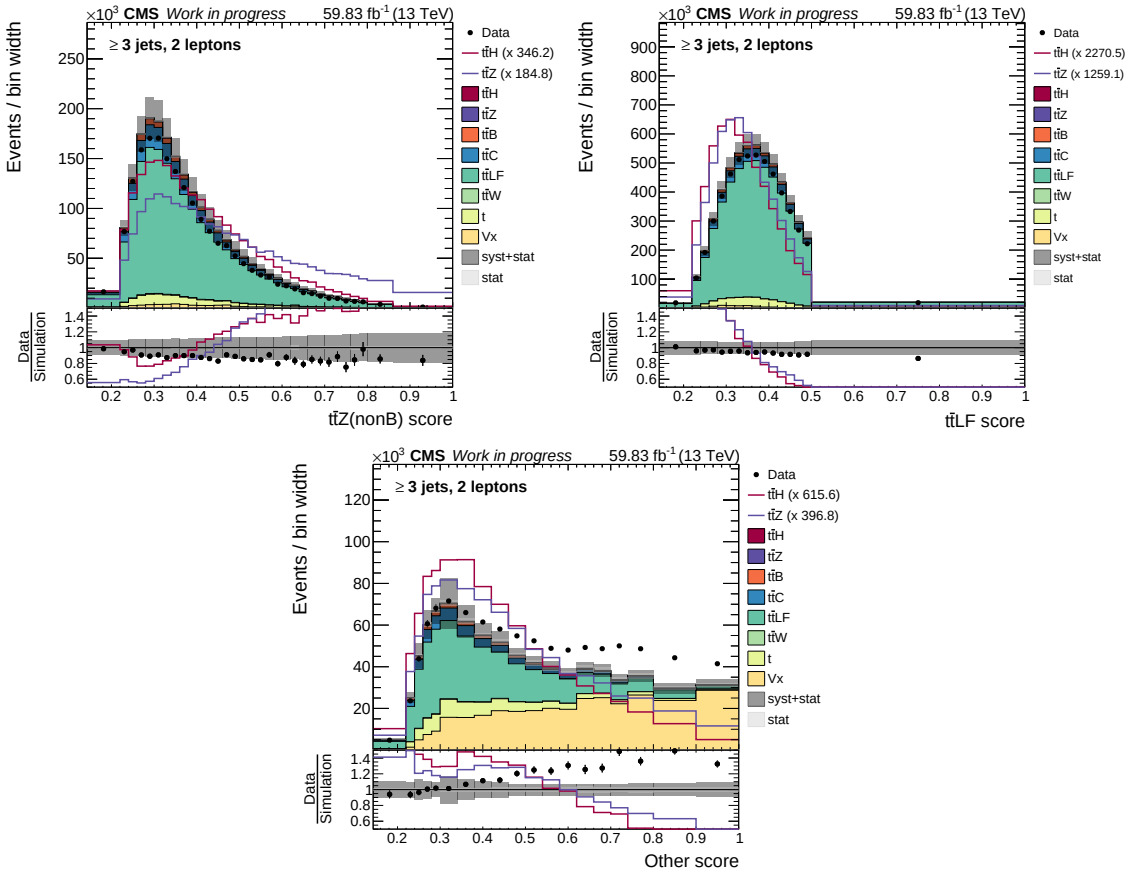


Figure 10.2: Pre-fit distributions of the GLP output score of all background classes. The contributions of all processes are displayed as stacked histograms, with data represented as black dots. The $t\bar{t}H$ and $t\bar{t}Z$ predictions are additionally displayed as a line scaled to the integral of the stacked histogram. The bottom panel depicts the ratio of data to the expected yields from event simulation. The gray bands represent the a-priori uncertainties described in Section 10.1.

investigating unwanted systematic and statistical fluctuations to increase the robustness of the statistical model. Hence, the risk of bias due to an insufficient model is minimized. The fits of the negative log-likelihood function ($-2\Delta\log(L)$) to pseudodata consists of the four POIs $\mu_{\mathrm{t\bar{t}B}}$, $\mu_{\mathrm{t\bar{t}C}}$, $\mu_{\mathrm{t\bar{t}H}}$, and $\mu_{\mathrm{t\bar{t}Z}}$, and additional 376 NPs in the fit representing the systematic uncertainties and the statistical uncertainties in the various bins of the discriminants.

The expected (exp) fit results determined in a fit of the statistical model to pseudodata result in

$$\mu^{\mathrm{exp}}_{\mathrm{t\bar{t}B}} = 1.00^{+0.35}_{-0.25} = 1.00 \pm 0.03\,(\mathrm{stat.})^{+0.35}_{-0.25}\,(\mathrm{syst.}) \quad , \tag{10.5}$$

$$\mu^{\mathrm{exp}}_{\mathrm{t\bar{t}C}} = 1.00^{+0.44}_{-0.37} = 1.00 \pm 0.03\,(\mathrm{stat.})^{+0.44}_{-0.37}\,(\mathrm{syst.}) \quad , \tag{10.6}$$

$$\mu^{\mathrm{exp}}_{\mathrm{t\bar{t}H}} = 1.00^{+0.94}_{-0.94} = 1.00^{+0.61}_{-0.59}\,(\mathrm{stat.})^{+0.72}_{-0.73}\,(\mathrm{syst.}) \quad , \tag{10.7}$$

$$\mu^{\mathrm{exp}}_{\mathrm{t\bar{t}Z}} = 1.00^{+1.13}_{-1.06} = 1.00^{+0.69}_{-0.68}\,(\mathrm{stat.})^{+0.89}_{-0.81}\,(\mathrm{syst.}) \quad , \tag{10.8}$$

including the statistical uncertainty (stat.) and systematic (syst.) uncertainty breakdowns of the total uncertainty. The expected significances of the four processes are

$$Z^{\mathrm{exp}}_{\mathrm{t\bar{t}B}} = 16, \quad Z^{\mathrm{exp}}_{\mathrm{t\bar{t}C}} = 2.9, \quad Z^{\mathrm{exp}}_{\mathrm{t\bar{t}H}} = 1.1, \quad Z^{\mathrm{exp}}_{\mathrm{t\bar{t}Z}} = 1.0 \quad . \tag{10.9}$$

The results indicate that the expected signal strength of $\mathrm{t\bar{t}B}$ processes exhibits the smallest total uncertainty. In this case, the systematic uncertainty drives the total uncertainty, while the contribution from statistical uncertainty is comparably small. Similarly for $\mu_{\mathrm{t\bar{t}C}}$, both signal strength parameters exhibit identical statistical uncertainties. However, compared to $\mu_{\mathrm{t\bar{t}B}}$, the systematic uncertainty is larger for the expected signal strength of $\mathrm{t\bar{t}C}$ processes. In contrast to $\mu_{\mathrm{t\bar{t}B}}$ and $\mu_{\mathrm{t\bar{t}C}}$, the total expected uncertainties on the signal strengths $\mu_{\mathrm{t\bar{t}H}}$ and $\mu_{\mathrm{t\bar{t}Z}}$ are significantly larger. For both processes, the total uncertainties are driven not only by the systematic uncertainty, but also by substantial contributions from statistical uncertainties. This behavior is consistent with the small predicted cross sections and correspondingly low expected yields, as shown in Table Table 9.7. Similarly, dedicated measurements of $\mathrm{t\bar{t}H}$ production by the ATLAS and CMS Collaborations report strong contributions of statistical uncertainty to the total uncertainty in the dilepton channel [126, 137].

Alternatively, instead of using the pure GLP output score as input for the fit, other discriminants can be constructed from the GLP model. In dedicated $\mathrm{t\bar{t}H(b\bar{b})}$ analyses, such as the measurement by the CMS Collaboration in Ref. [126], a ratio discriminant is constructed from the feedforward neural network prediction score for the $\mathrm{t\bar{t}H}$ and the $\mathrm{t\bar{t}B}$ classes. This approach improves the sensitivity on the $\mathrm{t\bar{t}H}$ process by 18% in the analysis [126]. To evaluate the effect of modified discriminant designs on the expected

sensitivity in this analysis, various discriminants are constructed. For this purpose, ratios from the GLP output score are calculated with

$$R_{\mathrm{t\bar{t}H\ vs.\ t\bar{t}B}} = \frac{p_{\mathrm{t\bar{t}H}}}{p_{\mathrm{t\bar{t}H}} + p_{\mathrm{t\bar{t}B}}} \quad \text{and} \quad R_{\mathrm{t\bar{t}H\ vs.\ t\bar{t}Z(B)}} = \frac{p_{\mathrm{t\bar{t}H}}}{p_{\mathrm{t\bar{t}H}} + p_{\mathrm{t\bar{t}Z(B)}}} \quad , \tag{10.10}$$

where p_i is the GLP output score of a given event in class i. Based on these ratios, numerous discriminants are constructed and tested, of which the results of three constructed statistical models are discussed in detail below. The first alternative statistical model employs the ratio $R_{\mathrm{t\bar{t}H\ vs.\ t\bar{t}B}}$ as discriminant instead of the pure GLP score of the $\mathrm{t\bar{t}H}$ and $\mathrm{t\bar{t}B}$ classes, following a similar strategy to Ref. [126]. Utilizing the ratio discriminant emphasizes the challenging distinction between the $\mathrm{t\bar{t}B}$ and $\mathrm{t\bar{t}H}$ processes. In addition to $R_{\mathrm{t\bar{t}H\ vs.\ t\bar{t}B}}$, the statistical model includes the pure GLP score of the other classes $\mathrm{t\bar{t}Z(B)}$, $\mathrm{t\bar{t}C}$, and $\mathrm{t\bar{t}Z(nonB)}$, analogous to the previously discussed statistical model. The second alternative statistical model uses $R_{\mathrm{t\bar{t}H\ vs.\ t\bar{t}Z(B)}}$ instead of the pure GLP score for $\mathrm{t\bar{t}H}$ and $\mathrm{t\bar{t}Z(B)}$ events, as well as the pure GLP score of the remaining classes $\mathrm{t\bar{t}B}$, $\mathrm{t\bar{t}C}$, and $\mathrm{t\bar{t}Z(nonB)}$. A third alternative statistical model is a combination of the first two statistical models, which is referred to as 2D unrolled statistical model in the following. In the 2D unrolled statistical model, both ratios are used, whereas the ratio discriminants are unfolded for the statistical model. To achieve this, $R_{\mathrm{t\bar{t}H\ vs.\ t\bar{t}Z(B)}}$ distributions are constructed in bins of $R_{\mathrm{t\bar{t}H\ vs.\ t\bar{t}B}}$. For the $\mathrm{t\bar{t}Z(nonB)}$ class, the pure GLP score is retained, as in the previous models. Consequently, all constructed statistical models are based on the same level of information from the GLP model. The results of the fits to pseudodata of the three alternative statistical models are summarized and compared to the statistical model using the pure GLP score in Table 10.1. The expected total uncertainties on the signal strength $\mu_{\mathrm{t\bar{t}H}}$ in the fits to pseudodata reveal that the statistical model employing the $R_{\mathrm{t\bar{t}H\ vs.\ t\bar{t}B}}$ ratio discriminant yields smaller uncertainties compared to the expected uncertainties on $\mu_{\mathrm{t\bar{t}H}}$ in the statistical model using the pure GLP score as discriminant. This is consistent with the findings in the $\mathrm{t\bar{t}H(b\bar{b})}$ analysis by the CMS Collaboration, which uses a similar discriminant [126]. However, the expected uncertainties on the signal strength parameters of the other signal processes suggest reduced precision compared to the statistical model with the pure GLP score, e.g. for $\mu_{\mathrm{t\bar{t}Z}}$. The statistical model utilizing the $R_{\mathrm{t\bar{t}H\ vs.\ t\bar{t}Z(B)}}$ discriminant shows a small improvement in the sensitivity on $\mu_{\mathrm{t\bar{t}Z}}$ compared to the statistical model that uses the pure GLP score. For all other processes, the expected precision is worse compared to the statistical model with the pure GLP score. The 2D unrolled statistical model reveals lower expected total uncertainties on the determination of the signal strength $\mu_{\mathrm{t\bar{t}Z}}$. However, a higher expected uncertainty on the determination of the other three signal strength parameters can also be seen in this statistical model compared to the statistical model with the pure GLP score as discriminant. In summary, this demonstrates the ability to construct statistical models that enhance the expected

Table 10.1: Expected fit results of the signal strength parameters obtained from statistical models based on the GLP score and constructed with different discriminants. The ratios $R_{\mathrm{t\bar{t}H\ vs.\ t\bar{t}B}}$ and $R_{\mathrm{t\bar{t}H\ vs.\ t\bar{t}Z(B)}}$ are defined according to Equation 10.10. The 2D unrolled statistical model is a combination of both ratio observables.

	Pure GLP score	$R_{\mathrm{t\bar{t}H\ vs.\ t\bar{t}B}}$	$R_{\mathrm{t\bar{t}H\ vs.\ t\bar{t}Z(B)}}$	2D unrolled
$\mu^{\mathrm{exp}}_{\mathrm{t\bar{t}B}}$	$1.00^{+0.35}_{-0.25}$	$1.00^{+0.36}_{-0.26}$	$1.00^{+0.36}_{-0.26}$	$1.00^{+0.38}_{-0.27}$
$\mu^{\mathrm{exp}}_{\mathrm{t\bar{t}C}}$	$1.00^{+0.44}_{-0.37}$	$1.00^{+0.52}_{-0.41}$	$1.00^{+0.47}_{-0.40}$	$1.00^{+0.47}_{-0.37}$
$\mu^{\mathrm{exp}}_{\mathrm{t\bar{t}H}}$	$1.00^{+0.94}_{-0.94}$	$1.00^{+0.90}_{-0.84}$	$1.00^{+1.32}_{-1.32}$	$1.00^{+1.11}_{-1.17}$
$\mu^{\mathrm{exp}}_{\mathrm{t\bar{t}Z}}$	$1.00^{+1.13}_{-1.06}$	$1.00^{+1.20}_{-1.13}$	$1.00^{+1.09}_{-1.02}$	$1.00^{+1.08}_{-1.00}$

precision of dedicated signal strength parameters using optimized discriminants, without specifically optimizing the AI model for classification. Since the statistical model with the pure GLP score exhibits the lowest overall expected uncertainties for all signal strength parameters simultaneously compared to all alternative models, this statistical model is used in the further analysis. This decision aligns with the strategy defined in Chapter 7, which emphasizes not favoring any of the four signal processes in the simultaneous cross section measurement.

In addition to the presented alternative statistical models, many other statistical models are investigated as well. In particular, splittings of the analysis phase space into sub-phase spaces are evaluated, e.g. by constructing signal enriched regions as well as background enriched regions through requirements on the number of jets and the number of b-tagged jets. For instance, regions with at least four jets and at least three or four b-tagged jets along with complementary orthogonal regions are studied. However, no improvement in terms of precision is achieved compared to the previously discussed statistical model, even when dedicated graph transformer models are trained in these phase space regions. Furthermore, it is also tested whether statistical models achieve lower uncertainties for the expected signal strength parameters if only events exceeding a predefined GLP output score are included in the statistical model. For example, discriminants are designed based on events whose GLP output scores p_i are greater than 0.5 in one or more GLP classes. Thus, this approach is based exclusively on events that have been classified with a comparably high output score, indicating greater confidence in the classification by the GLP. However, none of the statistical models constructed in these studies exhibited lower expected uncertainties for the signal strength parameters in the statistical tests than the statistical model with the full, pure GLP score discriminant information. In fact, almost all models suffered from large statistical uncertainties.

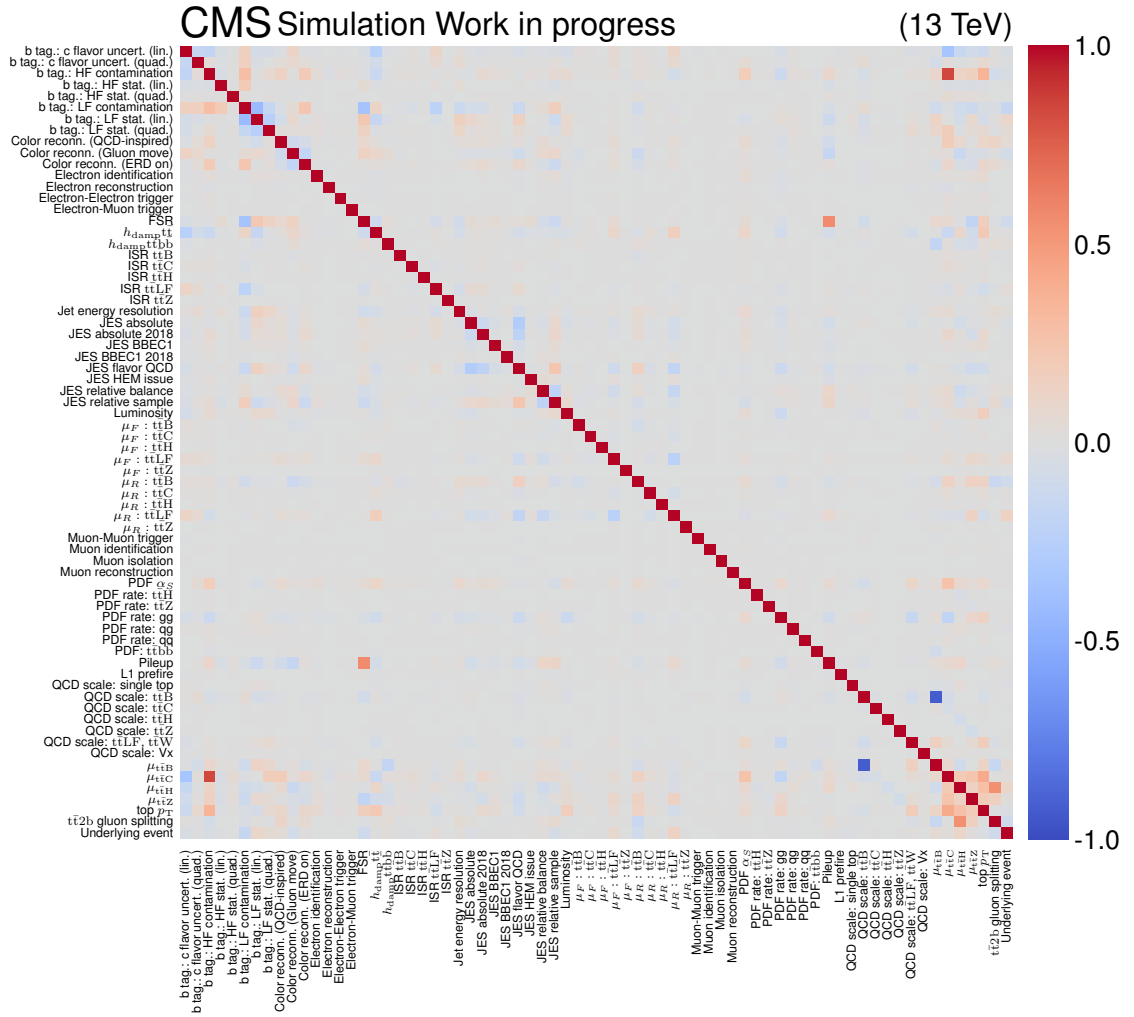


Figure 10.3: Correlation matrix of all POIs and NPs in a fit to pseudodata, except for the NPs representing the statistical uncertainties from event simulation. Positive correlations are marked in red, while anti-correlations are indicated in blue.

10.2.2 Pseudodata: correlations

The correlations of all parameters are determined from the covariance matrix as introduced in Section 6.2. The results of the correlations between all POIs and all NPs, except for the NPs representing the statistical uncertainties from event simulation, in a fit to the pseudodata set can be seen in Figure 10.3. In this figure, the correlations of the POIs and the NPs are shown as a symmetric correlation matrix. The vast majority of the parameters exhibit no or minor (anti-)correlations. However, some parameters display more pronounced correlations.

The strongest correlation, in terms of absolute magnitude, occurs between the QCD scale uncertainty on the $\mathrm{t\bar{t}B}$ processes and the signal strength parameter $\mu_{\mathrm{t\bar{t}B}}$. The uncertainty

incorporates the rate effect of the uncertainty in the choice of the renormalization scale μ_R of all $t\bar{t}B$ processes. The strong anti-correlation is evident when examining the relationship between μ_R and $\mu_{t\bar{t}B}$ in detail. Generally, the renormalization scale is the ancillary scale that controls the ultraviolet divergences that occur in QCD calculations. The renormalization eliminates these divergences by redefining the strong coupling constant α_S to yield finite values. This is done by the renormalization group equations (RGEs), with the solution for α_S given by

$$\alpha_S(\mu_R) = \frac{1}{\beta_0 \ln\left(\frac{\mu_R^2}{\Lambda_{\text{QCD}}^2}\right)} \quad , \tag{10.11}$$

where Λ_{QCD} denotes the energy scale of QCD and β_0 is a constant dependent on the number of quark flavors. Hence, the strong coupling constant decreases with an increasing energy scale μ_R, resulting in a reduced interaction rate between quarks and gluons. This effect is also known as asymptotic freedom. The dependency can be summarized as follows: if μ_R increases, α_S decreases, leading to a decrease in the interaction rate. This pattern is illustrated in Figure 10.4, which shows the GLP output score of simulated $t\bar{t}b\bar{b}$ events in the $t\bar{t}B$ class of the GLP model. In addition, the up and down variations of μ_R are shown, whereby the up variation is below the distribution with the nominal choice of renormalization scale. This behavior is consistent with the consideration from the RGEs. In the fit to pseudodata, the up variation μ_R must now be pulled back up towards the nominal value, compensated by the rate parameter of the freely floating POI $\mu_{t\bar{t}B}$. Since the NP describes the pure rate component of this effect, it directly opposes to the signal strength $\mu_{t\bar{t}B}$, resulting in an anti-correlation of nearly -1. In Section 10.3.2, it is demonstrated that this anti-correlation does not constitute any difficulties.

The correlation of the POIs is also investigated through two-dimensional scans of the negative log-likelihood function ($-2\Delta\log(L)$). All possible combinations of the four signal strength parameters in a fit to pseudodata are shown in Figure 10.5, with the correlations corresponding to the respective entries in the correlation matrix in Figure 10.3. Generally, the more circular the scans appear, the weaker the correlation between the two parameters under investigation, delivering a holistic perspective on the four POIs. The smallest correlation is expected between the two parameters $\mu_{t\bar{t}B}$ and $\mu_{t\bar{t}Z}$, the strongest correlation is exhibited between $\mu_{t\bar{t}C}$ and $\mu_{t\bar{t}Z}$. Moreover, the scans do not show any irregularities.

10.2.3 Bias test

Another important test of the statistical model is the study of systematic deviations caused by assumptions or properties in the model. To examine a possible systematic bias of the statistical model, toy data sets are generated. By uncovering a bias in advance, it is possible to check whether a bias is present before fitting to measured data. This allows the

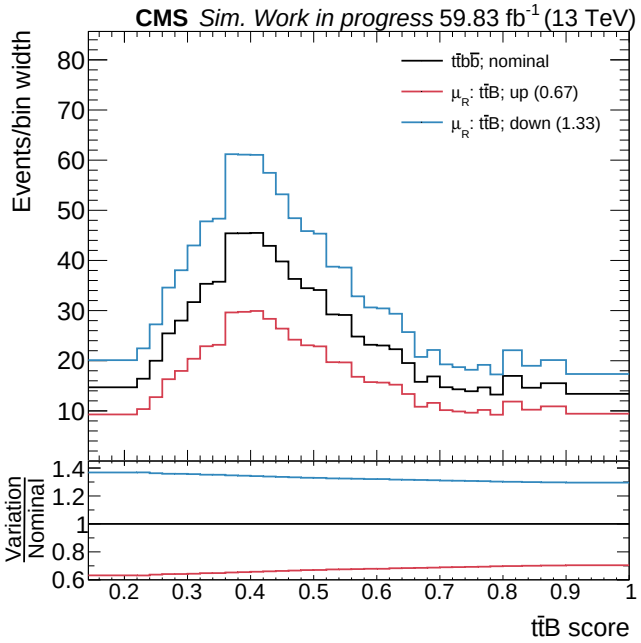


Figure 10.4: Distribution of simulated $\mathrm{t\bar{t}b\bar{b}}$ events in the $\mathrm{t\bar{t}B}$ class of the GLP model. The nominal distribution of $\mathrm{t\bar{t}b\bar{b}}$ events is drawn as black line. The up (down) variations of the renormalization scale μ_R on the $\mathrm{t\bar{t}B}$ processes are shown as red (blue) lines and the ratios of the integrals relative to the nominal value are given in brackets.

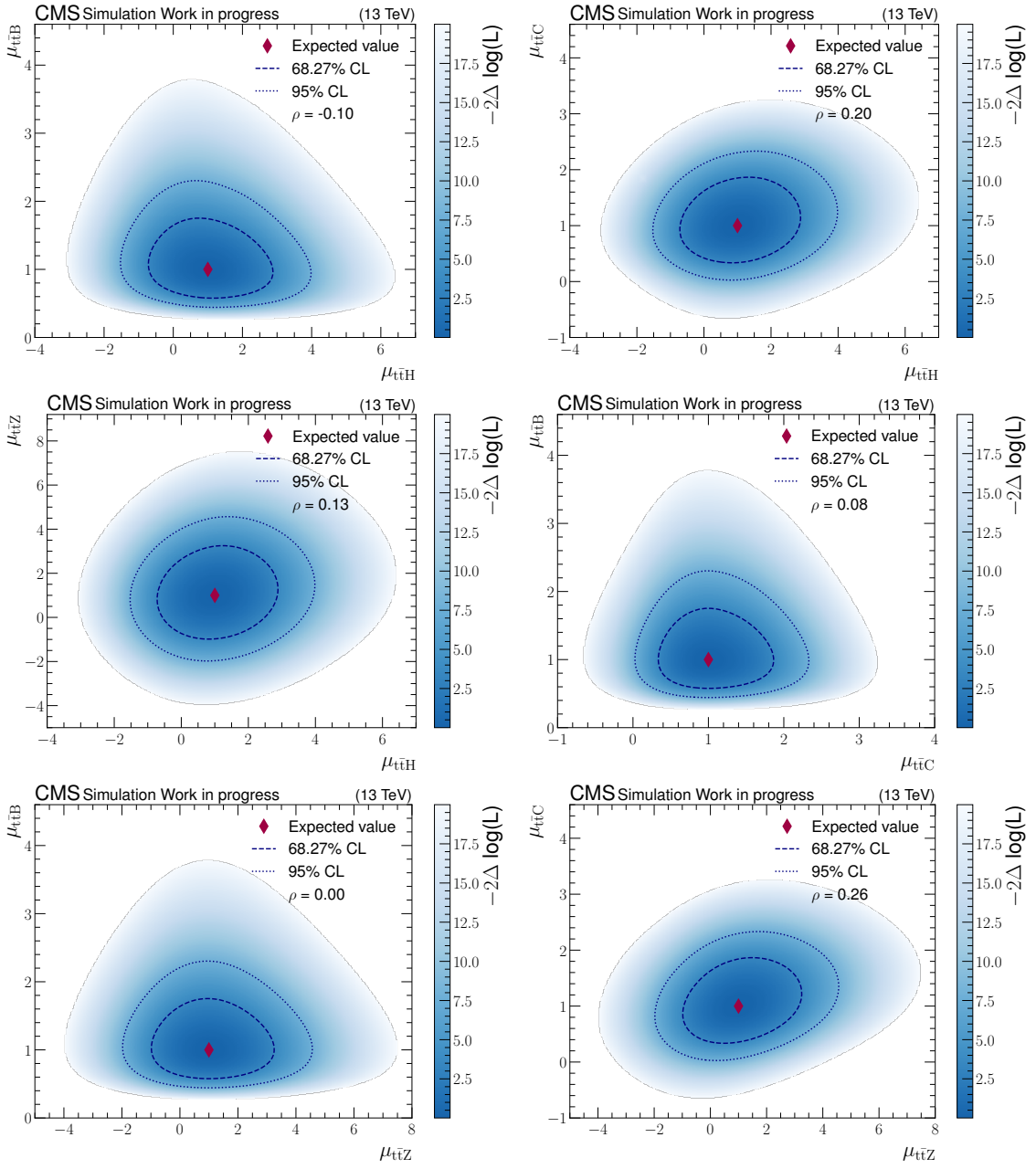


Figure 10.5: Expected negative log-likelihood test statistic (blue) as a function of the POIs, with two different combinations of POIs displayed in each panel. The remaining two POIs that are not displayed in a panel are freely floating in the fit, whereas the NPs are profiled. Additionally, the best fit values of the SM expectation (diamonds) are shown. The dashed lines represent the 68% and 95% confidence level (CL) regions. The correlation between the respective POIs is indicated in the legend of each panel.

model to be corrected beforehand if necessary while maintaining blindness with respect to data. Furthermore, an unbiased statistical model prevents the final result of the fitted POIs from being driven by properties of the statistical model, and instead reflects the true nature of the data.

The generated toy data set has no correlation with the measured data set but simulates possible behavior of the measured data set based on the statistical model. To generate the toy data set, both the background and the signal model are assumed to have a fixed signal strength parameter, referred to as the true value and set to $\mu_{\text{true}} = 1$. Next, random numbers are drawn from a Poisson distribution based on the expectation value of the statistical model. By repeating this process to generate N toy data sets, a series of randomly generated toy data sets is generated, which could, in principle, be observed in recorded data given the assumptions of the statistical model. Each of these toy data sets is then fitted to determine the fit parameter μ_{fit} and the corresponding uncertainty σ_{fit}. From this, a pull P_{model} is calculated for each toy data set with

$$P_{\text{model}} = \frac{\mu_{\text{truth}} - \mu_{\text{fit}}}{\sigma_{\text{fit}}} \quad . \tag{10.12}$$

The resulting distribution of P_{model} from the N pseudo-measurements is histogrammed and a potential bias can be identified. In case of no bias, a normal distribution centered around zero with a standard deviation of one will be obtained due to the central limit theorem. If a bias is present, it can be inferred from the mean of the distribution. A maximum value of 0.14 is set for the mean value prior to conducting the test. The reasoning behind the limit of 0.14 is based on the fact that a bias below this threshold would change the overall uncertainty by less than one percent[1]. The distribution of the pull for a total of $N = 5000$ toy data sets and the fit of a normal distribution $\mathcal{N}(\mu, \sigma^2)$ can be seen in Figure 10.6. The mean value of $\mu = 0.10$ and the standard deviation of $\sigma = 1.00$ indicate that there is no notable bias imposed by the statistical model in any of the four processes investigated in this thesis. In addition, the fit metric χ^2 divided by the number of degrees of freedom (ndf) is approximately one, indicating that the fit describes the distribution well.

10.2.4 Goodness of fit test

Another pivotal test of the statistical model is the goodness of fit test. In contrast to other tests in this section, the goodness of fit test depends on data. This test determines whether the observed data is compatible with the model's probability density function (PDF). Similar to the bias test in Section 10.2.3, toy data are generated and fitted afterwards. The signal strength parameters are freely floating in the fit, allowing for a measure that is independent from the presence or absence of the signals. For the goodness of fit test, a test

[1] A bias of 0.14 added in quadrature to the total uncertainty results in a relative change of $\sqrt{1^2 + 0.14^2} - 1 = 0.98\%$

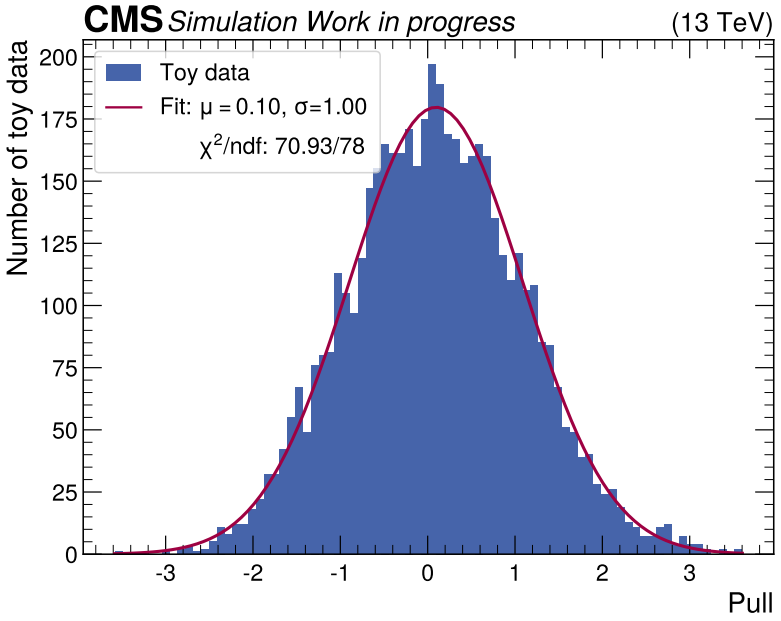


Figure 10.6: Distribution of the pull calculated from 5000 toy data sets (blue) based on the statistical model. A normal distribution $\mathcal{N}(\mu, \sigma^2)$ (red) is fitted to the pull distribution and the result of the fit parameters is given in the legend.

statistic t is calculated from which the discrepancy between the toy data and the prediction by the model can be quantified. The test statistic is based on the so called saturated model, a generalization of the χ^2 test statistic for likelihood fits to binned data [193, 194]. The test statistics are calculated for a total of $N = 10000$ toy data sets and histogrammed afterwards, resulting in a distribution $f(t)$. Additionally, the test statistic t_0 is determined for the fitted model given the data. The resulting p-value is derived from

$$p = \int_{t=t_0}^{\infty} f(t)\mathrm{d}t \tag{10.13}$$

and should be greater than 0.05. Figure 10.7 shows the distribution $f(t)$, the test statistic t_0, and the calculated p-value. The determined p-value of $p = 0.13$ exceeds the required threshold.

