Top quark pole mass from $t\bar{t}$+jet using a machine learning based reconstruction for $t\bar{t}$ kinematics

Dissertation

zur Erlangung des Doktorgrades
an der Fakultät für Mathematik, Informatik und Naturwisschenschaften
Fachbereich Physik
der Universität Hamburg

vorgelegt von

Sebastian Wuchterl

Hamburg

2022

Gutachter/innen der Dissertation:	Prof. Dr. Katerina Lipka Prof. Dr. Sven-Olaf Moch Dr. Marcel Vos
Zusammensetzung der Prüfungskommission:	Prof. Dr. Katerina Lipka Prof. Dr. Sven-Olaf Moch Prof. Dr. Elisabetta Gallo Prof. Dr. Dieter Horns Priv.-Doz. Dr. Olaf Behnke
Vorsitzender der Prüfungskommission:	Prof. Dr. Dieter Horns
Datum der Disputation:	10.02.2023
Vorsitzender Fach-Promotionsausschuss PHYSIK:	Prof. Dr. Günter H. W. Sigl
Leiter des Fachbereichs PHYSIK:	Prof. Dr. Wolfgang J. Parak
Dekan der Fakultät MIN:	Prof. Dr.-Ing. Norbert Ritter

Abstract

In this thesis, a measurement of the top quark mass as defined in the on-shell renormalization scheme, also known as pole-mass $m_{\mathrm{t}}^{\mathrm{pole}}$, is presented. The analysis is performed using events where a top quark-antiquark pair ($\mathrm{t\bar{t}}$) is produced in association with at least one additional jet ($\mathrm{t\bar{t}}$+jet) in proton-proton collisions at the CERN LHC at a center-of-mass energy of 13 TeV. The data are recorded by the Compact Muon Solenoid experiment in 2016, corresponding to a total integrated luminosity of $36.3\,\mathrm{fb}^{-1}$. Events with two opposite-sign leptons in the final state ($\mathrm{e}^{+}\mathrm{e}^{-}$, $\mu^{+}\mu^{-}$, $\mathrm{e}^{\pm}\mu^{\mp}$) are analyzed. The $\mathrm{t\bar{t}}$+jet production cross section is measured as a function of the inverse of the invariant mass of the $\mathrm{t\bar{t}}$+jet system, $\rho = 340\,\mathrm{GeV}/m_{\mathrm{t\bar{t}+jet}}$. A novel analysis technique based on machine learning is developed for the reconstruction of the main observable and for the event classification. The measurement is unfolded to the parton level and is compared to the theoretical prediction at next-to-leading order, using the dynamic renormalization and factorization scales. The theoretical predictions are obtained by using two alternative sets of parton distribution functions (PDFs). The top quark mass is extracted from the fit of the theoretical predictions to the data. For the ABMP16NLO PDF, this results in $m_{\mathrm{t}}^{\mathrm{pole}} = 172.93 \pm 1.36\,\mathrm{GeV}$. When using the CT18NLO PDF instead, the value is $m_{\mathrm{t}}^{\mathrm{pole}} = 172.13 \pm 1.43\,\mathrm{GeV}$.

Zusammenfassung

Diese Arbeit beschreibt eine Bestimmung der Masse des Top Quarks, definiert im Polmassenrenormierungsschema, auch bekannt als Polmasse $m_{\mathrm{t}}^{\mathrm{pole}}$. Es werden hierzu die in Proton-Proton Kollisionen am CERN LHC bei einer Schwerpunktsenergie von 13 TeV erzeugten Daten verwendet, die vom CMS-Detektor im Jahr 2016 aufgezeichnet wurden und einer integrierten Luminosität von $36.3\,\mathrm{fb}^{-1}$ entsprechen. Die Messung basiert auf der Sensitivität von Ereignissen, in denen ein Top Quark-Antiquark Paar ($\mathrm{t\bar{t}}$) im Zusammenhang mit einem zusaeztlichen Jet erzeugt wurde ($\mathrm{t\bar{t}}$+jet). Es werden Ereignisse, die zwei Leptonen unterschiedlicher Ladung ($\mathrm{e}^{+}\mathrm{e}^{-}$, $\mu^{+}\mu^{-}$, $\mathrm{e}^{\pm}\mu^{\mp}$) enthalten, untersucht. Der $\mathrm{t\bar{t}}$+jet Wirkungsquerschnitt wird als Funktion des Kehrwertsinvarianten der invarianten Masse des $\mathrm{t\bar{t}}$+jet Systems, $\rho = 340\,\mathrm{GeV}/m_{\mathrm{t\bar{t}+jet}}$, gemessen. Es werden für die Messung neue Analysemethoden basierend auf Maschniellem Lernen verwendet, um die Hauptbeobachtungsgröße zu rekonstruieren und Ereginisse zu kategorisieren. Die Ergebnisse der Messung werden durch Entfaltung auf Partonebene bestimmt, und anschliessend mit theoretischen Vorhersagen, die eine dynamische Renormierungs- und Faktorisierungsskala verwenden, in naechstführender Ordnung verglichen. Die theoretischen Vorhersagen werden auf Basis von zwei verschiedenen Partondichtefunktionen (PDFs) gewonnen. Die Top Quark Polmasse wird mit Hilfe einer Anpassung der theoretischen Vorhersagen zu den gemessenen Daten bestimmt. Für die APBM16NLO PDF, ist das Ergebnis $m_{\mathrm{t}}^{\mathrm{pole}} = 172.93 \pm 1.36\,\mathrm{GeV}$. Unter Verwendung der CT18NLO PDF ist der resultierende Wert $m_{\mathrm{t}}^{\mathrm{pole}} = 172.13 \pm 1.43\,\mathrm{GeV}$.

Contents

1 Introduction

The standard model (SM) of particle physics is a quantum field theory that provides the best understanding of the interactions between elementary particles, to date. The SM is characterized by the principle of gauge invariance, which requires the Lagrange density describing the dynamics of the fields to remain invariant under gauge transformations of the symmetry group $SU(3) \times SU(2) \times U(1)$. The interactions between fundamental particles are described by the strong and the electroweak interactions. The latter is an unification of the electromagnetic and weak interactions, as defined in the electroweak theory, based on the $SU(2) \times U(1)$ symmetry group. The strong interaction is described by quantum chromodynamics (QCD), based on the $SU(3)$ symmetry group. The masses of the elementary particles are generated through the Higgs–Brout–Englert mechanism [1, 2], which spontaneously breaks the electroweak symmetry. The discovery of the Higgs boson by the ATLAS and CMS experiments at the Large Hadron Collider at CERN in 2012 [3, 4] lead to the Nobel Prize in particle physics in 2013. The SM successfully describes all observations in particle physics experiment over a wide energy range.

A special role within the SM is taken by the top quark, which was discovered by the CDF and D0 experiments at the Fermilab Tevatron in 1995 [5, 6]. With a mass of about 173 GeV, the top quark is the most massive elementary particle known. As a consequence, the top quark decays before forming bound states and represents a unique possibility to study the properties of a bare quark. The mass of the top quark, m_{t}, is a fundamental parameter of the SM, however its value needs to be determined experimentally. The value and precision of m_{t} play a key role in understanding the electroweak symmetry breaking mechanism [7, 8]. The value of m_{t} is of high relevance for the self-consistency tests of the SM, via, e.g., global electroweak fits [9–12], which makes the precise measurement of m_{t} essential for the understanding of the fundamental structure and interactions of matter. One of the challenges is that there is no unique definition of the top quark mass. Due to vacuum polarization, all charges and couplings in QCD become subject to renormalization. Therefore, all quark masses depend on the renormalization scheme considered. Different renormalization strategies have been suggested, with a natural choice of a minimal-subtraction renormalization scheme. In this scheme, the quark masses become dependent on the renormalization scale, driven by the renormalization group equations, similar to the running of the strong coupling constant. The running mass of the top quark is well defined only above the top quark production threshold, which is, however, the region of largest sensitivity of current experimental measurements. Alternatively, the pole mass of the top quark is used ($m_{\mathrm{t}}^{\mathrm{pole}}$). The definition of $m_{\mathrm{t}}^{\mathrm{pole}}$ would correspond to a pole in the quark propagator as in the case of the mass of a free particle. Indeed, the pole mass definition for quarks has a

conceptual problem since quarks are confined in hadrons and are not free particles. The pole mass of the top quark is only valid in the perturbation theory and suffers from an ambiguity on the order of Λ_{QCD}, known as renormalon problem [13–15].

In the first experimental attempts to extract the value of m_{t}, the reconstructed invariant mass of the top quark decay products was used. These so-called direct measurements of m_{t} have recently reached a precision of 400 MeV [16–23]. However, measurements of this kind rely on the reconstruction of the top quark decay products and hence on the modeling by using multi-purpose Monte-Carlo (MC) event generators. These generators are based on parton shower models, while heuristic models tuned to the data are used to describe the color neutralization in the nonperturbative regime and the underlying event. This leads to an ambiguity in the relation of the resulting value of the parameter considered the top quark mass in the Monte Carlo simulation ($m_{\mathrm{t}}^{\mathrm{MC}}$) to the top quark mass as a parameter of the Lagrangian, defined in a certain renormalization scheme. This ambiguity leads to an additional uncertainty on the order of 0.5 GeV [24–31].

Alternatively, the value of m_{t} can be extracted by comparing measurements of the top quark production cross section with fixed-order theoretical predictions based on particular renormalization schemes. The value of $m_{\mathrm{t}}^{\mathrm{pole}}$ was extracted this way by both the ATLAS and CMS Collaborations [32–38], reaching an uncertainty of below 1 GeV [38, 39]. The top quark mass as defined in the modified minimal subtraction ($\overline{\mathrm{MS}}$) scheme, referred to as the top quark running mass, was also measured [31, 40, 41] and its scale dependence was recently investigated by the CMS experiment [42, 43].

Furthermore, novel observables are introduced to probe the top quark mass, which are less affected by the description of the non-perturbative effects. In this thesis, the top quark pole mass is extracted by using the normalized differential cross section of $\mathrm{t\bar{t}}$ production in association with at least one energetic jet [44]. In this thesis, such a novel observable ρ, related to the invariant mass of the $\mathrm{t\bar{t}}$+jet system, is explored as suggested in Refs. [41, 44, 45]. In particular, the normalized cross section of $\mathrm{t\bar{t}}$+jet production is measured as a function of ρ, where ρ is defined as

$$\rho = \frac{2\mathrm{m}_0}{m_{\mathrm{t\bar{t}+jet}}}. \tag{1.1}$$

Here, m_0 is a constant with a value of 170 GeV, as used in previous measurements, and $m_{\mathrm{t\bar{t}+jet}}$ is the invariant mass of the $\mathrm{t\bar{t}}$+jet system. The presence of an additional jet leads to increased sensitivity to $m_{\mathrm{t}}^{\mathrm{pole}}$ as compared to $\mathrm{t\bar{t}}$ production, since the kinematic properties of the radiated gluon off the top quark is driven by the top quark mass. As in the case of $\mathrm{t\bar{t}}$ production, highest sensitivity to $m_{\mathrm{t}}^{\mathrm{pole}}$ is expected close to the production threshold, corresponding to $\rho > 0.65$. The first measurement exploring the ρ observable was performed by the ATLAS collaboration using the proton-proton collision data at the LHC at $\sqrt{s} = 7$ TeV [35] and 8 TeV [46], and the values of $m_{\mathrm{t}}^{\mathrm{pole}}$ and $m_{\mathrm{t}}(m_{\mathrm{t}})$ were obtained.

In the analysis presented in this thesis, the extraction of $m_{\mathrm{t}}^{\mathrm{pole}}$ is performed using the measurement of the normalized cross section of $\mathrm{t\bar{t}}$+jet production as a function of ρ. The proton-proton collision data at $\sqrt{s} = 13$ TeV recorded by the CMS experiment at the LHC in 2016, corresponding to a total integrated luminosity of 36.3 fb^{-1}, are used. In the presented analysis, candidate $\mathrm{t\bar{t}}$+jet events are selected in the final state with two leptons of opposite charge ($\mathrm{e^+e^-}$,$\mathrm{e^\pm\mu^\mp}$,$\mu^+\mu^-$) and the normalized differential $\mathrm{t\bar{t}}$+jet cross section is measured directly at the parton level by using the maximum-likelihood unfolding method.

In this thesis, a novel method to reconstruct the kinematic properties of the $\mathrm{t\bar{t}}$+jet events is developed, based on a machine-learning approach. In particular, the value of ρ at the parton

level is directly reconstructed by a neural network, leading to a superior efficiency and resolution performance with respect to earlier analyses. In the likelihood fit, all systematic uncertainties are profiled using nuisance parameters, following the approach of Ref. [42]. The benefits of this method are that the signal and background processes are fitted simultaneously, and the systematic uncertainties and their correlations are determined from the experimental data. As a result, the systematic uncertainties are constrained in the fit, which leads to significantly improved experimental precision as compared to earlier approaches. Furthermore, the value of the assumption for the top quark mass parameter in the simulation is incorporated as a free parameter in the fit, mitigating the remaining dependence of the experimental acceptance on the value of $m_{\mathrm{t}}^{\mathrm{MC}}$. The value of $m_{\mathrm{t}}^{\mathrm{pole}}$ is extracted by comparing the measured differential cross section with the theoretical predictions at next-to-leading order QCD.

This thesis is organized as follows. A brief theoretical overview of the SM is provided in Chapter 2, with a particular focus on aspects relevant for the work presented in this thesis. This includes the role of the top quark in the SM. The CERN LHC and the CMS experiment are introduced in Chapter 3, and the main aspects of the CMS event and physics reconstruction methods are discussed in Chapter 4. Within the same chapter, studies and developments for the identification of b quark jets at the CMS High Level Trigger for the Run 3 and the High-Luminosity LHC, with significant contributions from the author of this thesis, are presented. These results are also documented in Refs. [47] and [48]. Chapter 5 introduces relevant statistical methods and the basic concepts of machine learning and neural networks that are employed for the novel reconstruction method of $\mathrm{t\bar{t}}$ kinematic properties. The measurement of the $\mathrm{t\bar{t}}$+jet differential cross section as a function of ρ is described in Chapter 6. In the same chapter, the extraction of the top quark mass $m_{\mathrm{t}}^{\mathrm{pole}}$ is discussed. The results obtained in this thesis are published in Ref. [49]. Finally, Chapter 7 presents a sensitivity study by projecting the measurement presented in this thesis when using the statistics of the full Run 2 data set, corresponding to the LHC data-taking years of 2016–2018.

2 The role of the top quark in particle physics

The standard model of particle physics (SM) is a quantum field theory that describes all fundamental particles known to date. It further explains their interactions via the strong, electromagnetic, and weak forces. Although the SM is successfully tested experimentally over a wide energy range to very high precision, it is known to be incomplete. The SM and its possible extensions are tested at the energy frontier using the proton-proton (pp) collisions at the Large Hadron Collider (LHC). The data of such pp collisions is used for the measurement of the production of top quark pairs associated with a jet to extract the top quark pole mass.

In this chapter, the aspects of the SM most relevant for this thesis are introduced. Extended reviews can be found in, e.g., Refs. [50–53]. In Section 2.1, the SM is introduced briefly and the concepts of the electroweak theory, the Higgs mechanism, and quantum chromodynamics (QCD) are provided. Specifics of the theoretical description of pp scattering events are given in Section 2.2, and the QCD factorization approach is discussed. In Section 2.3, the top quark and its properties are described. Finally, in Section 2.4, the theoretical prediction used for the work presented in this thesis is introduced.

2.1 The standard model of particle physics

As a quantum field theory, the Standard Model of fundamental particles and their interactions follows the rules of quantum mechanics and special relativity. A particle can be understood as an excitation of a quantum field, characterized by its quantum numbers such as charge and spin. The three interactions: electromagnetic, weak, and strong, are expressed as quantized fields. Gravitation, formulated in general relativity [54], is the only known fundamental interaction not described in the SM, however it applies only on macroscopic scales. The SM is represented by its underlying symmetries and can be formulated as a Lagrangian density $\mathscr{L}$ that is required to be invariant under local gauge transformations to preserve renormalizability [55]. The SM Lagrangian further is invariant under translational and rotational transformations, as well as Lorentz boosts. From the Noethers theorem [56] follows that symmetries are related with conservation laws, as energy, momentum, and angular momentum conservation follow from the

time invariance, rotation, and translation symmetries. The SM is built of a unified electroweak (EWK) sector, expressed by the $SU(2) \otimes U(1)$ symmetry group, and a strong sector based on the $SU(3)$ group.

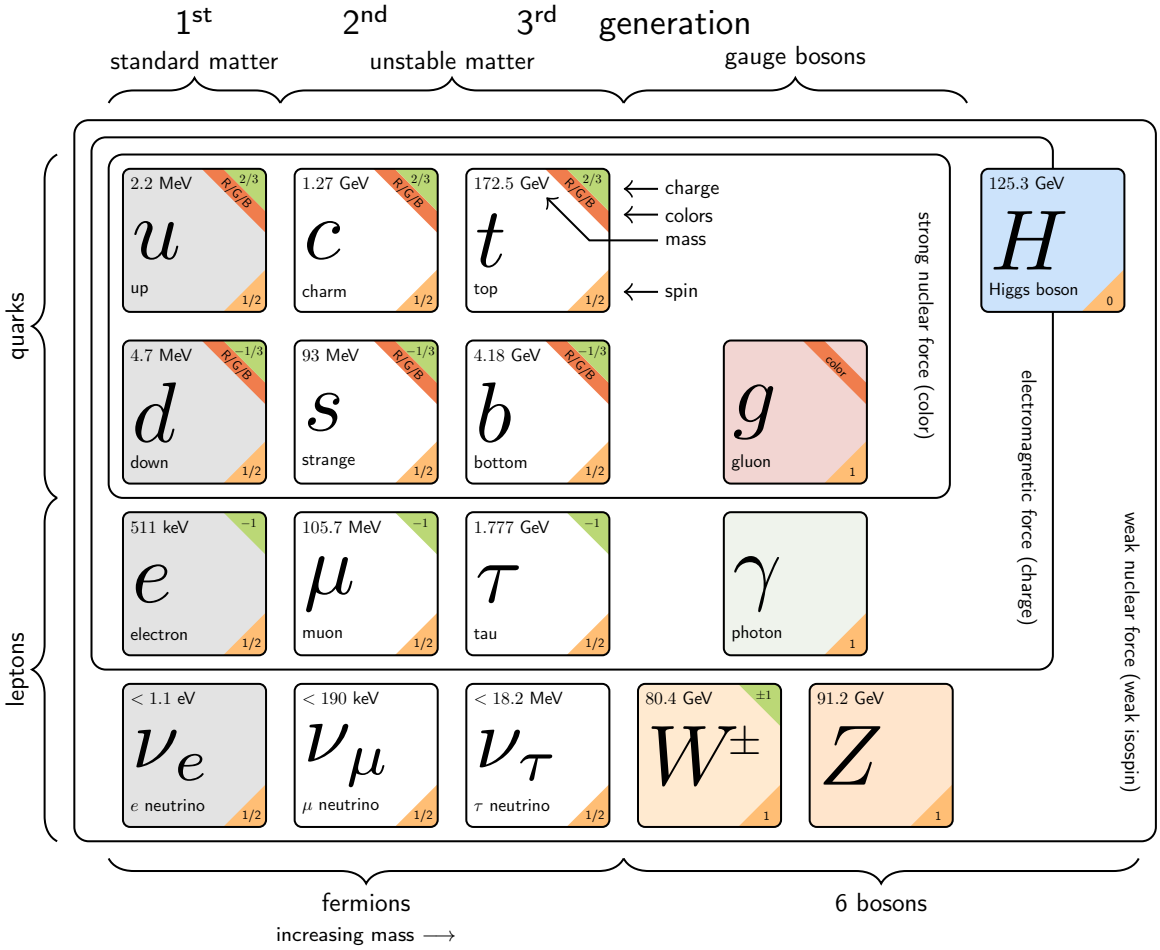

Figure 2.1: Summary of the fundamental particles of the standard model, showing fermions (leptons and quarks), gauge bosons (force carriers), and the Higgs boson. Based on [57], values from Ref. [58].

All elementary particles of the SM and their interactions are shown in Figure 2.1. Based on their spin, the particles can be subdivided into two categories. Particles with half-integer spin are called fermions and follow the Fermi–Dirac statistics. The fermions are leptons (e, ν_e, μ, ν_μ, τ, ν_τ) and quarks q (u, d, c, s, t, b) and represent the fundamental building blocks of matter in the universe. Quarks are subject to all fundamental interactions, while charged leptons do not interact via the strong interaction, and the neutrinos only interact weakly. For each fermion, a corresponding antiparticle exists, which has identical quantum numbers but with the opposite sign. Fermions are further categorized into three generations. While the quantum numbers of the particles in each generation are the same, their masses increase from one generation to the other. Each generation contains the up-type quarks (u, c, t) and down-type quarks (d, s, b), with an electrical charge of $2/3\,e$ and $1/3\,e$, respectively. Each generation of leptons consists of a charged and neutral lepton, with the latter called neutrinos. The charged leptons are the electron (e), muon (μ), and tau lepton (τ), and neutrinos of the same generation are denoted as electron- (ν_e), muon (ν_μ), and tau neutrino (ν_τ). Charged leptons have the charge of $1\,e$. Often,

the six types of quarks and leptons are referred to as "flavors". The origin of the number of generations is not explained in the SM. However, the number of neutrino generations, and hence the number of lepton generations, has been confirmed with very high precision at the CERN Large Electron Positron (LEP) Collider [59]. A fourth generation could be excluded with high probability as determined from decay rates of the Z boson [60].

The elementary particles mediating the interactions are called bosons: these carry integer spin and correspond to the gauge fields of the respective symmetry group. The weak interaction is mediated by the W and Z bosons, the electromagnetic interaction by the photon (γ), and the strong interaction by the gluon (g). The only scalar boson is the Higgs boson (H), which mediates a fundamentally different interaction, responsible for the generation of particle masses through the Higgs–Brout–Englert mechanism [1, 2].

The formalism to describe the SM can be demonstrated, e.g., for quantum electrodynamics (QED). Its Lagrangian density (or Lagrangian) $\mathscr{L}$, for the $U(1)$ group can be written as

$$\mathscr{L}_{\mathrm{QED}} = \overline{\Psi}(i\gamma^{\mu}\partial_{\mu} - m)\Psi - q\overline{\Psi}\gamma^{\mu}\Psi A_{\mu} - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \tag{2.1}$$

where m (q) is the electron mass (charge), γ^{μ} are the Dirac matrices, and Ψ is the wave function of spin 1/2 particles. The covariant variable A_{μ} is introduced to ensure invariance under local $U(1)$ gauge transformations. A covariant derivative is introduced as $\partial_{\mu} \rightarrow \partial_{\mu} + iqA_{\mu}$, where A_{μ} is the gauge field of the photon. In Equation 2.1, the second term corresponds to the interaction and the third one is the field. Now, the term $q\overline{\Psi}\gamma^{\mu}\Psi$ can be identified as the current j^{μ} of the electromagnetic interaction $-qj^{\mu}A_{\mu}$, where q takes the role of a coupling constant. Usually, it is transformed in terms of a coupling strength $\alpha = e^2/4\pi \approx 1/137$. As a mass contribution in the field term would violate $U(1)$ gauge symmetry, the photon is massless. This leads to an infinite range of the electromagnetic interaction. Interactions between two electrons can now be described as the exchange of "virtual" photons, mediating the electric current. The interactions can be illustrated by means of Feynman diagrams, cf. Figure 2.2.

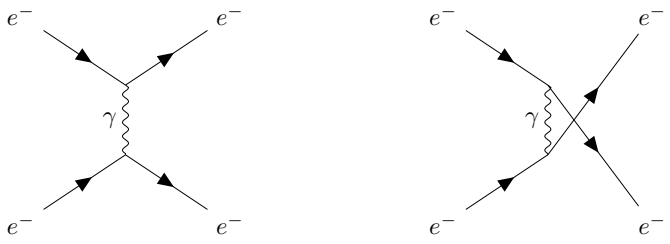

Figure 2.2: Feynman diagrams for electron–electron QED interactions.

Feynman diagrams allow calculating the transition probability according to Fermi's golden rule for a particular process, where every vertex, or interaction, adds a coupling strength to the calculation. If the coefficients are small, the interaction term can be considered a small perturbation of the free field term, and a perturbative approach can be applied to calculate scattering amplitudes.

Electroweak theory

The electroweak interaction is described by the unification of the electromagnetic and weak interactions, represented by the Glashow–Weinberg–Salam theory [61–63]. It is based on the $SU(2) \otimes U(1)$ symmetry group above an energy scale of about 246 GeV, and below is spon-

taneously broken into the weak and electromagnetic parts. The symmetry breaking will be discussed later in this section.

The gauge-invariant Lagrangian for the electroweak theory can be written as

$$\mathscr{L}_{EWK} = \underbrace{i\overline{L}\gamma^{\mu}D_{\mu}L}_{\text{doublets } L} + \underbrace{\overline{R}i\gamma^{\mu}D_{\mu}R}_{\text{singlets } R} - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} - \frac{1}{4}W_{a}^{\mu\nu}W_{\mu\nu}^{a}, \tag{2.2}$$

with

$$D_{\mu} = \partial_{\mu} - i\frac{g'}{2}\widehat{Y}B_{\mu} - ig\widehat{T}_{i}W_{\mu}^{i} \tag{2.3}$$

being the covariant derivative. The index i runs over the values 1–3. The generator $\widehat{Y}$ of the $U(1)$ group is the hypercharge operator, and $\widehat{T}_i$ are the three weak isospin generators of $U(1)$. This gives rise to one gauge boson B_μ for $U(1)$ and three gauge boson fields W_μ^i for $SU(3)$. They are defined as

$$B^{\mu\nu} = \partial_{\mu}B_{\nu} - \partial_{\nu}B_{\mu} \tag{2.4}$$

and

$$W_{\mu\nu}^{i} = \partial_{\mu}W_{\nu}^{i} - \partial_{\nu}W_{\mu}^{i} - g\epsilon^{ijk}W_{\mu}^{j}W_{\nu}^{k}. \tag{2.5}$$

The two constants g and g' are dimensionless coupling constants.

The fermion fields are split into left- and right-handed, where left-handed fermions and right-handed antifermions are organized in isospin doublets L of $SU(2)$ with a chirality of -1 and eigenvalues (T, T_3) of $(1/2, \pm 1/2)$. Instead, right-handed fermions and left-handed antifermions form isospin singlets R with a chirality of $+1$ and eigenvalues of $T = T_3 = 0$.

The weak interaction is observed to be maximal parity violating [64], meaning that only fermions with left chirality interact through the weak interaction. The charges and quantum numbers of the $SU(2)$ and $U(1)$ groups are the third component of the weak isospin T_3 and the weak hypercharge Y, respectively. They are related to the electric charge via the Gell–Mann–Nishijima formula [65, 66]:

$$Q = \frac{Y}{2} + T_3. \tag{2.6}$$

While the left-handed neutrinos participate in the weak interaction, the right-handed neutrinos do not.

The experimentally observed weak and charged current interactions can be expressed by a linear combination of W_μ^1 and W_μ^2 as

$$W_{\mu}^{\pm} = \frac{1}{\sqrt{2}}\left(W_{\mu}^{1} \mp W_{\mu}^{2}\right), \tag{2.7}$$

and performing a rotation of the kind

$$\begin{pmatrix} A_{\mu} \\ Z_{\mu} \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B_{\mu} \\ W_{\mu}^{3} \end{pmatrix} \tag{2.8}$$

to obtain the electromagnetic field A_μ and the neutral current field Z_μ corresponding to the Z boson. Here, θ_W is the electroweak mixing angle, which is also called Weinberg angle. The electroweak mixing is connected to the W and Z boson masses as $\cos(\theta_W) = m_{\mathrm{W}}/m_{\mathrm{Z}}$. Both coupling constants g and g' are related to the electromagnetic charge e as $e = g\sin\theta_W = g'\cos\theta_W$. The non-abelian structure of the $SU(2)$ symmetry group also gives rise to self-interactions of the gauge bosons, as illustrated in Figure 2.3. The effective mixing angle $\sin^2\theta_W^{\text{eff}}$

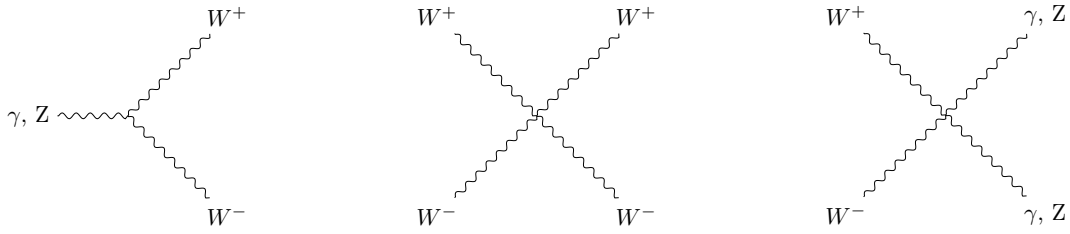

Figure 2.3: Feynman diagrams illustrating possible gauge boson self interactions.

was measured as 0.23122 ± 0.00002 [58]. A plethora of experimental precision tests of the EWK sector of the SM have been performed, e.g., at LEP [60].

The Higgs mechanism

The Lagrangian in Equation 2.2, without violating gauge symmetry, does not allow mass terms for fermions and electroweak bosons, which are however measured to be finite. In the SM, the masses are generated via the Higgs–Brout–Englert mechanism [1, 2] so that the symmetry is spontaneously broken. The Lagrangian of the new field after introducing a new doublet of scalar fields Φ can be written as

$$\mathscr{L}_H = \left(D_\mu \Phi^\dagger\right)\left(D^\mu \Phi\right) - V\left(\Phi\right), \tag{2.9}$$

with

$$D_\mu = \partial - \mu + ig\widehat{T}_i W_\mu^i + ig'\frac{\widehat{Y}}{2}B_\mu. \tag{2.10}$$

Here, $V(\Phi)$ is the Higgs potential dependent on two real parameters μ^2 and λ, defined as

$$V(\Phi) = -\mu^2 \Phi^\dagger \Phi + \lambda(\Phi^\dagger \Phi)^2. \tag{2.11}$$

This potential can be added to the EWK Lagrangian in Equation 2.2 as it is invariant under $SU(2) \otimes U(1)$ gauge transformation. The stability of the EWK vacuum requires the parameter λ to be positive [7], while the parameter μ^2 defines the minimum of the potential $V(\Phi)$ for a given value of λ. This effect and the shape of the profile of the potential are illustrated in Figure 2.4.

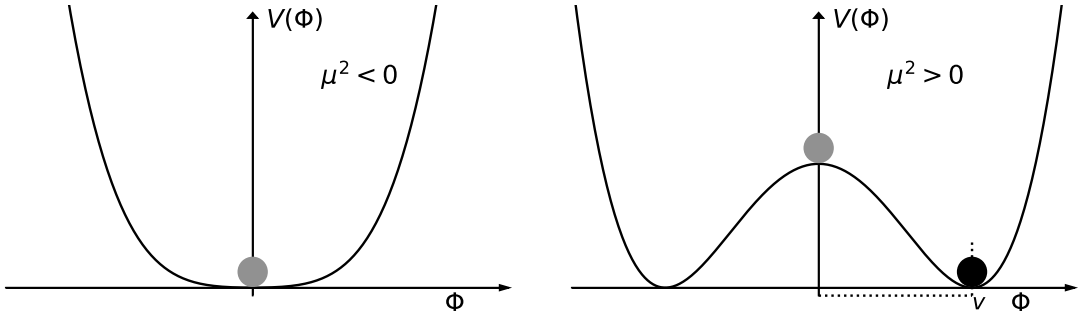

Figure 2.4: A sketch of the profile of the Higgs potential for values of $\mu^2 < 0$ before symmetry breaking (left), and $\mu^2 > 0$ after symmetry breaking (right).

For positive values of μ^2, the potential is still symmetric, but an infinite number of degenerate minima of the potential are generated. Consequently, this leads to ground-state fields not invariant under gauge transformation and a non-zero vacuum expectation value (VEV) v. This

state was chosen spontaneously when an energy below v was reached, while above that scale all particles are massless. After the electroweak symmetry breaking (EWSB), v is non-zero due to a change of μ^2. As physical properties do not depend on the exact minimum because of the symmetry of $V(\Phi)$, the minimum can be chosen such that $\Phi_{\text{min}} = (0, v/\sqrt{s})^{\text{T}}$ with $v = \sqrt{\mu^2\lambda}$. Expanding the ground-state field around the new minimum of the potential, the field can be written in terms of four degrees of freedom as

$$\Phi = \mathrm{e}^{i\xi_i \frac{\sigma_i}{2}} \begin{pmatrix} 0 \\ v + H \end{pmatrix}, \tag{2.12}$$

where H and ξ_i are real scalar fields. A transformation, denoted as unitary gauge, can be applied such that ξ_i vanish and H remains. This leads to the mass terms

$$m_{\text{Z}} = \frac{1}{2} v \sqrt{g + g'} \tag{2.13}$$

and

$$m_{\text{W}} = \frac{1}{2} v g, \tag{2.14}$$

while the photon stays massless. The value of v is determined to be about 246 GeV [58], and the large masses of the vector bosons are the origin for the short interaction range of the weak interaction.