10.3 Fit results

In the previous section, it is demonstrated that the constructed statistical model with four POIs for the processes $t\bar{t}B$, $t\bar{t}C$, $t\bar{t}H$, and $t\bar{t}Z$ together with the uncertainties introduced in Section 10.1, entering the model in the form of NPs, passes a comprehensive series of statistical tests on pseudodata. It is also demonstrated that a model using the pure GLP score as a discriminant achieves the best expected precision across all four signal strength

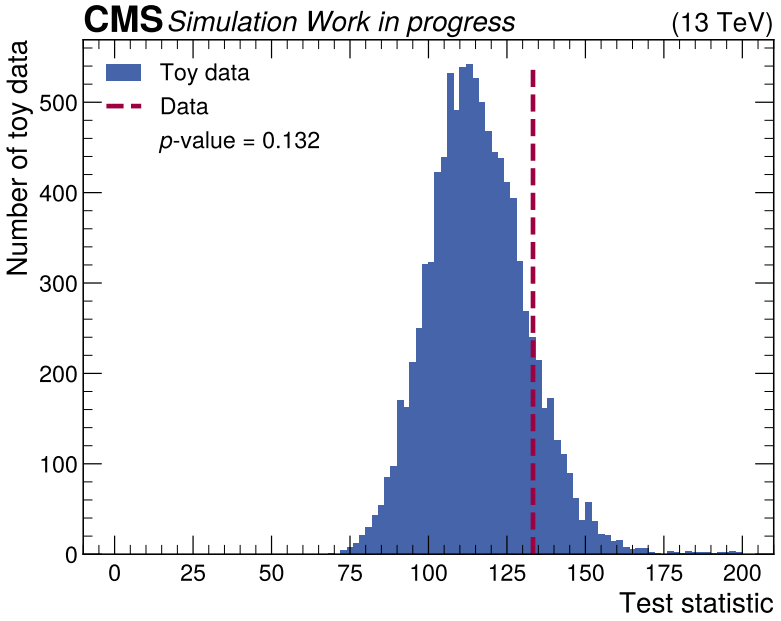


Figure 10.7: Distribution of the test statistic derived from 10000 toy data sets (blue) and the test statistic for the observed data (dashed red line). The resulting p-value is calculated with Equation 10.13

parameters simultaneously, without favoring selected parameters to the detriment of others. In this section, the statistical model is examined on measured data at the CMS experiment in the 2018 data-taking era. The signal strength parameters along with their uncertainties, significances, correlations, post-fit NP values, and the impacts on the POI are investigated. Furthermore, the results are compared to related measurements, which are discussed in Section 7.2.

The observed best-fit signal strength parameters for the four signal processes are

$$\mu^{\text{obs}}_{\text{t}\bar{\text{t}}\text{B}} = 0.98^{+0.34}_{-0.25}, \quad \mu^{\text{obs}}_{\text{t}\bar{\text{t}}\text{C}} = 0.74^{+0.41}_{-0.41}, \quad \mu^{\text{obs}}_{\text{t}\bar{\text{t}}\text{H}} = 0.89^{+0.95}_{-0.93}, \quad \mu^{\text{obs}}_{\text{t}\bar{\text{t}}\text{Z}} = 1.28^{+1.15}_{-1.06} \quad . \tag{10.14}$$

The results are also shown in Figure 10.8 with a breakdown of the total uncertainty into its statistical and theoretical contributions. The statistical uncertainty is determined in a maximum likelihood fit with all NP describing systematic uncertainties fixed to their best-fit values. The difference in quadrature between the total uncertainty and the statistical uncertainty defines the total systematic uncertainty.

The corresponding observed significances as defined in Section 6.3 for an excess of events over the expected background are

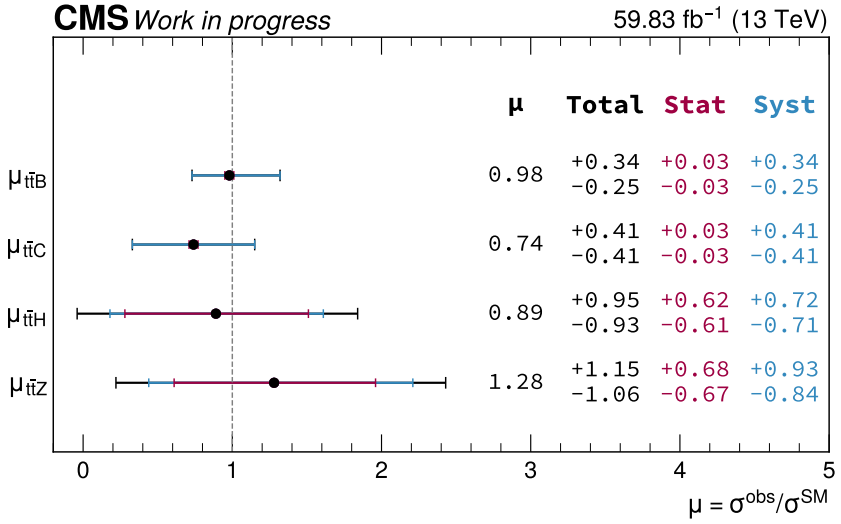


Figure 10.8: Best-fit signal strength parameters μ for the $t\bar{t}B$, $t\bar{t}C$, $t\bar{t}H$, and $t\bar{t}Z$ processes. The total uncertainty is broken down in a contribution from statistical and systematic uncertainty sources.

$$Z^{\text{obs}}_{t\bar{t}B} = 15, \quad Z^{\text{obs}}_{t\bar{t}C} = 1.8, \quad Z^{\text{obs}}_{t\bar{t}H} = 1.0, \quad Z^{\text{obs}}_{t\bar{t}Z} = 1.2 \quad . \tag{10.15}$$

The observed signal strength parameters reveal that the measured cross sections for the $t\bar{t}B$, $t\bar{t}C$, and $t\bar{t}H$ processes are smaller than the cross section predictions of the respective processes. Only the $t\bar{t}Z$ processes show a larger measured cross section relative to the cross section prediction by the SM. Despite these deviations, all four processes are compatible with the predicted cross sections of the SM. Consistent with the findings from the tests on pseudodata, it can be seen that the $t\bar{t}B$ and $t\bar{t}C$ processes are dominated by the systematic uncertainties. While the total uncertainties for the $t\bar{t}H$ and $t\bar{t}Z$ processes also have predominantly systematic contributions, they exhibit notable statistical uncertainty components as well. The observed significances, compared to the expected significances (cf. Equation 10.9) show small decreases for the $t\bar{t}B$ and $t\bar{t}H$ processes. A more pronounced decrease of the significance can be observed for $t\bar{t}C$ processes, whereas the observed significance of the $t\bar{t}Z$ processes show a slight increase compared to the expected significance. Such deviations between expected and observed significances are a common phenomenon in statistical analyses and does not constitute an anomaly in the measurement. A detailed discussion of the observed signal strength parameters is provided at the end of this chapter, after various factors of the measurement such as correlations, post-fit values of the NPs, and impacts on the signal strength parameters are examined.

Table 10.2 provides a detailed overview of the contributions to the total uncertainty for all four signal strength parameters. The QCD scale uncertainty is identified as the most

significant factor contributing to the signal strength parameters $\mu_{t\bar{t}B}$. The largest source of uncertainty on the signal strength parameters $\mu_{t\bar{t}C}$ is the uncertainty arising from b tagging with the DeepJet heavy flavor jet tagging algorithm. For the signal strength parameter of the $t\bar{t}H$ processes, the dominant contribution stems from the modeling uncertainty of the collinear gluon splitting in $t\bar{t}2b$ processes. The uncertainties reflecting the QCD scale, PS, damping scale h_{damp}, and JEC are the largest uncertainties contributing to the signal strength parameter $\mu_{t\bar{t}Z}$, each with roughly equal influence.

The GLP output score distributions after the maximum likelihood fit of the POIs and NPs are shown in Figure 10.9 for the signal classes and in Figure 10.10 for the background class considered in the fit. These distributions demonstrate a significantly improved agreement between predictions and data compared to the pre-fit distributions in Figures 10.1 and 10.2, particularly in bins with high signal purity, i.e. bins with large GLP scores in each class. For the background class $t\bar{t}Z$(nonB), the previously described over-prediction in simulation is mitigated by the fit to the data. However, some data points are not covered by the uncertainties, as the total uncertainties are reduced after the fit.

10.3.1 Correlations

The correlations between the parameters of the statistical model determined in a fit to data demonstrate similar patterns compared to the correlations determined in fits to pseudodata. All parameter correlations are depicted in Figure 10.11.

Some parameter correlations appear more pronounced in the fit to data compared to the parameter correlations in the fit to pseudodata, while others are comparatively weaker. The overall NP correlations in the fit to pseudodata are expected to be more pronounced than those in fits to measured data since pseudodata provides an ideal, noise-free, and systematic representation of the model predictions. In contrast, measured data includes statistical fluctuations and uncertainties, which tend to diminish the strength of the correlations. Hence, several factors can contribute to reduced correlations between POIs and NPs in fits to data compared to correlations visible in fits to pseudodata. In maximum likelihood fits to data, deviations in the background modeling or other effects from the measurement can directly constrain the NPs and reduce their allowed ranges. These constraints reduce their ability to vary in correlated ways. In contrast, NP correlations in fits to pseudodata are primarily driven by model assumptions and influenced by the chosen priors. In fits of the statistical model to data, the constraints from data encoded in the negative log-likelihood function can dominate and reduce the correlations imposed by the model. Another reason for reduced correlation arises when the POIs are smaller than expected, leading to a reduced contribution of the signal to the negative log-likelihood function. This is the case in this measurement, as signal strength parameters smaller than one are determined in fits to data for both $t\bar{t}C$ and $t\bar{t}H$. Therefore, the fit to data is dominated by the background model compared to fits on pseudodata, which constrains NPs more effectively, limits the

Table 10.2: Uncertainty breakdown for all observed best-fit signal strength parameters. The uncertainties under scrutiny are obtained by fixing all other uncertainties to their post-fit values and evaluating the difference relative to the full-fit result using quadratic subtraction. To estimate the statistical uncertainty, all NPs are fixed to their post-fit values and the fit is performed again. The total uncertainty differs from the quadratic sum of individual contributions due to rounding and correlations among the NPs. The miscellaneous group includes all NPs that do not belong to any of the other groups, such as the luminosity.

Source of uncertainty	$\mu_{t\bar{t}B}$	$\mu_{t\bar{t}C}$	$\mu_{t\bar{t}H}$	$\mu_{t\bar{t}Z}$
Experimental				
b tagging	$+0.04/-0.04$	$+0.37/-0.38$	$+0.15/-0.14$	$+0.25/-0.28$
JEC	±0.04	$+0.10/-0.09$	$+0.24/-0.20$	$+0.34/-0.35$
Lepton	±0.01	$+0.06/-0.05$	$+0.08/-0.11$	$+0.10/-0.15$
Theory				
QCD scale	$+0.32/-0.23$	$+0.06/-0.05$	$+0.13/-0.12$	$+0.39/-0.30$
$t\bar{t}$2b gluon splitting	$+0.06/-0.03$	±0.01	$+0.43/-0.52$	$+0.21/-0.19$
PS	±0.02	±0.04	$+0.12/-0.14$	$+0.33/-0.32$
h_{damp}	$+0.07/-0.05$	±0.08	$+0.06/-0.08$	$+0.39/-0.36$
top p_{T} correction	±0.01	±0.01	$+0.28/-0.18$	$+0.25/-0.16$
PDF	$+0.05/-0.04$	±0.05	$+0.07/-0.11$	$+0.28/-0.27$
Color reconnection	±0.01	±0.03	$+0.21/-0.20$	$+0.14/-0.30$
Miscellaneous	±0.01	$+0.06/-0.07$	$+0.22/-0.19$	$+0.18/-0.27$
Other POIs	±0.01	±0.01	±0.17	±0.18
Statistical	±0.03	±0.03	$+0.62/-0.61$	$+0.68/-0.67$
Total	$+0.34/-0.25$	$+0.41/-0.41$	$+0.95-0.93$	$+1.15/-1.06$

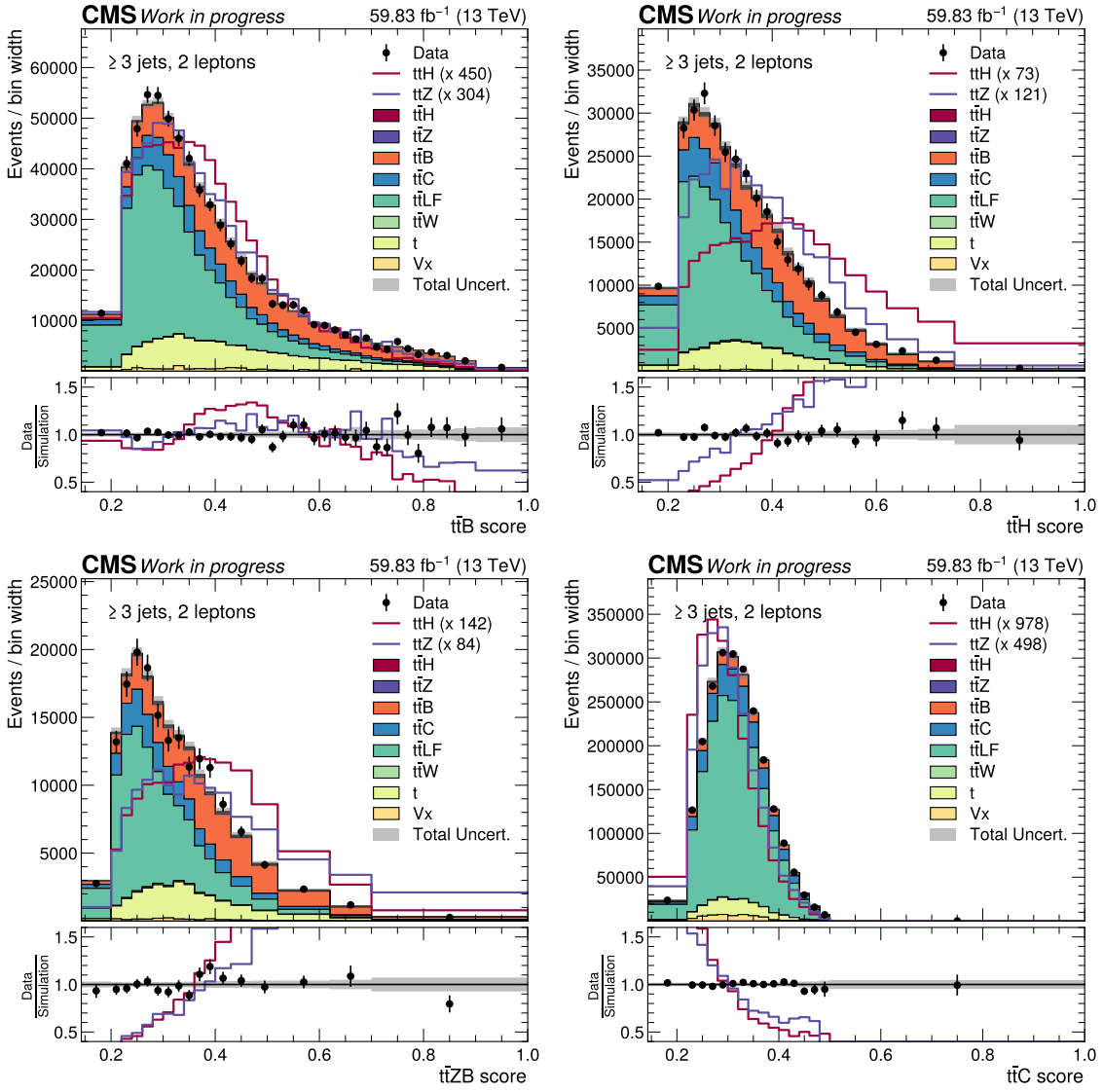


Figure 10.9: Post-fit distributions of the GLP output score of all signal classes. The contributions of all processes are displayed as stacked histograms, with data represented as black dots. The $t\bar{t}H$ and $t\bar{t}Z$ predictions are additionally displayed as a line scaled to the integral of the stacked histogram. The bottom panel depicts the ratio of data to the expected yields from event simulation. The gray bands include the post-fit uncertainties described in Section 10.1.

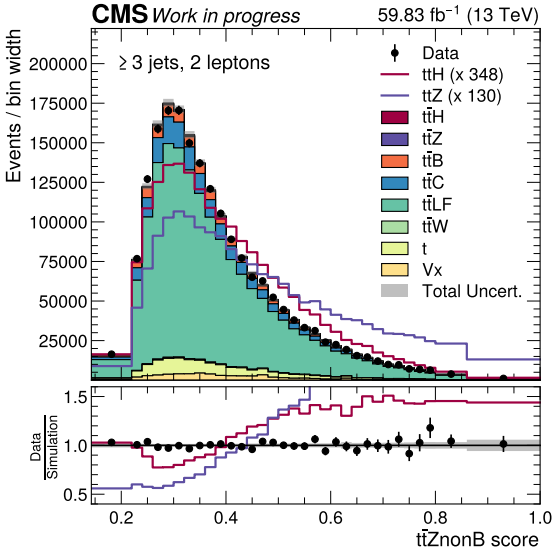


Figure 10.10: Post-fit distributions of the GLP output score of the background class used in the fit. The contributions of all processes are displayed as stacked histograms, with data represented as black dots. The $t\bar{t}H$ and $t\bar{t}Z$ predictions are additionally displayed as a line scaled to the integral of the stacked histogram. The bottom panel depicts the ratio of data to the expected yields from event simulation. The gray bands include the post-fit uncertainties described in Section 10.1.

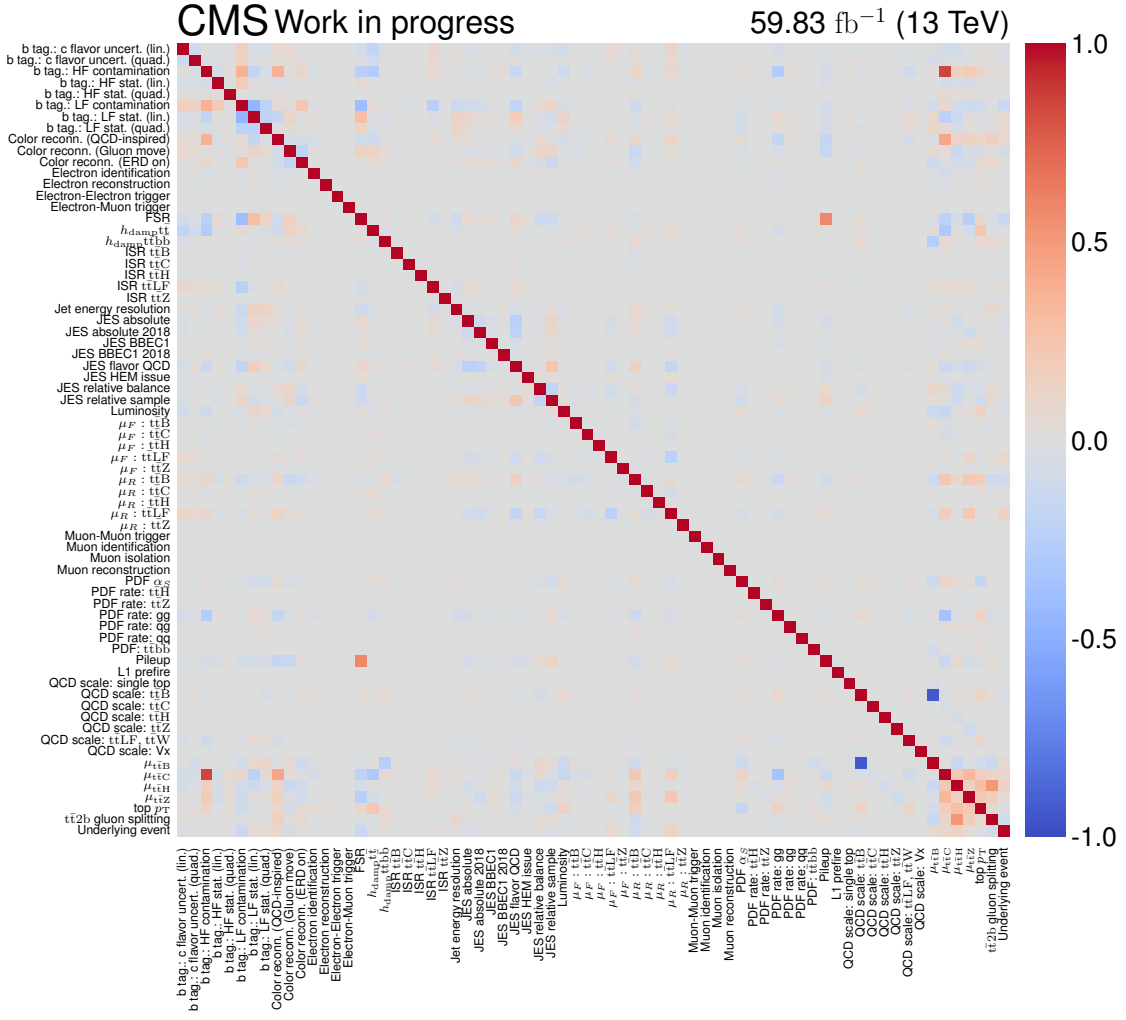


Figure 10.11: Correlation matrix of all POIs and NPs in a fit to data, except for the NPs representing the statistical uncertainties from event simulation. Positive correlations are marked in red, while anti-correlations are indicated in blue.

degrees of freedom and consequently reduces the correlation of the individual NPs. As an example, a lower parameter correlation in a fit to data, compared to a fit to pseudodata, can be seen in the correlation of the NP, which describes the c flavor uncertainty of the b tagging algorithm, with the signal strength parameter for the $t\bar{t}C$ processes. However, systematic effects in measured data can still introduce new correlations that are absent in the fit to pseudodata, potentially enhancing certain correlations between the model parameters. Examples of these cases are given in the discussion of the correlations between the POIs, which can be found at the end of this section.

The strongest correlation between parameters in the model is between the NP describing the uncertainty in the choice of the QCD scale for the $t\bar{t}B$ processes and the corresponding signal strength parameter $\mu_{t\bar{t}B}$. This relationship is elaborated in detail in Section 10.2.2 in the fit to pseudodata. The second largest correlation is the NP representing the b tagging uncertainty reflecting the heavy flavor contamination in the light flavor region with the $t\bar{t}C$ signal strength parameter. As shown in Chapter 9, the information of the heavy flavor tagger is crucial for the classification of $t\bar{t}C$ events. Since the GNN model relies exclusively on the b tagging discriminant provided by the DeepJet jet tagging algorithm and does not possess information specific to c jet tagging, this dependency is particularly pronounced.

The correlations between the POIs are examined in detail in scans of the observed negative log-likelihood test statistic as a function of the POIs. The negative log-likelihood for two POIs is scanned for each possible combination, while the remaining two POIs are left freely floating in the fit. The results of the six possible combinations are shown in Figure 10.12. The two-dimensional scans allow the correlations of the respective POIs to be discerned, the value of the parameter correlation is also stated in each panel in the figure. Additionally, the 68% and 95% CL regions are indicated. The signal strength parameters $\mu_{t\bar{t}Z}$ and $\mu_{t\bar{t}C}$ exhibit the largest correlation (39%), while the signal strength parameters $\mu_{t\bar{t}H}$ and $\mu_{t\bar{t}B}$ show the largest anti-correlation (−10%). Compared to the observed anti-correlation between the $\mu_{t\bar{t}H}$ signal strength and the $t\bar{t}B$ background normalization of −48% in the $t\bar{t}H$ analysis by the CMS Collaboration, the result of this thesis shows a reduced correlation of the two processes [126]. The axis scaling is chosen individually for each POI to represent the relevant range of the negative log-likelihood scan. No two-dimensional scan shows irregularities, all combinations demonstrate circular behavior as expected. The correlation between the POI combinations in pseudodata (data) is as follows: 0.08 (0.02) for $\mu_{t\bar{t}C}$ vs. $\mu_{t\bar{t}B}$, −0.10 (−0.10) for $\mu_{t\bar{t}H}$ vs. $\mu_{t\bar{t}B}$, 0.20 (0.27) for $\mu_{t\bar{t}H}$ vs. $\mu_{t\bar{t}C}$, 0.13 (0.17) for $\mu_{t\bar{t}H}$ vs. $\mu_{t\bar{t}Z}$, 0.0 (0.0) for $\mu_{t\bar{t}Z}$ vs. $\mu_{t\bar{t}B}$, and 0.26 (0.39) for $\mu_{t\bar{t}Z}$ vs. $\mu_{t\bar{t}C}$. Furthermore, the negative log-likelihood profiles of all NPs are scanned and examined. The scans of all NPs found to exhibit the anticipated parabolic structure with no substantial deviations observed.

In summary, correlations provide valuable insights into the statistical model's behavior and establish relationships between its parameters. However, non-linearities can lead to localized constraints on NPs, where the data impose tighter constraints in specific regions

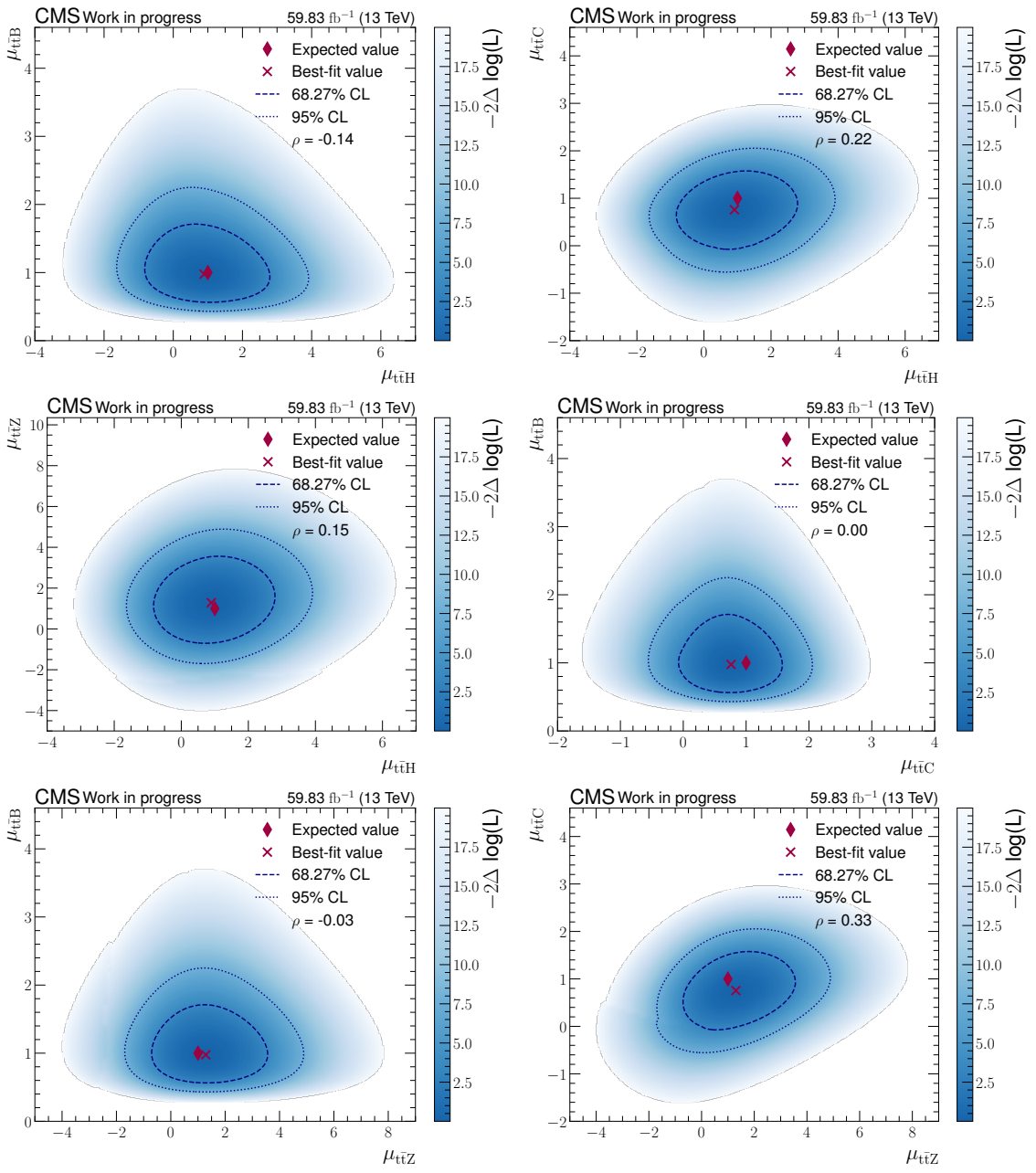


Figure 10.12: Observed negative log-likelihood test statistic (blue) as a function of the POIs, with two different combinations of POIs displayed in each panel. The remaining two POIs that are not displayed in a panel are freely floating in the fit, whereas the NPs are profiled. Additionally, the best-fit values of the SM expectation (diamonds) as well as the fit to data (crosses) are shown. The dashed lines represent the 68% and 95% confidence level (CL) regions. The correlation between the respective POIs is indicated in the legend of each panel.

of the parameter space on a set of NPs. For this reason, the following section examines not only correlations but also the impacts on the POIs, which are related to correlations but examine new aspects.

10.3.2 Post-fit parameters and impacts

A comprehensive assessment of the statistical model is given by the determination of the distance from the post-fit to the pre-fit NP value normalized to the pre-fit uncertainty and impacts of the NP on the signal strength parameters. The distances and impacts provide a comprehensive view of the parameter space and allow for an assessment of the robustness of the analysis. The distances of the NPs from the post-fit to the pre-fit values and the impacts are used to understand how strongly the data influence the NP and whether systematic uncertainties are reasonably estimated. Hence, significant shifts in parameter estimates and tensions with auxiliary measurements can be revealed, indicating potential issues with the statistical model.

The distance measure for a NP between the post-fit and the pre-fit model is defined as $(\hat{\theta} - \theta_{\mathrm{I}})/\sigma_{\mathrm{I}}$, where $\hat{\theta}$ is the post-fit value, θ_{I} is the pre-fit value, and σ_{I} is the pre-fit uncertainty of the NP. With this, the distance measure provides insights into how individual parameters deviate from their nominal value, offering a measure of agreement with the statistical model. This distance from the post-fit to the pre-fit value relative to the pre-fit uncertainty is a standard method of the COMBINE software package used for statistical analyses by the CMS Collaboration. For example, this measure is used in Ref. [126] and often referred to in technical jargon as a ‘‘pull’’. However, the definitions of pulls can vary. In constrained fits, Ref. [195] suggests incorporating the post-fit uncertainty into the calculation of a proper pull to quantify the compatibility between auxiliary measurements and the in-situ measurement from data. In contrast to other definitions such as the distance measure defined in this thesis, which do not include the post-fit uncertainty in the denominator, the pull definition in Ref. [195] takes the correlation between the uncertainties in the fit result $\hat{\theta}$ and the constraint θ_{I} into account. For this reason, the term ‘‘pull’’ is avoided for the distance measure of the post-fit value to the pre-fit value of a NP.

In the case of fits to pseudodata, the post-fit values correspond to the pre-fit values of the NP by design. Still, the asymmetric uncertainty bars on this post-fit value are a particularly important indicator. Due to the normalization of the distance measure to the pre-fit uncertainty, the uncertainty bars smaller than ± 1 state that the NP is constrained in the fit. The fit constraints of all NPs are calculated on pseudodata and measured data.

The impacts on a signal strength parameter provide valuable insights, allowing for a dynamic assessment of the relationship between the signal strength parameter and all other parameters. Impacts quantify the influence of a single NP or POI on the signal strength parameter under investigation, which is related to the determination of the correlation

between parameters, but also reflects non-linear relations between these. Since four POIs are considered simultaneously in this thesis, the impacts of the remaining three POIs on the signal strength parameter under scrutiny are also examined. The impacts are calculated by systematically shifting each parameter to its $\pm 1\sigma$ uncertainty and re-evaluating the negative log-likelihood. This reveals the relative change caused by the parameter variation on the POI under scrutiny. Through this method, impacts demonstrate the sensitivity of the maximum likelihood fit to specific parameters and enables to identify dominating contributors to the signal strength measurement. With this, the primary drivers of uncertainty can be identified, providing insights into areas that could be explored further to potentially refine precision, such as by improving specific systematic uncertainties.

The distance between the post-fit value $\hat{\theta}$ and the pre-fit value θ normalized to the pre-fit uncertainty σ_I along with the constraints, and the impacts on the observed signal strength parameters are shown in Figures 10.13 and 10.14. In each instance, one signal strength parameter of the multi-dimensional fit with four POIs is assessed. In these plots, the post-fit values of the NPs are identical across all examined signal strength parameters, as their fit values do not change when a different signal strength parameter is monitored. However, the impacts differ, as different NPs affect each signal strength parameter in distinct ways. The NPs are sorted top down from the largest impact to the smallest impact on the respective signal strength parameter under scrutiny. Each figure shows the 15 most important parameters simultaneously for expected and observed results. All other impacts can be found in Appendix C.

The parameter with the highest impact on the signal strength parameter $\mu_{\mathrm{t\bar{t}B}}$ is the QCD scale, whose effect is significantly more pronounced than that from all other parameters. The NP is not constrained in the fit and the post-fit value aligns with the pre-fit value. Since this NP represents a pure rate effect exclusively on the $\mathrm{t\bar{t}B}$ processes, which, as discussed in Section 10.2.2, strongly anti-correlates with the signal strength parameter $\mu_{\mathrm{t\bar{t}B}}$, it is expected that the NP parameter in the fit remains unchanged. Instead, the freely floating POI adjusts the rate effects. Therefore, the NP primarily reflects the uncertainty associated with the choice of the renormalization scale. Other parameters that impact the signal strength parameter $\mu_{\mathrm{t\bar{t}B}}$ are the uncertainty on the choice of the h_{damp} parameter, the signal strength parameter $\mu_{\mathrm{t\bar{t}H}}$, and the uncertainty on the $\mathrm{t\bar{t}2b}$ process due to the challenging description of the collinear gluon splitting. Still, these parameters and all others show only minor impacts on $\mu_{\mathrm{t\bar{t}B}}$.

The largest impact on the signal strength parameter of the $\mathrm{t\bar{t}C}$ processes is the NP reflecting the b tagging uncertainty of the heavy flavor contamination (cf. Section 10.1). A large impact of this NP on $\mu_{\mathrm{t\bar{t}C}}$ is expected, as heavy flavor jet tagging is particularly challenging for the reconstruction of c jets compared to b and light flavor jets. Other NPs that impact the signal strength parameter $\mu_{\mathrm{t\bar{t}C}}$ when shifting the parameter to its $\pm 1\sigma$ value are the

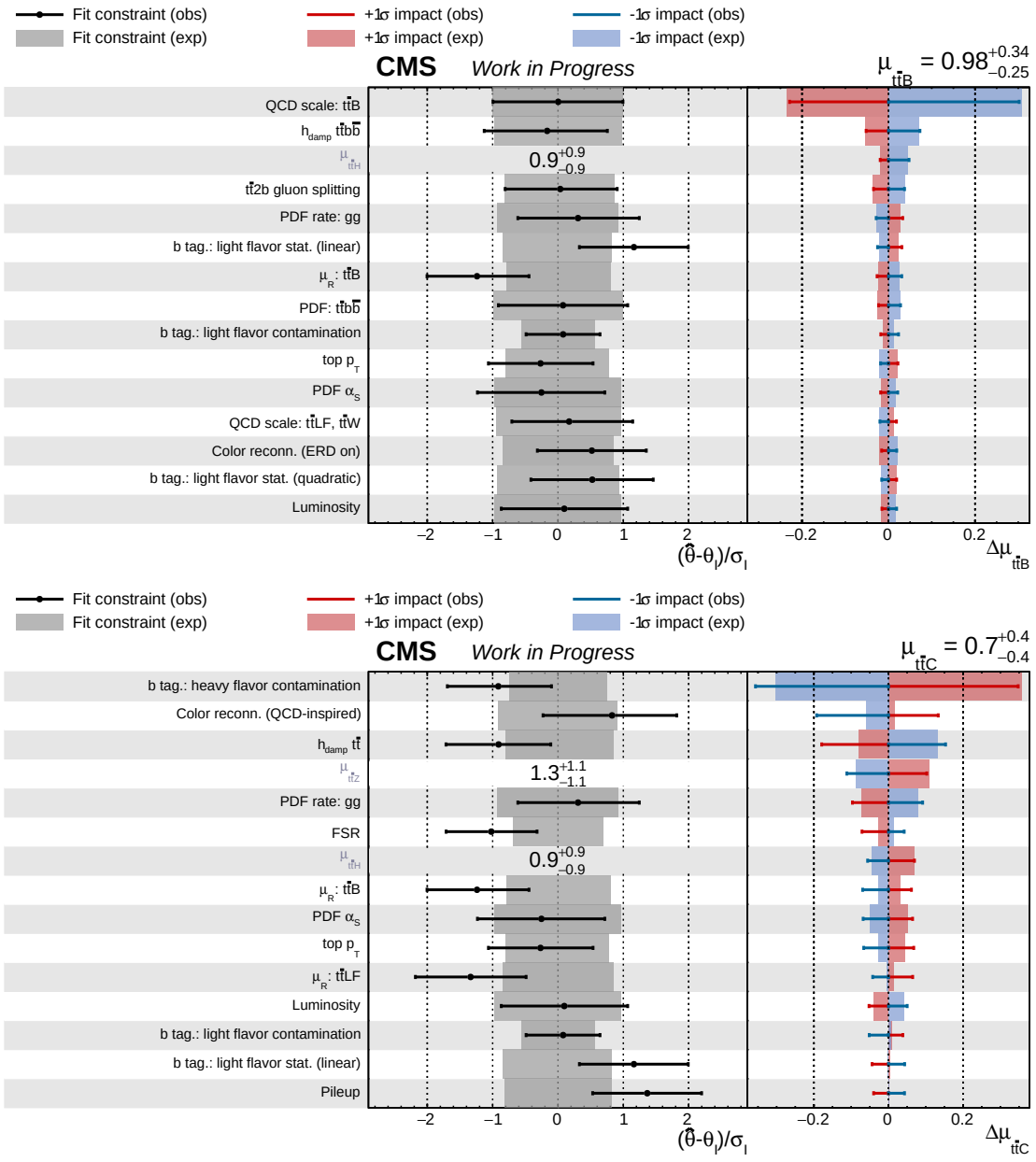


Figure 10.13: Pre-fit and post-fit results of the signal strength parameters $\mu_{t\bar{t}B}$ (top) and $\mu_{t\bar{t}C}$ (bottom). Parameter names are organized in the left panel. The distance from the post-fit value $\hat{\theta}$ to the pre-fit value θ normalized to the pre-fit uncertainty σ_I are shown in the middle panel. The right panel shows the relative change to the signal strength parameter μ under scrutiny if the NP under consideration is shifted to its $\pm 1\sigma$ value. The parameters are sorted according to the size of the impact, parameters one to 15 are shown. Created with [193].

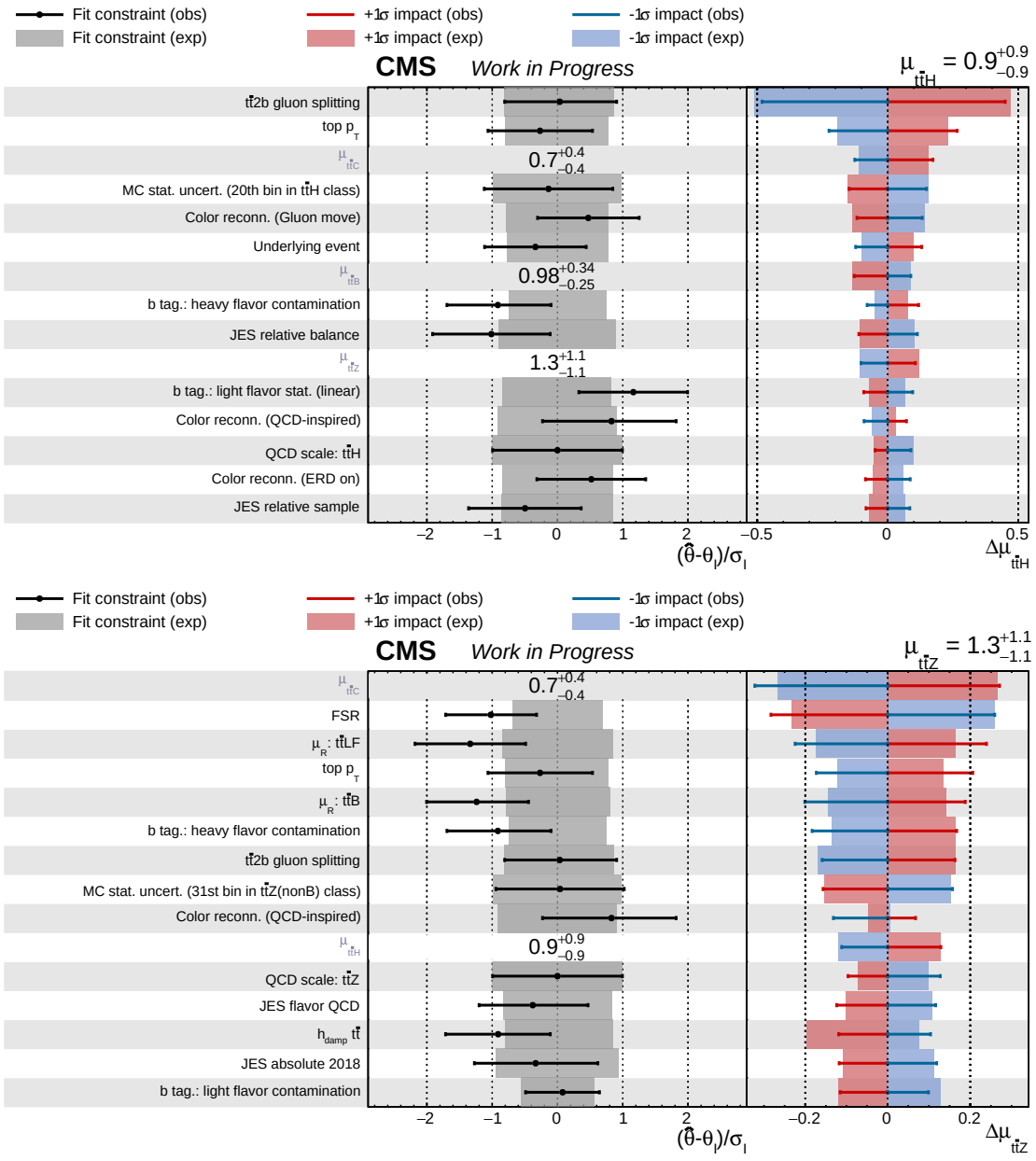


Figure 10.14: Pre-fit and post-fit results of the signal strength parameters $\mu_{t\bar{t}H}$ (top) and $\mu_{t\bar{t}Z}$ (bottom). Parameter names are organized in the left panel. The distance from the post-fit value $\hat{\theta}$ to the pre-fit value θ normalized to the pre-fit uncertainty σ_I are shown in the middle panel. The right panel shows the relative change to the signal strength parameter μ under scrutiny if the NP under consideration is shifted to its $\pm 1\sigma$ value. The parameters are sorted according to the size of the impact, parameters one to 15 are shown. Created with [193].

color reconnection in the QCD-inspired model, the uncertainty on the choice of the h_{damp} parameter, and the signal strength parameter $\mu_{\text{t}\bar{\text{t}}\text{Z}}$.

The gluon splitting uncertainty in the $\text{t}\bar{\text{t}}$2b process shows the largest impact on the signal strength parameter $\mu_{\text{t}\bar{\text{t}}\text{H}}$. This NP also has the largest impact on the $\text{t}\bar{\text{t}}$H signal strength parameter in the measurement published by the CMS Collaboration [126]. Further parameters that have a large impact on $\mu_{\text{t}\bar{\text{t}}\text{H}}$ are the top quark p_{T} modeling, the signal strength parameter $\mu_{\text{t}\bar{\text{t}}\text{C}}$, and the statistical uncertainty of a bin in the $\text{t}\bar{\text{t}}$H class of the GLP.

The largest impact on the signal strength parameter $\mu_{\text{t}\bar{\text{t}}\text{Z}}$ is the POI representing the $\text{t}\bar{\text{t}}$C signal strength in the fit. As previously discussed, the $\text{t}\bar{\text{t}}$Z and $\text{t}\bar{\text{t}}$C processes exhibit the strongest correlation among all POI combinations. Other parameters that impact the signal strength parameter $\mu_{\text{t}\bar{\text{t}}\text{Z}}$ when shifting the NP to their $\pm 1\sigma$ value are the uncertainty associated with FSR, the uncertainty on the choice of the renormalization scale of the $\text{t}\bar{\text{t}}$LF processes, and the uncertainty on the top quark p_{T} modeling.

The NPs with the largest difference between the post-fit and pre-fit values are the NPs representing the shape-changing effects of the uncertainty in the choice of the renormalization scale μ_R on the $\text{t}\bar{\text{t}}$B and $\text{t}\bar{\text{t}}$LF processes. Additionally, the NP representing the statistical fluctuations on the contamination from light and c flavor jets in the heavy flavor region of the b tagging algorithm in the fit shows an enhanced difference between the post-fit and pre-fit values. Also, the NP representing the pileup uncertainty in the fit demonstrates a difference greater than one between its post-fit and pre-fit values normalized to the pre-fit uncertainty. In Section 8.4.1, the correction of the pileup in simulated events due to a known discrepancy between event simulation and data is addressed. Although this correction reduced the discrepancy, it did not completely resolve the visible deviations (cf. Figure 8.2), and similar effects are observed in other analyses [132]. The NP reflecting the statistical uncertainty of the b tagging SF on heavy flavor jets shows the largest constraint of all NP.

10.3.3 Comparative analysis of the best-fit results

In this section, the measured signal strength parameters of this thesis are contextualized within the landscape of existing analyses, which are discussed in detail in Section 7.2.

The best-fit signal strength parameter $\mu_{\text{t}\bar{\text{t}}\text{B}}^{\text{obs}} = 0.98^{+0.34}_{-0.25}$ in this thesis aligns closely with the results of the inclusive and differential measurements of $\text{t}\bar{\text{t}}\text{b}\bar{\text{b}}$ in the semileptonic channel, as reported in Refs. [96, 132]. These analyses employed the same model to describe the $\text{t}\bar{\text{t}}\text{b}\bar{\text{b}}$ processes and observed a cross section lower than predicted, with the discrepancy increasing in phase spaces enriched with a high signal fraction. In the most inclusive phase space ≥ 5 jets, ≥ 3b jets of the measurement, the ratio of the measured cross section to the predicted cross section by the model is approximately one. Furthermore, the signal

strength parameter $\mu_{t\bar{t}B}$ obtained in this thesis is consistent with the $t\bar{t}B$ normalization in the $t\bar{t}H(b\bar{b})$ analysis by the CMS experiment, which also utilizes the same $t\bar{t}b\bar{b}$ modeling approach [126]. In Ref. [126], the post-fit $t\bar{t}B$ normalization is 1.19. However, the $t\bar{t}b\bar{b}$ cross section is scaled to the cross section prediction of the $t\bar{t}b\bar{b}$ processes described with $t\bar{t}$ at ME level, which in turn shows to predict a cross section of 10% to 24% lower than the observed cross sections in Refs. [96, 132]. Thus, due to the modified cross section, the value of the normalization $t\bar{t}B$ parameter is in agreement with the measured signal strength parameter $\mu_{t\bar{t}B}$ in this thesis. The uncertainty on the $\mu_{t\bar{t}B}$ value in this thesis is dominated by the uncertainty on the choice of the renormalization scale. In the joint study with the ATLAS Collaboration comparing different $t\bar{t}$+jets and $t\bar{t}b\bar{b}$ event simulation approaches, it is shown that larger μ_R variation uncertainties arise in $t\bar{t}b\bar{b}$ event simulations with a definition of $t\bar{t}b\bar{b}$ at ME level compared to a definition of $t\bar{t}$+jets at ME level [124]. The variation on μ_R is determined to be up to 70% Ref. [124]., which is also observed in Refs. [96, 132].

A direct comparison of the best-fit signal strength parameter $\mu^{\text{obs}}_{t\bar{t}C} = 0.74^{+0.41}_{-0.41}$ with other measurements is not feasible, as no measurement employing an identical model exists (cf. Section 7.2). The signal strength parameter measured in this thesis exhibits a lower observed cross section compared to the cross section prediction by the $t\bar{t}c\bar{c}$ modeling. However, the measured cross section is compatible with the SM prediction within the 68% confidence interval. A signal strength parameter lower than one for the $t\bar{t}C$ process is also plausible due to challenges in the event classification caused by sparse c jet tagging information in the GLP. In particular, there is an increased mis-classification rate of $t\bar{t}C$ events with $t\bar{t}LF$ events. A slight overestimation of $t\bar{t}$ events with additional light jet radiation is observed in the analysis phase space, consistent with findings reported in Ref. [132]. This suggests that the POI for the $t\bar{t}C$ processes partially compensates for these effects, especially given that the $t\bar{t}LF$ contribution dominates the analysis phase space.

The best-fit signal strength parameter $\mu^{\text{obs}}_{t\bar{t}H} = 0.89^{+0.95}_{-0.93}$ measured in this thesis indicates a lower observed cross section compared to the SM prediction. The signal strength parameters of related analyses including their uncertainties can be found in Figure 10.15 and are based on Refs. [126, 134, 136, 137]. In this figure, the measured signal strength parameters of all analyzed channels are provided, along with the results specific to the dilepton channel for comparison with this thesis, if it was investigated. Related $t\bar{t}H$ analyses exhibit significantly varying central values for the best-fit parameters, particularly in the dilepton channel. Some analyses published best-fit signal strength parameter values of $\mu < 0$ in the dilepton channel, not compatible with $\mu = 1$ within the 68% confidence interval. Furthermore, it is evident that in measurements in the dilepton channel, the statistical uncertainty contributes substantially to the total uncertainty and, in certain cases, surpasses all other sources of uncertainty.

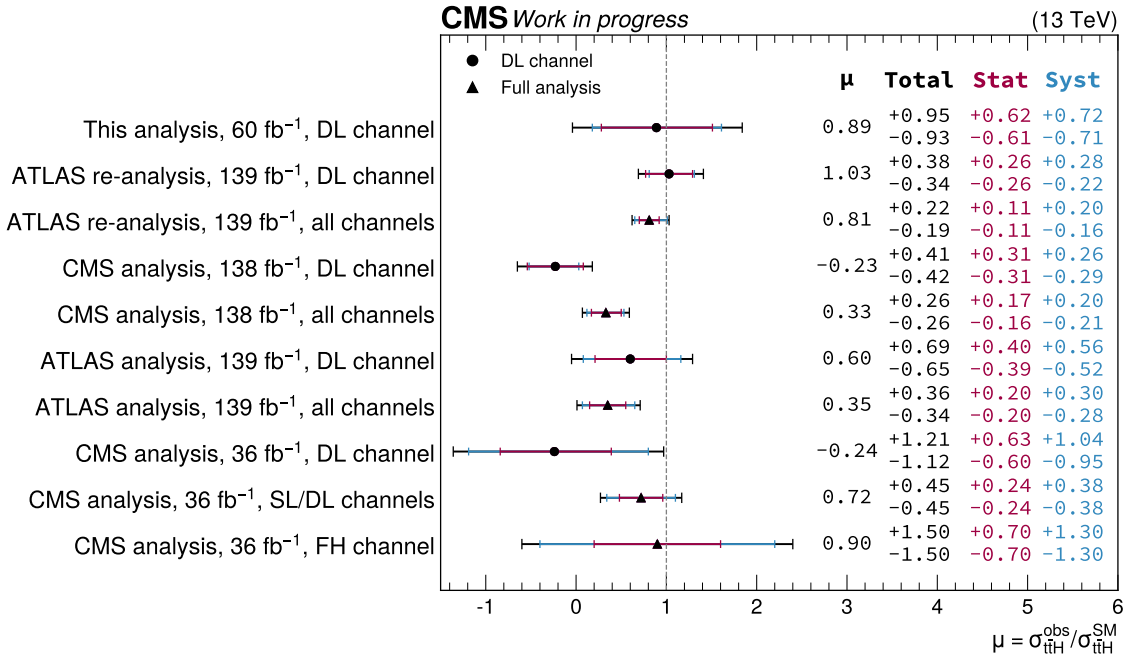


Figure 10.15: Best-fit values of the signal strength parameters $\mu_{\mathrm{t\bar{t}H}}$ in measurements of the ATLAS and CMS Collaborations [126, 134, 136, 137]. The results in the dilepton (DL) channel (if applicable) and the combined results of the measurements in all channels of the respective analysis are stated. The semileptonic (fully hadronic) channel is denoted with SL (FH). The total uncertainty is broken down in a contribution from statistical uncertainties and systematic uncertainties.

The best-fit signal strength parameter $\mu^{\text{obs}}_{\text{t}\bar{\text{t}}\text{Z}} = 1.28^{+1.15}_{-1.06}$ in this analysis shows a larger measured cross section relative to the SM prediction. A comparison of $\mu_{\text{t}\bar{\text{t}}\text{Z}}$ with other measurements is not possible, as no comparable measurements from the ATLAS or CMS Collaborations in the $\text{Z} \to \text{b}\bar{\text{b}}$ decay channel exist. The $\text{t}\bar{\text{t}}\text{Z}$ cross section measurement by the CMS Collaboration in the $\text{Z} \to \ell\ell$ decay channel, which is the measurement most comparable with this thesis, observed a 13% larger cross section of $\sigma^{\text{obs}}_{\text{t}\bar{\text{t}}\text{Z}} = 0.95 \pm 0.08\,\text{pb}$ compared to the predicted value of $\sigma^{\text{obs}}_{\text{t}\bar{\text{t}}\text{Z}} = 0.84 \pm 0.10\,\text{pb}$. The thesis in Ref. [139] analyzing data recorded in the 2018 era at the CMS experiment found a signal strength parameter of $\mu_{\text{t}\bar{\text{t}}\text{Z}} = -0.77^{+1.19}_{-1.26}$ in the semileptonic $\text{t}\bar{\text{t}}$ channel.

To summarize, in this chapter the simultaneous measurement of the four signal strength parameters $\mu_{\text{t}\bar{\text{t}}\text{B}}$, $\mu_{\text{t}\bar{\text{t}}\text{C}}$, $\mu_{\text{t}\bar{\text{t}}\text{H}}$, and $\mu_{\text{t}\bar{\text{t}}\text{Z}}$ is presented and discussed in detail. The construction of the statistical model is based on the output score of the GLP multi-class classifier, which is described in Chapter 9. The observed signal strength parameters, summarized in Figure 10.8, exhibit good agreement with the cross sections predicted by the SM. Furthermore, all influencing factors, i.e. NPs, on the signal strength parameters are thoroughly investigated.

Part IV

Conclusion

11 Summary and outlook

In this thesis, a simultaneous cross section measurement of four different top quark-antiquark pair production ($t\bar{t}$) processes with additional heavy flavor jets is performed. The analyzed processes include the associated production of $t\bar{t}$ with an additional pair of b quarks ($t\bar{t}b\bar{b}$), or an additional pair of c quarks ($t\bar{t}c\bar{c}$). These processes lead to different event topologies and are consolidated into $t\bar{t}B$ and $t\bar{t}C$, respectively. Likewise, the production of $t\bar{t}$ in association with a Higgs ($t\bar{t}H$) or Z boson ($t\bar{t}Z$), where the boson subsequently decays into a pair of b quarks ($H/Z \to b\bar{b}$). All four processes feature an equivalent final state posing a significant challenge for differentiation.

The cross section measurement analyzes the data recorded by the Compact Muon Solenoid (CMS) experiment at the Large Hadron Collider (LHC) at Conseil Européen pour la Recherche Nucléaire (CERN). The recorded and analyzed data of proton-proton collisions at a center-of-mass energy at $\sqrt{s} = 13\,\text{TeV}$ corresponds to an integrated luminosity of about $60\,\text{fb}^{-1}$. Collision events are analyzed if they contain exactly two electrons, two muons, or one electron and one muon, targeting the decay channel of the $t\bar{t}$ system where both top quark decays include leptonic processes. Furthermore, at least three jets are required in an event.

The cross sections are measured in a parameterization and are expressed as signal strength parameters relative to the cross section predicted by the Standard Model of particle physics (SM). The signal strength parameters are defined as $\mu_i = \sigma^{\text{obs}}{}_i/\sigma_i^{\text{SM}}$, where σ_i^{obs} denotes the observed cross section and σ_i^{SM} refers to the cross section of process i predicted by the SM. With this approach, the cross section is measured in the phase space of three jets and two leptons and extrapolated to the full phase space using Monte Carlo event simulation (MC). The results correspond to an observed (expected) significance of 15 standard deviations (σ) (16σ) for $t\bar{t}B$, 1.8σ (2.9σ) for $t\bar{t}C$, 1.0σ (1.1σ) for $t\bar{t}H$, and 1.2σ (1.0σ) for $t\bar{t}Z$ compared to the SM expectation without these processes included. The signal strength parameters

are simultaneously extracted a in maximum likelihood fit to binned distributions of data in all regions and result to

$$\mu^{\text{obs}}_{\text{ttB}} = 0.98^{+0.34}_{-0.25} = 0.98 \pm 0.03\,(\text{stat.})^{+0.34}_{-0.25}\,(\text{syst.}) \quad , \tag{11.1}$$
$$\mu^{\text{obs}}_{\text{tt}\bar{\text{c}}} = 0.74^{+0.41}_{-0.41} = 0.74 \pm 0.03\,(\text{stat.})^{+0.41}_{-0.41}\,(\text{syst.}) \quad , \tag{11.2}$$
$$\mu^{\text{obs}}_{\text{ttH}} = 0.89^{+0.95}_{-0.93} = 0.89^{+0.62}_{-0.61}\,(\text{stat.})^{+0.72}_{-0.71}\,(\text{syst.}) \quad , \tag{11.3}$$
$$\mu^{\text{obs}}_{\text{ttZ}} = 1.28^{+1.15}_{-1.06} = 1.28^{+0.68}_{-0.67}\,(\text{stat.})^{+0.93}_{-0.84}\,(\text{syst.}) \quad . \tag{11.4}$$

The analysis strategy to simultaneously measure four processes in the $\text{t}\bar{\text{t}}$ + heavy flavor phase space leverages state-of-the-art Artificial Intelligence (AI) models for jet and event classification. Collision events are transformed into a mathematical graph structure and processed with graph transformer neural network architectures based on multi-head attention mechanisms. The advantages of a graph structure in graph neural networks (GNNs) compared to deep neural networks (DNNs) are discussed in detail in Ref. [111] and are applied in this analysis. A two-stage training procedure is implemented. In the first stage, a classification algorithm is employed which is tasked with identifying the origin of jets in an event. A binary jet classification is implemented to identify additional jets in an event that do not originate from the $\text{t}\bar{\text{t}}$ system, as these jets show the highest discrimination power based on an input feature analysis. The classification method is employed as node level prediction (NLP) and benchmarked against an established method which identifies the additional jets through the distance measure ΔR. It is shown that this algorithm shows a significant performance improvement compared to the ΔR method. The jet origin information is subsequently fed into the second stage, which is a graph neural network designed for multi-class event classification, enabling the generation of multiple signal regions. The performance of the network is extensively studied and tuned. Additionally, the dependence of event classification on jet identification as an input is evaluated, highlighting its exceptional importance.

The signal strength parameters are inferred by fitting the negative log-likelihood function of the statistical model to the observed data. This approach ensures an optimal determination of the signal hypotheses, as predicted by the statistical model and observed in the data. In contrast to other analyses, the strategy of this thesis is a simultaneous measurement of all four processes. This ensures a consistent approach to handling all processes and their associated uncertainties, with each process being treated with equal importance.

The statistical model incorporates a comprehensive set of systematic uncertainties arising from both experimental and theoretical sources. These uncertainties are included as nuisance parameters (NPs) in the fit. Before being applied to the data, the statistical model is rigorously validated. This involves assessing potential biases, performing a

goodness-of-fit test, determining all relevant correlations, and examining the influence of each signal strength parameter in detail. Moreover, it is shown how the design of various discriminants based on the GNN can be strategically optimized to enhance the sensitivity on specific processes. The resulting signal strength parameters from the fit to recorded data are thoroughly explored. Together with an uncertainty breakdown, the impact of all parameters on each individual signal strength parameter has been analyzed. This includes, in particular, the quantum chromodynamics (QCD) scale choice as the determining factor for $\mu_{t\bar{t}B}$, the b tagging uncertainty as the primary driver for $\mu_{t\bar{t}C}$, the gluon splitting into a pair of b quarks in $t\bar{t}$ processes as the key influence for $\mu_{t\bar{t}H}$, and the signal strength parameter of the $t\bar{t}H$ process as the dominant factor for $\mu_{t\bar{t}Z}$. The results are contextualized both in relation to SM predictions and in comparison to other analyses with comparable measurements. The findings of this thesis demonstrate good agreement with the predictions of the SM and particularly with dedicated measurements of the $t\bar{t}b\bar{b}$ processes. The measurement of the $t\bar{t}C$ processes demonstrates good compatibility with the SM, while also highlighting potential areas for improvement, such as b vs. c jet separation. For the $t\bar{t}H$ process, which has shown significant deviations from the SM prediction in the dilepton $t\bar{t}$ channel in previous measurements, the central value exhibits good compatibility with the expectations of the SM in this thesis. In the case of $t\bar{t}Z$ processes, where no related measurements in $Z \to b\bar{b}$ exist, the results show agreement with the closest related analysis, revealing a slightly larger measured cross section compared to the cross section prediction by the SM. To summarize, this thesis presents a precision measurement of the SM and reports the parameterized observed cross sections of four parameters in the $t\bar{t}$ + heavy flavor phase space for future research. This work also outlines necessary improvements in future precision measurements to enhance the sensitivity.

Future analyses will benefit significantly from improved heavy flavor jet tagging algorithms, enhancing the precision of all processes, especially $t\bar{t}C$. Additionally, exploring $t\bar{t}H$ production in the $H \to c\bar{c}$ decay channel may constrain or discover the coupling of the Higgs boson to the charm quark and provide deeper insights into potential new physics. The inclusion of data from LHC Run-III is expected to improve the precision of all processes, and in particular reduce the large statistical uncertainties that currently have a significant impact on $t\bar{t}H$ and $t\bar{t}C$. Also, a combination with other $t\bar{t}$ decay channels such as the semileptonic decay channel would significantly increase the amount of data and thus lower the statistical uncertainty. Furthermore, breaking down the $t\bar{t}B$ process into its sub-processes $t\bar{t}b\bar{b}$, $t\bar{t}2b$, and $t\bar{t}b$ will allow for a more detailed understanding of the underlying production mechanisms. This could pave the way for measurements with improved precision, offering deeper insights into the SM and potential indications of beyond the Standard Model (BSM) physics.

Acronyms

4FS four flavor scheme. 23, 66, 74, 75, 136, 137

5FS five flavor scheme. 23, 66, 74, 75, 136, 137

AI Artificial Intelligence. 4, 5, 42, 71, 97, 106, 120, 133, 146, 174

AUC Area Under the Curve. 115

B background. 127, 128

BatchNorm batch normalization. 111, 122

BSM beyond the Standard Model. 3, 4, 9, 17, 28, 42, 59--61, 175

CERN Conseil Européen pour la Recherche Nucléaire. v, 4, 25, 26, 173

CKM Cabibbo-Kobayashi-Maskawa. 19, 38

CMS Compact Muon Solenoid. v, 3, 4, 28, 29, 33, 35--38, 41--43, 52, 56, 64--70, 74, 75, 80--85, 87, 89--91, 97, 98, 104, 117, 134--138, 140, 144, 145, 153, 160, 162, 166--169, 173, 229

CP charge conjugation parity symmetry. 28

CvB c vs. b. 40, 93, 95, 103, 104, 126

CvL c vs. light. 40, 93, 95, 103, 104, 126

DGLAP Dokshitzer-Gribov-Lipatov-Altarelli-Parisi. 20

DNN deep neural network. 38, 39, 46, 48, 98, 99, 174

ECAL electromagnetic calorimeter. 30, 31, 36, 37, 62, 63, 80

ERD early resonance decays. 138

FSR final-state radiation. 43, 137, 140, 166

GLP graph level prediction. 49, 105, 106, 120, 122--131, 140--146, 148, 149, 152, 155, 157, 158, 166, 167, 169, 201--203

GNN graph neural network. 40, 45, 48--50, 72, 97--99, 105, 107, 109, 111, 120, 133, 160, 174, 175

HCAL hadronic calorimeter. 31, 36, 37, 62, 63, 135

HLT high-level trigger. 32, 76, 78, 79, 86, 87, 134

ISR initial-state radiation. 43, 137, 139, 140

itFit iterative Fit. 90--92, 135

JEC jet energy correction. 87, 89, 90, 134, 135, 140, 155, 156

JER jet energy resolution. 89, 134, 135

JES jet energy scale. 89, 134, 135

KL Kullback-Leibler. 129, 201--203

L1 Level-1. 32, 86, 134

LHC Large Hadron Collider. v, 4, 17--19, 25--28, 30, 33, 39, 41, 59--61, 66, 68, 98, 134, 136, 137, 173, 175

LO leading order. 42, 60, 61, 64, 73, 75, 76, 78, 83

LSTM Long Short-Term Memory. 39

MC Monte Carlo event simulation. 67, 84, 127, 128, 173

ME matrix element. 22, 23, 42, 43, 69, 73--76, 89, 136, 137, 167

MET missing transverse energy. 34, 35, 42, 83, 87, 93, 94, 99--101, 135

miniAOD mini Analysis Object Data. 98

MLP multilayer perceptron. 40, 46, 49, 107, 108, 111--113, 120--123, 125

nanoAOD nano Analysis Object Data. 97--99

ndf number of degrees of freedom. 82, 151

NLO Next-to-leading order. 42, 66, 70, 71, 73, 75, 76, 89, 90, 136, 137, 139

NLP node level prediction. 49, 105, 111--114, 116--120, 123, 129--131, 140, 174

NNLO Next-to-next-to-leading order. 74, 76, 89, 90, 136, 139

NP nuisance parameter. 53--55, 133--136, 139, 140, 144, 147, 148, 150, 152--156, 159--166, 169, 174, 205--228

PairNorm pair normalization. 111, 113

PDF parton distribution function. 17, 20--23, 74, 136, 137, 156

PDF probability density function. 51, 55, 151

PF Particle Flow. 82

POI parameter of interest. 51, 53, 54, 140, 144, 147, 148, 150--153, 155, 156, 159--163, 166, 167

PS parton shower. 23, 43, 66, 74, 137, 138, 155, 156

QCD quantum chromodynamics. 4, 11--13, 19, 20, 22, 37, 42, 60, 66, 71, 73--76, 90, 136, 138, 147, 148, 154--156, 160, 163, 166, 175

QED quantum electrodynamics. 10--13

RGE renormalization group equation. 20, 22, 148

ROC Receiver Operating Characteristic. 40, 41, 115

ROC AUC Receiver Operating Characteristic - Area Under the Curve. 112--114, 124, 125, 130

S signal. 127, 128

SC supercluster. 80, 81

SF scale factor. 85--91, 104, 134, 135, 166

SM Standard Model of particle physics. v, 3, 4, 9, 13--17, 25, 28, 42, 52, 59--61, 65, 68, 70, 72, 140, 141, 150, 154, 161, 167, 169, 173, 175

WP working point. 41, 79--83, 90, 93, 95

Bibliography

[1] P. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13 (16 Oct. 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508.