The Higgs boson field itself obtains a mass, too, which is given by

$$m_{\text{H}} = \sqrt{2\lambda}\mu. \tag{2.15}$$

In addition, interaction terms between the Higgs and vector bosons, as well as trilinear and quartic self-interactions of the Higgs boson are generated, which exhibit coupling strengths proportional to the Higgs boson mass of $\mu^2\lambda$ and λ, respectively. The Higgs boson was discovered by the ATLAS and CMS Collaborations in 2012 [3, 4]. Its mass is determined to be $m_{\text{H}} = 125.25$ with a precision of 0.14% [58].

Fermion masses and quark mixing

The mechanism described above does not generate masses for fermions, but missing couplings can be introduced by a further extension of the Lagrangian. The so-called Yukawa interaction term, which reads

$$\mathscr{L}_Y^{l+dq} = -y_f \left[\overline{L}\Phi R + \left(\overline{L}\Phi, R \right)^\dagger \right] \tag{2.16}$$

generates masses for down-type quarks and charged leptons. The conjugate field Φ_c can be used via the term

$$\mathscr{L}_Y^{uq} = y_f \left[\overline{L}\Phi_c R + \left(\overline{L}\Phi_c R \right)^\dagger \right] \tag{2.17}$$

to generate masses for up-type quarks. The constant y_f is called Yukawa coupling and is defined as

$$y_f = \frac{m_f}{v}, \tag{2.18}$$

with m_f being the mass of the fermion. Additional terms are also generated that describe the interaction between the Higgs and fermion fields, and give rise to the coupling with a strength proportional to the fermion mass.

The couplings of the Higgs boson to fermions have been studied experimentally at the LHC [67, 68]. The production of a top quark-antiquark pair associated with a Higgs boson [69, 70], the decays of the Higgs boson to a pair of bottom quarks [71, 72] and tau leptons [73, 74] were measured. These serve as verification of the coupling to the third-generation fermions. Evidence was found for the Higgs coupling to muons [75, 76] and limits are obtained for the Higgs coupling to charm quarks [77, 78]. However, the couplings to first generation fermions are still inaccessible.

To describe the nature of the experimentally observed flavor-changing charged current, where the W boson couples to quarks of two generations and to explain the violation of charge and parity (CP) conservation in the SM [79], a mixing between EWK q' and mass eigenstates q is introduced. The corresponding mixing can be written as

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V_{CKM} \begin{pmatrix} d \\ s \\ b \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}, \tag{2.19}$$

where V_{CKM} is the Cabibbo–Kobayashi-Maskawa (CKM) matrix [80, 81]. In the SM, the CKM matrix is required to be unitary, and under this assumption all CKM elements have been determined with very high precision [58]. Furthermore, tests of unitarity are performed using data from different experiments [82]. The off-diagonal elements of the CKM matrix are small.

A similar mixing in the leptonic sector could be introduced in the same manner, but as neutrinos with right chirality are not described by the SM, the extension in Equation 2.16 suffices to generate masses for charged leptons. In contradiction to the SM expectation, neutrinos are required to have masses, since neutrino oscillations have been observed experimentally [83–85]. This could be solved by the addition of the Pontecorvo–Maki–Nakagawa–Sakata (PMNS) matrix [86, 87] that would be characterized by different phases depending on whether neutrinos are Dirac or Majorana particles [88]. The values of neutrino masses are currently still unknown, but upper limits are set experimentally to below 0.8 eV [89].

Quantum chromodynamics

The strong interaction between quarks and gluons is described by Quantum Chromodynamics (QCD). Strong interaction is responsible for the emerging of bound states of quarks and the formation of nuclei. QCD is a gauge theory based on the $SU(3)$ symmetry group, operating with the color charge, taking the values "red", "green", and "blue". While quarks carry color, the antiquarks carry anti-color. This concept was initially introduced to solve the problem of Pauli-principle violation by the wave function of the observed lightest baryons consisting of three quarks of the same flavor, e.g., Δ^{++} [90, 91]. Quark fields are organized in color triplets of the $SU(3)$ group as $Q = (q_{\text{red}}, q_{\text{green}}, q_{\text{blue}})^T$, based on the fermion fields of the quarks with a given color. The dynamics of the field can be written as a Lagrangian term, which reads

$$\mathscr{L}_{QCD} = -\frac{1}{4} F^a_{\mu\nu} F_a^{\mu\nu} + \sum_{flavors} \overline{Q}(i\gamma^\mu D_\mu - m_q)Q \tag{2.20}$$

with the covariant derivative

$$D_\mu = \partial_\mu - i g_s t_a G^a_\mu \tag{2.21}$$

and the field strength tensor $F^a_{\mu\nu}$ of the gluon fields G^a_μ

$$F^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s f^{abc} G^b_\mu G^c_\nu. \tag{2.22}$$

Here, f^{abc} are the structure constants of the $SU(3)$ group defined by the commutation relation $[T^a, T^b] = if^{abc}T^c$ of its eight generators with $a, b, c = 1 \dots 8$, and g_s is a dimensionless parameter related to the coupling strength of the strong interaction $\alpha_S = g_s^2/(4\pi)$. As a consequence of non-abelian structure of the color-$SU(3)$ in QCD, the structure constants are non-zero and the gluon self-interaction is possible, as shown in Figure 2.5. The existence of the gluon

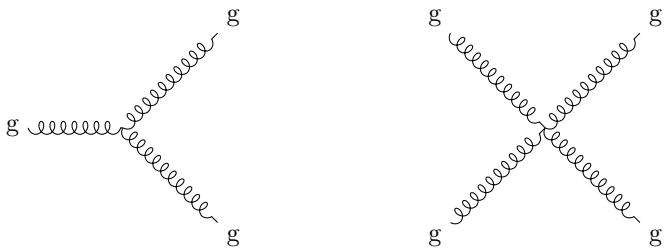

Figure 2.5: Feynman diagrams corresponding to the triple (left) and quadruple (right) self-couplings of the gluon fields.

was confirmed by observing three-jet events in e^+e^- collisions at the PETRA storage ring at DESY [92]. Also, the gluon spin was measured, establishing QCD as a proper theory based on the local gauge invariance.

As a result of the gluon self-coupling, beyond leading order in perturbation theory, corrections to the quark and gluon propagators have to be considered. Examples are illustrated in Figure 2.6, where "virtual" particle loops lead to ultraviolet divergences in the calculation of the scattering amplitudes. Using renormalization [55], these infinities are cut off at a certain renormalization

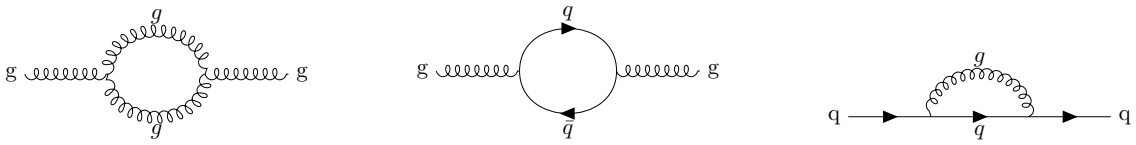

Figure 2.6: Feynman diagrams representing loop corrections to the quark and gluon propagators.

scale μ_R. The renormalized parameters are experimentally measurable at a fixed μ_R.

The modified minimal subtraction ($\overline{\mathrm{MS}}$) scheme is one of the most widely used renormalization procedures. In this scheme, the subtraction of ultraviolet divergences is performed at a fixed μ_R, and as a result, α_S and the quark masses depend on the choice of μ_R. Any physical quantity at all orders however should not be dependent on the artificially-introduced scale μ_R. Considering a physical observable $\mathcal{O}$, which depends on an energy scale Q, this can be formulated in terms of renormalized quantities as [53]

$$\left[\mu^2 \frac{\partial}{\partial \mu^2} + \beta(\alpha_S)\frac{\partial}{\partial \alpha_S} - \gamma_m(\alpha_S) m \frac{\partial}{\partial m}\right] \mathcal{O}(\alpha_S, m) = 0, \tag{2.23}$$

which is known as the so-called renormalization group equations (RGEs) for the strong coupling constant and for the quark mass. The universal functions $\beta(\alpha_S)$ and $\gamma_m(\alpha_S(\mu))$ are defined through the relations

$$\mu^2 \frac{d\alpha_S}{d\mu^2} = \beta(\alpha_S), \tag{2.24}$$

and

$$\mu^2 \frac{dm(\mu)}{d\mu^2} = -\gamma_m(\alpha_S(\mu))m(\mu), \tag{2.25}$$

where $\beta(\alpha_S)$ is known as the QCD beta function and $-\gamma_m(\alpha_S(\mu))$ as the mass anomalous dimension. In QCD, this leads to the strong coupling constant α_S being energy dependent, which is also referred to as "running". If α_S is small, perturbation theory can be applied, and the dependence of $\beta(\alpha_S)$ and $\gamma_m(\alpha_S)$ can be expanded in powers of α_S as

$$\beta(\alpha_S) = -\alpha_S^2 \sum_{n=0} \beta_n \alpha_S^n \tag{2.26}$$

and

$$\gamma_m(\alpha_S) = \alpha_S \sum_{n=0} c_n \alpha_S^n. \tag{2.27}$$

Here, the coefficients β_n and c_n represent the $n+1$-loop corrections. The coefficients of the anomalous dimension and the beta function are currently known up to five loops [93–104]. For the coupling constant, the explicit energy dependence can be written at the lowest order as

$$\alpha_S(\mu^2) \approx \frac{\alpha_S(\mu_0^2)}{1 + \beta_0 \alpha_S(\mu_0^2) \ln\left(\frac{\mu^2}{\mu_0^2}\right)}. \tag{2.28}$$

The parameter β_0 is a dimensionless constant defined as $\beta_0 = (33 - 2n_f)/12\pi$ with n_f being the number of flavors considered massless at the energy scale μ. The scale μ_0 represents the initial scale at which α_S is evaluated. However, the value of $\alpha_S(\mu_0)$ has to be determined experimentally. With increasing energy scales, the value decreases logarithmically. As a result, at high energies, i.e., small distances, quarks and gluons can be described as a free particles. This concept is referred to as "asymptotic freedom" [93, 95], and perturbative calculations can be applied. The running of α_S was confirmed by many experiments in a large energy range [105].

At low energies or large distances between two quarks, the potential of the strong interaction grows and becomes infinite. Therefore, it is energetically preferable to create a new quark-antiquark pair and form colorless hadrons. This phenomenon is referred to as "confinement", and the process is referred to as hadronization. Hadrons are classified into mesons and baryons, whereas mesons consist of a quark- antiquark pair, and baryons of three quarks. Also bound states of four (tetraquarks) or five (pentaquarks) quarks were observed by different experiments [106–108]. The only quark which does not hadronize is the top quark. Its mass is so large that its lifetime is shorter than the time of the hadronization [58]. Thus, the top quark decays via the electroweak interaction into a b-type quark and a W boson.

Similar to the strong coupling constant, also quark masses are subject to renormalization. Using the $\overline{\text{MS}}$ scheme, the RGE for the dependence of the quark mass m on the scale μ is given by Equation 2.25. Usually, the $\overline{\text{MS}}$ mass is reported at the scale of the mass value itself as $m(m)$.

Another solution of how to renormalize the quark masses is the on-shell, or pole mass, scheme. Here, the mass parameter is defined such that it corresponds to the pole in the quark propagator analogous to the masses of free particles in QED. All self-energy corrections up to scales of the mass are absorbed into its definition. Nonetheless, due to nonperturbative effects in QCD, the definition of the top quark pole mass has an ambiguity of Λ_{QCD} [13–15], which is referred to

as the renormalon problem. The pole and $\overline{\text{MS}}$ masses can be translated into each other. For example, at one-loop level this relation reads

$$m_{\text{pole}} = m(m)\left(1 + \frac{4}{3\pi}\alpha_S(m)\right). \tag{2.29}$$

An alternative short-distance quark mass definition, providing a smooth transition between the pole and the $\overline{\text{MS}}$ mass, is the MSR scheme [109–111]. The pole and MSR masses are related by the equation

$$m_{\text{pole}} = m_{\text{MSR}} + R\sum_{n=1}^{\infty} a_n^{\text{MSR}}\left(\frac{\alpha_S(R)}{4\pi}\right)^n, \tag{2.30}$$

where a_n^{MSR} are decoupling coefficients and R is an additional arbitrary scale. For scales of $R \to 0$, the MSR mass approaches the pole mass, as for the $\overline{\text{MS}}$ mass at $R = m(m)$. Similar to the $\overline{\text{MS}}$ mass, the MSR mass is renormalon free, and all self-energy corrections for scales between R and m are absorbed into its definition.

2.2 Hadron collider physics

In the high-energy pp collisions at the LHC, the structure of both protons is resolved. As a result of an interaction between the partons (quarks or gluons) from both protons, referred to as a hard interaction, further particles are produced. The cross section of the particle production is factorized into a convolution of the structure of both protons and a partonic cross section, which is calculable perturbatively. In the following, the QCD factorization and the proton structure are described. Further, the concepts of Monte-Carlo (MC) event generation are illustrated and the modeling of additional radiation, fragmentation, and hadronization are explained.

QCD factorization and parton distribution functions

The QCD factorization theorem is introduced in [112] and represents the possibility to separate parton dynamics into the short-distance (asymptotic freedom) and long-distance (confinement) parts at a certain factorization scale μ_{F}. The production cross section for a certain process in the pp interaction, $\sigma_{\text{pp}\to ab+X}$, can be written as

$$\sigma_{\text{pp}\to ab+X} = \sum_{i,j} f_i(x_1, \mu_{\text{F}}^2) f_j(x_2, \mu_{\text{F}}^2) \widehat{\sigma}_{ij\to ab}(\hat{s}, \mu_{\text{F}}^2, \mu_{\text{R}}^2, \alpha_S) dx_1 dx_2. \tag{2.31}$$

Here, the $f_i(x, \mu_{\text{F}}^2)$ are the parton distribution functions (PDFs) of both colliding protons. These describe the probability for a parton of the flavor i (j) to carry the momentum fraction x_1 (x_2) of the proton 1 (2) at a scale of μ_{F}^2. The partonic center-of-mass energy is denoted by $\hat{s}$ and is defined as $\hat{s} = x_1 x_2 s$, and $\widehat{\sigma}_{ij\to ab}$ is the partonic cross section.

The partonic cross section $\widehat{\sigma}_{ij\to ab}$ can be calculated perturbatively via the scattering amplitude, which is often referred to as matrix element (ME). Usually, calculations are performed at a certain order in perturbation theory, corresponding to different orders of the EWK and QCD couplings.

Similar to μ_{R}, μ_{F} is not a physical parameter and to all orders, the calculation should not depend on its choice. However, since the calculations are available to a certain fixed order, μ_{F} and μ_{R} are typically chosen close to a physical scale of the process. The possible effects of missing order corrections are mimicked by varying μ_{R} and μ_{F} by a factor of two up and down.

The scale μ_F usually represents the minimal scale for which a physics process can be resolved, and is introduced to remove infrared divergences arising from collinear initial state radiation.

While the scale-dependence of the PDFs can be calculated in perturbative QCD, the dependence on x has to be determined experimentally. The dependence of the PDFs on μ_F is described by the Dokshitzer–Gribov–Lipatov–Altarelli–Parisi (DGLAP) equations [113–115], which are a set of integral-differential-equations and can be used to evolve a PDF from one scale to another.

For the gluon (g) and the singlet distributions (q_s), these are given by

$$\frac{\partial}{\partial \ln(\mu^2)} \begin{pmatrix} q_s \\ g \end{pmatrix} = \begin{pmatrix} P_{qq} & P_{qg} \\ P_{gq} & P_{gg} \end{pmatrix} \otimes \begin{pmatrix} q_s \\ g \end{pmatrix}, \tag{2.32}$$

where $\otimes$ stands for the Mellin convolution [116]. For the non-singlet ($q_{\mathrm{ns}}^{\pm}$) and valence distributions ($q_{\mathrm{ns}}^{\mathrm{v}}$) they are given by

$$\frac{\partial}{\partial \ln(\mu^2)} q_{\mathrm{ns}}^i = P_{\mathrm{ns}}^i \otimes q_{\mathrm{ns}}^i, \qquad \text{where } i \in [\mathrm{v}, \pm]. \tag{2.33}$$

Here, P_{ij} are the splitting functions that represent the probability for a parton i to emit a parton j carrying a momentum fraction z of its longitudinal momentum in the collinear approximation. The splitting functions itself are calculable using perturbation theory, as it is illustrated with the Feynman representation shown in Figure 2.7 at LO, and are currently known up to three loops in QCD [117, 118].

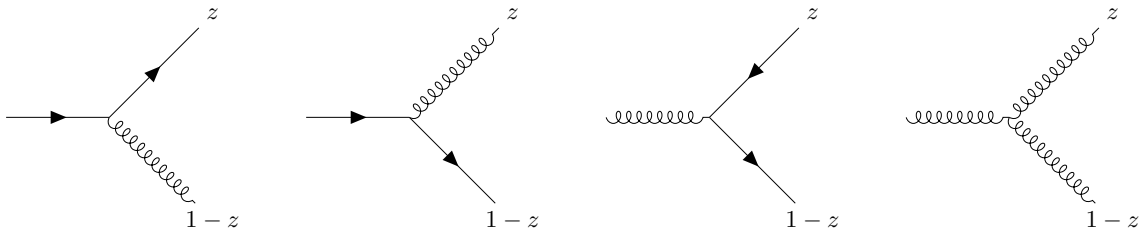

Figure 2.7: Sample Feynman diagrams contributing to the splitting functions $P_{qq}(z)$ (left), $P_{gq}(z)$ (middle left), $P_{qg}(z)$ (middle right), and $P_{gg}(z)$ (right) at LO QCD.

Since the x-dependence of the PDFs cannot yet be calculated from first principles, these are determined by using the experimental data. The data of deep inelastic scattering (DIS) experiments, in particular neutrino-nucleon [119, 120] and electron-nucleon scattering [121–123], provide major information about the proton structure and the PDFs. In addition to the DIS measurements, the LHC data provide further information about the proton PDFs, in particular on the gluon distribution at large x [38, 39, 124]. In Figure 2.8, the PDFs obtained by using the CMS measurements of jet production [39] in addition to the DIS measurements are shown in compared to the result of the fit using only DIS data. A significant improvement in the gluon distribution at high x is observed.

Monte Carlo simulation

Precise tests of the SM are usually performed through extraction of its fundamental parameters. For this purpose, the experimentally recorded stable particles (detector-level particles) have to be associated with the original partons produced in the hard interaction (parton-level objects).

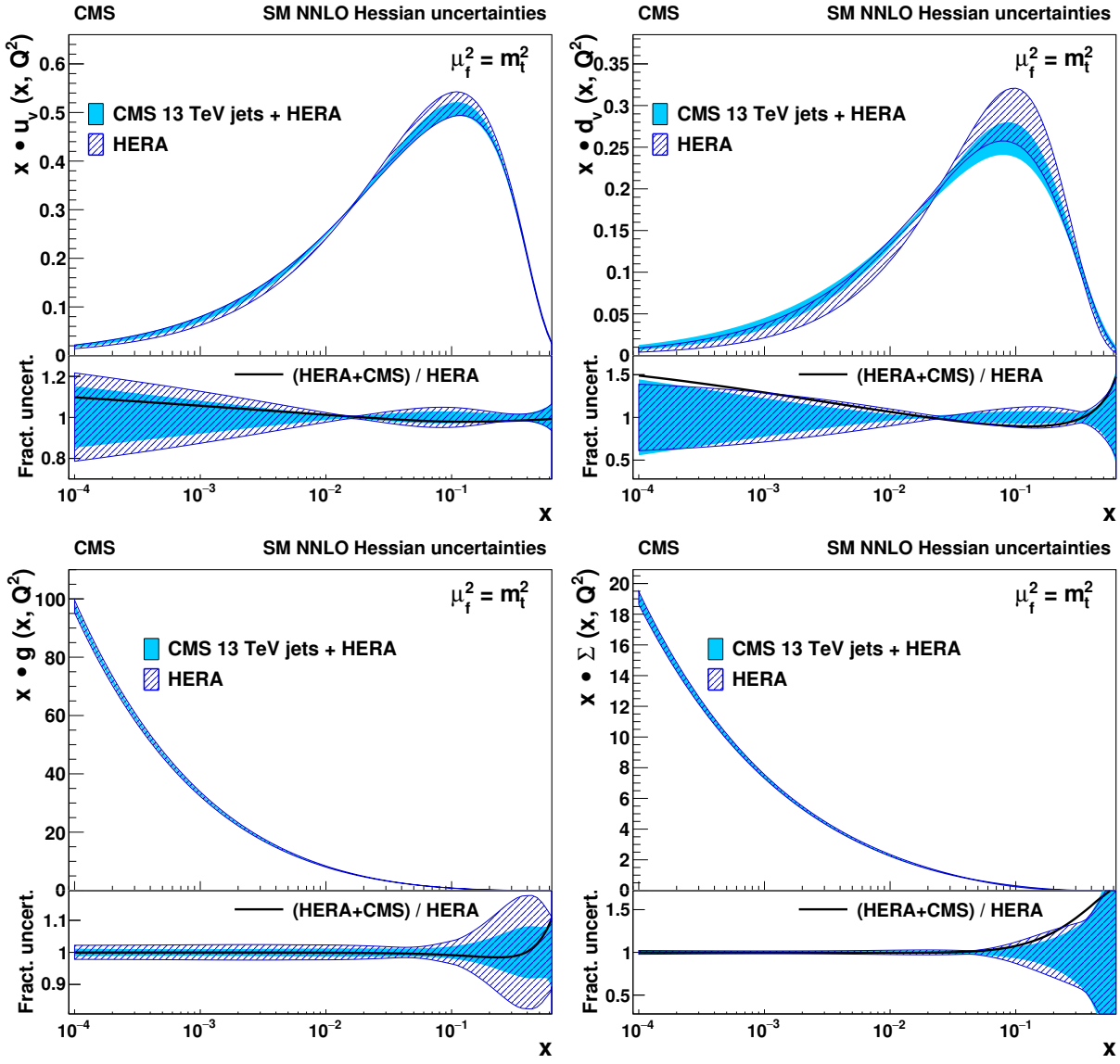

Figure 2.8: The u_v-valence (upper left), d_v-valence (upper right), gluon (lower left), and sea quark (lower right) PDF distributions, shown as a function of x at the scale $\mu_F = 2m_t$. The filled (hatched) band shows the results of a fit using HERA DIS and CMS inclusive jet cross section data at $\sqrt{s} = 13\,\text{TeV}$ (HERA DIS data only) with their total uncertainty. The lower panels show the comparison of the relative PDF uncertainties for each distribution, and the dashed line corresponds to the ratio of the central PDF values of the two variants of the fit. The plots are taken from Ref. [39].

This requires good description of the way from the partons to stable hadrons, including the fragmentation, hadronization, and the decay of unstable particles. For this purpose, MC event generators have been developed. They simulate this complete chain from the hard interaction at high scales down to low scales for hadronization and particle decays. Also, the detector response can be fully simulated. The output of such generators can usually directly be compared to the experimental data.

In the MC event generator, the partonic scattering is derived using perturbation theory. In addition, massive particles such as the Higgs, W, and Z bosons and top quarks can be treated as resonances due to generally small interference effects that could lead to the same final state. Their decays can be externalized, that simplifies the simulation. To account for higher order terms in the scattering amplitude calculation, and to simulate soft and collinear emission of partons, so-called parton shower (PS) models are used. Additional soft interactions, hadronization, hadronic decays, and fragmentation are approximated by phenomenological models. An illustration of a hadron collision event as could be simulated by a MC generator is depicted in Figure 2.9.

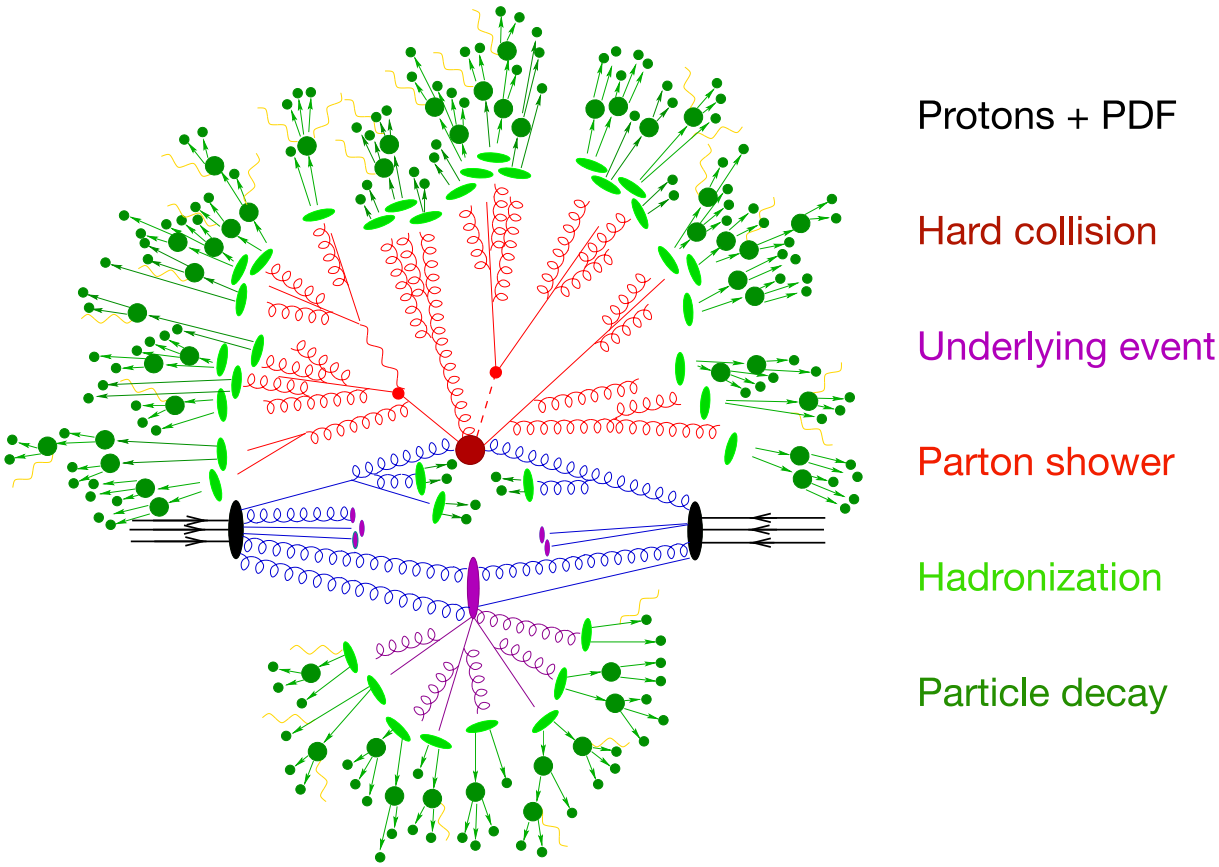

Figure 2.9: Illustration of the MC simulation of a pp collision [125] (modified). Partons (blue) from the initial protons (black) contribute to the hard interaction (dark red) according to their structure as given by the PDFs. An underlying-event interaction is visible in purple. Additional emissions and the parton shower branching is shown in red, and additional photon radiations are presented in yellow. After the hadronization of partons into hadrons (light green), the decay into final state particles takes place (dark green).

Parton shower models are based on the splitting functions as introduced in the context of the PDFs. The cross section of a given process involving the parton j and a subsequent parton branching i can be represented in the collinear limit as [126]

$$d\sigma \approx \sigma_0 \sum_{\text{partons j},i} \frac{\alpha_S}{2\pi} \frac{d\theta^2}{\theta^2} dz P_{ji}(z). \tag{2.34}$$

Here, z is the momentum fraction of the emitted parton, $P_{ji}(z)$ are the splitting functions, θ is the angle of the emission, and σ_0 is the cross section of the process without the additional branching. Based on Equation 2.34, using an iterative procedure starting from the hard scattering ME, parton branchings can be added to each iteration, and the result is considered as the hard interaction for the next evaluation cycle. To avoid the divergences in the limit of $\theta \to 0$, an additional cutoff scale Q_0 has to be introduced. It resembles the scale at which additional parton emissions cannot be resolved anymore. The probability $\Delta_i(Q^2, q^2)$ that there is no parton branching with $k^2 > q^2$ is introduced as

$$\Delta_i(Q^2, q^2) = \exp\left(- \int_{q^2}^{Q^2} \frac{dk^2}{k^2} \frac{\alpha_S}{2\pi} \int_{Q_0^2/k^2}^{1-Q_0^2/k^2} dz P_{ij}(z) \right). \tag{2.35}$$

Here, q is the virtuality of the emitted parton and k is an ordering variable related to the transverse momentum of the emitted parton. The function $\Delta_i(Q^2, q^2)$ is also denoted as Sudakov form factor [127, 128], and Q^2 is the starting scale of the shower. In the MC, a $\Delta_i(Q^2, q^2)$ is generated using random numbers, and Equation 2.35 is solved for q^2. Given a solution for $q^2 > Q^2$, a parton branching is generated at the scale q^2. Otherwise, the shower terminates. The values for z are chosen according to the splitting functions $P_{ij}(z)$. The procedure is sequentially performed until the cutoff scale Q_0 is reached. As the Sudakov form factor sums all terms of the kind $(\alpha_S \ln^2(Q^2/Q_0^2))^k$ with the greatest power of logarithms to all orders, it is called to be of leading logarithmic accuracy [126]. Another important effect to consider is the running of the strong coupling constant. With decreasing q^2, α_S increases, and hence the parton branching probabilities significantly increase. This results in a much faster multiplication of the shower. Furthermore, the originally arbitrary cutoff scale Q_0 has to be chosen much larger than Λ_{QCD} to ensure the applicability of the perturbative calculation. The cutoff also becomes a parameter with significant impact on the final states of the PS generator. Similar to the description of the final state radiation, the initial state radiation can be generated by using the DGLAP equations for the PDF evolution and the corresponding momentum fraction for each parton.

Since the emission is present in both, fixed-order calculation of the ME and in the PS, a potential double counting needs to be avoided. This is resolved via subtracting terms in the matching of the PS to the ME calculation. Because PSs are designed to model soft emission at lower scales in the collinear limit and matrix-element calculations precisely determine hard emissions, consequently the hardest emission of the PS has to be corrected.

One popular matching procedure is the POWHEG method [129, 130]. In this method, corrections are applied to the Sudakov form factor and the matching scale h_{damp} is introduced. By suitable weighting or damping the first emission, a smooth transition between the ME and PS is maintained. This damping function depends on the hardness scale of the emission and h_{damp}. The scale h_{damp} can be understood as a resummation scale [131] at which the transition happens. The value of h_{damp} is usually tuned using the experimental measurements [132]. This matching procedure preserves NLO accuracy for the integrated quantities and the ME, while the leading-logarithmic accuracy of the PS is kept.

In pp collisions, not only the hard interaction and subsequent parton branchings contribute to each event but also secondary parton–parton interactions and effects of beam remnants have to be taken into account. These multi-parton interactions (MPIs) lead to soft QCD processes and interference with particles emerging from the hard scattering. These effects are collectively denoted as underlying event (UE) and are described by phenomenological models as they cannot be calculated using perturbative QCD. Typically, dedicated tunes are derived by fitting models to the experimental data [132]. Additionally, the interference between final-state radiation and the hard-interaction partons due to color correlations have to be considered [133]. Various methods haven been developed [134–136] of how colored object interference has to be treated in color reconnection (CR) models, which are currently tested by the LHC experiments.

For the energy scales below Λ_{QCD} and the PS cutoff scale, the colored final-state objects recombine and form color-neutral hadrons. This process is described by the hadronization models. Two major classes of hadronization models are the so-called string [137] and cluster [138] models. In the string model, interactions between quark–antiquark pairs via gluons are described by a linear rising potential through uniform color-flux tubes, called strings. With increasing distance between quarks, and thus increasing potential energy, the strings break and quark–antiquark pairs are produced until neutral hadrons are formed. In this model, gluons are described as kinks in the strings affecting the kinematic properties of the hadrons. In the MC simulation used for this thesis, the Lund string model as implemented in PYTHIA8 [139] is employed. The momentum of the hadrons are related to the momentum of the parent quarks via phenomenological fragmentation functions. The commonly used Peterson [140] or Kartvelishvili [141] functions depend on a single parameter, which is tuned by using the experimental data. In the Bowler–Lund fragmentation function [142], the effects of quark masses are additionally considered.