[2] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (9 Aug. 1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321.

[3] G. S. Guralnik, C. R. Hagen, and T. W. Kibble. "Global Conservation Laws and Massless Particles". In: *Phys. Rev. Lett.* 13 (20 Nov. 1964), pp. 585–587. DOI: 10.1103/PhysRevLett.13.585.

[4] Y. Nambu and G. Jona-Lasinio. "Dynamical Model of Elementary Particles Based on an Analogy with Superconductivity. I". In: *Phys. Rev.* 122 (1 Apr. 1961), pp. 345–358. DOI: 10.1103/PhysRev.122.345.

[5] J. Goldstone, A. Salam, and S. Weinberg. "Broken Symmetries". In: *Phys. Rev.* 127 (3 Aug. 1962), pp. 965–970. DOI: 10.1103/PhysRev.127.965.

[6] T. W. Kibble. "Symmetry Breaking in Non-Abelian Gauge Theories". In: *Phys. Rev.* 155 (5 Mar. 1967), pp. 1554–1561. DOI: 10.1103/PhysRev.155.1554.

[7] The ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 1–29. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2012.08.020.

[8] The CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 30–61. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2012.08.021.

[9] R. Wolf. *The Higgs Boson Discovery at the Large Hadron Collider*. Springer, 2015. ISBN: 978-3-319-18511-8. DOI: 10.1007/978-3-319-18512-5.

[10] S. Tomonaga. "On a Relativistically Invariant Formulation of the Quantum Theory of Wave Fields". In: *Progress of Theoretical Physics* 1.2 (Aug. 1946), pp. 27–42. ISSN: 0033-068X. DOI: 10.1143/PTP.1.27.

[11] J. Schwinger. "Quantum Electrodynamics. I. A Covariant Formulation". In: *Phys. Rev.* 74 (10 Nov. 1948), pp. 1439–1461. DOI: 10.1103/PhysRev.74.1439.

[12] J. Schwinger. "On Quantum-Electrodynamics and the Magnetic Moment of the Electron". In: *Phys. Rev.* 73 (4 Feb. 1948), pp. 416–417. DOI: 10.1103/PhysRev.73.416.

[13] R. P. Feynman. "Mathematical Formulation of the Quantum Theory of Electromagnetic Interaction". In: *Phys. Rev.* 80 (3 Nov. 1950), pp. 440–457. DOI: 10.1103/PhysRev.80.440.

[14] R. P. Feynman. "Space-Time Approach to Quantum Electrodynamics". In: *Phys. Rev.* 76 (6 Sept. 1949), pp. 769–789. DOI: 10.1103/PhysRev.76.769.

[15] F. J. Dyson. "The S Matrix in Quantum Electrodynamics". In: *Phys. Rev.* 75 (11 June 1949), pp. 1736–1755. DOI: 10.1103/PhysRev.75.1736.

[16] P. A. M. Dirac. "The quantum theory of the electron". In: *Proc. Roy. Soc. Lond. A* 117 (1928), pp. 610–624. DOI: 10.1098/rspa.1928.0023.

[17] E. Noether. "Invariant variation problems". In: *Transport Theory and Statistical Physics* 1.3 (1971), pp. 186–207. DOI: 10.1080/00411457108231446.

[18] H. Weyl. "A New Extension of Relativity Theory". In: *Annalen Phys.* 59 (1919), pp. 101–133. DOI: 10.1002/andp.19193641002.

[19] M. Gell-Mann. "Symmetries of Baryons and Mesons". In: *Phys. Rev.* 125 (3 Feb. 1962), pp. 1067–1084. DOI: 10.1103/PhysRev.125.1067.

[20] M. Gell-Mann. "A schematic model of baryons and mesons". In: *Physics Letters* 8.3 (1964), pp. 214–215. ISSN: 0031-9163. DOI: 10.1016/S0031-9163(64)92001-3.

[21] G. Zweig. "An SU_3 model for strong interaction symmetry and its breaking; Version 2". In: (1964). URL: https://cds.cern.ch/record/570209.

[22] D. J. Gross and F. Wilczek. "Ultraviolet Behavior of Non-Abelian Gauge Theories". In: *Phys. Rev. Lett.* 30 (26 June 1973), pp. 1343–1346. DOI: 10.1103/PhysRevLett.30.1343.

[23] D. J. Gross and F. Wilczek. "Asymptotically Free Gauge Theories. I". In: *Phys. Rev. D* 8 (10 Nov. 1973), pp. 3633–3652. DOI: 10.1103/PhysRevD.8.3633.

[24] D. J. Gross and F. Wilczek. "Asymptotically free gauge theories. II". In: *Phys. Rev. D* 9 (4 Feb. 1974), pp. 980–993. DOI: 10.1103/PhysRevD.9.980.

[25] E. Fermi. "Versuch einer Theorie der β-Strahlen. I". In: *Zeitschrift für Physik* 88.3 (Mar. 1934), pp. 161–177. ISSN: 0044-3328. DOI: 10.1007/BF01351864.

[26] C. S. Wu et al. "Experimental Test of Parity Conservation in Beta Decay". In: *Phys. Rev.* 105 (4 Feb. 1957), pp. 1413–1415. DOI: 10.1103/PhysRev.105.1413.

[27] M. Goldhaber, L. Grodzins, and A. W. Sunyar. "Helicity of Neutrinos". In: *Phys. Rev.* 109 (3 Feb. 1958), pp. 1015–1017. DOI: 10.1103/PhysRev.109.1015.

[28] S. L. Glashow. "Partial-symmetries of weak interactions". In: *Nuclear Physics* 22.4 (1961), pp. 579–588. ISSN: 0029-5582. DOI: 10.1016/0029-5582(61)90469-2.

[29] A. Salam and J. C. Ward. "Weak and electromagnetic interactions". In: *Il Nuovo Cimento (1955-1965)* 11.4 (1959), pp. 568–577. ISSN: 1827-6121. DOI: 10.1007/BF02726525.

[30] S. Weinberg. "A Model of Leptons". In: *Phys. Rev. Lett.* 19 (21 Nov. 1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264.

[31] T. Nakano and K. Nishijima. "Charge Independence for V-particles". In: *Progress of Theoretical Physics* 10.5 (Nov. 1953), pp. 581–582. ISSN: 0033-068X. DOI: 10.1143/PTP.10.581.

[32] K. Nishijima. "Charge Independence Theory of V Particles". In: *Progress of Theoretical Physics* 13.3 (Mar. 1955), pp. 285–304. ISSN: 0033-068X. DOI: 10.1143/PTP.13.285.

[33] M. Gell-Mann. "The interpretation of the new particles as displaced charge multiplets". In: *Il Nuovo Cimento (1955-1965)* 4.2 (Apr. 1956), pp. 848–866. ISSN: 1827-6121. DOI: 10.1007/BF02748000.

[34] The UA1 Collaboration. "Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s} = 540\,\text{GeV}$". In: *Physics Letters B* 122.1 (1983), pp. 103–116. ISSN: 0370-2693. DOI: 10.1016/0370-2693(83)91177-2.

[35] The UA2 Collaboration. "Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider". In: *Physics Letters B* 122.5 (1983), pp. 476–485. ISSN: 0370-2693. DOI: 10.1016/0370-2693(83)91605-2.

[36] The UA1 Collaboration. "Experimental observation of lepton pairs of invariant mass around $95\,\text{GeV}/c^2$ at the CERN SPS collider". In: *Physics Letters B* 126.5 (1983), pp. 398–410. ISSN: 0370-2693. DOI: 10.1016/0370-2693(83)90188-0.

[37] J. Ellis. *Higgs Physics*. 2013. DOI: 10.48550/arXiv.1312.5672. arXiv: 1312.5672 [hep-ph].

[38] H. Yukawa. "On the Interaction of Elementary Particles. I". In: *Progress of Theoretical Physics Supplement* 1 (Jan. 1955), pp. 1–10. ISSN: 0375-9687. DOI: 10.1143/PTPS.1.1.

[39] W. Pauli. "Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren". In: *Zeitschrift für Physik* 31.1 (Feb. 1925), pp. 765–783. ISSN: 0044-3328. DOI: 10.1007/BF02980631.

[40] P. A. M. Dirac. "On the Theory of quantum mechanics". In: *Proc. Roy. Soc. Lond. A* 112 (1926), pp. 661–677. DOI: 10.1098/rspa.1926.0133.

[41] A. Einstein. "Quantentheorie des einatomigen idealen Gases". In: *Albert Einstein: Akademie-Vorträge*. John Wiley & Sons, Ltd, 2005, pp. 237–244. ISBN: 9783527608959. DOI: 10.1002/3527608958.ch27.

[42] S. Bose. "Plancks Gesetz und Lichtquantenhypothese". In: *Zeitschrift für Physik* 26.1 (Dec. 1924), pp. 178–181. ISSN: 0044-3328. DOI: 10.1007/BF01327326.

[43] S. Navas et al. "Review of particle physics". In: *Phys. Rev. D* 110.3 (2024), p. 030001. DOI: 10.1103/PhysRevD.110.030001.

[44] The CDF Collaboration. "Observation of Top Quark Production in $\overline{p}p$ Collisions with the Collider Detector at Fermilab". In: *Phys. Rev. Lett.* 74 (14 Apr. 1995), pp. 2626–2631. DOI: 10.1103/PhysRevLett.74.2626.

[45] The D0 Collaboration. "Observation of the Top Quark". In: *Phys. Rev. Lett.* 74 (14 Apr. 1995), pp. 2632–2637. DOI: 10.1103/PhysRevLett.74.2632.

[46] U. Husemann. "Top-quark physics: Status and prospects". In: *Progress in Particle and Nuclear Physics* 95 (July 2017), pp. 48–97. ISSN: 0146-6410. DOI: 10.1016/j.ppnp.2017.03.002.

[47] The CMS Collaboration. *Review of top quark mass measurements in CMS*. 2024. arXiv: 2403.01313 [hep-ex].

[48] The ATLAS Collaboration. "Observation of quantum entanglement with top quarks at the ATLAS detector". In: *Nature* 633.8030 (Sept. 2024), pp. 542–547. ISSN: 1476-4687. DOI: 10.1038/s41586-024-07824-z.

[49] The CMS Collaboration. "Observation of quantum entanglement in top quark pair production in proton–proton collisions at 13 TeV". In: *Reports on Progress in Physics* 87.11 (Oct. 2024), p. 117801. ISSN: 1361-6633. DOI: 10.1088/1361-6633/ad7e4d.

[50] M. Czakon, P. Fiedler, and A. Mitov. "Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $\mathcal{O}(\alpha_S^4)$". In: *Phys. Rev. Lett.* 110 (25 June 2013), p. 252004. DOI: 10.1103/PhysRevLett.110.252004.

[51] The ATLAS Collaboration. *Top cross section summary plots - April 2024*. Tech. rep. Geneva: CERN, 2024. URL: https://cds.cern.ch/record/2896104.

[52] N. Cabibbo. "Unitary Symmetry and Leptonic Decays". In: *Phys. Rev. Lett.* 10 (12 June 1963), pp. 531–533. DOI: 10.1103/PhysRevLett.10.531.

[53] M. Kobayashi and T. Maskawa. "CP-Violation in the Renormalizable Theory of Weak Interaction". In: *Progress of Theoretical Physics* 49.2 (Feb. 1973), pp. 652–657. ISSN: 0033-068X. DOI: 10.1143/PTP.49.652.

[54] L. Wolfenstein. "Parametrization of the Kobayashi-Maskawa Matrix". In: *Phys. Rev. Lett.* 51 (21 Nov. 1983), pp. 1945–1947. DOI: 10.1103/PhysRevLett.51.1945.

[55] I. Neutelings. *Branching fraction matrix of pair decays*. https://tikz.net/axis3d_cms. Accessed: 12-2024.

[56] Y. L. Dokshitzer. "Calculation of the Structure Functions for Deep Inelastic Scattering and e+ e- Annihilation by Perturbation Theory in Quantum Chromodynamics." In: *Sov. Phys. JETP* 46 (1977), pp. 641–653.

[57] V. Gribov and L. Lipatov. "Deep inelastic electron scattering in perturbation theory". In: *Physics Letters B* 37.1 (1971), pp. 78–80. ISSN: 0370-2693. DOI: 10.1016/0370-2693(71)90576-4.

[58] G. Altarelli and G. Parisi. "Asymptotic freedom in parton language". In: *Nuclear Physics B* 126.2 (1977), pp. 298–318. ISSN: 0550-3213. DOI: 10.1016/0550-3213(77)90384-4.

[59] H. David Politzer. "Asymptotic freedom: An approach to strong interactions". In: *Physics Reports* 14.4 (1974), pp. 129–180. ISSN: 0370-1573. DOI: 10.1016/0370-1573(74)90014-3.

[60] R. D. Ball et al. "Parton distributions from high-precision collider data: NNPDF Collaboration". In: *The European Physical Journal C* 77.10 (Oct. 2017). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-017-5199-5.

[61] L. Evans and P. Bryant. "LHC Machine". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08001. DOI: 10.1088/1748-0221/3/08/S08001.

[62] E. Lopienska. "The CERN accelerator complex, layout in 2022. Complexe des accélérateurs du CERN en janvier 2022". In: (2022). General Photo. URL: https://cds.cern.ch/record/2800984.

[63] The ALICE Collaboration. "The ALICE experiment at the CERN LHC". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08002–S08002. DOI: 10.1088/1748-0221/3/08/s08002.

[64] The LHCb Collaboration. "The LHCb Detector at the LHC". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08005–S08005. DOI: 10.1088/1748-0221/3/08/s08005.

[65] The ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08003–S08003. DOI: 10.1088/1748-0221/3/08/s08003.

[66] The CMS Collaboration. "The CMS experiment at the CERN LHC". In: *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.

[67] T. Sakuma. *Cutaway diagrams of CMS detector*. 2019. URL: https://cds.cern.ch/record/2665537.

[68] The CMS Collaboration. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". In: *JINST* 9 (2014), P10009. DOI: 10.1088/1748-0221/9/10/P10009. arXiv: 1405.6569 [physics.ins-det].

[69] The CMS Collaboration. "Strategies and performance of the CMS silicon tracker alignment during LHC Run 2". In: *Nucl. Instrum. Meth. A* 1037 (2022), p. 166795. DOI: 10.1016/j.nima.2022.166795. arXiv: 2111.08757 [physics.ins-det].

[70] The Tracker Groupof the CMS Collaboration. *The CMS Phase-1 Pixel Detector Upgrade*. Tech. rep. Geneva: CERN, 2020. URL: https://cds.cern.ch/record/2745805.

[71] The CMS Collaboration. *The CMS tracker system project: Technical Design Report*. Technical design report. CMS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/368412.

[72] The CMS Collaboration. *The CMS tracker: addendum to the Technical Design Report*. Technical design report. CMS. Geneva: CERN, 2000. URL: https://cds.cern.ch/record/490194.

[73] The CMS Collaboration. "Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC". In: *JINST* 16 (2021), P05014. DOI: 10.1088/1748-0221/16/05/P05014. arXiv: 2012.06888 [hep-ex].

[74] The CMS Collaboration. *ECAL 2016 refined calibration and Run2 summary plots*. CMS Detector Performance Summary CMS-DP-2020-021. 2020. URL: https://cds.cern.ch/record/2717925.

[75] The CMS Collaboration. *Technical proposal for the upgrade of the CMS detector through 2020*. Tech. rep. 2011. URL: https://cds.cern.ch/record/1355706.

[76] The CMS Collaboration. *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical design report. CMS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/349375.

[77] The CMS Collaboration. *Changes to CMS ECAL electronics: addendum to the Technical Design Report*. Technical design report. CMS. Geneva: CERN, 2002. URL: https://cds.cern.ch/record/581342.

[78] The CMS Collaboration. *CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter*. Tech. rep. 2012. URL: https://cds.cern.ch/record/1481837.

[79] The CMS Collaboration. *The CMS hadron calorimeter project: Technical Design Report*. Technical design report. CMS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/357153.

[80] The CMS Collaboration. *The CMS magnet project: Technical Design Report*. Technical design report. CMS. Geneva: CERN, 1997. DOI: 10.17181/CERN.6ZU0.V4T9.

[81] The CMS Collaboration. "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$". In: *JINST* 13 (2018), P06015. DOI: 10.1088/1748-0221/13/06/P06015. arXiv: 1804.04528 [physics.ins-det].

[82] The CMS Collaboration. *The CMS muon project: Technical Design Report*. Technical design report. CMS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/343814.

[83] The CMS Collaboration. "Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$". In: *JINST* 15 (2020), P10017. DOI: 10.1088/1748-0221/15/10/P10017.

[84] The CMS Collaboration. "The CMS trigger system". In: *Journal of Instrumentation* 12.01 (Jan. 2017), P01020. DOI: 10.1088/1748-0221/12/01/P01020.

[85] The CMS Collaboration. "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ using the CMS detector". In: *JINST* 14 (2019), P07004. DOI: 10.1088/1748-0221/14/07/P07004. arXiv: 1903.06078 [hep-ex].

[86] The CMS Collaboration. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". In: *Journal of Instrumentation* 9.10 (Oct. 2014), P10009. DOI: 10.1088/1748-0221/9/10/P10009.

[87] P. Billoir. "Progressive track recognition with a Kalman-like fitting procedure". In: *Computer Physics Communications* 57.1 (1989), pp. 390–394. ISSN: 0010-4655. DOI: 10.1016/0010-4655(89)90249-X.

[88] W. Waltenberger, R. Frühwirth, and P. Vanlaer. "Adaptive vertex fitting". In: *Journal of Physics G: Nuclear and Particle Physics* 34.12 (Nov. 2007), N343. DOI: 10.1088/0954-3899/34/12/N01.

[89] The CMS Collaboration. "Particle-flow reconstruction and global event description with the CMS detector". In: *JINST* 12 (2017), P10003. DOI: 10.1088/1748-0221/12/10/P10003.

[90] M. Cacciari, G. P. Salam, and G. Soyez. "The anti-kt jet clustering algorithm". In: *Journal of High Energy Physics* 2008.04 (Apr. 2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063.

[91] The CMS Collaboration. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV". In: *Journal of Instrumentation* 13.05 (May 2018), P05011–P05011. ISSN: 1748-0221. DOI: 10.1088/1748-0221/13/05/p05011.

[92] E. Bols et al. "Jet flavour classification using DeepJet". In: *Journal of Instrumentation* 15.12 (Dec. 2020), P12012–P12012. ISSN: 1748-0221. DOI: 10.1088/1748-0221/15/12/p12012.

[93] H. Qu and L. Gouskos. "Jet tagging via particle clouds". In: *Phys. Rev. D* 101 (5 Mar. 2020), p. 056019. DOI: 10.1103/PhysRevD.101.056019.

[94] H. Qu, C. Li, and S. Qian. *Particle Transformer for Jet Tagging*. 2024. arXiv: 2202.03772 [hep-ph].

[95] E. Pfeffer et al. "Heavy flavor jet tagging algorithm developments at CMS for HL-LHC". In: *PoS* LHCP2022 (2023), p. 270. DOI: 10.22323/1.422.0270.

[96] The CMS Collaboration. "Inclusive and differential cross section measurements of $t\bar{t}b\bar{b}$ production in the lepton+jets channel at $\sqrt{s} = 13\,\text{TeV}$". In: *Journal of High Energy Physics* 2024.5 (May 2024). ISSN: 1029-8479. DOI: 10.1007/jhep05(2024)042.

[97] The CMS Collaboration. "Performance summary of AK4 jet b tagging with data from proton-proton collisions at 13 TeV with the CMS detector". In: (2023). URL: https://cds.cern.ch/record/2854609.

[98] S. Frixione, P. Nason, and C. Oleari. "Matching NLO QCD computations with parton shower simulations: the POWHEG method". In: *Journal of High Energy Physics* 2007.11 (Nov. 2007), pp. 070–070. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2007/11/070.

[99] S. Alioli et al. "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX". In: *Journal of High Energy Physics* 2010.6 (June 2010). ISSN: 1029-8479. DOI: 10.1007/jhep06(2010)043.

[100] T. Sjöstrand et al. "An introduction to PYTHIA 8.2". In: *Computer Physics Communications* 191 (2015), pp. 159–177. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2015.01.024.

[101] V. V. Sudakov. "Vertex parts at very high-energies in quantum electrodynamics". In: *Sov. Phys. JETP* 3 (1956), pp. 65–71.

[102] B. Andersson et al. "Parton fragmentation and string dynamics". In: *Physics Reports* 97.2 (1983), pp. 31–145. ISSN: 0370-1573. DOI: 10.1016/0370-1573(83)90080-7.

[103] S. Agostinelli et al. "Geant4 - a simulation toolkit". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: 10.1016/S0168-9002(03)01368-8.

[104] J. de Favereau et al. "DELPHES 3: a modular framework for fast simulation of a generic collider experiment". In: *Journal of High Energy Physics* 2014.2 (Feb. 13, 2014), p. 57. ISSN: 1029-8479. DOI: 10.1007/JHEP02(2014)057.

[105] S. Lelyakin. "Monte-Carlo-Simulations for tt+X-Events at the CMS-Experiment". Bachelor thesis. Karlsruhe Institute of Technology (KIT), 2022.

[106] M. Cacciari and G. P. Salam. "Pileup subtraction using jet areas". In: *Physics Letters B* 659.1 (2008), pp. 119–126. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2007.09.077.

[107] The CMS Collaboration. *Object definitions for top quark analyses at the particle level*. Tech. rep. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2267573.

[108] D. P. Kingma and J. Ba. *Adam: A Method for Stochastic Optimization*. 2017. arXiv: 1412.6980 [cs.LG].

[109] N. Srivastava et al. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: *Journal of Machine Learning Research* 15 (June 2014), pp. 1929–1958.

[110] P. Hartmann. *Mathematik für Informatiker*. Springer Vieweg Wiesbaden, Mar. 2020. ISBN: 978-3-658-26524-3. DOI: 10.1007/978-3-658-26524-3.

[111] E. Pfeffer et al. "A Case Study of Sending Graph Neural Networks Back to the Test Bench for Applications in High-Energy Particle Physics". In: *Computing and Software for Big Science* 8.1 (July 2024). ISSN: 2510-2044. DOI: 10.1007/s41781-024-00122-3.

[112] A. Vaswani et al. *Attention Is All You Need*. 2023. arXiv: 1706.03762 [cs.CL].

[113] The LHC Higgs Combination Group. *Procedure for the LHC Higgs boson search combination in Summer 2011*. Tech. rep. Geneva: CERN, 2011. URL: https://cds.cern.ch/record/1379837.

[114] G. Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *The European Physical Journal C* 71.2 (Feb. 2011). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-011-1554-0.

[115] G. Cowan. *Topics in statistical data analysis for high-energy physics*. 2010. arXiv: 1012.3589 [physics.data-an].

[116] J. Neyman and E. S. Pearson. "On the Problem of the Most Efficient Tests of Statistical Hypotheses". In: *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 231.694-706 (1933), pp. 289–337. DOI: 10.1098/rsta.1933.0009.

[117] S. S. Wilks. "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses". In: *Annals Math. Statist.* 9.1 (1938), pp. 60–62. DOI: 10.1214/aoms/1177732360.

[118] T. McCauley. "A browser-based event display for the CMS Experiment at the LHC using WebGL". In: *Journal of Physics: Conference Series* 898.7 (Oct. 2017), p. 072030. DOI: 10.1088/1742-6596/898/7/072030.

[119] The CMS Collaboration. "Evidence for four-top quark production in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$". In: *Physics Letters B* 844 (2023), p. 138076. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2023.138076.

[120] The CMS Collaboration. "Observation of four top quark production in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$". In: *Physics Letters B* 847 (Dec. 2023), p. 138290. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2023.138290.

[121] The CMS Collaboration. *Stairway to discovery: a report on the CMS programme of cross section measurements from millibarns to femtobarns*. 2024. arXiv: 2405.18661 [hep-ex].

[122] The LHC Higgs Cross Section Working Group. *CERN Yellow Reports: Monographs, Vol 2 (2017): Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector*. en. 2017. DOI: 10.23731/CYRM-2017-002.

[123] A. Kulesza et al. "Associated production of a top quark pair with a heavy electroweak gauge boson at NLO+NNLL accuracy". In: *The European Physical Journal C* 79.3 (Mar. 2019). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-019-6746-z.

[124] L. Ferencz et al. *Study of ttbb and ttW background modelling for ttH analyses*. 2023. arXiv: 2301.11670 [hep-ex].

[125] T. Ježo et al. "New NLOPS predictions for $t\bar{t} + b$-jet production at the LHC". In: *The European Physical Journal C* 78.6 (June 2018). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-018-5956-0.

[126] The CMS Collaboration. *Measurement of the $\mathrm{t\bar{t}}H$ and tH production rates in the $H \to \mathrm{b\bar{b}}$ decay channel using proton-proton collision data at $\sqrt{s} = 13\ TeV$*. 2024. arXiv: 2407.10896 [hep-ex].

[127] The ATLAS Collaboration. "Measurements of inclusive and differential fiducial cross-sections of $t\bar{t}$ production with additional heavy-flavour jets in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector". In: *Journal of High Energy Physics* 2019.4 (Apr. 2019). ISSN: 1029-8479. DOI: 10.1007/jhep04(2019)046.

[128] The CMS Collaboration. "Measurements of $\mathrm{t\bar{t}}$ cross sections in association with b jets and inclusive jets and their ratio using dilepton final states in pp collisions at $\sqrt{s} = 13\,\text{TeV}$". In: *Physics Letters B* 776 (2018), pp. 355–378. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2017.11.043.

[129] The CMS Collaboration. "Measurement of the $\mathrm{t\bar{t}b\bar{b}}$ production cross section in the all-jet final state in pp collisions at $\sqrt{s} = 13\,\text{TeV}$". In: *Physics Letters B* 803 (2020), p. 135285. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2020.135285.

[130] The ATLAS Collaboration. *Measurement of $t\bar{t}$ production in association with additional b-jets in the $e\mu$ final state in proton-proton collisions at $\sqrt{s} = 13\ TeV$ with the ATLAS detector*. 2024. arXiv: 2407.13473 [hep-ex].

[131] E. Pfeffer. *Inclusive and differential cross section measurements of $t\bar{t}b\bar{b}$ production in the lepton+jets channel at $\sqrt{s} = 13$ TeV*. 2024. arXiv: 2401.06743 [hep-ex].

[132] J. van der Linden. "Inclusive and differential cross section measurement of $t\bar{t}b\bar{b}$ production and studies of $t\bar{t}$ production with additional jet radiation". PhD thesis. Karlsruher Institut für Technologie (KIT), 2023. 221 pp. DOI: 10.5445/IR/1000160616.

[133] The CMS Collaboration. *Inclusive and differential cross section measurements of $\mathrm{t\bar{t}b\bar{b}}$ production in the lepton+jets channel at $\sqrt{s} = 13$ TeV*. HEPData (collection). 2022. DOI: 10.17182/hepdata.138416.

[134] The CMS Collaboration. "Search for $\mathrm{t\bar{t}H}$ production in the $\mathrm{H} \to \mathrm{b\bar{b}}$ decay channel with leptonic $\mathrm{t\bar{t}}$ decays in proton-proton collisions at $\sqrt{s} = 13\,\mathrm{TeV}$". In: *Journal of High Energy Physics* 2019.3 (Mar. 2019). ISSN: 1029-8479. DOI: 10.1007/jhep03(2019)026.

[135] The CMS Collaboration. "Search for $\mathrm{t\bar{t}H}$ production in the all-jet final state in proton-proton collisions at $\sqrt{s} = 13\,\mathrm{TeV}$". In: *Journal of High Energy Physics* 2018.6 (June 2018). ISSN: 1029-8479. DOI: 10.1007/jhep06(2018)101.

[136] The ATLAS Collaboration. "Measurement of Higgs boson decay into b-quarks in associated production with a top-quark pair in pp collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ with the ATLAS detector". In: *Journal of High Energy Physics* 2022.6 (June 2022). ISSN: 1029-8479. DOI: 10.1007/jhep06(2022)097.

[137] The ATLAS Collaboration. *Measurement of the associated production of a top-antitop-quark pair and a Higgs boson decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*. 2024. arXiv: 2407.10904 [hep-ex].

[138] E. Pfeffer. *A Rivet routine for the Study of ttbb and ttW background modelling for ttH analyses*. github. https://github.com/emanuelpf/rivet3-analyses. 2024.

[139] J. van der Linden. "Limit on $t\bar{t} + Z$ production in the $Z \to b\bar{b}$ channel at the CMS experiment". Master thesis. Karlsruhe Institute of Technology, 2019. URL: https://cds.cern.ch/record/2703587.

[140] The CMS Collaboration. "Inclusive and differential cross section measurements of single top quark production in association with a Z boson in proton-proton collisions at $\sqrt{s} = 13\,\mathrm{TeV}$". In: *Journal of High Energy Physics* 2022.2 (Feb. 2022). ISSN: 1029-8479. DOI: 10.1007/jhep02(2022)107.

[141] The CMS Collaboration. "Search for new physics using effective field theory in 13 TeV pp collision events that contain a top quark pair and a boosted Z or Higgs boson". In: *Phys. Rev. D* 108 (3 Aug. 2023), p. 032008. DOI: 10.1103/PhysRevD.108.032008.

[142] The CMS Collaboration. "First measurement of the cross section for top quark pair production with additional charm jets using dileptonic final states in pp collisions at $\sqrt{s} = 13\,\text{TeV}$". In: *Physics Letters B* 820 (2021), p. 136565. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2021.136565.

[143] R. Frederix and S. Frixione. "Merging meets matching in MC at NLO". In: *Journal of High Energy Physics* 2012.12 (2012), p. 61. ISSN: 1029-8479. DOI: 10.1007/JHEP12(2012)061.

[144] The ATLAS Collaboration. *Measurement of top-quark pair production in association with charm quarks in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*. 2024. arXiv: 2409.11305 [hep-ex].

[145] The CMS Collaboration. "Extraction and validation of a new set of CMS pythia8 tunes from underlying-event measurements". In: *The European Physical Journal C* 80.1 (Jan. 2020). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-019-7499-4.

[146] T. Ježo and P. Nason. "On the treatment of resonances in next-to-leading order calculations matched to a parton shower". In: *Journal of High Energy Physics* 2015.12 (Dec. 2015), pp. 1–47. ISSN: 1029-8479. DOI: 10.1007/jhep12(2015)065.

[147] F. Buccioni et al. "OpenLoops 2". In: *The European Physical Journal C* 79.10 (Oct. 2019). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-019-7306-2.

[148] 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". In: *Journal of High Energy Physics* 2014.7 (July 2014). ISSN: 1029-8479. DOI: 10.1007/jhep07(2014)079.

[149] P. Artoisenet et al. "Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations". In: *Journal of High Energy Physics* 2013.3 (Mar. 2013). ISSN: 1029-8479. DOI: 10.1007/jhep03(2013)015.

[150] M. Czakon and A. Mitov. "Top++: A program for the calculation of the top-pair cross-section at hadron colliders". In: *Computer Physics Communications* 185.11 (2014), pp. 2930–2938. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2014.06.021.

[151] N. Kidonakis and N. Yamanaka. "Higher-order corrections for tW production at high-energy hadron colliders". In: *Journal of High Energy Physics* 2021.5 (May 2021). ISSN: 1029-8479. DOI: 10.1007/jhep05(2021)278.

[152] J. Campbell, T. Neumann, and Z. Sullivan. "Single-top-quark production in the t-channel at NNLO". In: *Journal of High Energy Physics* 2021.2 (Feb. 2021). ISSN: 1029-8479. DOI: 10.1007/jhep02(2021)040.

[153] The CMS Collaboration. "W^+W^- boson pair production in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$". In: *Phys. Rev. D* 102 (9 Nov. 2020), p. 092001. DOI: 10.1103/PhysRevD.102.092001.

[154] J. M. Campbell, R. K. Ellis, and C. Williams. "Vector boson pair production at the LHC". In: *Journal of High Energy Physics* 2011.7 (July 2011). ISSN: 1029-8479. DOI: 10.1007/jhep07(2011)018.

[155] The CMS Collaboration. *Measurement of the pp → ZZ production cross section at $\sqrt{s} = 13$ TeV with the Run 2 data set*. Tech. rep. Geneva: CERN, 2019. URL: https://cds.cern.ch/record/2668717.

[156] Y. Li and F. Petriello. "Combining QCD and electroweak corrections to dilepton production in the framework of the FEWZ simulation code". In: *Phys. Rev. D* 86 (9 Nov. 2012), p. 094034. DOI: 10.1103/PhysRevD.86.094034.

[157] W. Adam, R. Frühwirth, and T. Strandlie Are and Todor. *Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC*. Tech. rep. Geneva: CERN, 2005. URL: https://cds.cern.ch/record/815410.

[158] The CMS Collaboration. "Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC". In: *Journal of Instrumentation* 16.05 (May 2021), P05014. DOI: 10.1088/1748-0221/16/05/P05014.