2.3 The top quark

The existence of the third-generation quarks was predicted by Kobayashi and Maskawa in 1973 to explain observations of CP violations in the quark sector [81]. With the observation of the b quark in 1977 at Fermilab [143], this idea was confirmed and the quest for the top quark has started. After 20 years of searches, based on different mass assumptions, the top quark was discovered by the CDF and D0 Collaborations at the Fermilab Tevatron using $\mathrm{p\overline{p}}$ collisions with a center of mass energy of 1.8 TeV [5, 6]. With a pole mass $m_{\mathrm{t}}^{\mathrm{pole}}$ of about 172.5 ± 0.7 GeV [58], it is the most massive elementary particle known, and hence has a total width of about 1.42 GeV [58] and a lifetime of around 5×10^{-25} s [58], which is substantially smaller than the time of the hadronization of $1/\Lambda_{\mathrm{QCD}} \approx 3 \times 10^{-24}$ s. As a result, top quarks decay before forming hadronic bound states and yield a unique possibility to study properties of a bare top quark as for example the charge and polarization. The large mass of the top quark provides a natural scale to study perturbative QCD and is further of particular relevance for EWK physics. For example, the Higgs boson mass receives large perturbative corrections from diagrams that depend on the top quark Yukawa coupling, and in turn on m_{t} [144], as illustrated in Figure 2.10. To first order, the contribution $\delta\mu^2$ of this correction to the parameter μ^2 in the Higgs potential can be calculated as

$$\delta\mu^2 = -\frac{3m_{\mathrm{t}}^2}{4\pi^2}\frac{\Lambda^2}{v^2}, \tag{2.36}$$

where $\Lambda \gg 1$ TeV is the cut-off scale [144]. To obtain the measured value of the Higgs boson mass, and thus μ^2, this contribution has to be canceled by a fine-tuning of the tree-level parameter of μ. This problem is also known as the hierarchy or naturalness problem.

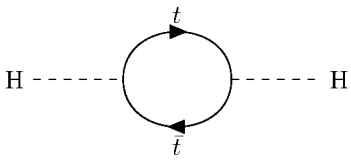

Figure 2.10: Feynman representations for the one-loop correction to the Higgs boson propagator from a top quark loop.

In a similar fashion, loop corrections involving top quarks lead to corrections to the quartic coupling parameter λ in the Higgs potential. As a result, extrapolations of λ to higher energy scales depend also on the top quark mass, as well as on α_S and the Higgs boson mass [7, 8, 145, 146]. This is illustrated by the two plots in Figure 2.11. The result also shows that the

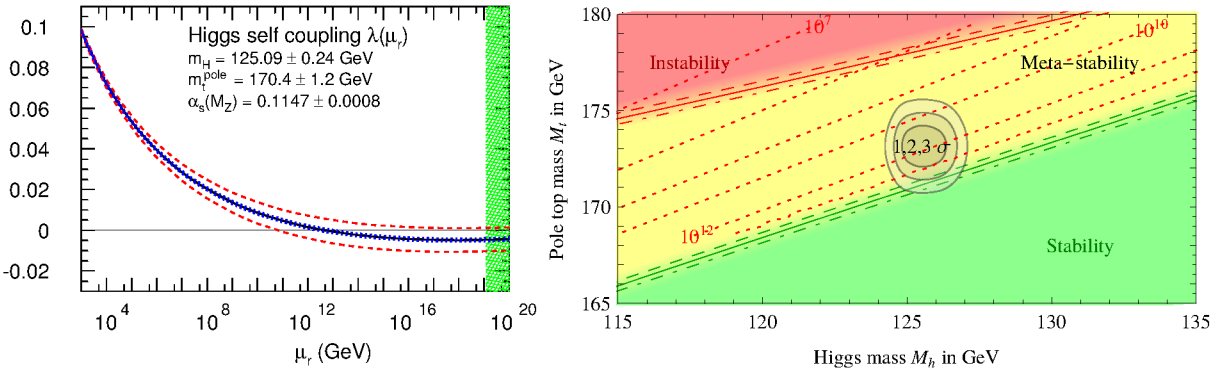

Figure 2.11: Left: evolution of the parameter λ as a function of the scale μ_{R}. The dashed red (blue) line indicates the one standard variation of the values of m_{t} and α_S (the Higgs boson mass) [11]. Right: regions of stability, meta-stability, and instability of the SM EWK vacuum as a function of the top quark ans Higgs boson masses [7].

Higgs boson self-coupling, and in turn the stability of the EWK vacuum, crucially depend on the value of m_{t}. Based on the current experimental precision on m_{t} and the Higgs boson mass, only the instability can be excluded. Future investigations can shed light on the question between stability and meta-stability [7]. In case of a significant deviation of m_{t} from the value needed for a stable vacuum assuming the SM, possible extensions of the SM can be tested.

The top quark mass can also be obtained as a result of the so-called global electroweak fits [9–11, 147]. In the SM, the masses of the top quark, W and H bosons are related to each other via vacuum corrections. Having two of these parameters, the third one can be unambiguously determined and its deviations from the SM expectation would consequently hint to new physics. Recent $m_{\mathrm{t}}^{\mathrm{pole}}$ results from the electroweak fits, obtained by using the experimental measurements of m_{W} and m_{H}, differ from the value of $m_{\mathrm{t}}^{\mathrm{pole}}$ obtained by the ATLAS and CMS collaborations [148, 149]. This fact motivates further detailed measurements of the top quark mass using orthogonal methods in order to probe our understanding of the SM.

Top quark production and decay

In pp collisions, top quarks are predominantly produced as quark-antiquark pairs ($\mathrm{t\bar{t}}$), as illustrated at LO in Figure 2.12. The dominant production mechanism is the gluon-gluon fusion process. The gluon-gluon fusion production mechanism, as illustrated by the first two diagrams, contributes to about 90% of the $\mathrm{t\bar{t}}$ cross section for pp collisions [58].

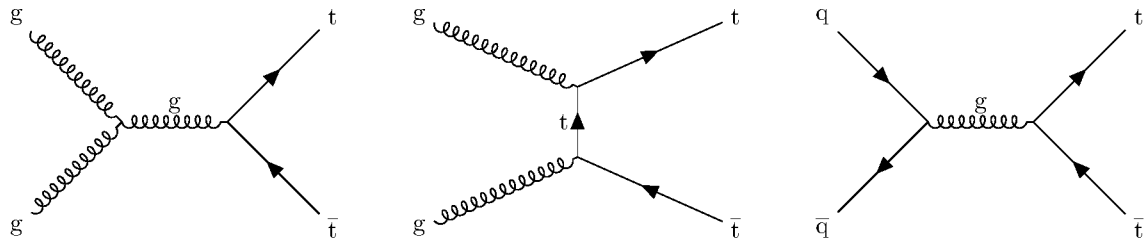

Figure 2.12: Feynman diagram representations for $t\bar{t}$ production at leading order. The first (second) diagram shows the gluon-gluon fusion production channel in the s (t) channel. The right diagram shows the production via quark-antiquark annihilation.

The $t\bar{t}$ production cross section in proton collisions was measured at various center of mass energies, as illustrated in Figure 2.13. The measurements are compared to the theory prediction at NNLO+next-to-leading-log (NNLL) [150] using NNPDF3.0 [151], and assuming $\alpha_S = 0.118$ and $m_t = 172.5\,\text{GeV}$. For $\sqrt{s} = 13\,\text{TeV}$, the predicted $t\bar{t}$ cross section is $839.5\,^{+26.9}_{-34.2}\,\text{pb}$, and the agreement between measurements and calculations in NNLO+NNLL QCD is observed to be very good. The $t\bar{t}$ production is one of the abundant processes at the LHC, allowing for very precise measurements.

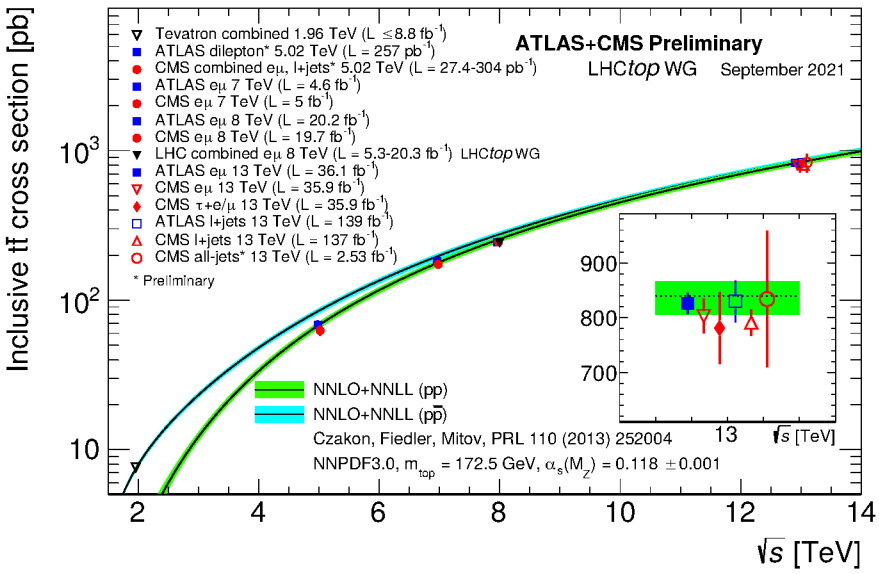

Figure 2.13: Measurements for the pp and $p\bar{p}$ production cross section by Tevatron (black), ATLAS (red), and CMS (blue) for center of mass energies ranging from about 2 to 13 TeV [152]. The measurements are compared to the theoretical predictions calculated using the TOP++ framework [153].

The top quark decays electromagnetically into a W boson and a b-type quark. The branching ratio is given by [58]

$$\mathcal{B}_{\text{bW}} = \frac{|V_{\text{tb}}|^2}{|V_{\text{tb}}|^2 + |V_{\text{ts}}|^2 + |V_{\text{td}}|^2} = 0.998, \tag{2.37}$$

where V_{tq} are the coefficients of the CKM matrix. The value of $|V_{tb}| = 0.088 \pm 0.024$ is measured by CMS [154] and the upper limit of 0.955 at 95% confidence level on $\mathcal{B}_{bW}$ is obtained. The decays of the top quarks are categorized by the decays of the W boson as dileptonic, semileptonic, and fully hadronic, as sketched in Figure 2.14.

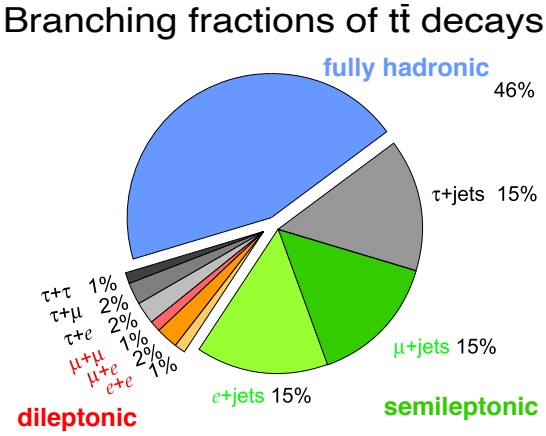

Figure 2.14: Branching ratios and final states of $t\bar{t}$ production [155].

Despite the small branching ratio of the dileptonic decay channel, it yields the highest experimental purity because of clear identification of the final-state leptons. The fully hadronic channel receives large background from QCD multijet production. The semileptonic channel has intermediate properties.

Measurements of the top quark mass

The value of the top quark mass can be measured by reconstructing the invariant mass of the top quark's decay products, or by measuring observables that are strongly correlated to it. Measurements of this kind are often referred to as direct mass measurements, and they reach a precision on the order of 0.4 GeV [23]. A summary of the results obtained in such direct measurements is given in Figure 2.15. In spite of the high accuracy, these measurements heavily rely on the multi-purpose MC generators [126, 156], which allow inferring the value of m_t from comparisons of the simulated final-state distributions to the experimental measurements. These simulations are based on the parton shower models and heuristic models, tuned to the data to describe the UE, hadronization and CR. Consequently, this leads to an ambiguity in the interpretation of the result of the direct measurements in terms of the top quark mass in the Lagrangian. Various studies suggest [24–31], that the top quark mass parameter assumed in the MC simulation, m_t^{MC}, resulting from the direct measurement, can be related to the top quark pole mass m_t^{pole}, with an additional uncertainty on the order of ± 0.5 GeV, depending on the MC generator used. Similar studies show that m_t^{MC} is close to the MSR mass for a scale of $R = 1$ GeV [25, 157] with a similar uncertainty.

Alternatively, the value of the top quark mass defined in a certain renormalization scheme can be obtained by comparing measured observables, sensitive to m_t, to the fixed-order theoretical predictions. Such observables might be either inclusive cross sections [31–33, 158] or m_t-sensitive shapes of the differential cross sections [37, 38, 41, 44, 45]. A summary of the results of such indirect mass measurements is shown in Figure 2.16. The value of m_t^{pole} was obtained with an uncertainty below 1 GeV by using inclusive and differential cross sections of $t\bar{t}$ production at the LHC [11, 32–39]. The top quark running mass $m_t(m_t)$ defined in the $\overline{MS}$ scheme was also measured [31, 40, 41] and its energy scale dependence was recently investigated at NLO [42] and NNLO [159].

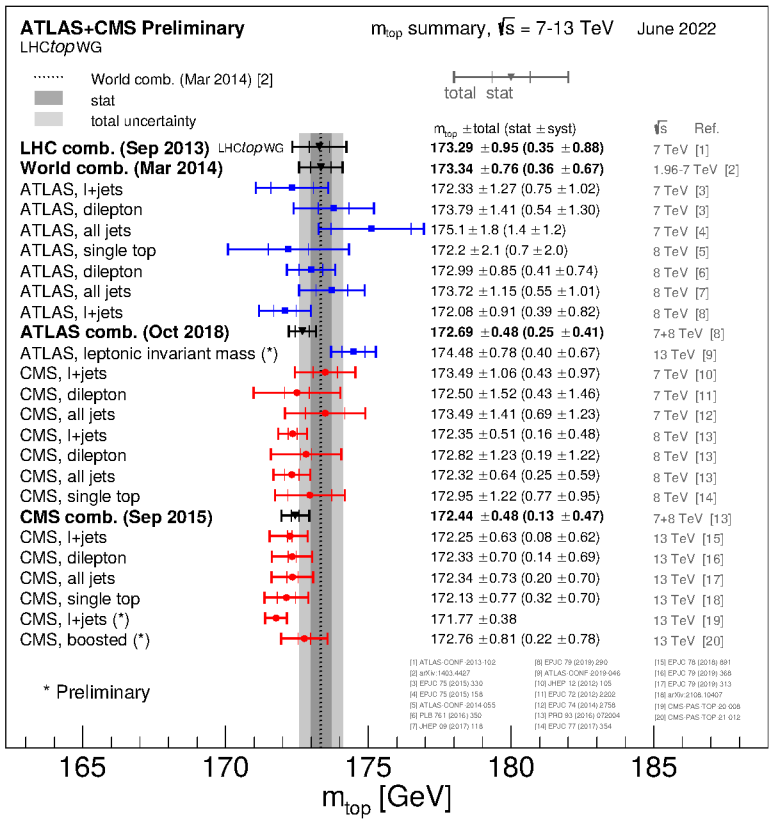

Figure 2.15: Recent results obtained by the LHC experiments on the value of $m_{\mathrm{t}}^{\mathrm{MC}}$ [152].

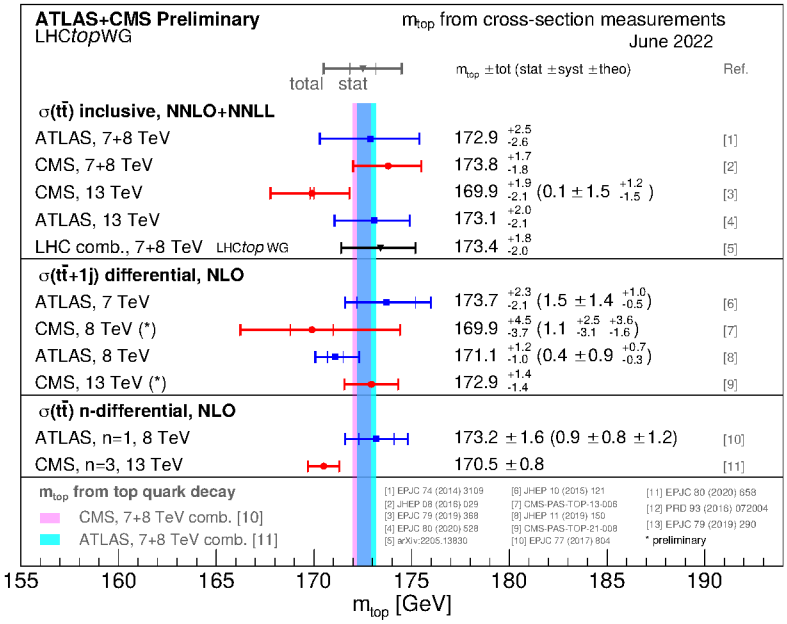

Figure 2.16: Recent results obtained by the LHC experiments on the value of the top quark pole mass [152].