[159] The CMS Collaboration. "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Journal of Instrumentation* 13.06 (June 2018), P06015. DOI: 10.1088/1748-0221/13/06/P06015.

[160] The CMS Collaboration. "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV". In: *Journal of Instrumentation* 12.02 (Feb. 2017), P02014. DOI: 10.1088/1748-0221/12/02/P02014.

[161] The CMS Collaboration. *Jet algorithms performance in 13 TeV data*. Tech. rep. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2256875.

[162] The CMS Collaboration. "Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV". In: *Journal of High Energy Physics* 2018.7 (July 2018). ISSN: 1029-8479. DOI: 10.1007/jhep07(2018)161.

[163] D. Meuser. "Measurement of the differential dileptonic $t\bar{t}$ cross section in a BSM phase space with the CMS detector using the full LHC Run 2 data set". Presented 08 Dec 2023. PhD thesis. RWTH Aachen U., 2023. DOI: 10.18154/RWTH-2023-11969.

[164] The CMS Collaboration. *Measurement of the dineutrino system kinematics in dileptonic top quark pair events in pp collisions at $\sqrt{s} = 13$ TeV*. Tech. rep. Geneva: CERN, 2024. URL: https://cds.cern.ch/record/2905837.

[165] The CMS Collaboration. *Differential cross section measurements for the production of top quark pairs and of additional jets using dilepton events from pp collisions at $\sqrt{s} = 13$ TeV*. 2024. arXiv: 2402.08486 [hep-ex].

[166] M. Czakon et al. "Top-pair production at the LHC through NNLO QCD and NLO EW". In: *Journal of High Energy Physics* 2017.10 (Oct. 2017). ISSN: 1029-8479. DOI: 10.1007/jhep10(2017)186.

[167] E. Pfeffer. "Studies on $t\bar{t} + b\bar{b}$ production at the CMS experiment". Master thesis. Karlsruhe Institute of Technology, 2021. URL: https://cds.cern.ch/record/2919021.

[168] T. Halenke. "Studien zu Graph Neural Networks in tt + bb- Prozessen am CMS-Experiment". Bachelor thesis. Karlsruhe Institute of Technology (KIT), 2021.

[169] C. Wolter. "Bayesian Optimization of Graph Neural Networks with Hypergraph Inputs in tt+bb Events at the CMS Experiment". Bachelor thesis. Karlsruhe Institute of Technology (KIT), 2022.

[170] Y.-Y. C. Cung. "Feasibility and Reliability Studies of Graph Neural Networks for Multivariate tt+X Event Classification at the CMS Experiment at CERN". Master thesis. Karlsruhe Institute of Technology (KIT), 2022.

[171] M. Erhart. "Simultaneous Cross Section Measurements of tt + X processes using Graph Neural Networks." Master thesis. Karlsruhe Institute of Technology (KIT), 2022.

[172] P. Oßwald. "Generating Explanations for Graph Neural Networks for tt+X events at the CMS Experiment". Bachelor thesis. Karlsruhe Institute of Technology (KIT), 2023.

[173] Y. Shi et al. *Masked Label Prediction: Unified Message Passing Model for Semi-Supervised Classification*. 2021. arXiv: 2009.03509 [cs.LG].

[174] J. L. Ba, J. R. Kiros, and G. E. Hinton. *Layer Normalization*. 2016. arXiv: 1607.06450 [stat.ML].

[175] S. Ioffe and C. Szegedy. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. 2015. arXiv: 1502.03167 [cs.LG].

[176] L. Zhao and L. Akoglu. *PairNorm: Tackling Oversmoothing in GNNs*. 2020. arXiv: 1909.12223 [cs.LG].

[177] H. Gao and S. Ji. *Graph U-Nets*. 2019. arXiv: 1905.05178 [cs.LG].

[178] C. Cangea et al. *Towards Sparse Hierarchical Graph Classifiers*. 2018. arXiv: 1811.01287 [stat.ML].

[179] B. Knyazev, G. W. Taylor, and M. R. Amer. *Understanding Attention and Generalization in Graph Neural Networks*. 2019. arXiv: 1905.02850 [cs.LG].

[180] T.-Y. Lin et al. *Focal Loss for Dense Object Detection*. 2018. arXiv: 1708.02002 [cs.CV].

[181] The CMS Collaboration. *CMS luminosity measurement for the 2018 data-taking period at* $\sqrt{s} = 13\ TeV$. Tech. rep. Geneva: CERN, 2019. URL: https://cds.cern.ch/record/2676164.

[182] The CMS Collaboration. "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV". In: *JINST* 12 (2017), P02014. DOI: 10.1088/1748-0221/12/02/P02014.

[183] F. Demartin et al. "Higgs production in association with a single top quark at the LHC". In: *The European Physical Journal C* 75.6 (2015), p. 267. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-015-3475-9.

[184] J. Butterworth et al. "PDF4LHC recommendations for LHC Run II". In: *Journal of Physics G: Nuclear and Particle Physics* 43.2 (Jan. 2016), p. 023001. ISSN: 1361-6471. DOI: 10.1088/0954-3899/43/2/023001.

[185] The CMS Collaboration. *Reweighting simulated events using machine-learning techniques in the CMS experiment*. 2024. arXiv: 2411.03023 [hep-ex].

[186] The CMS Collaboration. "Extraction and validation of a new set of CMS pythia8 tunes from underlying-event measurements". In: *The European Physical Journal C* 80.1 (Jan. 2020). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-019-7499-4.

[187] S. Argyropoulos and T. Sjöstrand. "Effects of color reconnection on $t\bar{t}$ final states at the LHC". In: *Journal of High Energy Physics* 2014.11 (Nov. 2014). ISSN: 1029-8479. DOI: 10.1007/jhep11(2014)043.

[188] J. R. Christiansen and P. Z. Skands. "String formation beyond leading colour". In: *Journal of High Energy Physics* 2015.8 (Aug. 2015). ISSN: 1029-8479. DOI: 10.1007/jhep08(2015)003.

[189] J. Conway. "Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra". In: (2011), pp. 115–120. DOI: 10.5170/CERN-2011-006.115. arXiv: 1103.0354.

[190] R. Barlow and C. Beeston. "Fitting using finite Monte Carlo samples". In: *Computer Physics Communications* 77.2 (1993), pp. 219–228. ISSN: 0010-4655. DOI: 10.1016/0010-4655(93)90005-W.

[191] W. S. Cleveland. "Robust Locally Weighted Regression and Smoothing Scatterplots". In: *Journal of the American Statistical Association* 74.368 (1979), pp. 829–836. DOI: 10.1080/01621459.1979.10481038.

[192] W. S. Cleveland and S. J. Devlin. "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting". In: *Journal of the American Statistical Association* 83.403 (1988), pp. 596–610. DOI: 10.1080/01621459.1988.10478639.

[193] The CMS Collaboration. *The CMS statistical analysis and combination tool: COMBINE*. 2024. arXiv: 2404.06614 [physics.data-an].

[194] R. D. Cousins. *Generalization of Chisquare Goodness-of-Fit Test for Binned Data Using Saturated Models*. Tech. rep. Mar. 2013. URL: https://www.physics.ucla.edu/~cousins/stats/cousins_saturated.pdf.

[195] L. Demortier and L. Lyons. *Everything you always wanted to know about pulls*. Tech. rep. Aug. 2002. URL: https://hep-physics.rockefeller.edu/luc/technical_reports/cdf5776_pulls.pdf.

Appendix

A Input feature distributions

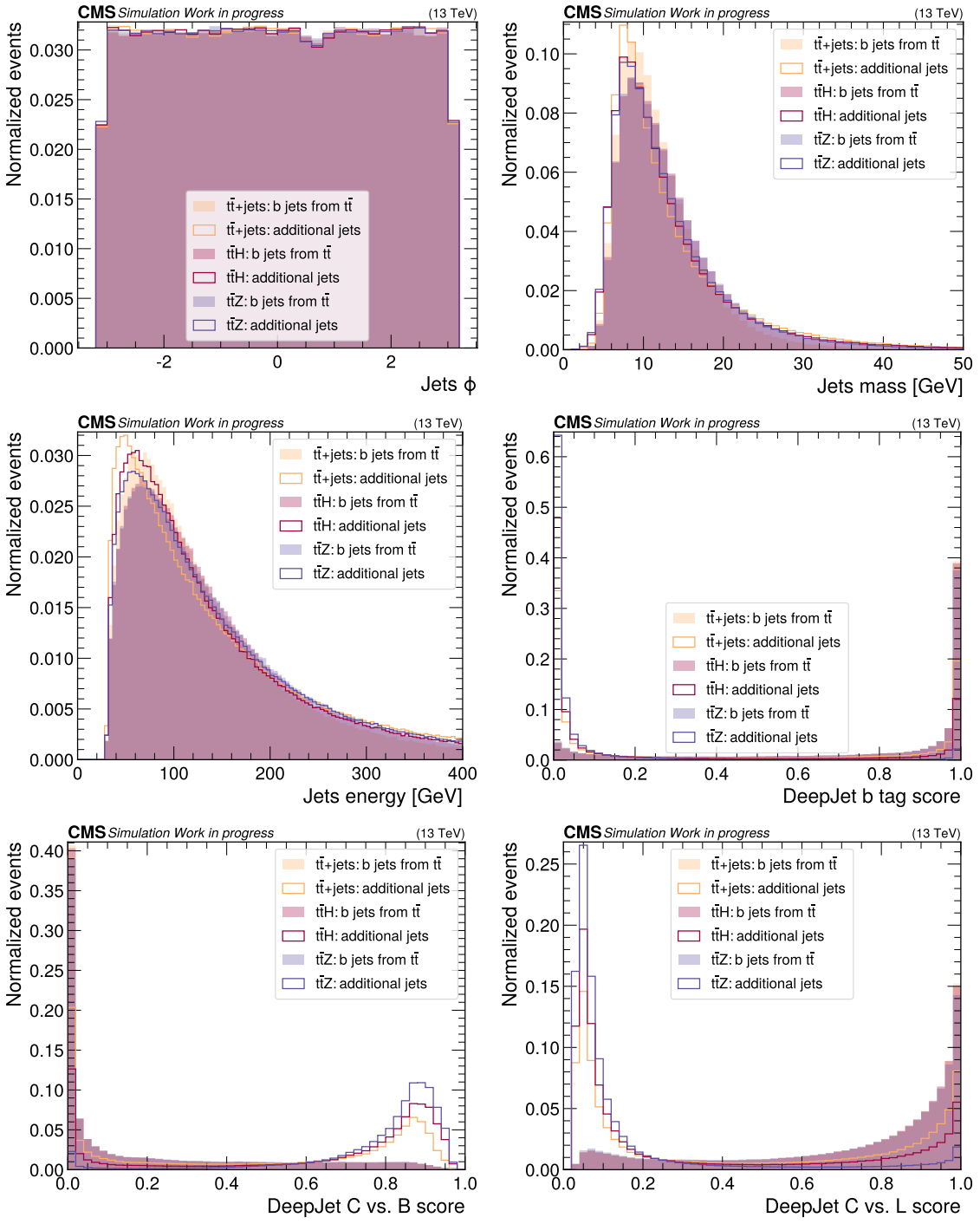


Figure A.1: Input feature distributions of jet ϕ (top left), jet mass (top right), jet energy (middle left) and jet b tagging value (middle right) embedded as node features for simulated $t\bar{t}$+jets, $t\bar{t}H$, and $t\bar{t}Z$ events. The tagger values c vs. b and c vs. light are shown at the bottom left and bottom right respectively. These inputs are not used in the training process. Distributions are separated into b jets from the $t\bar{t}$ system and additional jets.

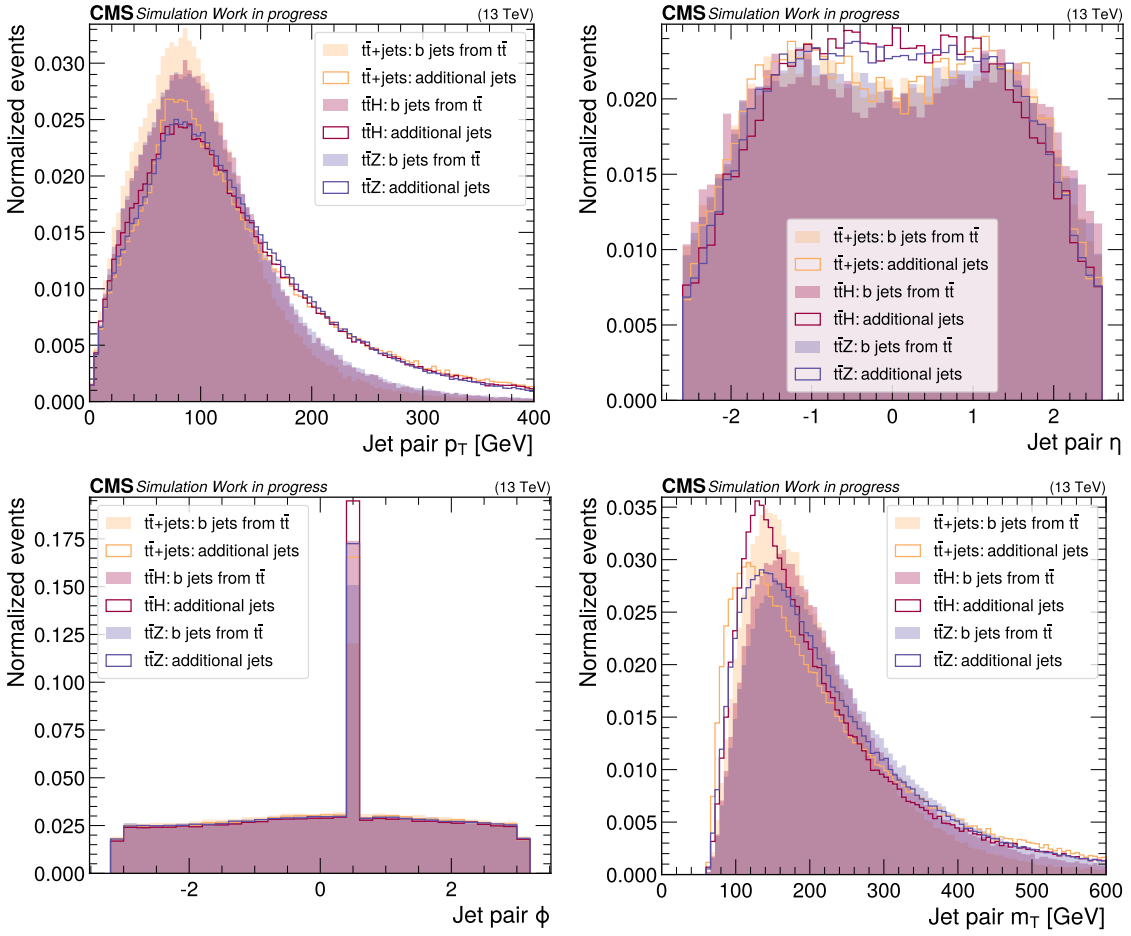


Figure A.2: Input feature distributions of jet pair p_T (top left), jet pair η (top right), jet pair ϕ (bottom left) and jet pair m_T (bottom right) embedded as edge features for simulated $t\bar{t}$+jets, $t\bar{t}H$, and $t\bar{t}Z$ events. Distributions are separated into b jets from the $t\bar{t}$ system and additional jets.

B GLP evaluation on simulated events

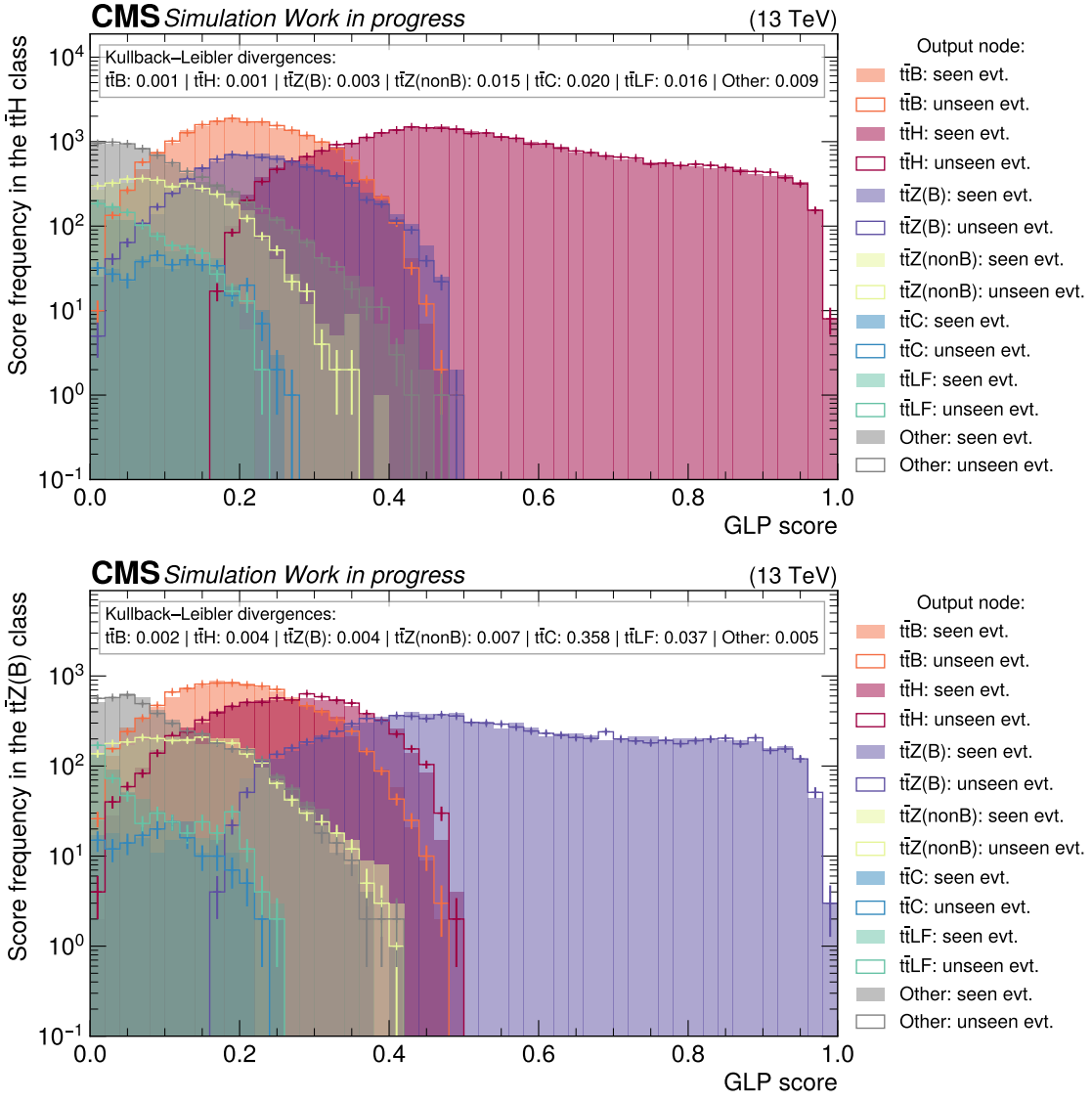


Figure B.3: Evaluation of the graph level prediction (GLP) model on events seen during the training versus events not seen during the training (evt.). The distributions of GLP scores on events assigned to the $t\bar{t}H$ class (top) and $t\bar{t}Z(B)$ (bottom) are shown for each individual node of the multi-class classifier. The Kullback-Leibler (KL) divergence for the distributions of the GLP node scores for seen (filled histograms) and unseen events (enveloping distribution) is calculated and given at the top.

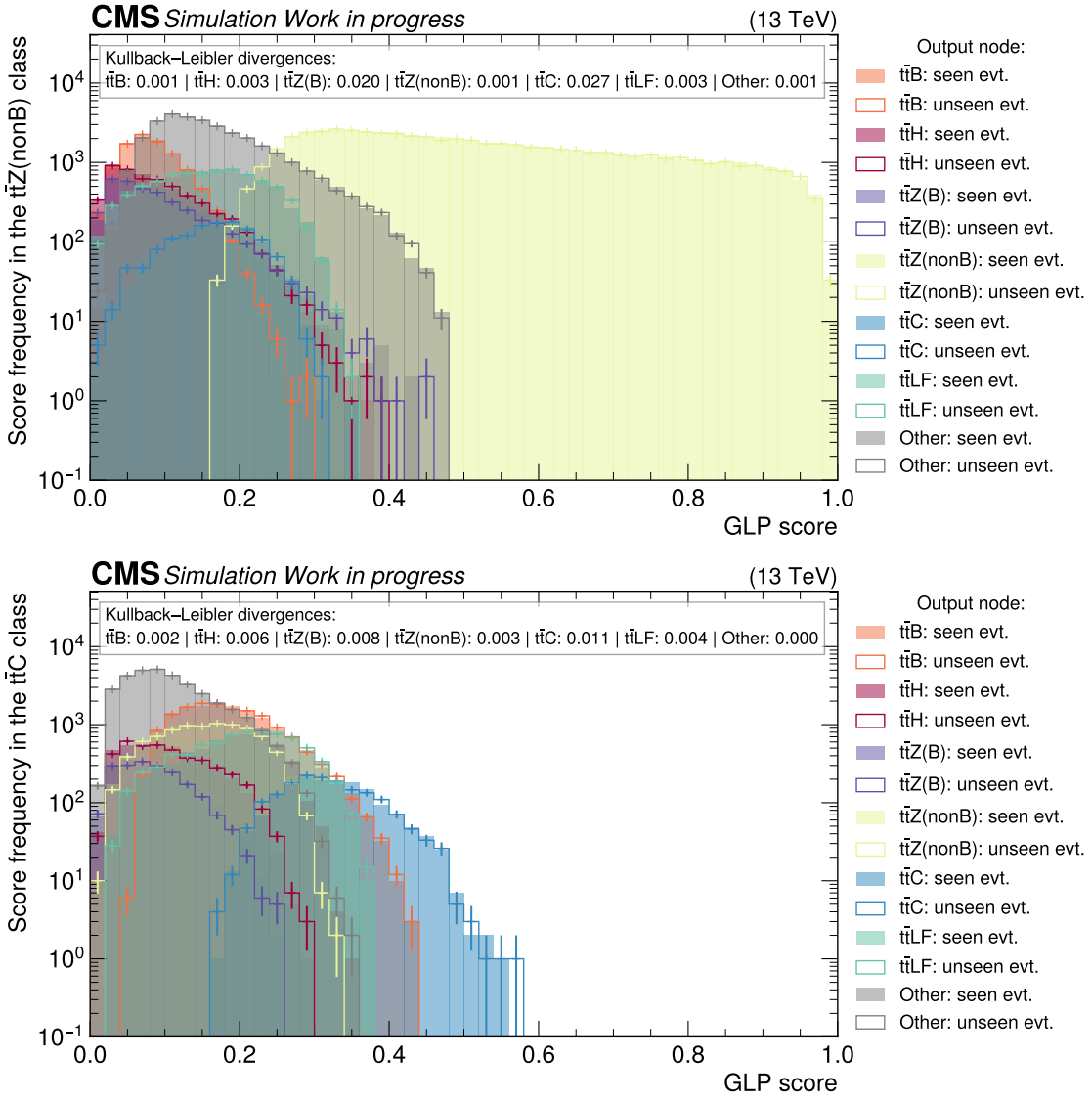


Figure B.4: Evaluation of the GLP model on events seen during the training versus events not seen during the training (evt.). The distributions of GLP scores on events assigned to the $t\bar{t}Z$(nonB) class (top) and $t\bar{t}C$ (bottom) are shown for each individual node of the multi-class classifier. The KL divergence for the distributions of the GLP node scores for seen (filled histograms) and unseen events (enveloping distribution) is calculated and given at the top.

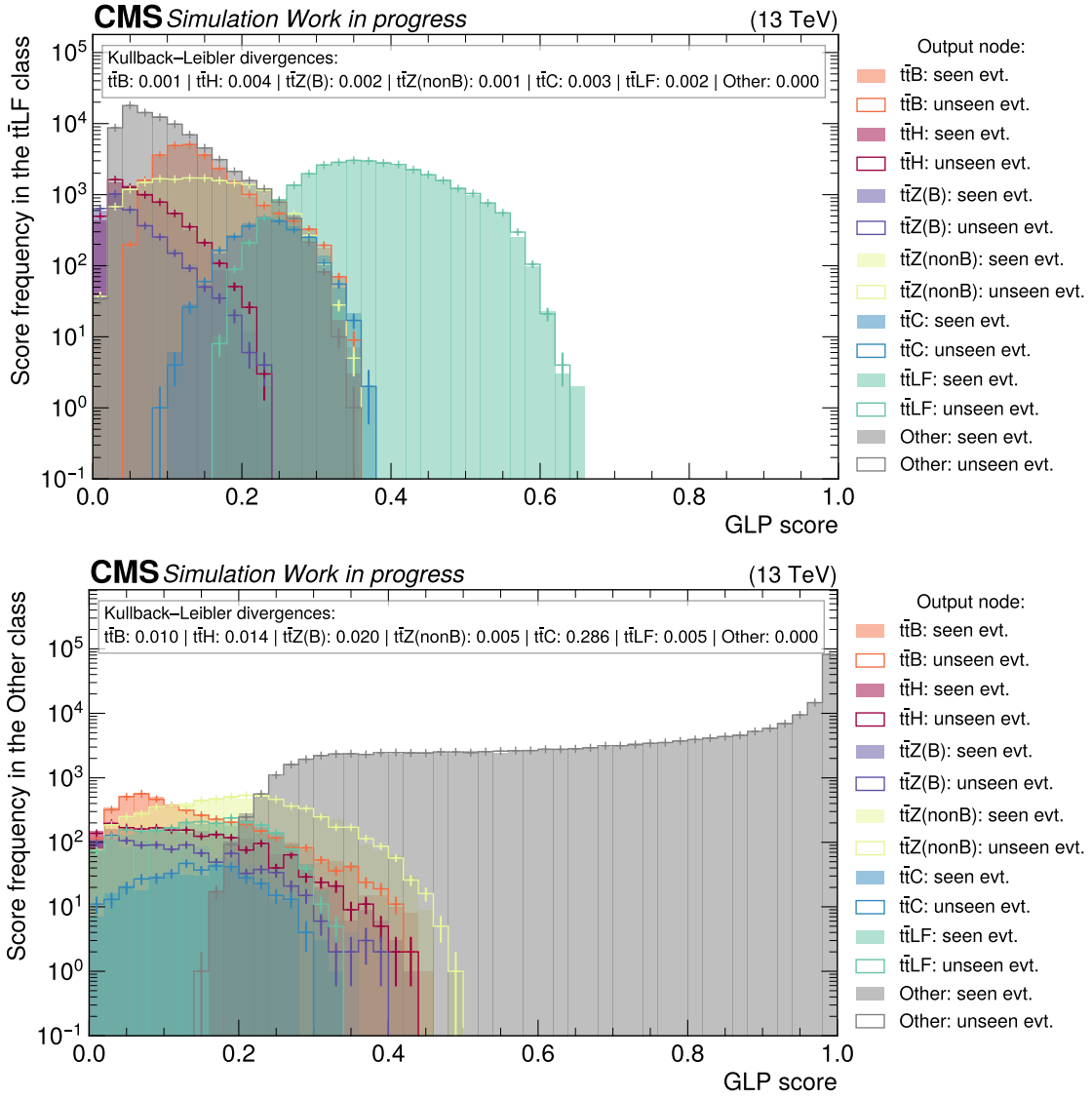


Figure B.5: Evaluation of the GLP model on events seen during the training versus events not seen during the training (evt.). The distributions of GLP scores on events assigned to the tt̄LF class (top) and Other (bottom) are shown for each individual node of the multi-class classifier. The KL divergence for the distributions of the GLP node scores for seen (filled histograms) and unseen events (enveloping distribution) is calculated and given at the top.

C Post-fit parameters and impacts

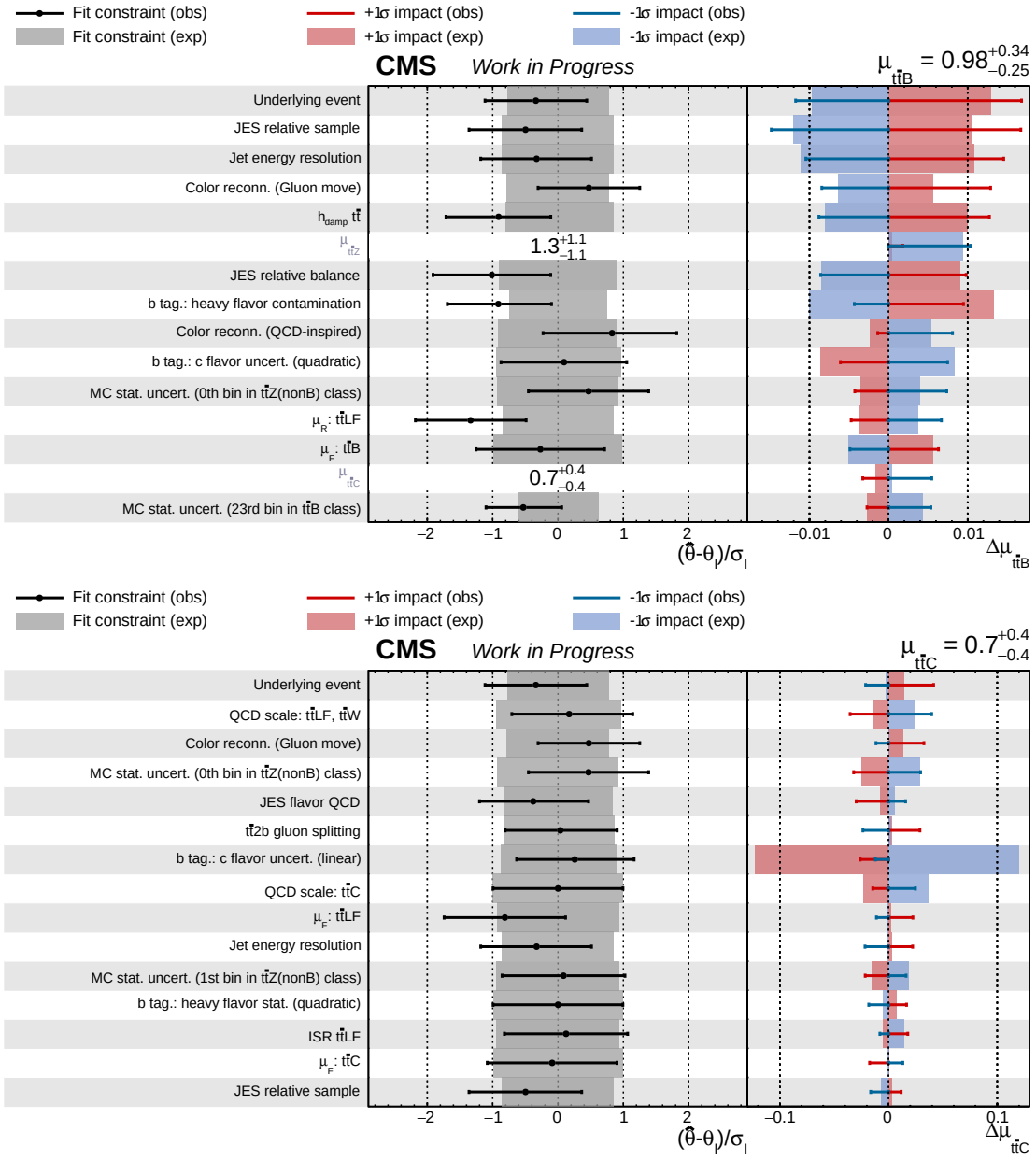


Figure C.6: Pre-fit and post-fit results of the signal strength parameters $\mu_{t\bar{t}B}$ (top) and $\mu_{t\bar{t}C}$ (bottom). Parameter names are organized in the left panel. The distance from the post-fit value $\hat{\theta}$ to the pre-fit value θ normalized to the pre-fit uncertainty σ_I are shown in the middle panel. The right panel shows the relative change to the signal strength parameter μ under scrutiny if the NP under consideration is shifted to its $\pm 1\sigma$ value. The parameters are sorted according to the size of the impact, parameters 16 to 30 are shown. Created with [193].

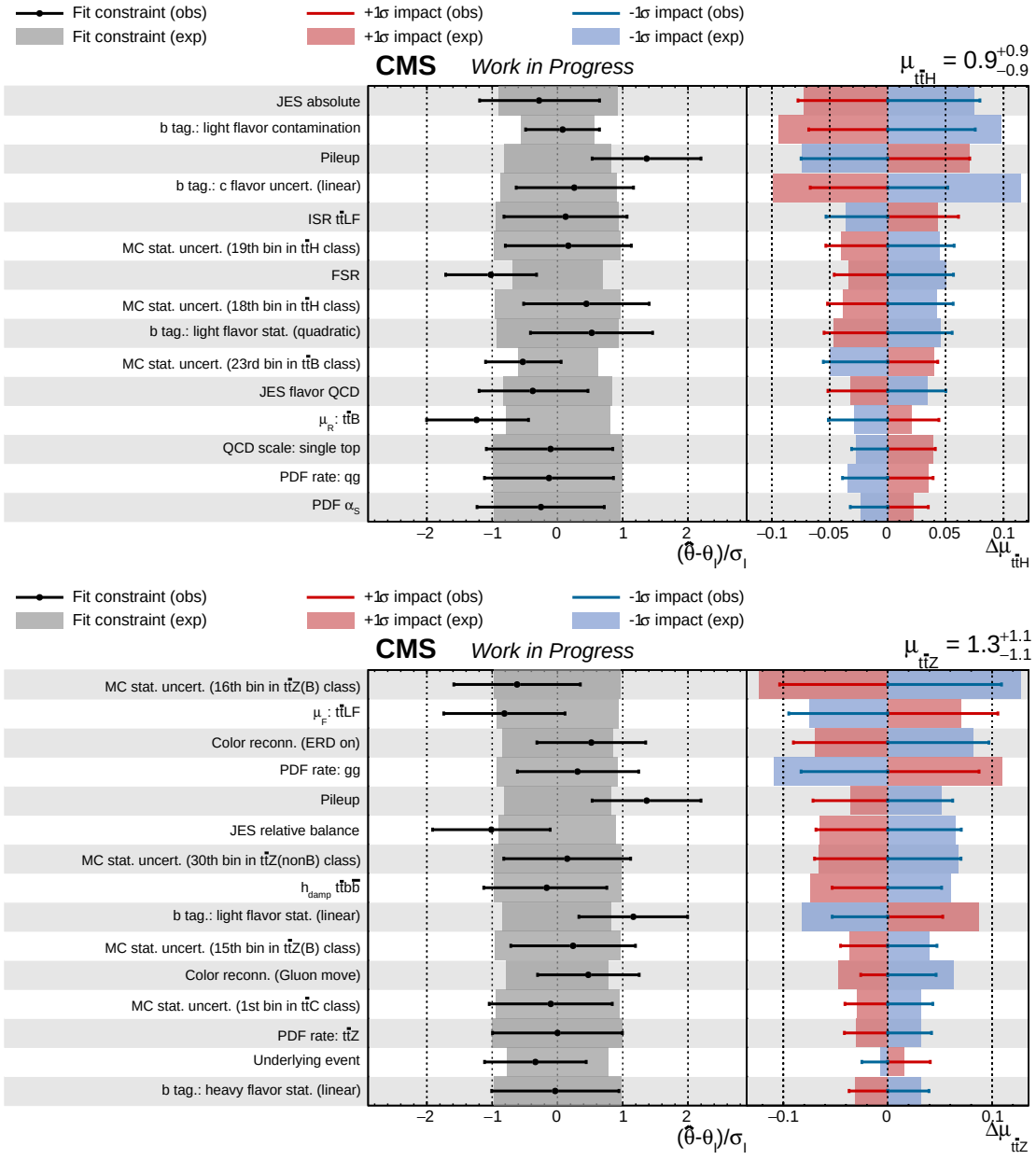


Figure C.7: Pre-fit and post-fit results of the signal strength parameters $\mu_{t\bar{t}H}$ (top) and $\mu_{t\bar{t}Z}$ (bottom). Parameter names are organized in the left panel. The distance from the post-fit value $\hat{\theta}$ to the pre-fit value θ normalized to the pre-fit uncertainty σ_I are shown in the middle panel. The right panel shows the relative change to the signal strength parameter μ under scrutiny if the NP under consideration is shifted to its $\pm 1\sigma$ value. The parameters are sorted according to the size of the impact, parameters 16 to 30 are shown. Created with [193].

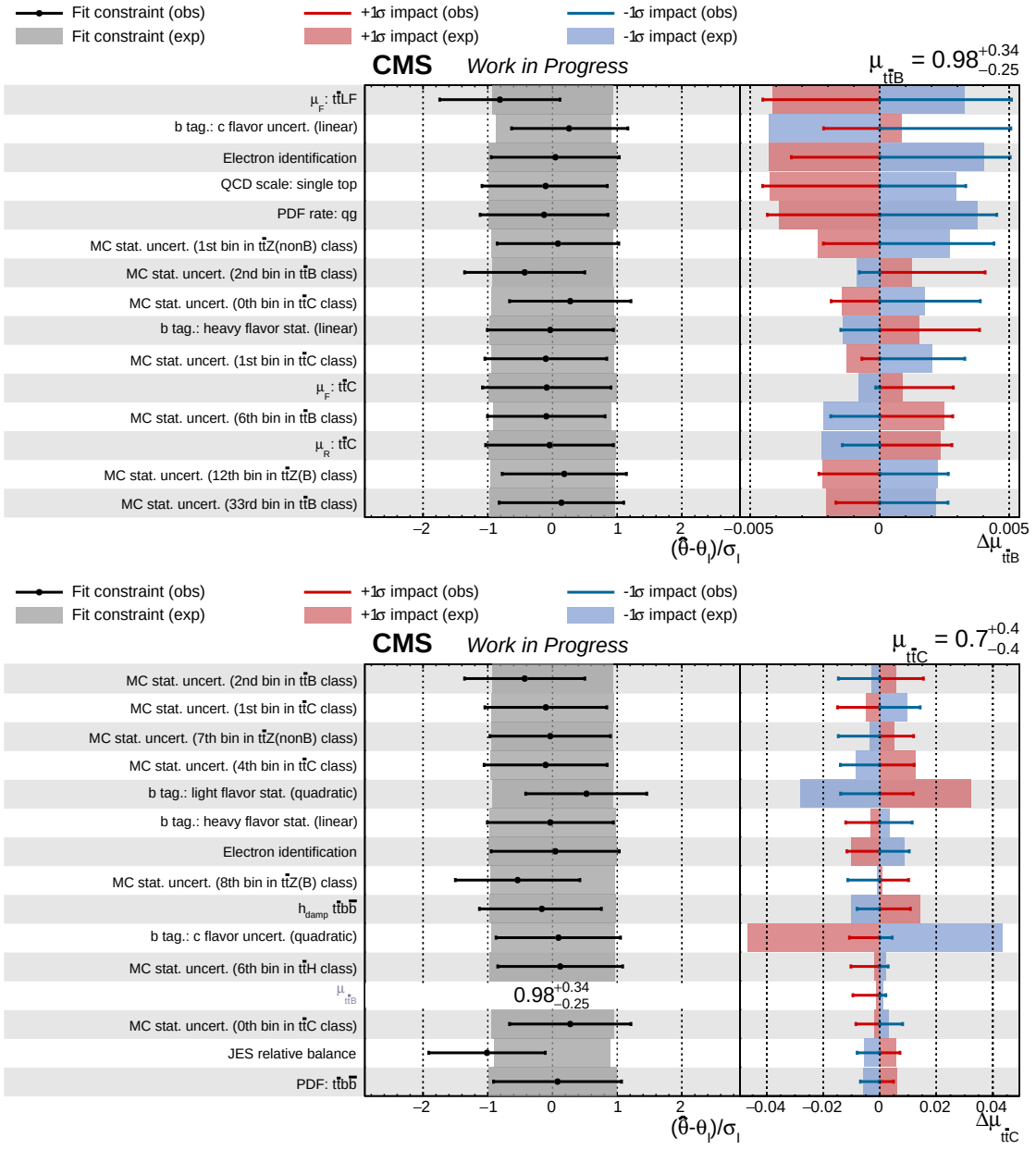


Figure C.8: Pre-fit and post-fit results of the signal strength parameters $\mu_{t\bar{t}B}$ (top) and $\mu_{t\bar{t}C}$ (bottom). Parameter names are organized in the left panel. The distance from the post-fit value $\hat{\theta}$ to the pre-fit value θ normalized to the pre-fit uncertainty σ_I are shown in the middle panel. The right panel shows the relative change to the signal strength parameter μ under scrutiny if the NP under consideration is shifted to its $\pm 1\sigma$ value. The parameters are sorted according to the size of the impact, parameters 31 to 45 are shown. Created with [193].

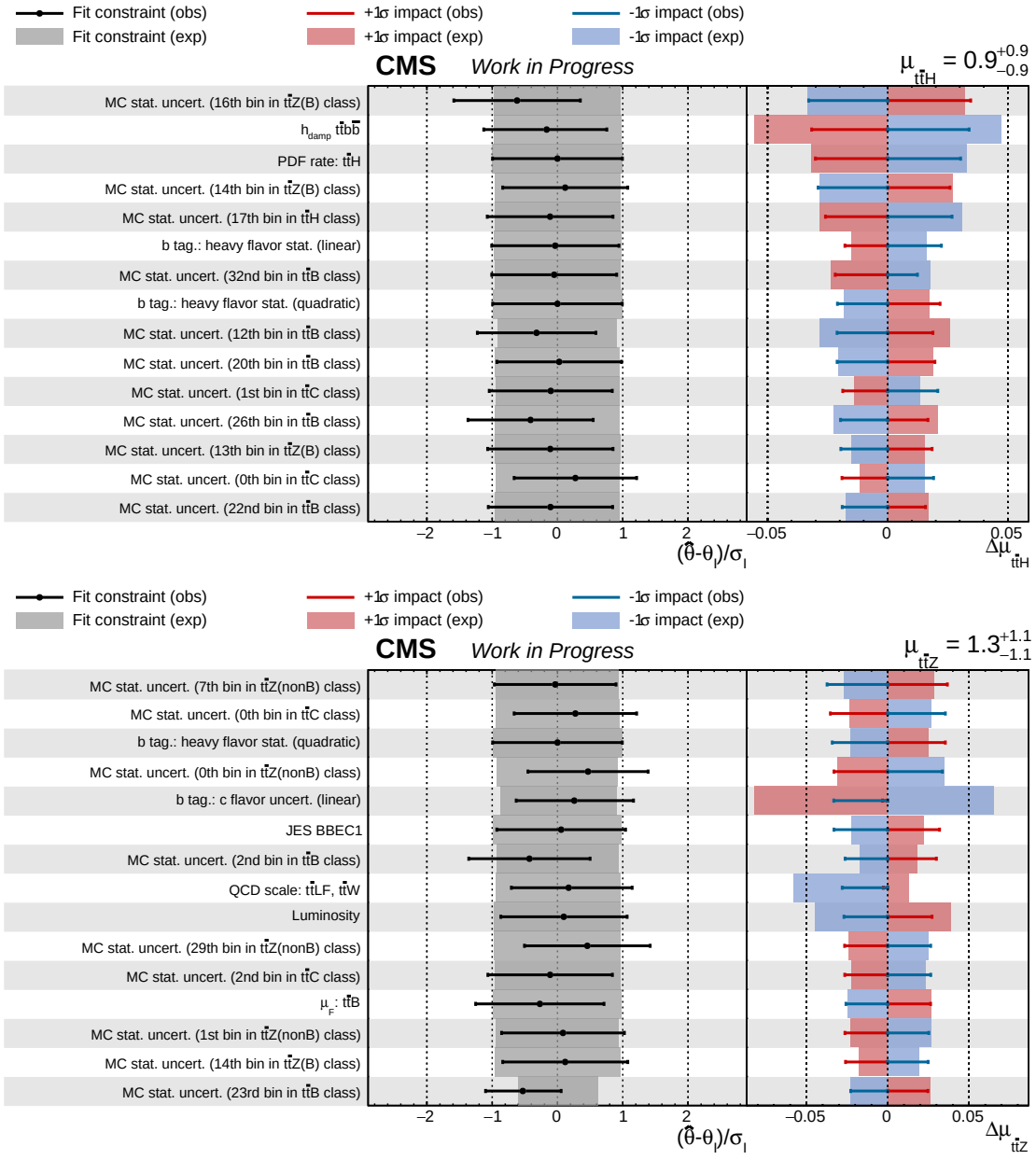


Figure C.9: Pre-fit and post-fit results of the signal strength parameters $\mu_{t\bar{t}H}$ (top) and $\mu_{t\bar{t}Z}$ (bottom). Parameter names are organized in the left panel. The distance from the post-fit value $\hat{\theta}$ to the pre-fit value θ normalized to the pre-fit uncertainty σ_I are shown in the middle panel. The right panel shows the relative change to the signal strength parameter μ under scrutiny if the NP under consideration is shifted to its $\pm 1\sigma$ value. The parameters are sorted according to the size of the impact, parameters 31 to 45 are shown. Created with [193].

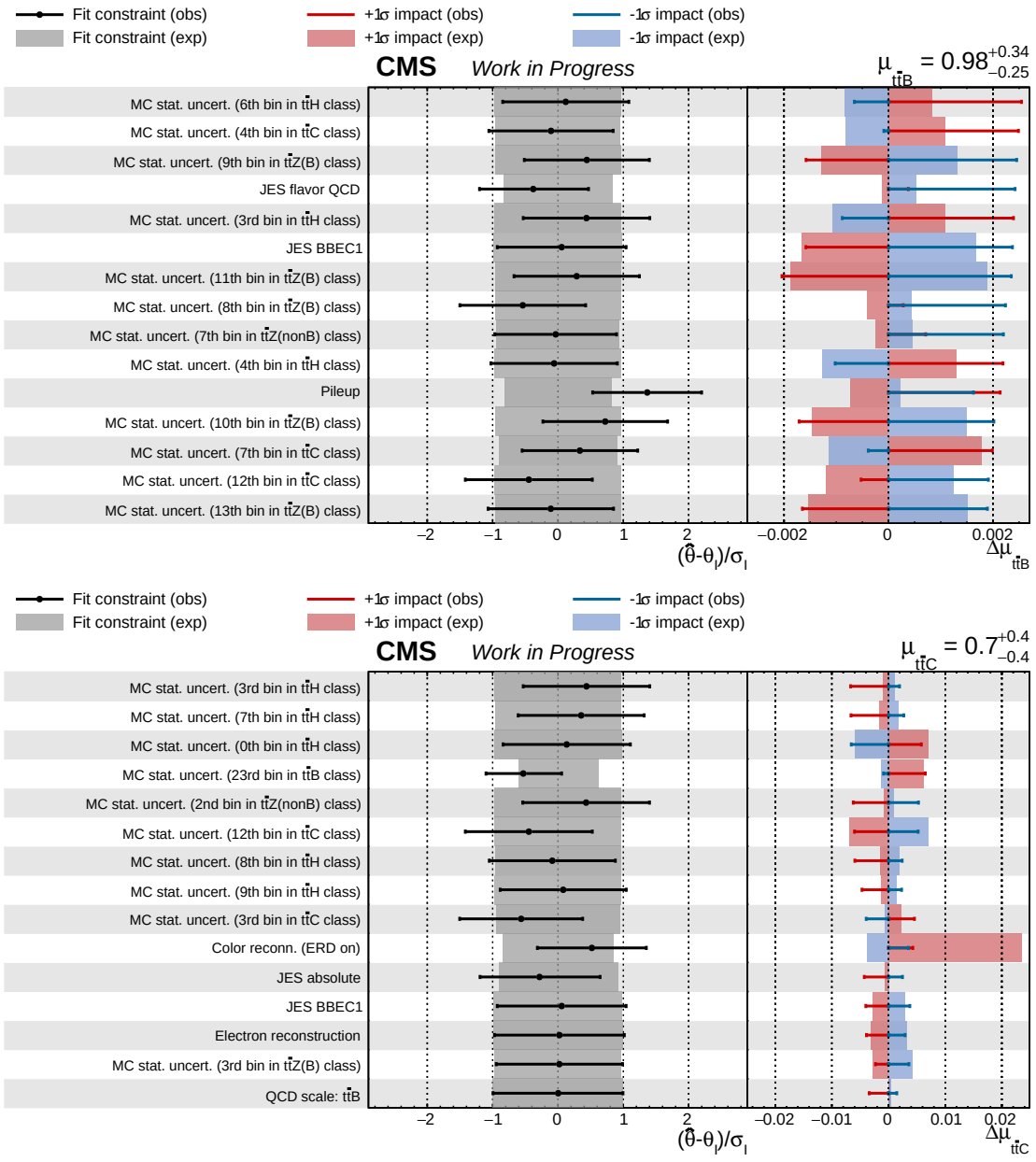


Figure C.10: Pre-fit and post-fit results of the signal strength parameters $\mu_{t\bar{t}B}$ (top) and $\mu_{t\bar{t}C}$ (bottom). Parameter names are organized in the left panel. The distance from the post-fit value $\hat{\theta}$ to the pre-fit value θ normalized to the pre-fit uncertainty σ_I are shown in the middle panel. The right panel shows the relative change to the signal strength parameter μ under scrutiny if the NP under consideration is shifted to its $\pm 1\sigma$ value. The parameters are sorted according to the size of the impact, parameters 46 to 60 are shown. Created with [193].

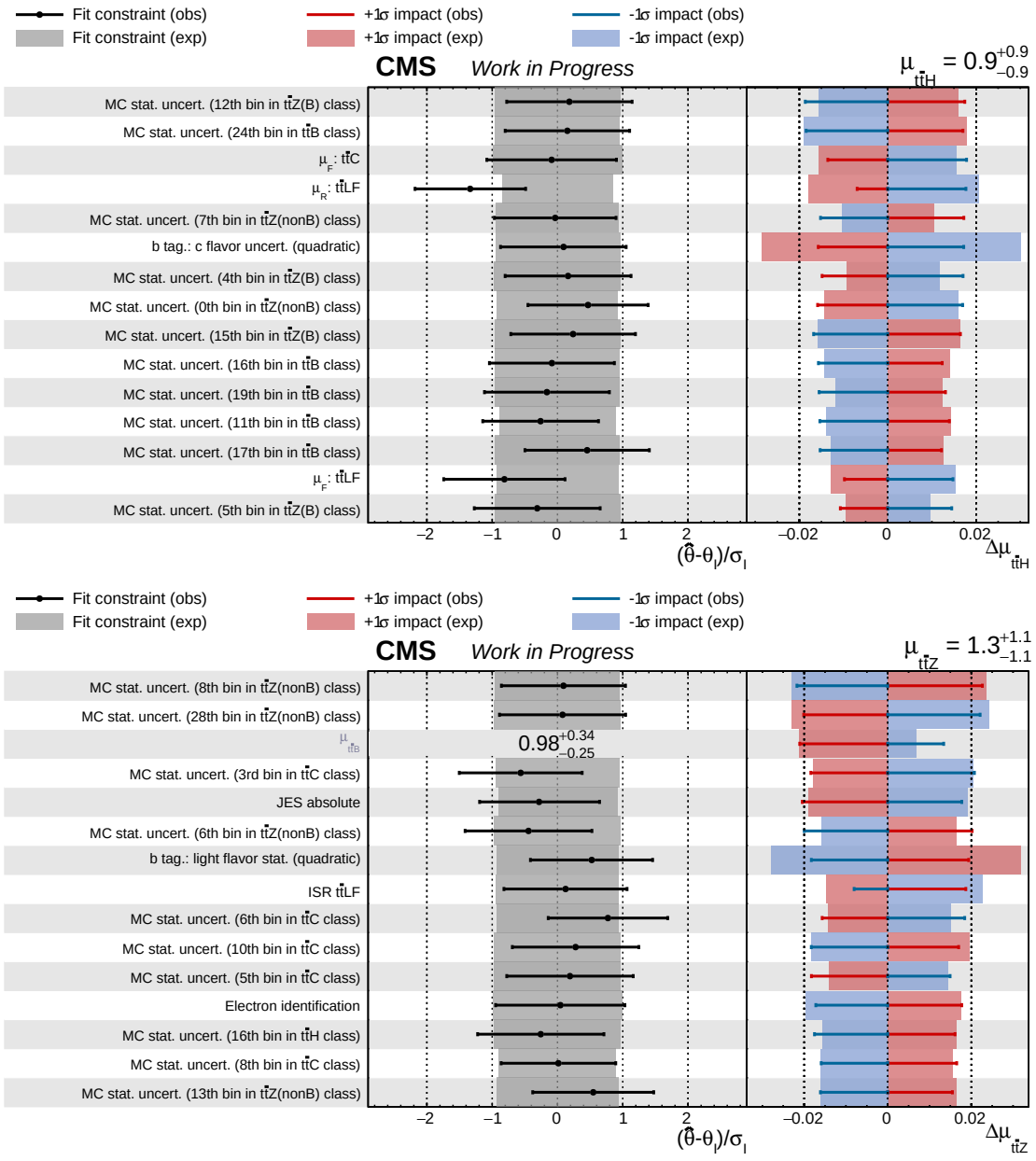


Figure C.11: Pre-fit and post-fit results of the signal strength parameters $\mu_{t\bar{t}H}$ (top) and $\mu_{t\bar{t}Z}$ (bottom). Parameter names are organized in the left panel. The distance from the post-fit value $\hat{\theta}$ to the pre-fit value θ normalized to the pre-fit uncertainty σ_I are shown in the middle panel. The right panel shows the relative change to the signal strength parameter μ under scrutiny if the NP under consideration is shifted to its $\pm 1\sigma$ value. The parameters are sorted according to the size of the impact, parameters 46 to 60 are shown. Created with [193].

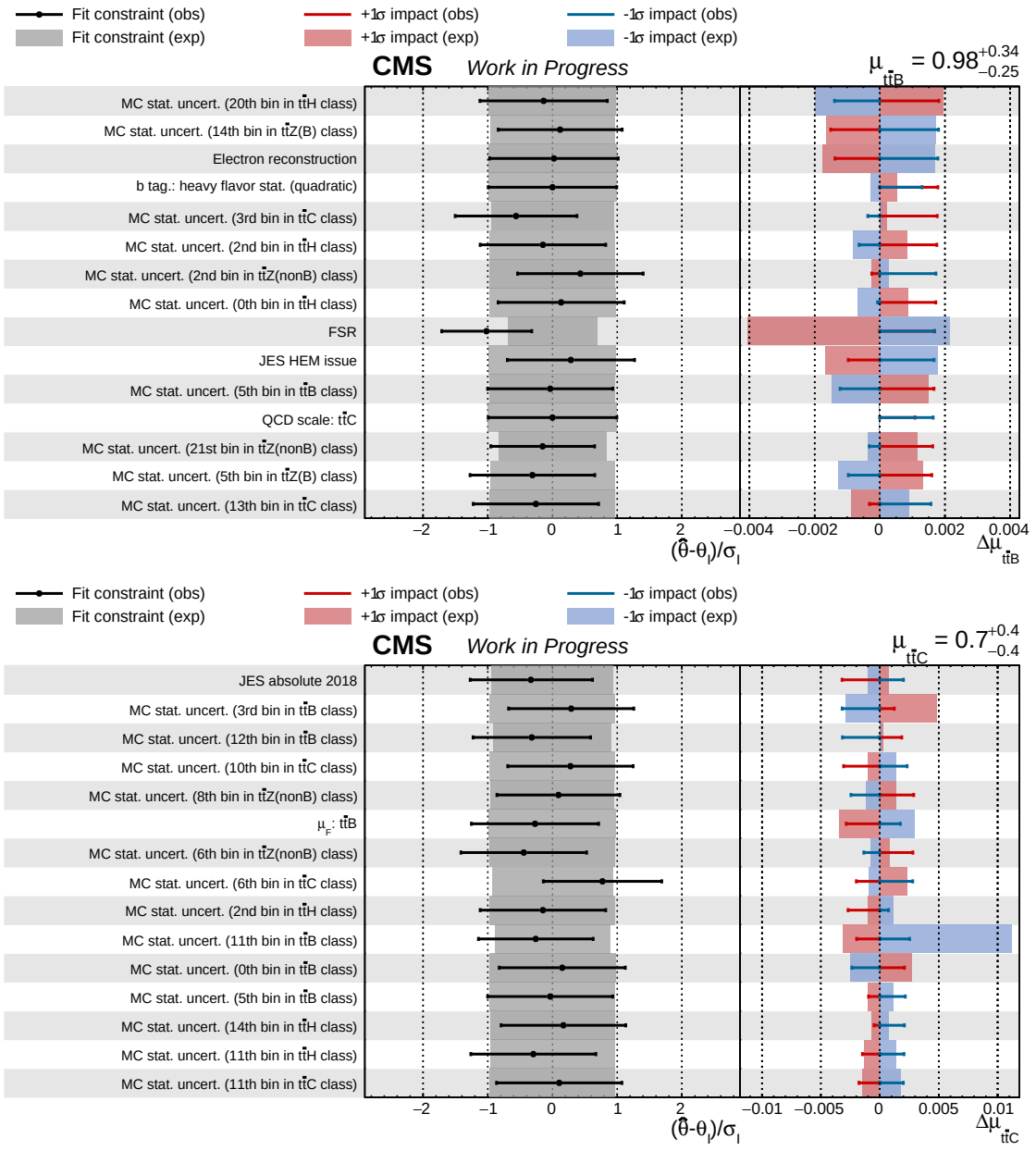


Figure C.12: Pre-fit and post-fit results of the signal strength parameters $\mu_{t\bar{t}B}$ (top) and $\mu_{t\bar{t}C}$ (bottom). Parameter names are organized in the left panel. The distance from the post-fit value $\hat{\theta}$ to the pre-fit value θ normalized to the pre-fit uncertainty σ_I are shown in the middle panel. The right panel shows the relative change to the signal strength parameter μ under scrutiny if the NP under consideration is shifted to its $\pm 1\sigma$ value. The parameters are sorted according to the size of the impact, parameters 61 to 75 are shown. Created with [193].

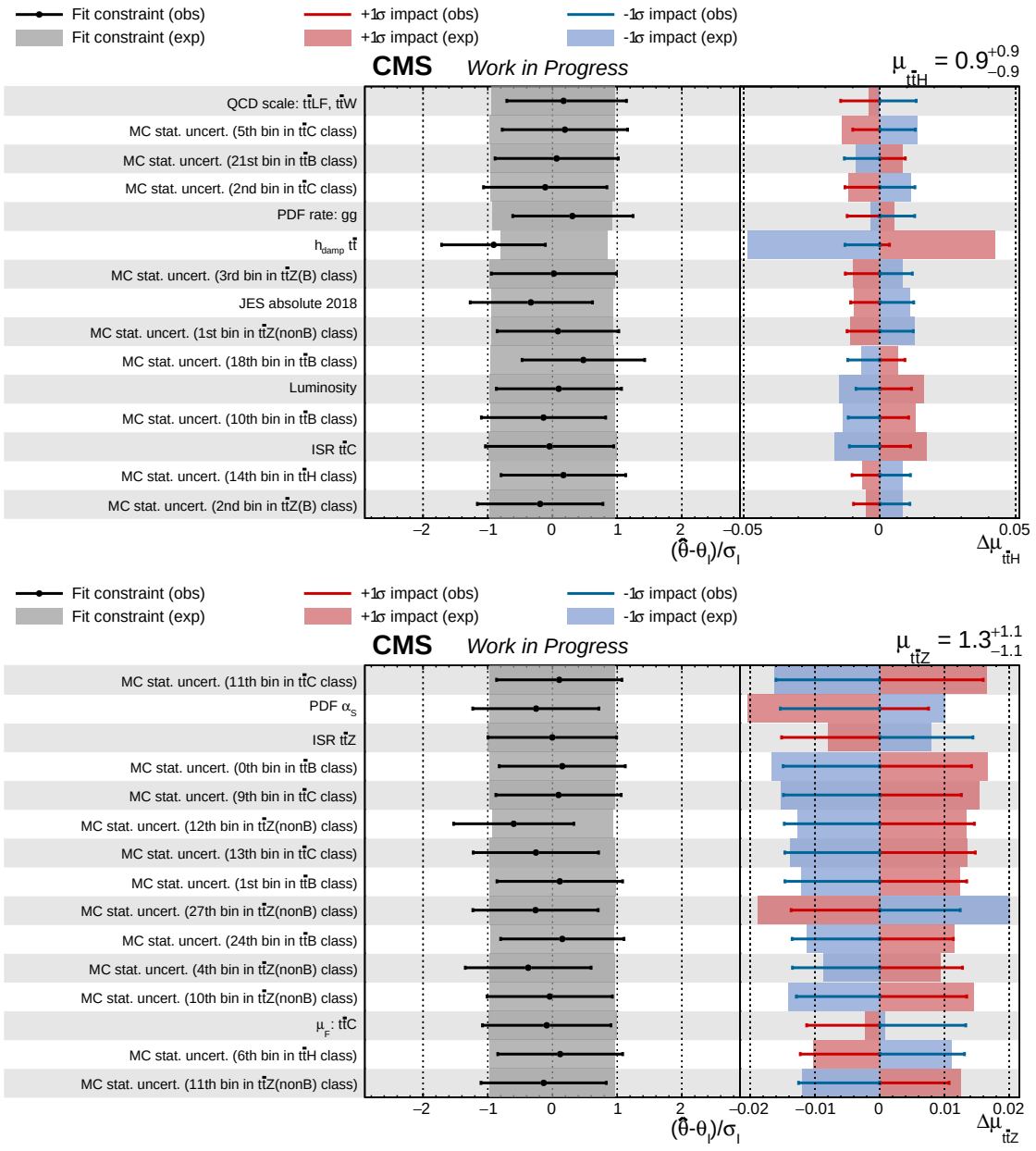


Figure C.13: Pre-fit and post-fit results of the signal strength parameters $\mu_{t\bar{t}H}$ (top) and $\mu_{t\bar{t}Z}$ (bottom). Parameter names are organized in the left panel. The distance from the post-fit value $\hat{\theta}$ to the pre-fit value θ normalized to the pre-fit uncertainty σ_I are shown in the middle panel. The right panel shows the relative change to the signal strength parameter μ under scrutiny if the NP under consideration is shifted to its $\pm 1\sigma$ value. The parameters are sorted according to the size of the impact, parameters 61 to 75 are shown. Created with [193].

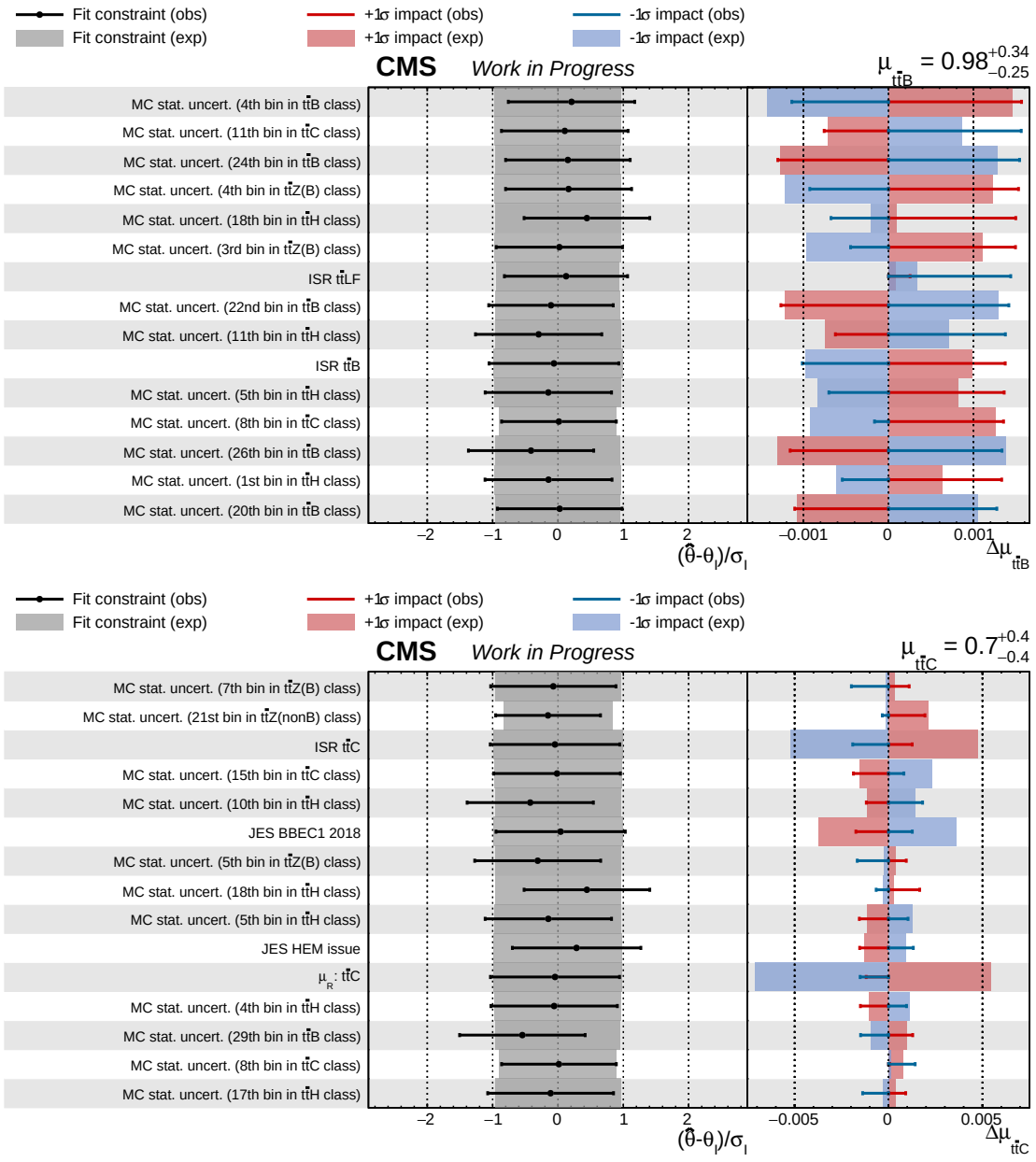


Figure C.14: Pre-fit and post-fit results of the signal strength parameters $\mu_{t\bar{t}B}$ (top) and $\mu_{t\bar{t}C}$ (bottom). Parameter names are organized in the left panel. The distance from the post-fit value $\hat{\theta}$ to the pre-fit value θ normalized to the pre-fit uncertainty σ_I are shown in the middle panel. The right panel shows the relative change to the signal strength parameter μ under scrutiny if the NP under consideration is shifted to its $\pm 1\sigma$ value. The parameters are sorted according to the size of the impact, parameters 76 to 90 are shown. Created with [193].