2.4 Top quark-antiquark production in association with one jet

In this thesis, the value of $m_{\mathrm{t}}^{\mathrm{pole}}$ is extracted by using events where the top quark-antiquark pair is produced in association with at least one energetic jet ($\mathrm{t\bar{t}}$+jet). It was demonstrated [44], that the invariant mass of the $\mathrm{t\bar{t}}$+jet system, $m_{\mathrm{t\bar{t}+jet}}$, exhibits high sensitivity to the value of $m_{\mathrm{t}}^{\mathrm{pole}}$. A novel observable, ρ, was introduced, defined as

$$\rho = \frac{2m_0}{m_{\mathrm{t\bar{t}+jet}}}, \tag{2.38}$$

where m_0 is an arbitrary scaling constant, set to $m_0 = 170\,\mathrm{GeV}$, close to the $m_{\mathrm{t}}^{\mathrm{pole}}$ value. This observable was used by the ATLAS Collaboration to measure the top quark pole mass at center of mass energies of 7 and 8 TeV [35, 46]. In the measurement presented in this thesis, the ρ observable is used to extract $m_{\mathrm{t}}^{\mathrm{pole}}$ using the proton collision data at the LHC at a center of mass energy of 13 TeV, following the general strategy of the approach developed in Refs. [41, 44, 45].

The $m_{\mathrm{t}}^{\mathrm{pole}}$-sensitivity measure $\mathcal{S}$ is introduced as

$$\mathcal{S}(\rho) = \sum_{\Delta = \pm 3\,\mathrm{GeV}} \left(\mathcal{R}(\rho, m_{\mathrm{t}}^{\mathrm{pole}}) - \mathcal{R}(\rho, m_{\mathrm{t}}^{\mathrm{pole}} + \Delta)\right) \;/\; 2|\Delta|\mathcal{R}(\rho, m_{\mathrm{t}}^{\mathrm{pole}}), \tag{2.39}$$

where $\mathcal{R}$ is the normalized differential cross section of $\mathrm{t\bar{t}}$+jet production as a function of ρ. The value of $\mathcal{S}$ quantifies how the differential cross section changes as a function of $m_{\mathrm{t}}^{\mathrm{pole}}$. To compare the sensitivities in $\mathrm{t\bar{t}}$+jet and $\mathrm{t\bar{t}}$ production, the invariant mass of $\mathrm{t\bar{t}}$+jet in the ρ definition is replaced by the invariant mass of the $\mathrm{t\bar{t}}$ pair. The resulting sensitivities for $\sqrt{s} = 13\,\mathrm{TeV}$ are compared in Figure 2.17. In both cases, the sensitivity becomes large close to the threshold of the $\mathrm{t\bar{t}}$ production, however in the case of $\mathrm{t\bar{t}}$+jet the sensitivity is significantly enhanced above the threshold region.

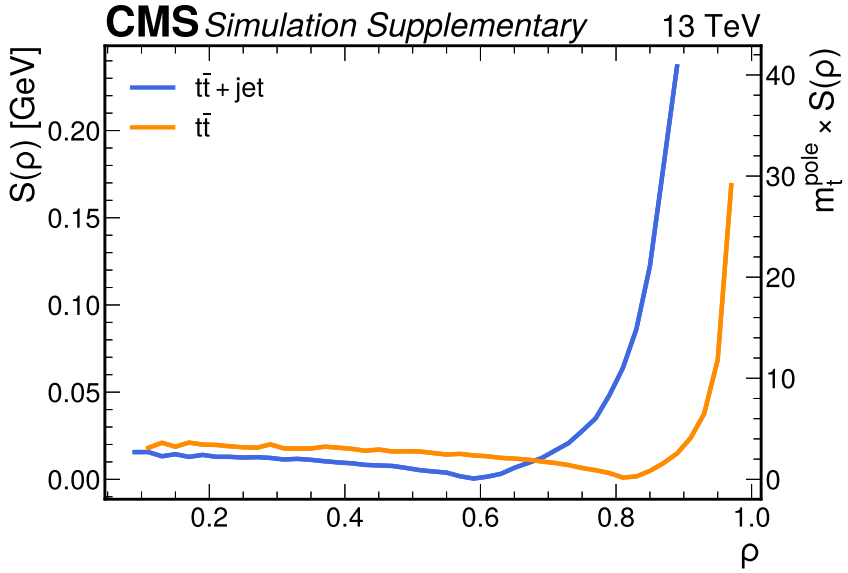

Figure 2.17: The sensitivity of $\mathcal{R}(m_{\mathrm{t}}^{\mathrm{pole}}, \rho)$ to the top quark mass for $\mathrm{t\bar{t}}$+jet (blue) and $\mathrm{t\bar{t}}$ production (orange). The variable $\mathcal{S}$ is defined as given in the text.

The mass sensitivity of the shape of the $\mathcal{R}$ distribution is illustrated in Figure 2.18. The curves are normalized to unity and cross for a value of $\rho \approx 0.65$. At this point, the observable is insensitive to the top quark mass. High sensitivity to $m_{\mathrm{t}}^{\mathrm{pole}}$ is expected around the $\mathrm{t\bar{t}}$+jet production threshold, corresponding to $\rho > 0.65$. In turn, for large values of $m_{\mathrm{t\bar{t}+jet}}$, i.e., values of $\rho < 0.55$, lower mass sensitivity is expected.

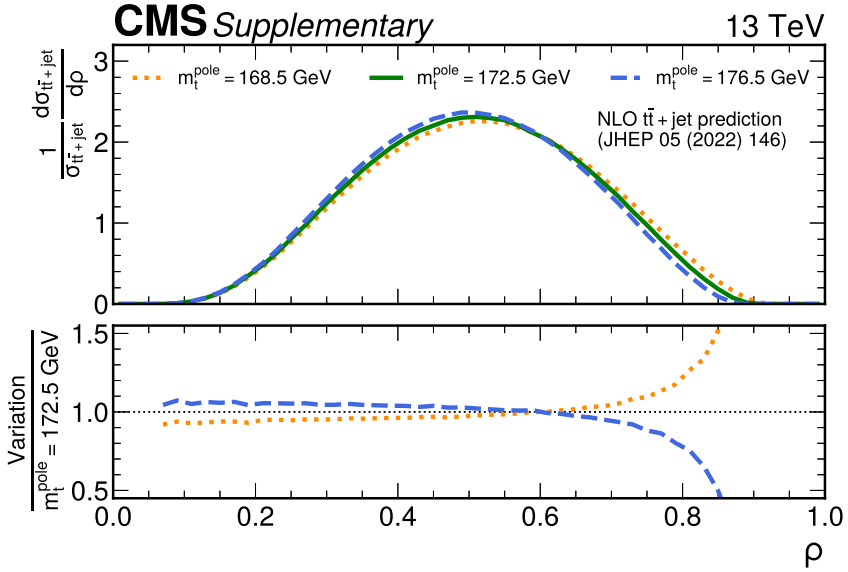

Figure 2.18: The normalized differential $\mathrm{t\bar{t}}$+jet cross section as a function of ρ at parton level for three different top quark mass assumptions: 168.5 (orange), 172.5 (green), and 176.5 GeV (blue). The lower panel shows the ratio to the central value of 172.5 GeV. The predictions are provided as given in Ref. [160].

The additional jet is reconstructed with the anti-k_{T} algorithm [161], using a distance parameter $R = 0.4$, and is required to have a transverse momentum larger than 30 GeV [44]. In addition to high sensitivity of the $\mathcal{R}$ shape to the value ot $m_{\mathrm{t}}^{\mathrm{pole}}$, the rate of $\mathrm{t\bar{t}}$+jet is rather large being about a third of the inclusive $\mathrm{t\bar{t}}$ production. The data sample used for the main measurement of this thesis contains about 12 million $\mathrm{t\bar{t}}$+jet events. Moreover, the measurement of the normalized cross section allows for cancellation of certain experimental uncertainties and therefore better precision.

The theoretical prediction for $\mathrm{t\bar{t}}$+jet is available in QCD to NLO precision [162, 163], implying the pole or $\overline{\mathrm{MS}}$ schemes for the top quark mass. In the analysis of this thesis, the prediction in the pole mass scheme is chosen, considering it a better definition at the threshold of the top quark production. The $\mathrm{t\bar{t}}$+jet process has been implemented in the POWHEG ME generator [164], which allows for fast calculation of the fixed-order prediction as well as for generation of events with subsequent showering via PYTHIA. In Ref. [160], among other studies, the renormalization and factorization scale choice for $\mathrm{t\bar{t}}$+jet production in the fixed-order calculation (and in turn in the POWHEG prediction) was revisited. Using the dynamic scale resulted in a better perturbative convergence of the calculation and a flat k-factors for different observables. Following the suggestion of Ref. [165], the scale $H_{\mathrm{T}}^{B}/2$ is considered, which is defined as

$$H_{\mathrm{T}}^{B} = \left(\sqrt{(p_{\mathrm{T}}^{\mathrm{t},B})^2 + m_{\mathrm{T}}^2} + \sqrt{(p_{\mathrm{T}}^{\bar{\mathrm{t}},B})^2 + m_{\mathrm{T}}^2} + p_{\mathrm{T}}^{j,B} \right). \tag{2.40}$$

Here, the index B refers to the Born level, $p_{\mathrm{T}}^{\mathrm{t},B}$ and $p_{\mathrm{T}}^{\bar{\mathrm{t}},B}$ are the transverse momenta of the top quark and antiquark, respectively, and $p_{\mathrm{T}}^{j,B}$ is the transverse momentum of the additional jet. Theoretical uncertainties due to missing higher order effects, referred in the following as scale uncertainties, are estimated by varying the μ_{R} and μ_{F} scales by a factor of two up and down, independently. As a result, the choice of $H_{\mathrm{T}}^{B}/2$ results in symmetric scale uncertainties in contrast to the strongly asymmetric scale uncertainty, obtained for the scale choice of $m_{\mathrm{t}}^{\mathrm{pole}}$. Furthermore, a reduction of the scale uncertainty for the $\mathcal{R}(\rho)$ distribution, was observed [160].

In $\mathrm{t\bar{t}}$ production, the cross section is driven to a similar extent by the gluon PDF and the values of α_S and $m_{\mathrm{t}}^{\mathrm{pole}}$. Each variation in one of these parameters can be compensated in the cross section by a variation of the others. In contrast, the $\mathcal{R}$ distribution has only a small dependence on the value of α_S, as illustrated in Figure 2.19. A variation α_S from 0.114 to 0.123, alters the ρ shape only for $\rho < 0.4$, while leaving the threshold range of $\mathcal{R}$ unaffected. However, the

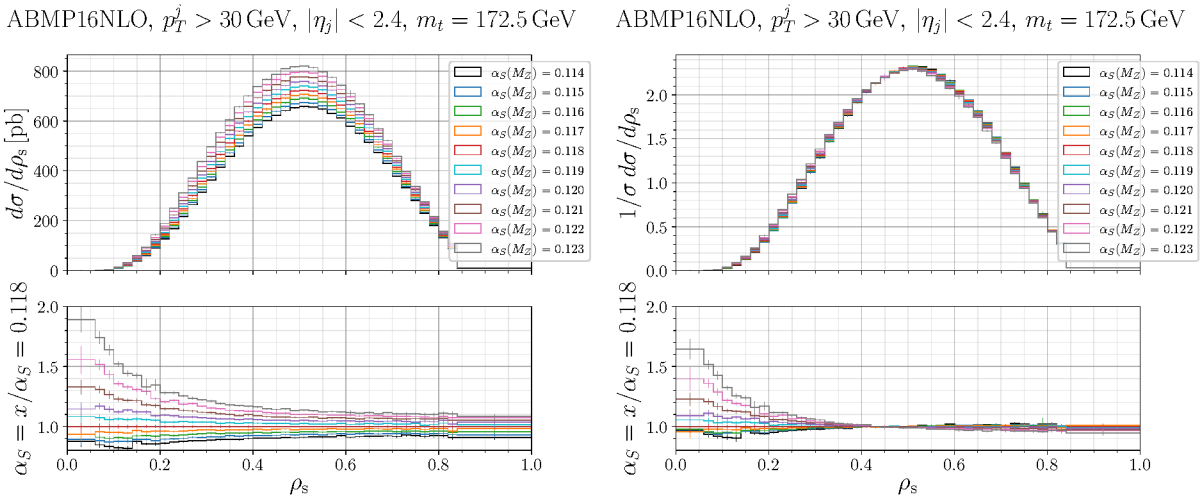

Figure 2.19: The absolute (left) and normalized (right) differential $\mathrm{t\bar{t}}$+jet cross section as a function of ρ for varying values of α_S [160]. The bottom panel shows the ratio to a central value of $\alpha_S = 0.118$.

PDFs play an important role in the $\mathcal{R}$ prediction, as illustrated in Figure 2.20. While most of the PDFs lead to the very similar predictions within the PDF uncertainties of about 10%, a significant difference is observed once the ABMP16NLO is used [160]. Since the α_S dependence of $\mathcal{R}$ was shown to be small, the origin of the shape difference must be a difference in the gluon distribution at high x among different PDFs used.

Total theoretical uncertainties in the normalized $\mathrm{t\bar{t}}$+jet cross section of up to 20% are estimated, as shown in Figure 2.20. Both PDF and scale uncertainties are of similar size and mainly affect the tails of the distribution.

For aforementioned calculations, the five-flavor number scheme is used, which means that all quarks except the top quark are considered massless. Consistently, this is taken into account also in the choice of the PDF sets.

Previous measurements and the effect of the center of mass energy

The mass sensitivity to the top quark mass of the normalized $\mathrm{t\bar{t}}$+jet cross section as a function of ρ was investigated previously in two analyses by the ATLAS Collaboration using 7 and 8 TeV pp collision data, corresponding to an integrated luminosity of 4.6 and 20.2 fb^{-1}, respectively [35,

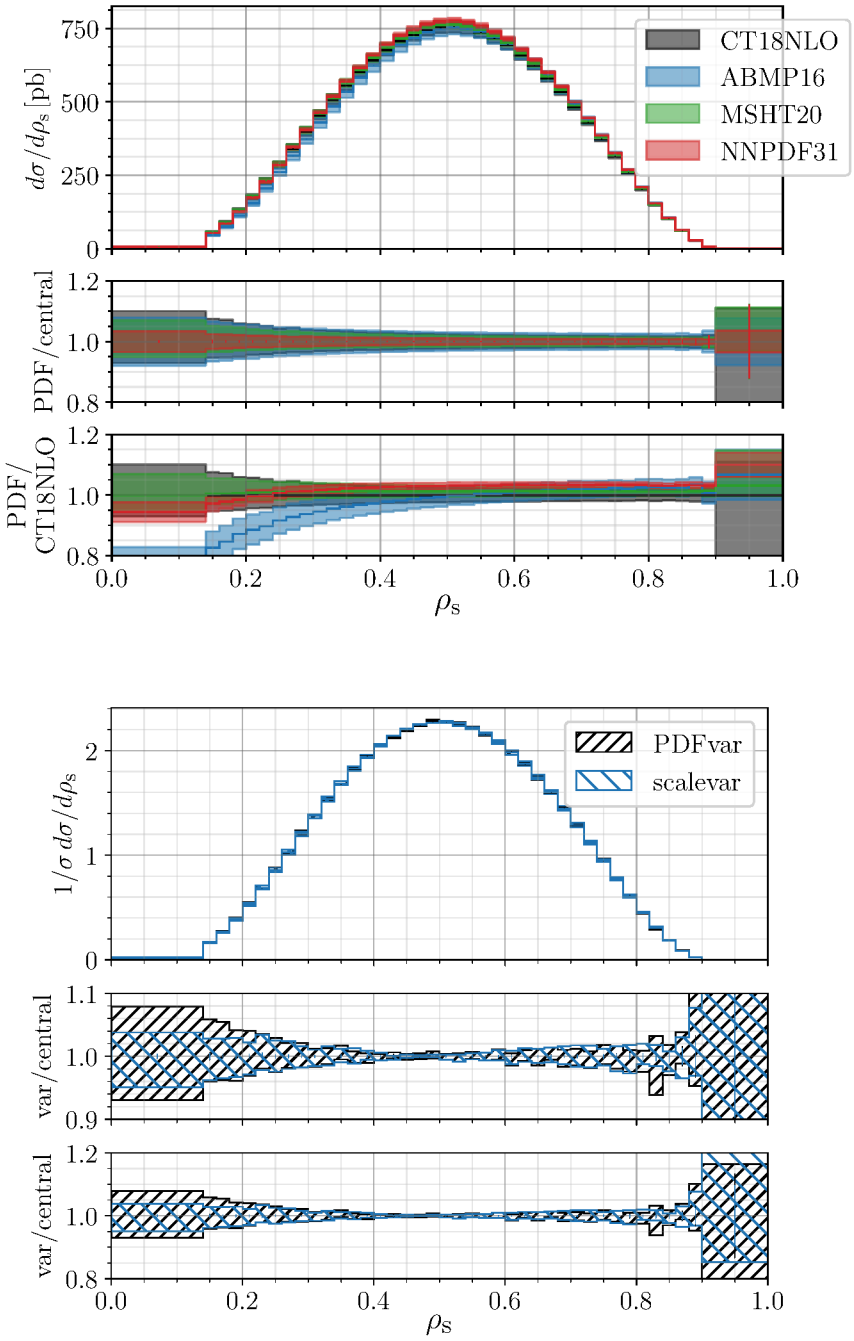

Figure 2.20: Top: the absolute differential $t\bar{t}$+jet cross section as a function of ρ for varying PDF sets [160]. The CT18NLO [166] (gray), ABMP16NLO [167] (blue), MSHT20 [168] (green), and NNPDF3.1 [169] (red) are compared. The upper panel shows the ratio to a central prediction for each PDF set while the lower panel shows the ratio to the prediction using the CT18NLO prediction. Bottom: the normalized differential cross section including the total PDF (black) and scale uncertainties (blue) when using the CT18NLO PDF. The bottom panel show the ratio of the uncertainties to the central prediction in two scales. The uncertainties in the cross section due to the PDF set are determined as given by the individual PDF prescriptions.