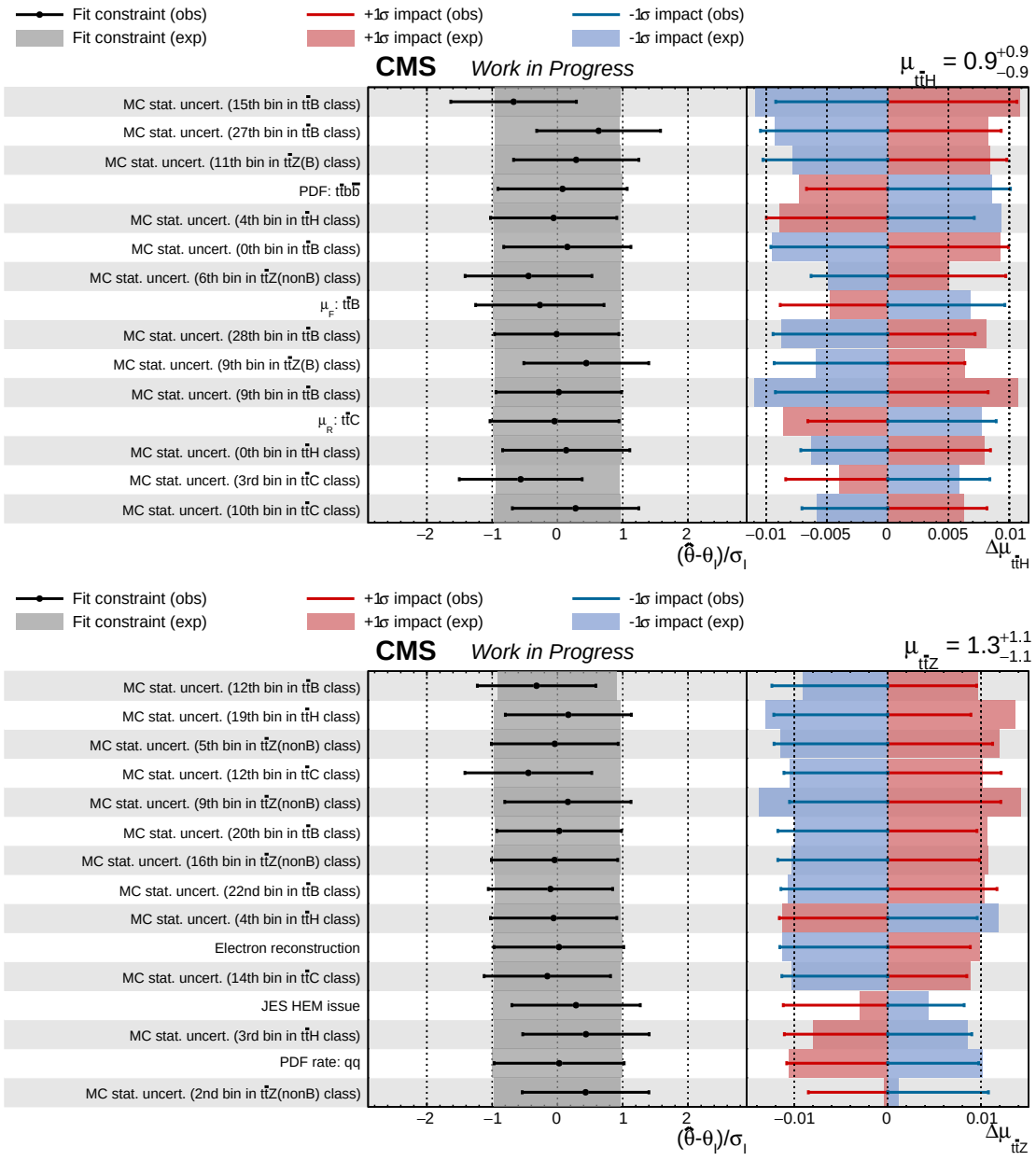


Figure C.15: Pre-fit and post-fit results of the signal strength parameters $\mu_{t\bar{t}H}$ (top) and $\mu_{t\bar{t}Z}$ (bottom). Parameter names are organized in the left panel. The distance from the post-fit value $\hat{\theta}$ to the pre-fit value θ normalized to the pre-fit uncertainty σ_I are shown in the middle panel. The right panel shows the relative change to the signal strength parameter μ under scrutiny if the NP under consideration is shifted to its $\pm 1\sigma$ value. The parameters are sorted according to the size of the impact, parameters 76 to 90 are shown. Created with [193].

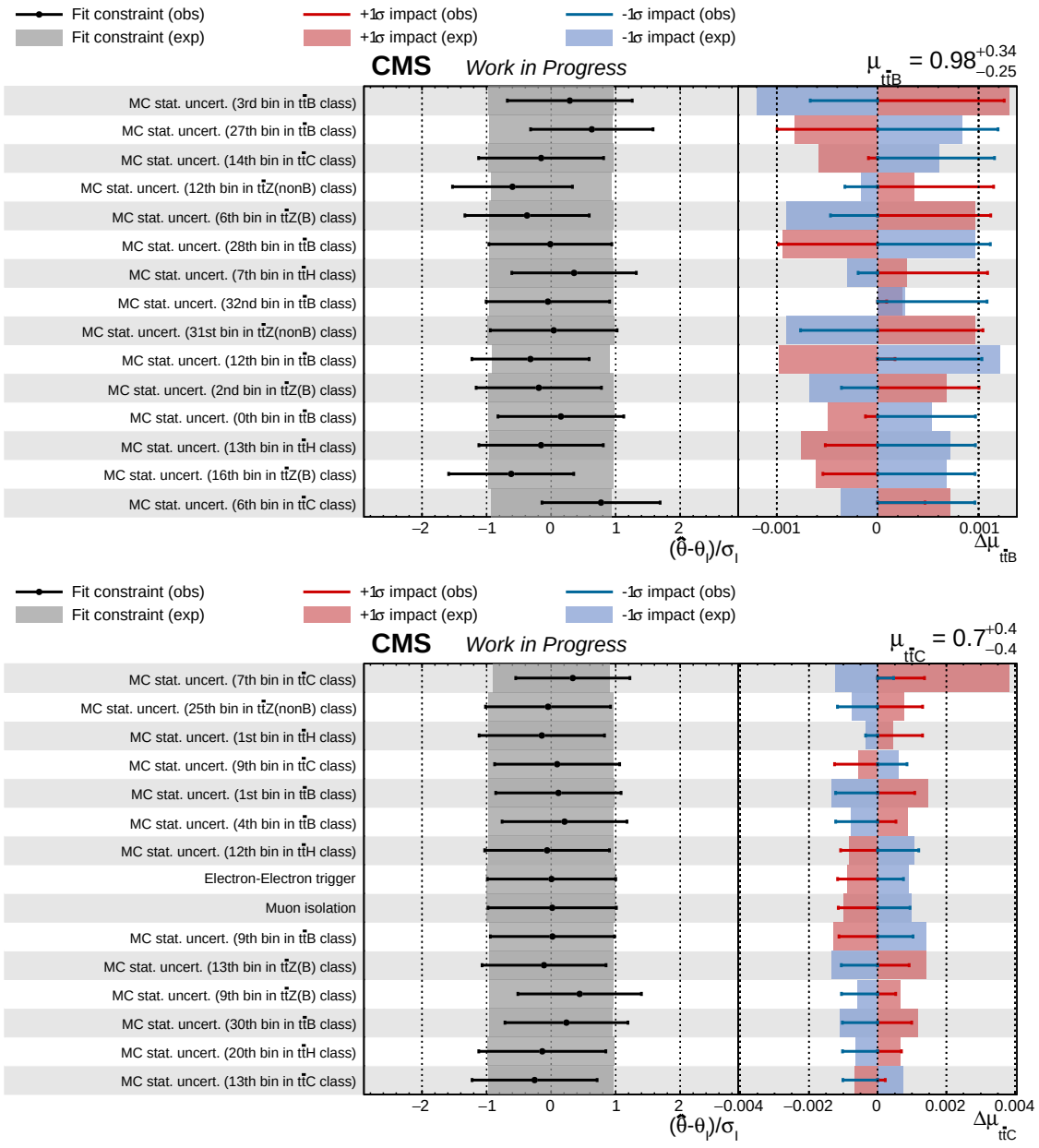


Figure C.16: Pre-fit and post-fit results of the signal strength parameters $\mu_{t\bar{t}B}$ (top) and $\mu_{t\bar{t}C}$ (bottom). Parameter names are organized in the left panel. The distance from the post-fit value $\hat{\theta}$ to the pre-fit value θ normalized to the pre-fit uncertainty σ_I are shown in the middle panel. The right panel shows the relative change to the signal strength parameter μ under scrutiny if the NP under consideration is shifted to its $\pm 1\sigma$ value. The parameters are sorted according to the size of the impact, parameters 91 to 105 are shown. Created with [193].

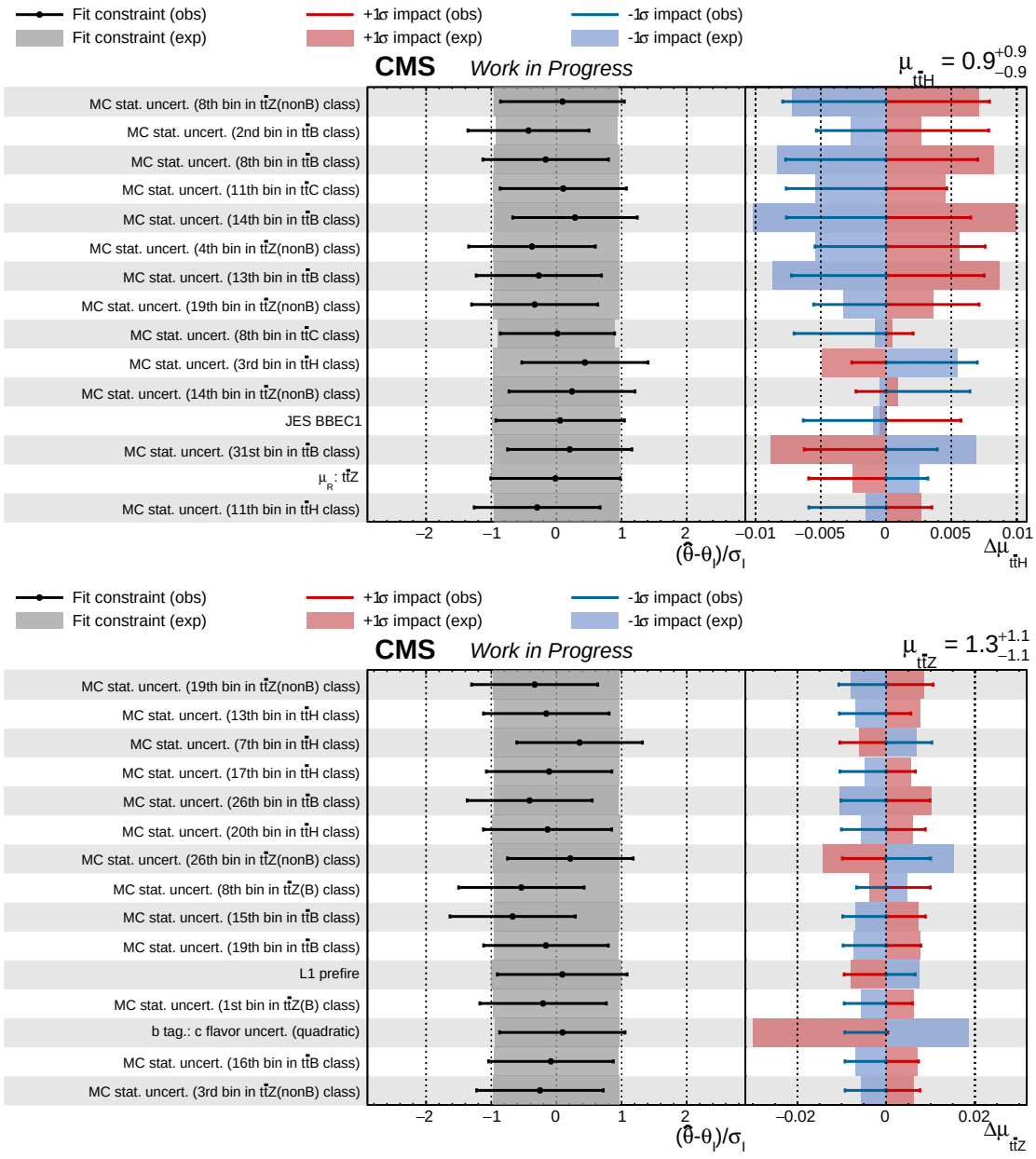


Figure C.17: Pre-fit and post-fit results of the signal strength parameters $\mu_{t\bar{t}H}$ (top) and $\mu_{t\bar{t}Z}$ (bottom). Parameter names are organized in the left panel. The distance from the post-fit value $\hat{\theta}$ to the pre-fit value θ normalized to the pre-fit uncertainty σ_I are shown in the middle panel. The right panel shows the relative change to the signal strength parameter μ under scrutiny if the NP under consideration is shifted to its $\pm 1\sigma$ value. The parameters are sorted according to the size of the impact, parameters 91 to 105 are shown. Created with [193].

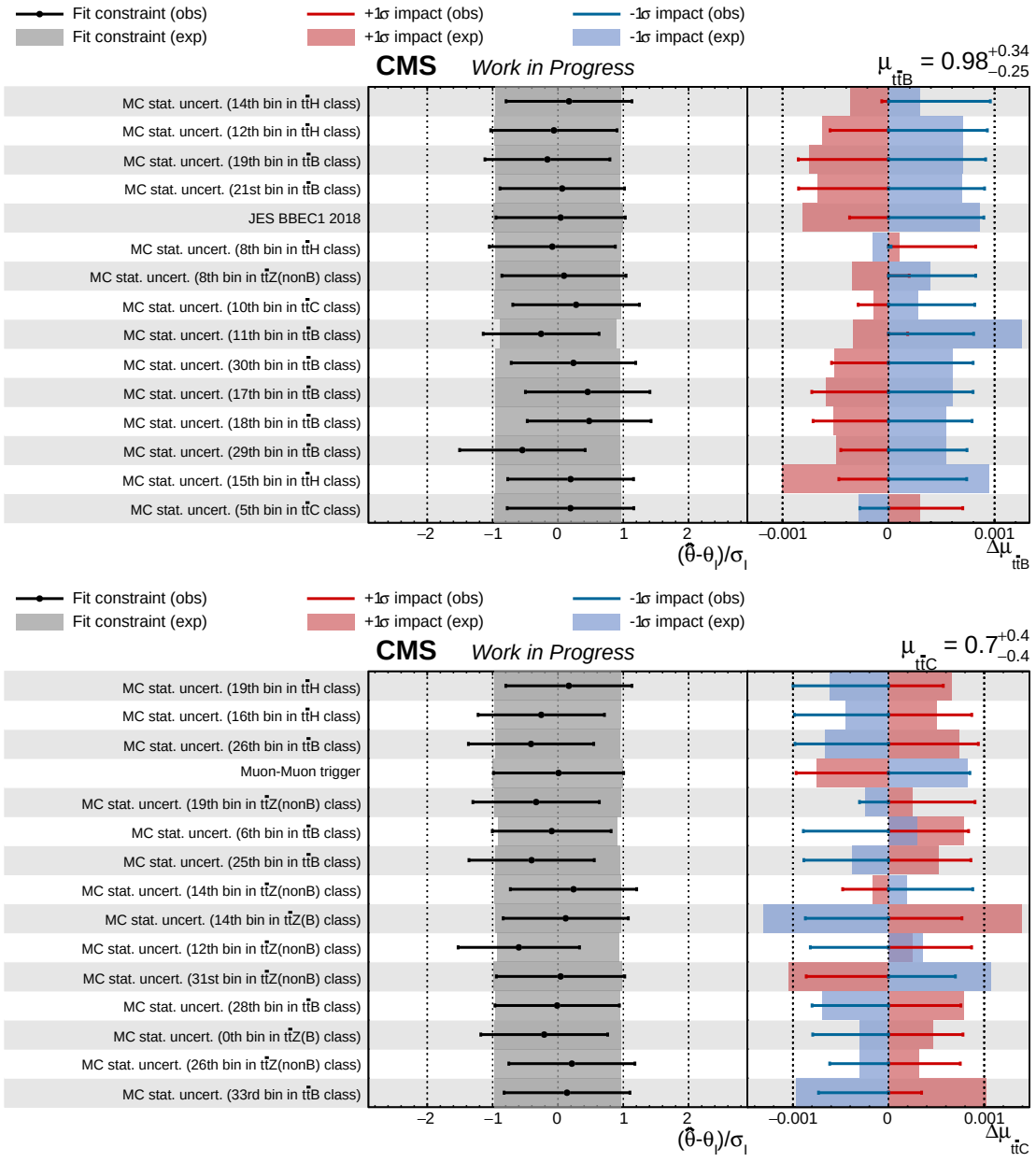


Figure C.18: Pre-fit and post-fit results of the signal strength parameters $\mu_{t\bar{t}B}$ (top) and $\mu_{t\bar{t}C}$ (bottom). Parameter names are organized in the left panel. The distance from the post-fit value $\hat{\theta}$ to the pre-fit value θ normalized to the pre-fit uncertainty σ_I are shown in the middle panel. The right panel shows the relative change to the signal strength parameter μ under scrutiny if the NP under consideration is shifted to its $\pm 1\sigma$ value. The parameters are sorted according to the size of the impact, parameters 106 to 120 are shown. Created with [193].

4

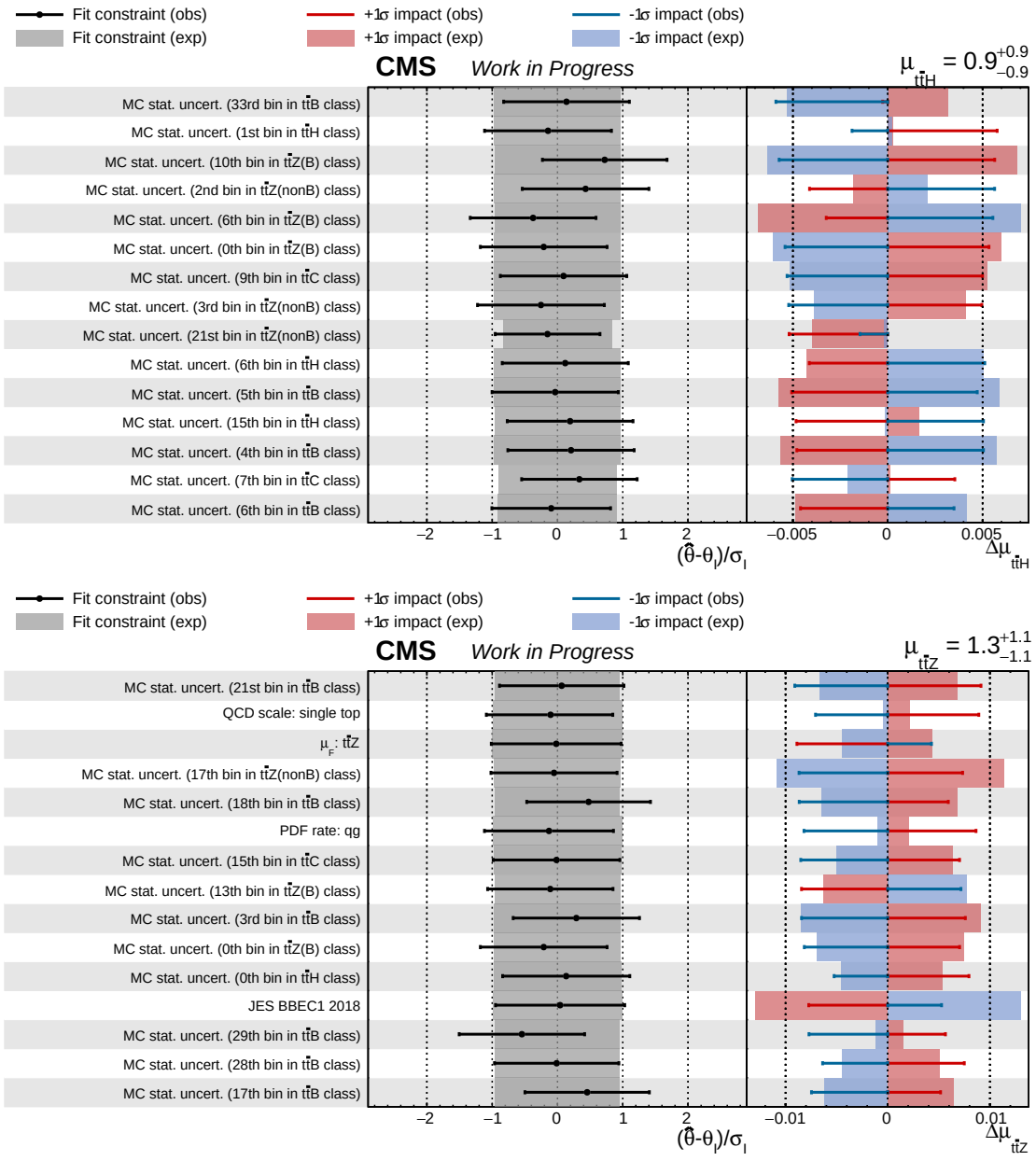


Figure C.19: Pre-fit and post-fit results of the signal strength parameters $\mu_{t\bar{t}H}$ (top) and $\mu_{t\bar{t}Z}$ (bottom). Parameter names are organized in the left panel. The distance from the post-fit value $\hat{\theta}$ to the pre-fit value θ normalized to the pre-fit uncertainty σ_I are shown in the middle panel. The right panel shows the relative change to the signal strength parameter μ under scrutiny if the NP under consideration is shifted to its $\pm 1\sigma$ value. The parameters are sorted according to the size of the impact, parameters 106 to 120 are shown. Created with [193].

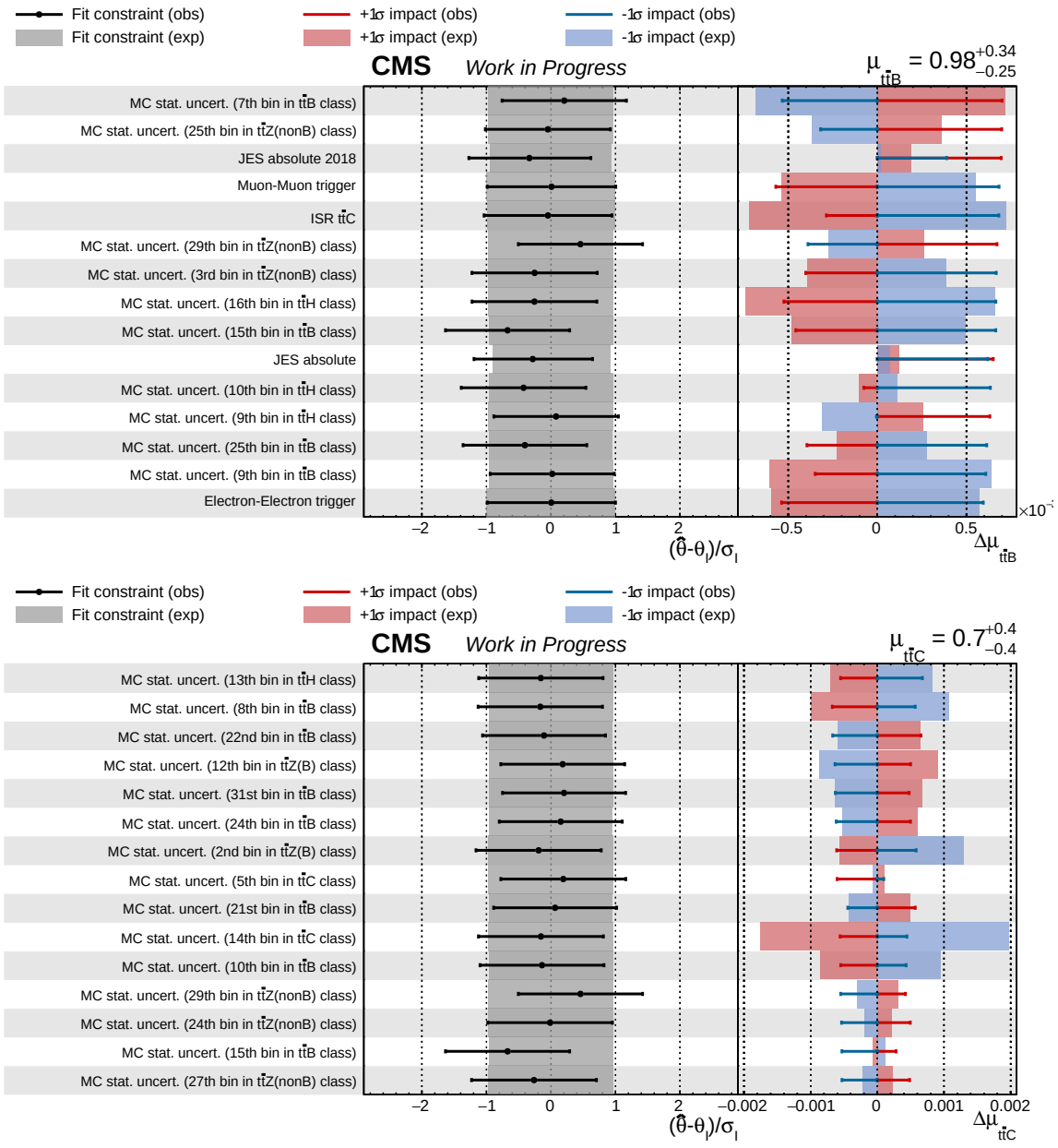


Figure C.20: Pre-fit and post-fit results of the signal strength parameters $\mu_{t\bar{t}B}$ (top) and $\mu_{t\bar{t}C}$ (bottom). Parameter names are organized in the left panel. The distance from the post-fit value $\hat{\theta}$ to the pre-fit value θ normalized to the pre-fit uncertainty σ_I are shown in the middle panel. The right panel shows the relative change to the signal strength parameter μ under scrutiny if the NP under consideration is shifted to its $\pm 1\sigma$ value. The parameters are sorted according to the size of the impact, parameters 121 to 135 are shown. Created with [193].

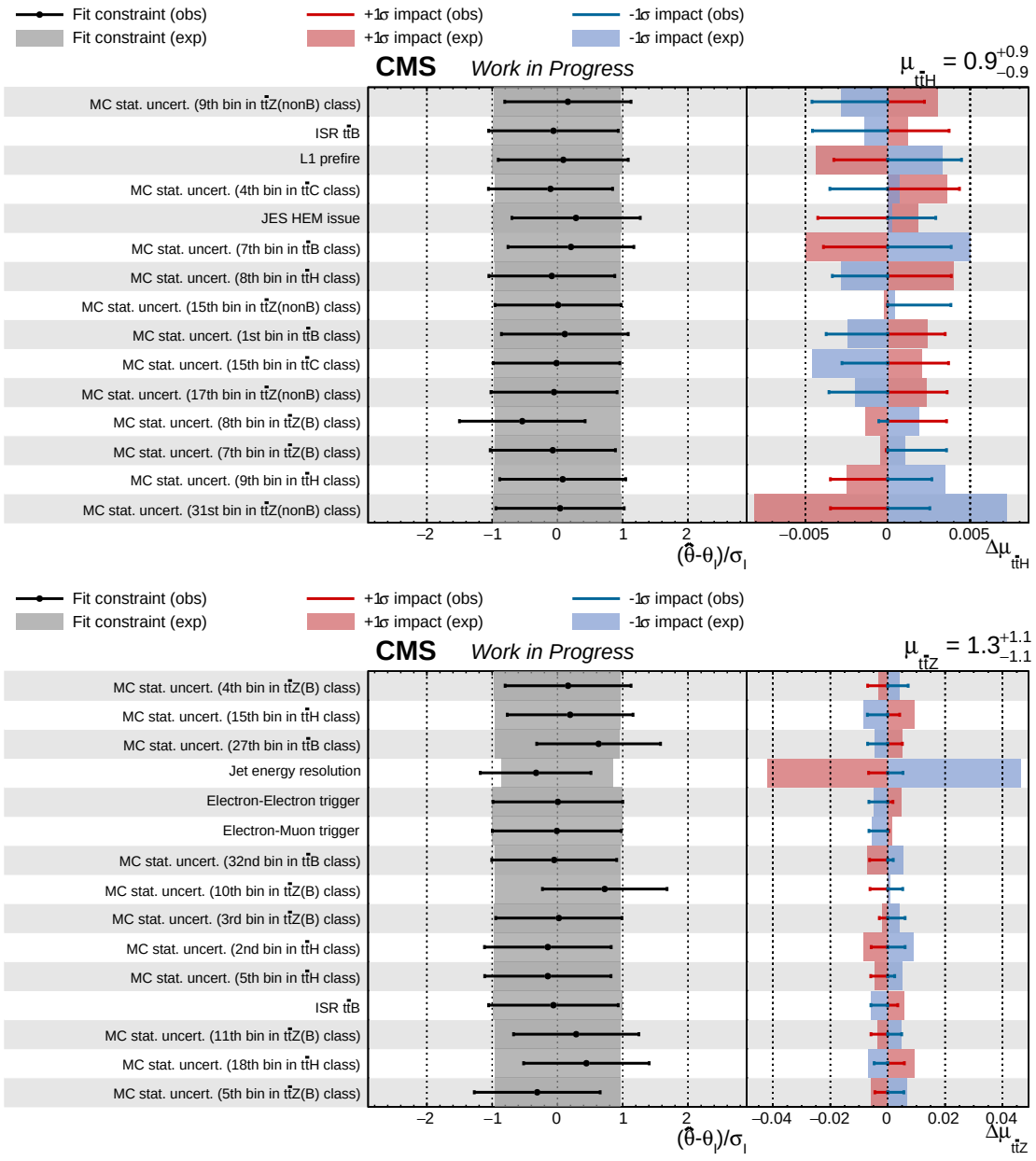


Figure C.21: Pre-fit and post-fit results of the signal strength parameters $\mu_{t\bar{t}H}$ (top) and $\mu_{t\bar{t}Z}$ (bottom). Parameter names are organized in the left panel. The distance from the post-fit value $\hat{\theta}$ to the pre-fit value θ normalized to the pre-fit uncertainty σ_I are shown in the middle panel. The right panel shows the relative change to the signal strength parameter μ under scrutiny if the NP under consideration is shifted to its $\pm 1\sigma$ value. The parameters are sorted according to the size of the impact, parameters 121 to 135 are shown. Created with [193].

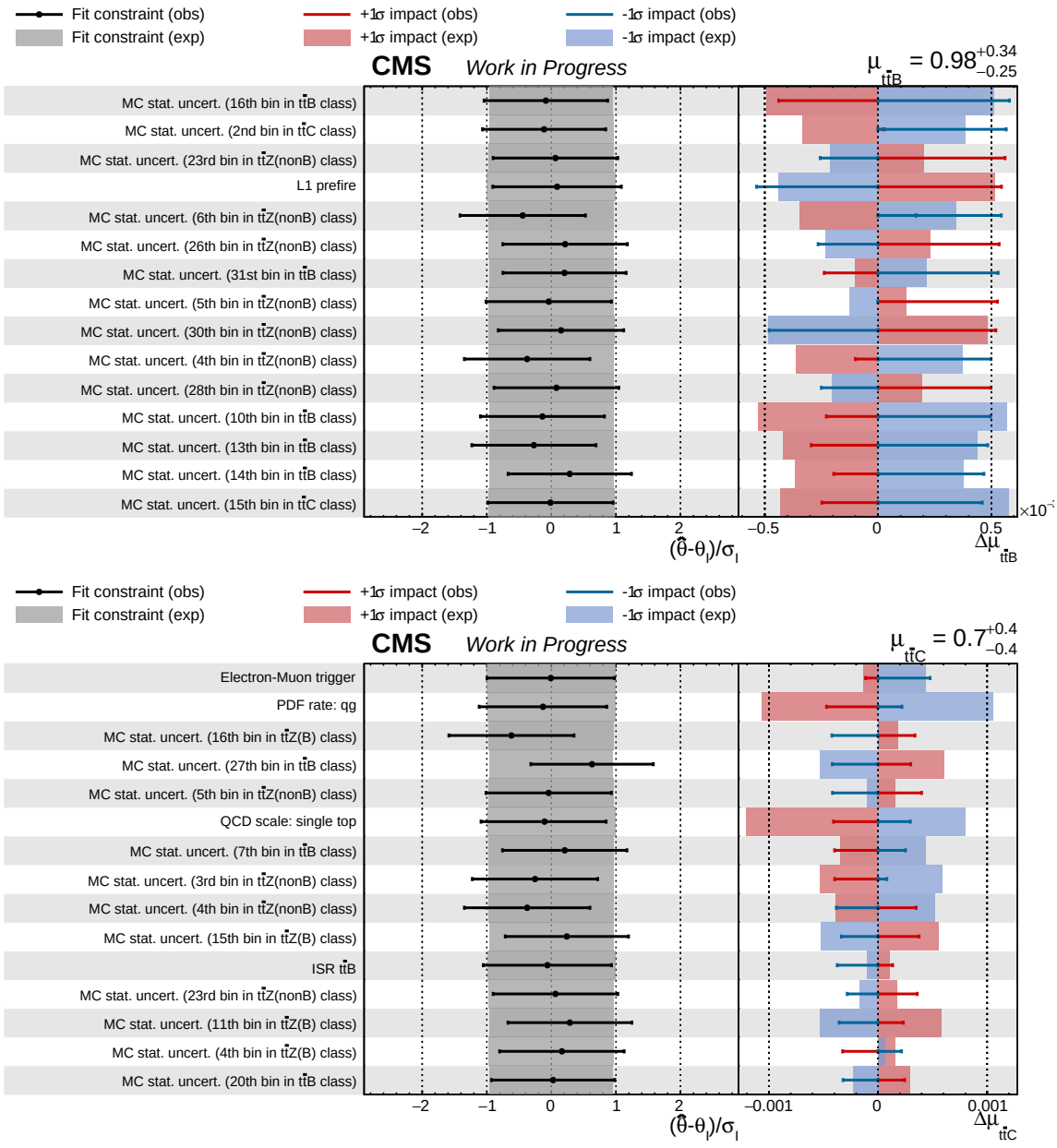


Figure C.22: Pre-fit and post-fit results of the signal strength parameters $\mu_{t\bar{t}B}$ (top) and $\mu_{t\bar{t}C}$ (bottom). Parameter names are organized in the left panel. The distance from the post-fit value $\hat{\theta}$ to the pre-fit value θ normalized to the pre-fit uncertainty σ_I are shown in the middle panel. The right panel shows the relative change to the signal strength parameter μ under scrutiny if the NP under consideration is shifted to its $\pm 1\sigma$ value. The parameters are sorted according to the size of the impact, parameters 136 to 150 are shown. Created with [193].

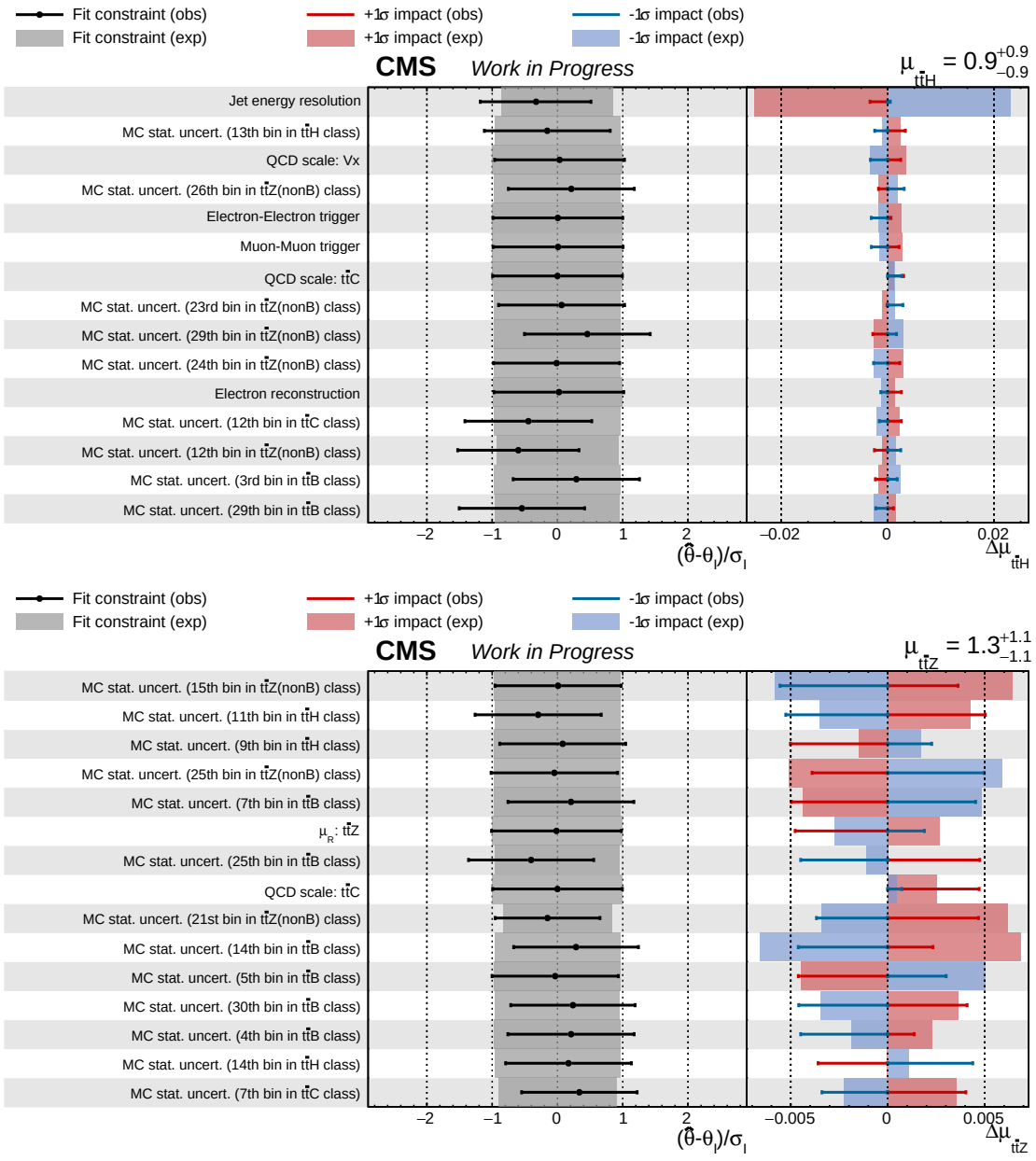


Figure C.23: Pre-fit and post-fit results of the signal strength parameters $\mu_{t\bar{t}H}$ (top) and $\mu_{t\bar{t}Z}$ (bottom). Parameter names are organized in the left panel. The distance from the post-fit value $\hat{\theta}$ to the pre-fit value θ normalized to the pre-fit uncertainty σ_I are shown in the middle panel. The right panel shows the relative change to the signal strength parameter μ under scrutiny if the NP under consideration is shifted to its $\pm 1\sigma$ value. The parameters are sorted according to the size of the impact, parameters 136 to 150 are shown. Created with [193].

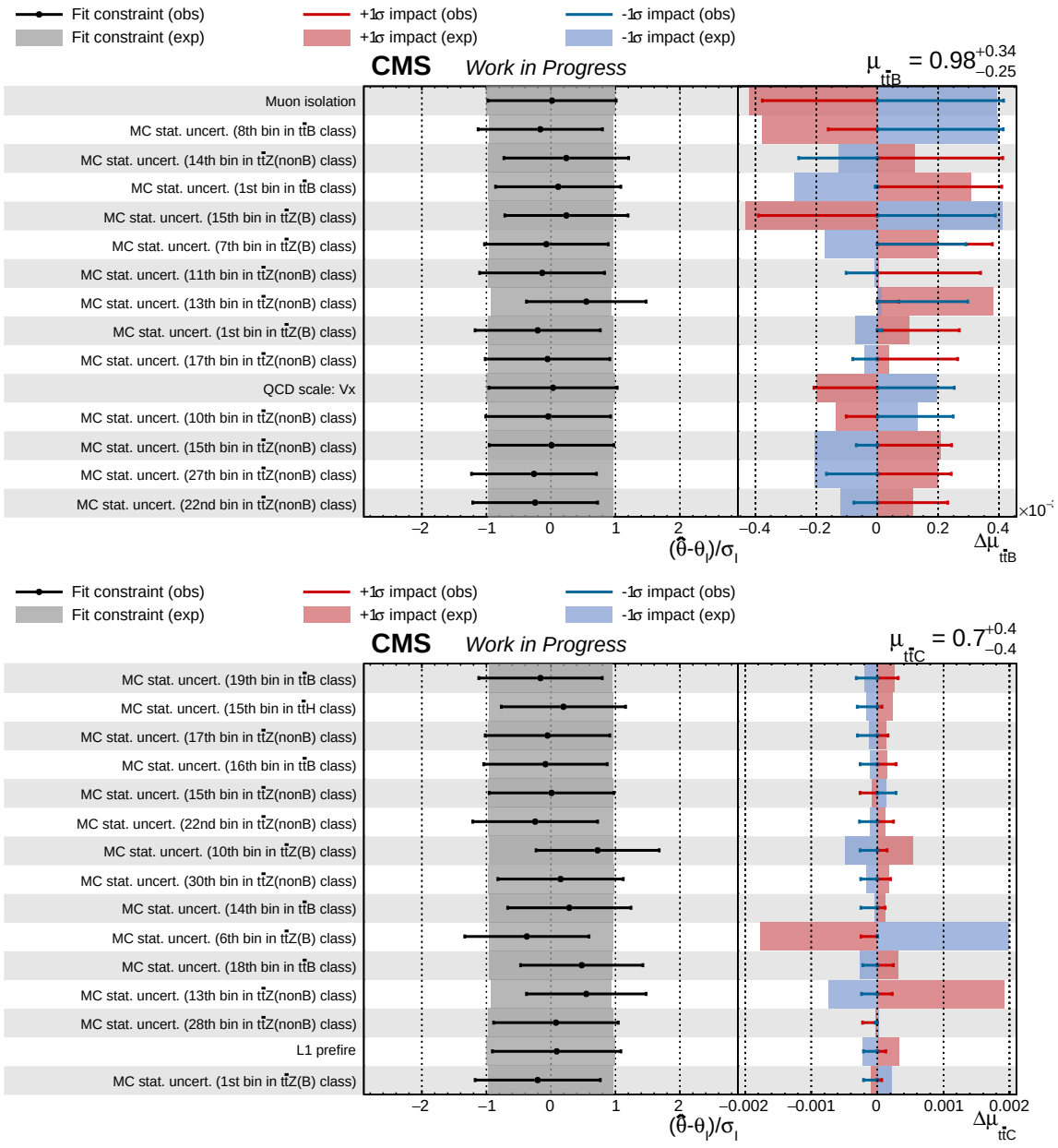


Figure C.24: Pre-fit and post-fit results of the signal strength parameters $\mu_{t\bar{t}B}$ (top) and $\mu_{t\bar{t}C}$ (bottom). Parameter names are organized in the left panel. The distance from the post-fit value $\hat{\theta}$ to the pre-fit value θ normalized to the pre-fit uncertainty σ_I are shown in the middle panel. The right panel shows the relative change to the signal strength parameter μ under scrutiny if the NP under consideration is shifted to its $\pm 1\sigma$ value. The parameters are sorted according to the size of the impact, parameters 151 to 165 are shown. Created with [193].

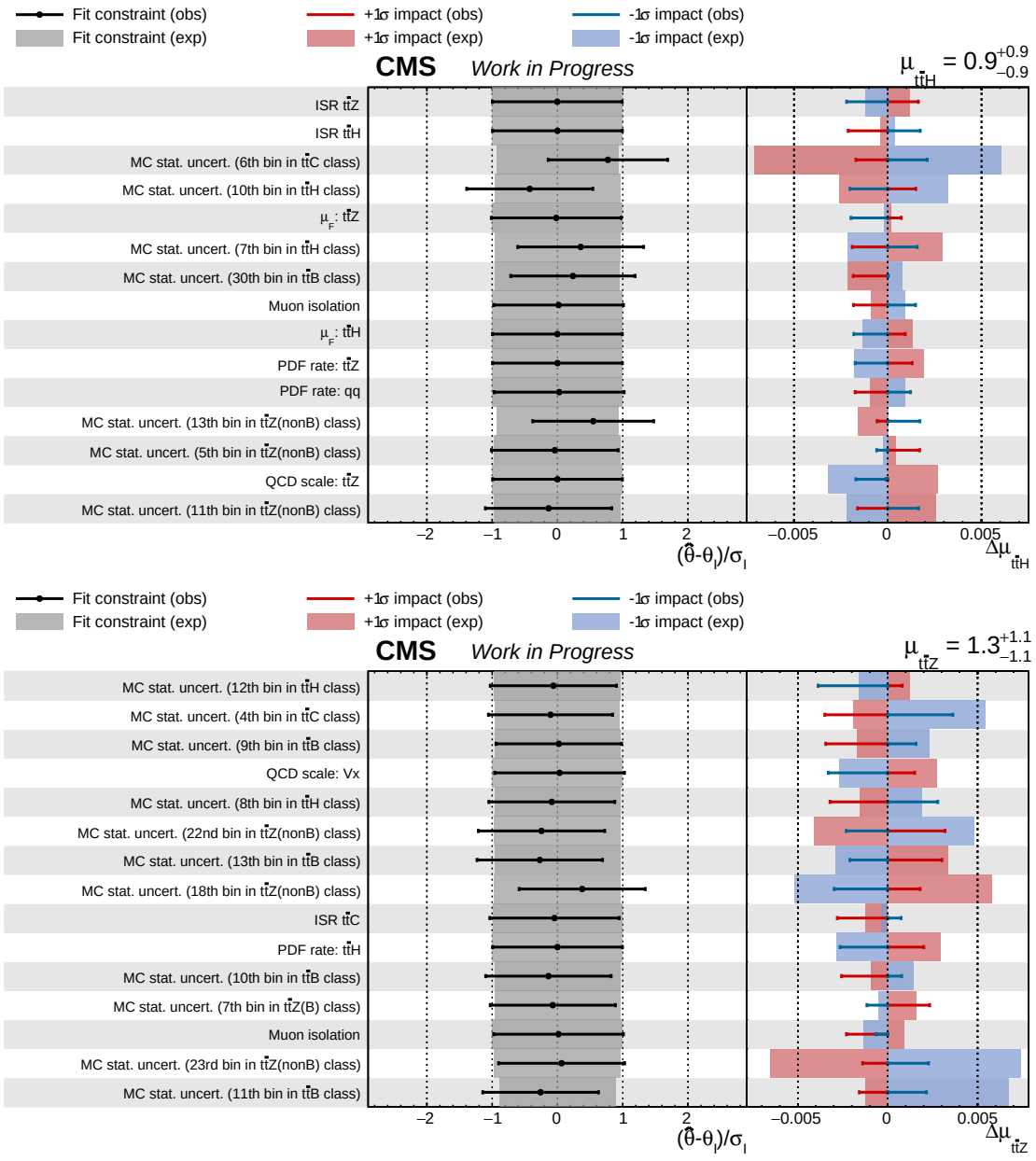


Figure C.25: Pre-fit and post-fit results of the signal strength parameters $\mu_{t\bar{t}H}$ (top) and $\mu_{t\bar{t}Z}$ (bottom). Parameter names are organized in the left panel. The distance from the post-fit value $\hat{\theta}$ to the pre-fit value θ normalized to the pre-fit uncertainty σ_I are shown in the middle panel. The right panel shows the relative change to the signal strength parameter μ under scrutiny if the NP under consideration is shifted to its $\pm 1\sigma$ value. The parameters are sorted according to the size of the impact, parameters 151 to 165 are shown. Created with [193].

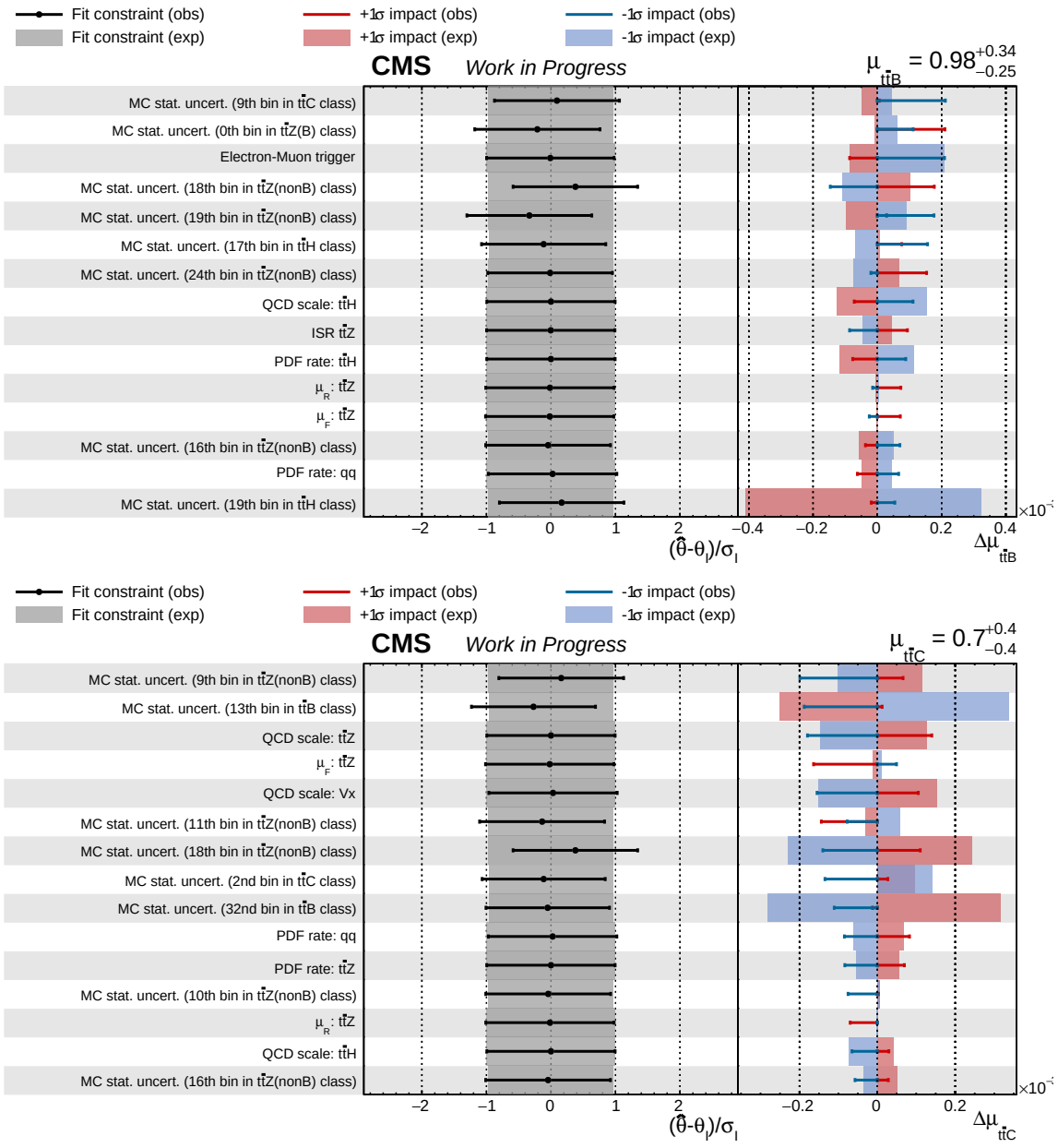


Figure C.26: Pre-fit and post-fit results of the signal strength parameters $\mu_{t\bar{t}B}$ (top) and $\mu_{t\bar{t}C}$ (bottom). Parameter names are organized in the left panel. The distance from the post-fit value $\hat{\theta}$ to the pre-fit value θ normalized to the pre-fit uncertainty σ_I are shown in the middle panel. The right panel shows the relative change to the signal strength parameter μ under scrutiny if the NP under consideration is shifted to its $\pm 1\sigma$ value. The parameters are sorted according to the size of the impact, parameters 166 to 180 are shown. Created with [193].

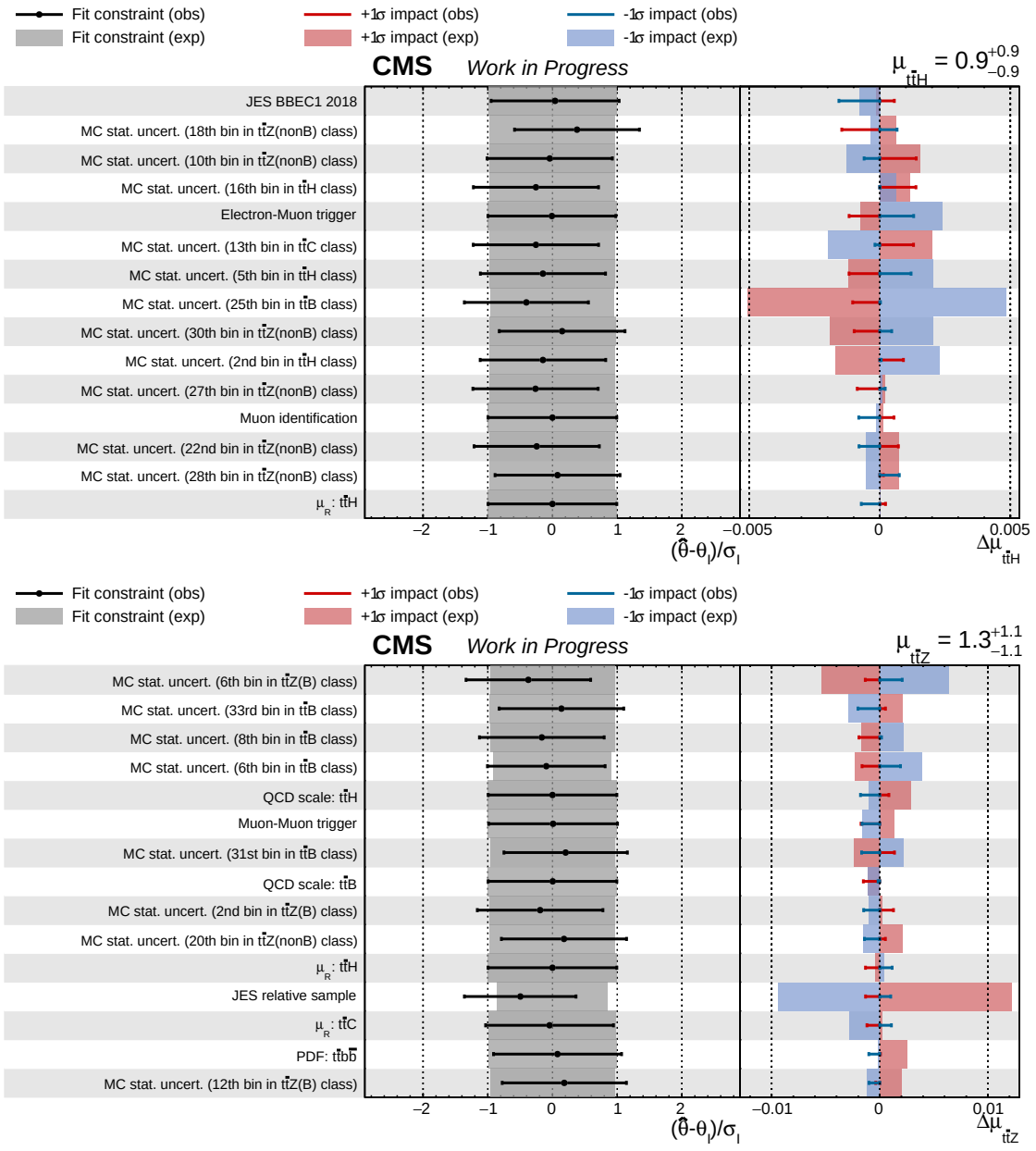


Figure C.27: Pre-fit and post-fit results of the signal strength parameters $\mu_{t\bar{t}H}$ (top) and $\mu_{t\bar{t}Z}$ (bottom). Parameter names are organized in the left panel. The distance from the post-fit value $\hat{\theta}$ to the pre-fit value θ normalized to the pre-fit uncertainty σ_I are shown in the middle panel. The right panel shows the relative change to the signal strength parameter μ under scrutiny if the NP under consideration is shifted to its $\pm 1\sigma$ value. The parameters are sorted according to the size of the impact, parameters 166 to 180 are shown. Created with [193].

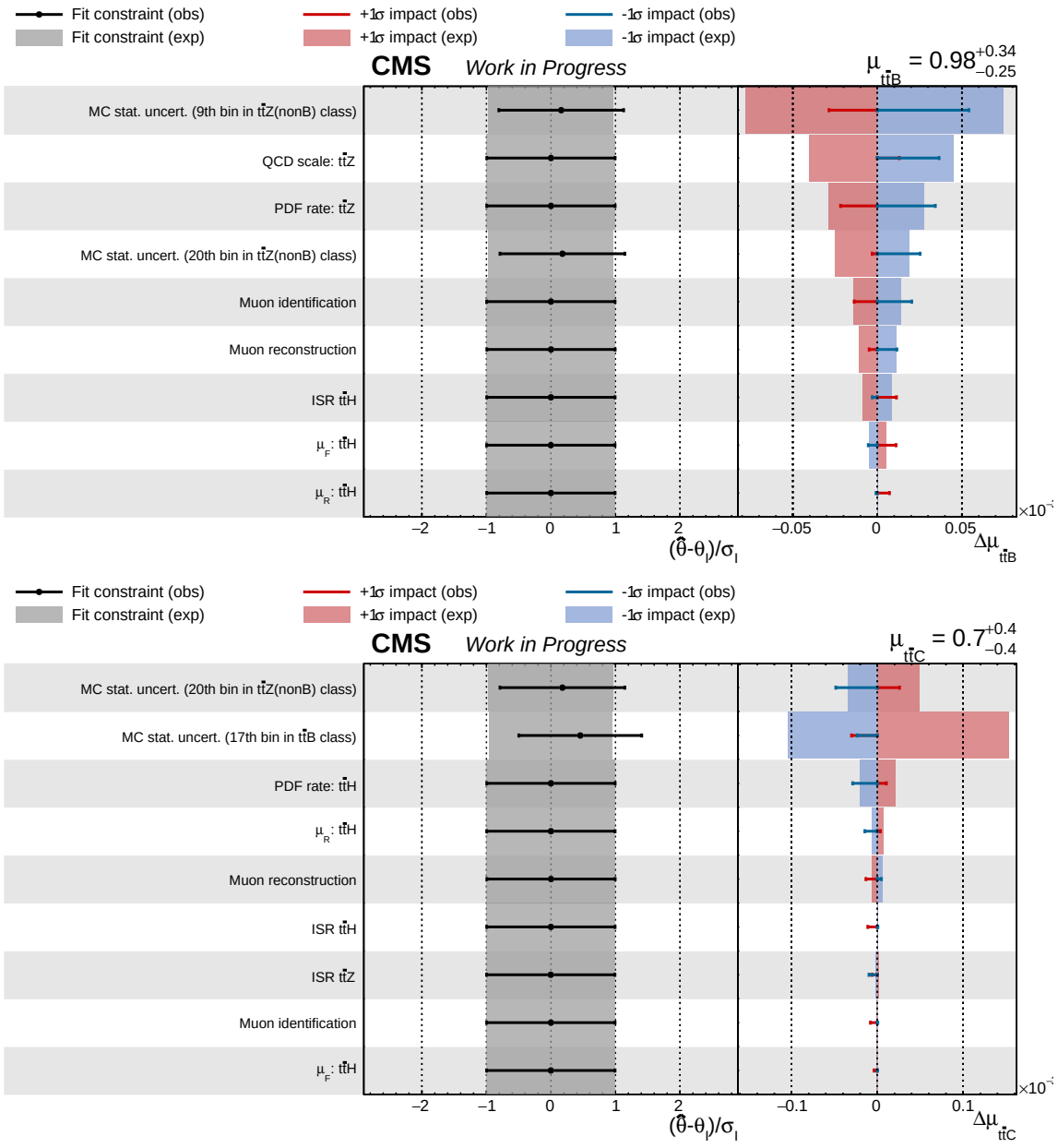


Figure C.28: Pre-fit and post-fit results of the signal strength parameters $\mu_{t\bar{t}B}$ (top) and $\mu_{t\bar{t}C}$ (bottom). Parameter names are organized in the left panel. The distance from the post-fit value $\hat{\theta}$ to the pre-fit value θ normalized to the pre-fit uncertainty σ_I are shown in the middle panel. The right panel shows the relative change to the signal strength parameter μ under scrutiny if the NP under consideration is shifted to its $\pm 1\sigma$ value. The parameters are sorted according to the size of the impact, parameters 180 to 189 are shown. Created with [193].

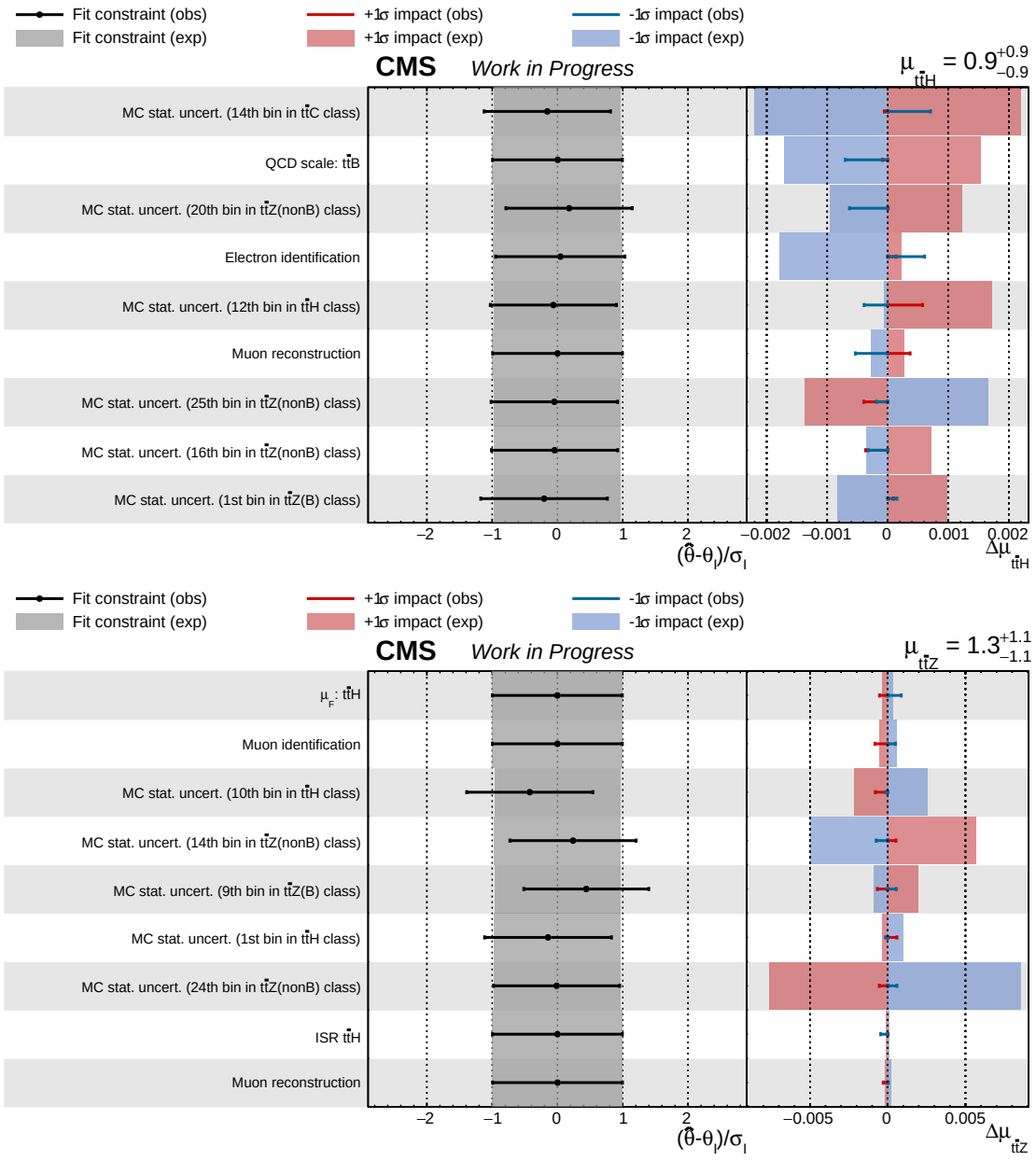


Figure C.29: Pre-fit and post-fit results of the signal strength parameters $\mu_{t\bar{t}H}$ (top) and $\mu_{t\bar{t}Z}$ (bottom). Parameter names are organized in the left panel. The distance from the post-fit value $\hat{\theta}$ to the pre-fit value θ normalized to the pre-fit uncertainty σ_I are shown in the middle panel. The right panel shows the relative change to the signal strength parameter μ under scrutiny if the NP under consideration is shifted to its $\pm 1\sigma$ value. The parameters are sorted according to the size of the impact, parameters 180 to 189 are shown. Created with [193].

Acknowledgments

Zuallererst möchte ich Prof. Dr. Ulrich Husemann meinen herzlichen Dank für die wertvolle Unterstützung und die intensive Betreuung während der gesamten Doktorandenzeit aussprechen. Besonders dankbar bin ich für die gewährten Gestaltungsfreiheiten und die Möglichkeit, unterschiedliche Themenfelder aktiv gestalten zu können. Diese Erfahrungen haben meinen Horizont erweitert und entscheidend dazu beigetragen, dass Ideen und Erkenntnisse in diese Arbeit einfließen konnten.

Darüber hinaus möchte ich Prof. Dr. Thomas Müller für die Unterstützung und Übernahme des Korreferats danken.

Ein weiterer Dank gilt dem Institut für Experimentelle Teilchenphysik für die Ermöglichung dieser Dissertation sowie die Teilnahme an nationalen und internationalen Konferenzen, die den Wissensaustausch und die Präsentation von Ergebnissen förderten. Mein Dank gilt auch dem Sekretariat, Diana Fellner und Bärbel Bräunling, für die zuverlässige Unterstützung und die reibungslose Organisation. Darüber hinaus danke ich dem Bundesministerium für Bildung und Forschung für die Förderung der Grundlagenforschung.

Einen ganz besonderen Dank möchte ich Dr. Jan van der Linden und Dr. Michael Wassmer für ihre technische Expertise und ihre wertvolle Unterstützung bei der Umsetzung dieser Arbeit aussprechen. Ebenfalls danke ich den Kollegen der CMS Collaboration, insbesondere in TOP sowie BTV, für interessanten Input und wertvolle Anregungen. Im Rahmen von BTV möchte ich vor allem Dr. Soureek Mitra und Dr. Max Neukum meinen Dank aussprechen.

Ebenso möchte ich meinen Arbeitskollegen danken, insbesondere Nikita Shadskiy, Moritz Molch, Rufa Rafeek und Michele Mormile für bereichernde Perspektiven. Ein weiterer Dank gilt den Bachelor- und Masteranden, die ich betreut habe und deren Studien wertvoll zu dieser Arbeit beigetragen haben: Yee-Ying Cung, Timo Halenke, Clemens Wolter, Sergey Lelyakin, sowie Paul Osswald.

Ein herzlicher Dank gilt meiner Familie und meinen Freunden für ihre unermüdliche Unterstützung während der gesamten Zeit dieser Arbeit.