46]. Both measurements were performed by using the semileptonic decay of the top quarks. In the 7 TeV measurement, a value of

$$m_{\mathrm{t}}^{\mathrm{pole}} = 173.7 \pm 1.5\,(\mathrm{stat.}) \pm 1.4\,(\mathrm{syst.})^{+1.0}_{-0.5}\,(\mathrm{theory}) \tag{2.41}$$

was found using the parton-level differential cross section. An additional reinterpretation of the data was performed to extract the top quark mass as defined in the $\overline{\mathrm{MS}}$ scheme [41], resulting in a value of

$$m_{\mathrm{t}}(m_{\mathrm{t}}) = 165.9 \pm 1.4\,(\mathrm{stat.}) \pm 1.3\,(\mathrm{syst.})^{+1.5}_{-0.6}\,(\mathrm{theory}). \tag{2.42}$$

In 8 TeV measurement, the top quark mass was extracted in both the pole and $\overline{\mathrm{MS}}$ scheme. Given the data set being larger by a factor of around four, the statistical uncertainty is significantly reduced. This further results in an increase of the overall precision of the measurement. The results are

$$m_{\mathrm{t}}^{\mathrm{pole}} = 171.1 \pm 0.4\,(\mathrm{stat.}) \pm 0.9\,(\mathrm{syst.})^{+0.7}_{-0.3}\,(\mathrm{theory}) \tag{2.43}$$

and

$$m_{\mathrm{t}}(m_{\mathrm{t}}) = 162.9 \pm 0.5\,(\mathrm{stat.}) \pm 1.0\,(\mathrm{syst.})^{+2.1}_{-1.2}\,(\mathrm{theory}). \tag{2.44}$$

In both the $m_{\mathrm{t}}^{\mathrm{pole}}$ and the $m_{\mathrm{t}}(m_{\mathrm{t}})$ measurements, the theory uncertainty includes the scale uncertainties and variations of different PDF sets. The experimental uncertainties are of similar size, while the theoretical uncertainties are up to factor of two larger for the $\overline{\mathrm{MS}}$ mass. The larger scale uncertainties in the $m_{\mathrm{t}}(m_{\mathrm{t}})$ might be attributed to the fact that the $\overline{\mathrm{MS}}$ mass is not well-defined close to the $\mathrm{t\bar{t}}$ production threshold [170].

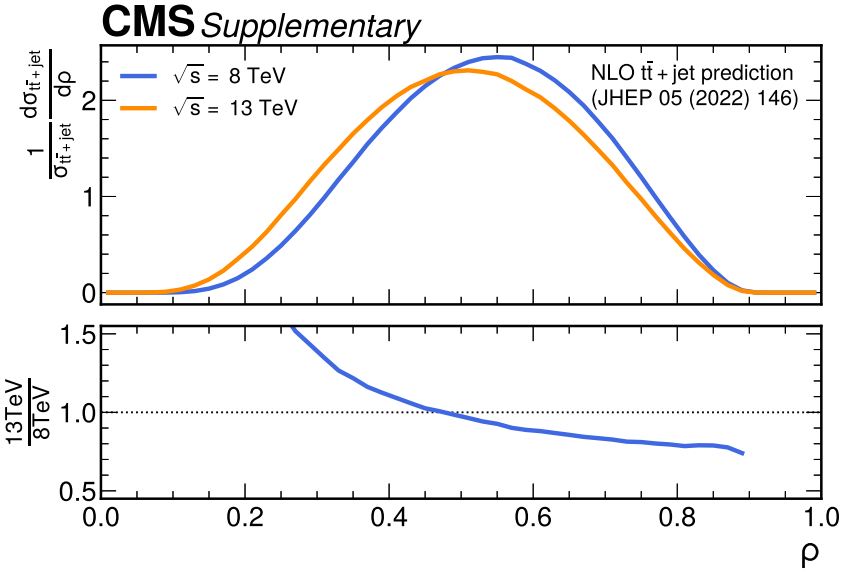

Figure 2.21: The normalized differential $\mathrm{t\bar{t}}$+jet cross section as a function of ρ for varying center of mass energies. The predictions for $\sqrt{s}$ of 8 TeV (13 TeV) are shown in blue (orange). The predictions are obtained as given in Ref. [160].

Increased center of mass energies have a non-negligible effect on the shape of the $\mathcal{R}$ distribution, and consequently on measurements of $m_{\mathrm{t}}^{\mathrm{pole}}$. This is illustrated in Figure 2.21. For a higher collision energy, the ρ distribution is observed to be shifted towards smaller values. As a result, the phase space region close to the production threshold shrinks by 25%. Hence, the total

sensitivity to $m_{\mathrm{t}}^{\mathrm{pole}}$ decreases for $\sqrt{s} = 13\,\mathrm{TeV}$ with respect to measurements at $\sqrt{s} = 7$ and $8\,\mathrm{TeV}$ significantly.

3 The experiment

This chapter introduces the experimental facilities used to collect the data which are analyzed in this thesis. In Section 3.1, the Large Hadron Collider (LHC) with its experiments is described. In Section 3.2, the Compact Muon Solenoid (CMS) experiment is discussed with its most important subdetectors. Additionally, the CMS trigger system and the measurement of the integrated luminosity are explained. Finally, future upgrades of the LHC and CMS are described in Section 3.3.

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [171] is a facility located at the European Organization of Nuclear Research (CERN) near Geneva in Switzerland. Built as a circular particle accelerator and storage ring, it is placed in a tunnel 100 m underground with a total circumference of 27 km, that has been previously used for the LEP collider [59]. In Figure 3.1, a schematic view of the LHC accelerator complex with the preaccelerators is shown. The proton beams are preaccelerated in the Booster, Proton Synchrotron (PS), and the Super Proton Synchrotron (SPS) to an energy of up to 450 GeV before entering the main storage ring. After injecting the beam's bunches into the two LHC beam pipes, the beams rotate in opposite directions to provide head-on-head collisions. The proton beams are accelerated using superconducting cavities operating at 400 MHz, and are bent by the dipole magnets, which reach a magnetic field of up to 8.3 T. The quadrupole (sextupole) magnets are used to perform a geometrical (momentum dependent and orbital) correction. The design energy per proton beam is 7 TeV, but the LHC is also operated with heavy ions, for example lead and xenon ions are accelerated to an energy of around 2.76 TeV per nucleon. In the analysis presented in this thesis, only proton-proton (pp) collision data are analyzed. Four main experiments at the LHC surround each of the interaction points. These are: ALICE (A Large Ion Collider Experiment) [172], ATLAS (A Toroidal LHC Apparatus) [173], CMS (Compact Muon Solenoid) [174], and LHCb (LHC Beauty) [175].

The LHC schedule and timeline is shown in Figure 3.2. So far, the LHC successfully accomplished two data taking periods: Run 1 (2010-2012), with a center-of-mass energy of 7 TeV (2010-2011) and 8 TeV (2012), and Run 2 (2015-2018), with a center-of mass energy increased to 13 TeV.

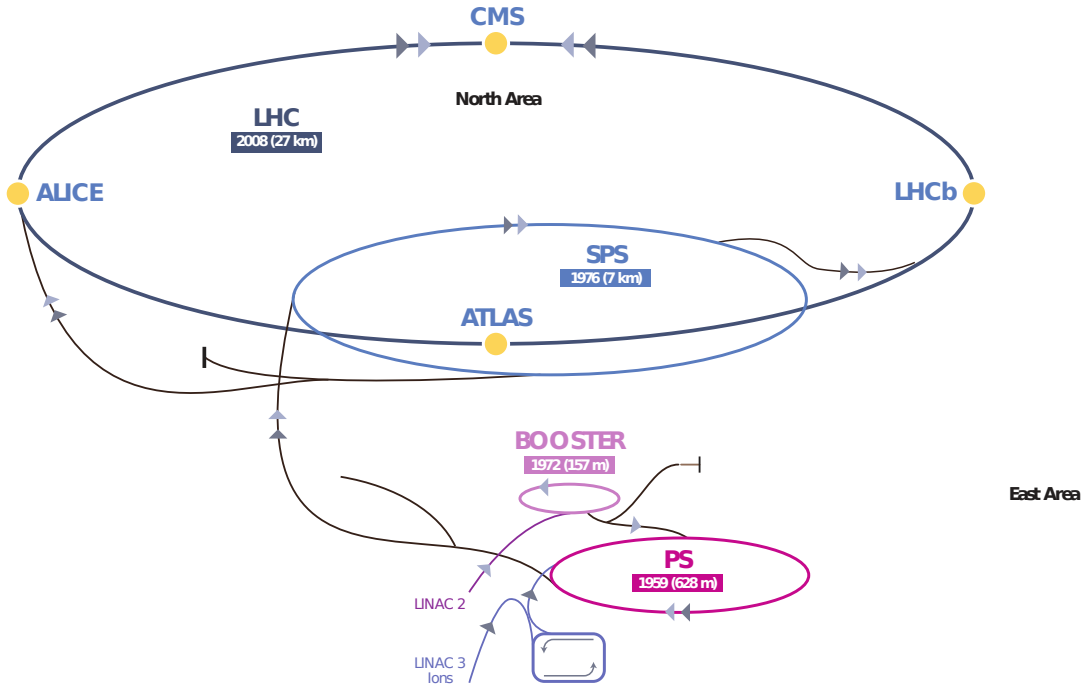

Figure 3.1: A schematic sketch of the LHC accelerator complex [176] including preaccelerators. The four main experiments are marked as yellow dots.

Special low energy runs have been carried out in 2013 at a center-of-mass energy of 2.76 TeV and 2015 and 2017 at 5.02 TeV. From 2022 on, the operation of the LHC with the design energy of 13.6 TeV and 14 TeV is planned for Run 3. During the long shutdown three (LS3), the LHC and the experiments are planned to be upgraded for the High Luminosity LHC (HL-LHC) planned to start from 2027.

The instantaneous luminosity $d\mathcal{L}/dt$ relates the event rate for a given process to the cross section σ as

$$\frac{dN}{dt} = \frac{d\mathcal{L}}{dt}\sigma. \tag{3.1}$$

The luminosity depends on the machine-specific parameters and can be determined as

$$\mathcal{L} = \frac{N_\mathrm{b}^2 n_\mathrm{b} f_\mathrm{rev}}{4\pi\sigma_x\sigma_y} F, \tag{3.2}$$

where N_b is the number of particles per bunch, n_b the number of bunches, f_rev the revolution frequency, σ_x and σ_y are the widths of Gaussian distributed beam profiles in x- and y-direction, and F is a geometrical factor accounting for the crossing angle of the beams [178]. The integrated luminosity $\mathcal{L}_\mathrm{int}$ is obtained from the instantaneous luminosity $d\mathcal{L}/dt$ as

$$\mathcal{L}_\mathrm{int} = \int \frac{d\mathcal{L}}{dt} dt = \frac{N}{\sigma}, \tag{3.3}$$

where N is the production rate of a certain process with a production cross section of σ. With $N_\mathrm{b} = 1.15 \cdot 10^{11}$ particles per bunch, a bunch spacing of 25 ns, and $n_\mathrm{b} = 2808$ bunches, the design instantaneous luminosity of the LHC is $10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. In 2016, the resulting instantaneous luminosity at its peak has been $0.77 \cdot 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ [179]. A large instantaneous luminosity consequently also leads to a large average number of simultaneous pp interactions per bunch crossing, also referred to as pileup. These additional collisions lead to major challenges in

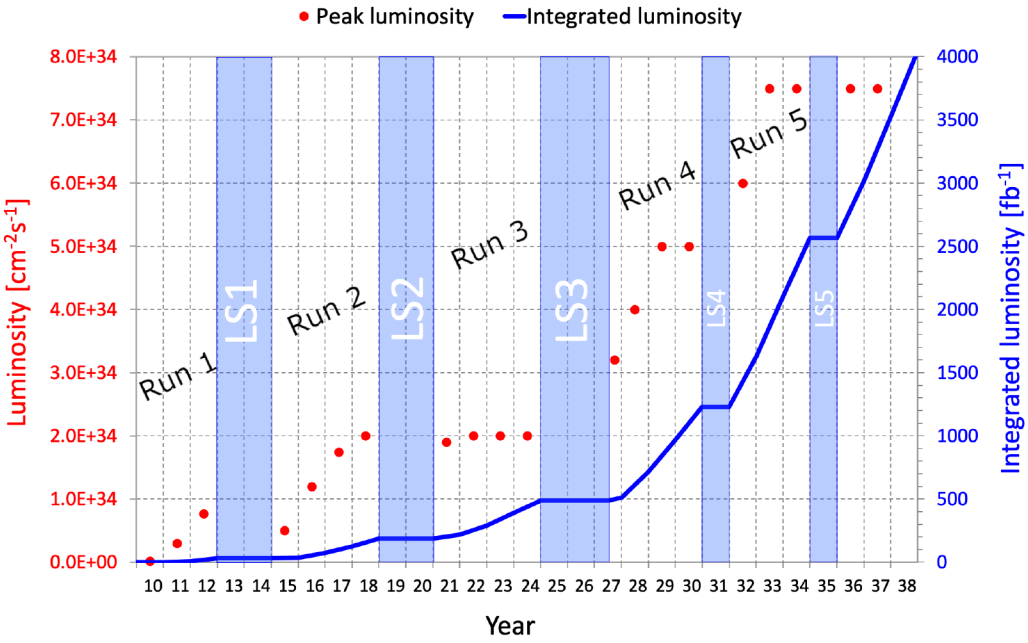

Figure 3.2: The LHC schedule and ultimate design values for the instantaneous and integrated luminosity [177]. The run periods are indicated together with the long shutdowns (LS).

the reconstruction as contributions of different interactions overlap in the readout of detectors. Mitigation techniques of pileup effects in CMS effects are further discussed in Chapter 4. The dedicated procedure of the luminosity measurement is described in Section 3.2.

3.2 The Compact Muon Solenoid

The CMS [174, 180] detector is a multi-purpose experiment at the CERN LHC encompassing various subdetectors that are designed to measure mass and momenta of the particles produced in the pp collisions. It follows a cylindrical layer design around the beam pipe. A sketch of the CMS detector is shown in Figure 3.3. From inside to outside, the first major component is the tracker system including pixel and silicon strip detectors. The tracker is surrounded by the electromagnetic calorimeter, followed by the hadronic calorimeter, the solenoid yoke, and the muon system. Geometrically the detector is split into a central "barrel" and two "endcaps" on either side.

The CMS event reconstruction relies on the reconstruction and identification of all particles produced in a pp collision event by taking advantage of the particle-specific interaction with matter and corresponding signature in the detector. Particle identification and measurement of its kinematic properties are based on the combination of the information from different subdetectors. The general approach is illustrated in Fig 3.4. Charged particles, such as charged leptons and hadrons, produce hits in the tracking system of the CMS detector, and their trajectories are bent by the magnetic field. By combining the hit information, tracks are reconstructed and their curvature and, in turn, their momenta are measured. The energy of electrons, positrons or charged hadrons is determined by measuring the deposited energy in electromagnetic and hadronic calorimeter, respectively. This information can be combined with the reconstructed tracks in the tracking system to increase the precision in the momentum estimation. Photons

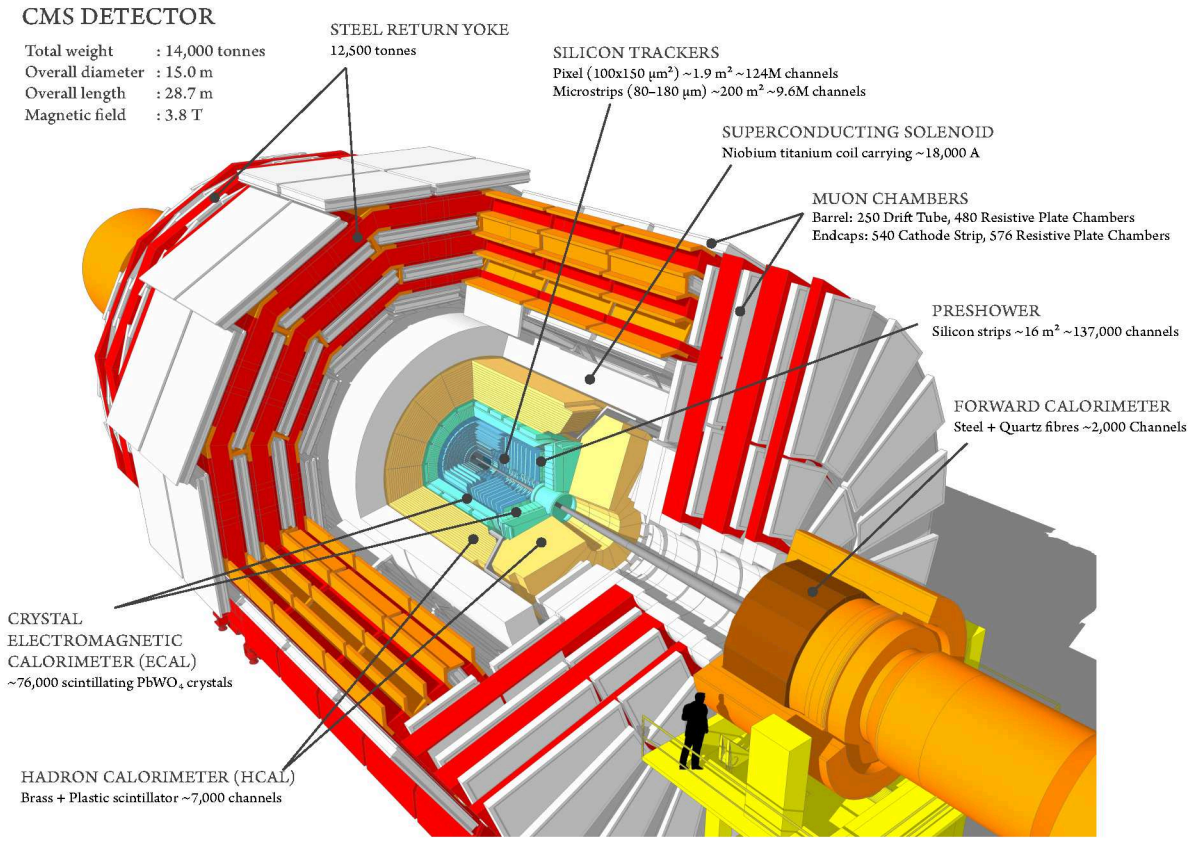

Figure 3.3: A schematic representation of the CMS detector showing all subdetectors [181]. The innermost part of the detector is the tracking system, composed of the pixel and silicon strip detectors. Outside the tracking system the electromagnetic and hadron calorimeters are installed, both within the solenoid magnet. The outermost part of the detector is the muon system.

and neutral hadrons are also absorbed in the calorimeters but cannot be reconstructed as tracks. Muons leave the CMS detector without being fully stopped in the material, but through combination of their hits in the tracker and their trajectory as measured in the muon chambers, their momenta can be determined. Neutrinos are the only particles with no signature in the detector. Nevertheless, by measuring the missing transverse momentum in an event, conclusions about the presence of neutrinos can be drawn. By combining the information of all reconstructed particles, tracks, vertices, hypothesis can be built for stable particles in terms of four-momenta, trajectory, and particle type.

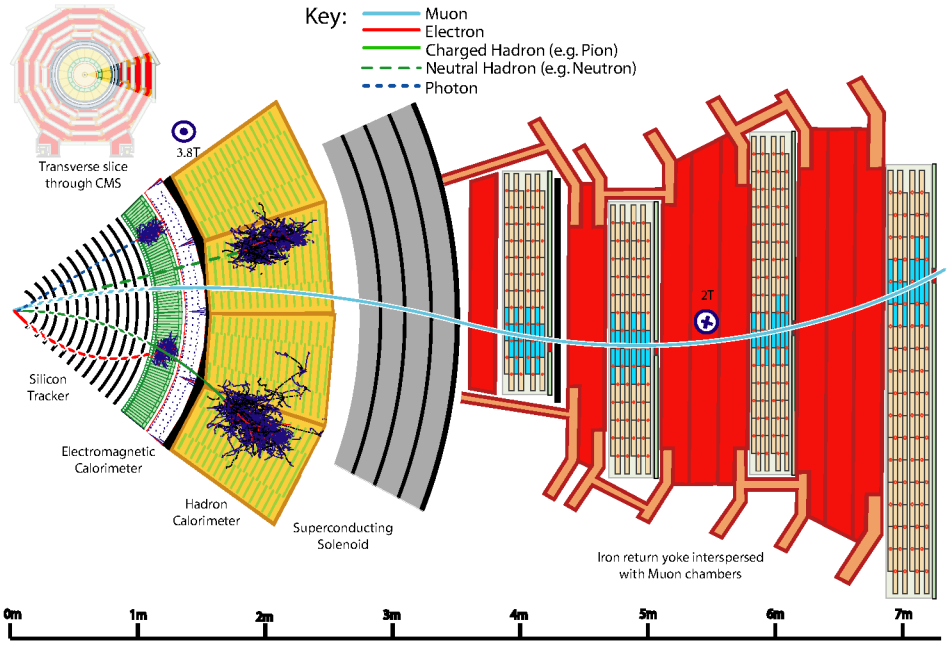

Figure 3.4: Schematic view of a transverse slice of the CMS detector, illustrating interactions of final-state particles with all subdetectors. The muon and charged hadron shown are positively charged, whereas the electron is negatively charged [182].

The coordinate system of CMS has its origin located at the interaction point, and the z-axis points in the direction of the beam axis. The y-axis points vertically upwards, while the x-axis is directed towards the center of the LHC ring, respectively. An illustration of the coordinate system is shown in Figure 3.5. The particle kinematic properties are expressed in terms of the azimuthal angle ϕ, the transverse momentum p_{T}, and the rapidity y (or pseudo-rapidity η). The angle ϕ is measured in the x–y plane, where $\phi = 0$ corresponds to the direction of the x-axis, and ϕ ranges from $-\pi$ to π. Similarly, the polar angle θ is measured from the positive z-axis. The transverse momentum p_{T} and the rapidity y are defined as

$$p_{\mathrm{T}} = |\vec{p}_{\mathrm{T}}| = |\vec{p} \cdot \sin(\theta)| \tag{3.4}$$

and

$$y = \frac{1}{2} \ln \frac{E + p_{\mathrm{z}}}{E - p_{\mathrm{z}}}, \tag{3.5}$$

where E and p_{z} are the energy and momentum in direction of the z-axis of the particle. Differences in y are invariant under Lorentz boosts in beam (z) direction. Similarly, the pseudorapidity η is defined as

$$\eta = -\ln \left(\tan \left(\frac{\theta}{2} \right) \right), \tag{3.6}$$

which is equivalent to y in the relativistic limit where particles are considered massless. In addition, the flux of particles is constant as a function of η. The geometric distance between two objects in the detector is further defined as

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}, \tag{3.7}$$

which is also known as the boost-invariant distance.

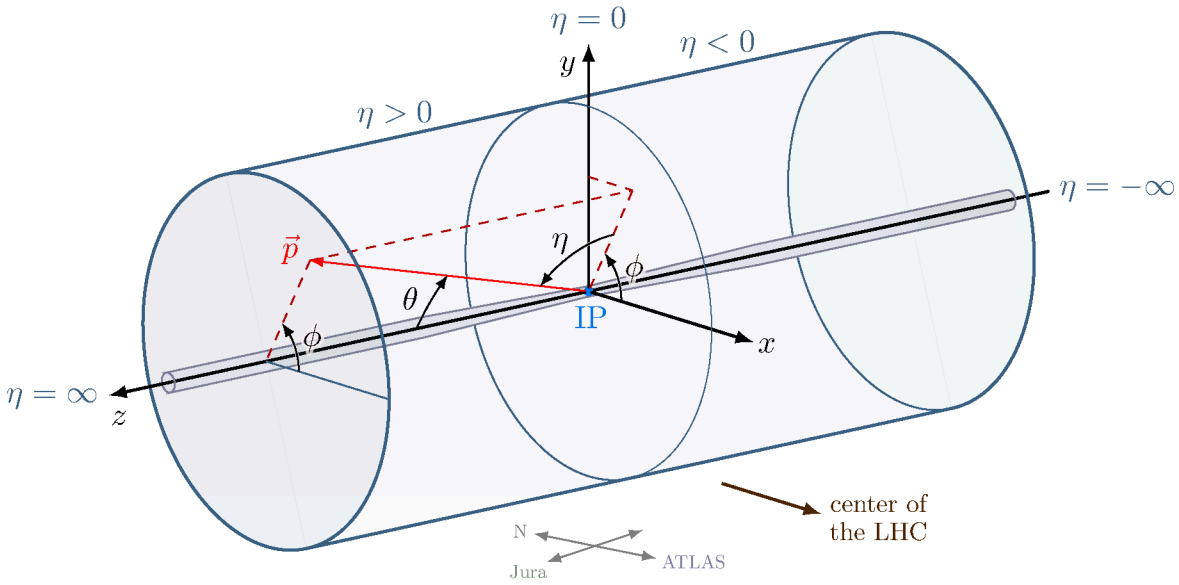

Figure 3.5: The coordinate system used by the CMS experiment. The x-axis points to the LHC center, the y-axis is defined upwards, and the z-axis is oriented to be parallel to the beam line in counter-clockwise direction. Based on [183].

The solenoid magnet

To bend the trajectories of charged particles, a solenoid magnet with a magnetic field of 3.8 T is used, which allows for a precise measurement of particle momenta using the tracking system. With a total size of 12.5 m in length and a diameter of 6.3 m, the CMS solenoid is the largest ever constructed. The coil is made of Niobium-titanium (NbTi) and is cooled down to 4.8 K to operate in a superconducting regime.

The tracking system

The CMS tracker consists of two main components, a silicon pixel and a silicon strip detector. It is designed to perform precise measurements of the momenta of the charged particles in the magnetic field and allow for an identification of primary and secondary vertices. The latter is especially important for an accurate identification of heavy-flavor jets as discussed in Section 4.2, as well as to associate each reconstructed track and secondary vertex to the correct bunch crossing and vertex. Therefore, both high granularity and fast response are essential. The barrel and the endcap parts of the tracker cover a total length of 5.8 m and a diameter of 2.5 m. A sketch of the whole CMS tracker with its subdetectors is shown in Figure 3.6. The silicon pixel tracker is composed of three barrel layers (BPIX) (the closest at a radial distance of 4.4 cm to the center of the beam pipe) and of two disks in the endcaps (FPIX), covering the total area of around 1 m^2. Each of the about 66 million silicon pixel cells has a size of $100 \times 150\,\mu\mathrm{m}^2$, which

provides good granularity in any direction. The silicon strip detector consists of four strip layers in the inner (TIB) and six layers in the outer barrels (TOB). In the endcaps, it is built of three inner disk layers (TID) and nine layers in the outer part (TEC), each. The strip thickness ranges from 320 μm to 500 μm with increasing radius while the distances between the strips range from 80 μm to 140 μm. The inner tracker in total covers a range of $|\eta| < 2.5$, and the sensoring system provides an active detector area of around 200 m^2 with approximately 10 million strip sensors. At the end of the 2016 data taking period, the silicon pixel detector has been replaced [184]. The upgrade includes an additional fourth layer of the pixel detector in the barrel region and one additional disk per endcap. Together with a closer position of the innermost disks and layers in the barrel to the beam pipe, a significant improvement in the precision of the track and vertex reconstruction could be achieved. The tracker efficiency is estimated with the tag-and-probe method [185], by using the muons originating from the Z boson decays and results in 96–99% depending on the pileup condition [186]. The resolution in the collision vertex reconstruction is better than 14 (19) μm in x (z) direction [187].

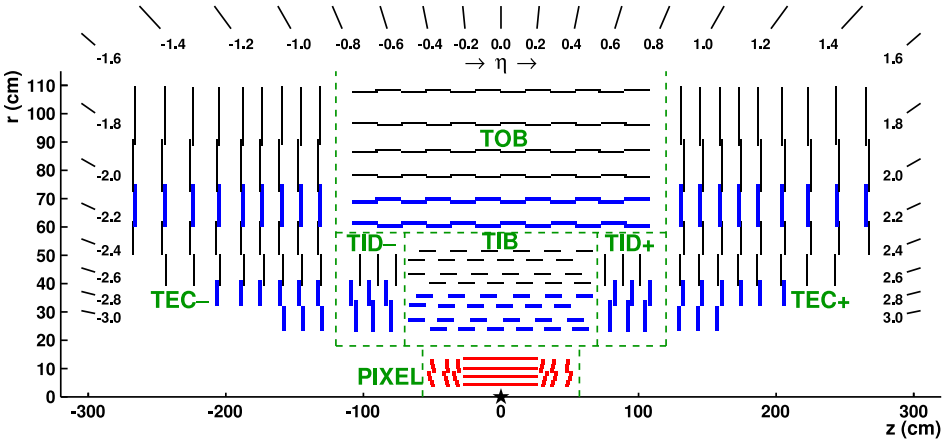

Figure 3.6: A sketch of one quadrant of the CMS tracker after the upgrade [185]. It is composed of the inner pixel (red) and of the strip detectors (blue). The silicon strip detector is divided into the TIB, TOB, TID, and TEC. The star marks the interaction point.

The electromagnetic calorimeter

The electromagnetic calorimeter (ECAL) of CMS is built around the tracker, and its main purpose is to measure the energy of electrons and photons, which result in similar detector signatures due to emergence of the electromagnetic showers in the calorimeter material. Together with combined measurements in the tracker, photons can be distinguished from the electrons. A schematic view of the ECAL is shown in Figure 3.7. The ECAL consists of lead-tungstate ($PbWO_4$) scintillator crystals and is divided into a barrel (EB: $|\eta| < 1.479$) and an endcap part (EE: $1.479 < |\eta| < 3.0$). The crystals act as absorber and scintillation material at the same time, making the ECAL a fully homogeneous calorimeter. Electromagnetic showers initiated by Bremsstrahlung through the absorption in the crystals lead to photon emission via scintillation. The emitted light is detected by photodiodes and is proportional to the deposited energy. The properties of $PbWO_4$ are high density with respectively short radiation length of 0.89 cm, allowing the ECAL to be rather compact with a large sensitive volume. The $PbWO_4$ crystals have short scintillation time such that in a time window of 25 ns most of the visible light (approximately 80%) is emitted, being especially suitable for a bunch spacing of 25 ns. The crystal length in the EB (EE) is 23 cm (22 cm), corresponding to 26 (25) radiation lengths. To prevent losses of

signal along the crystals, those in EE and EB are tilted in an angle of 2–8 degrees with respect to their orientation towards the interaction point in dependence of η. An additional preshower detector (ES) is installed in front of the endcaps to improve the identification of photons coming from, e.g., π^0 decays.

The energy resolution of the ECAL is determined in dedicated test beam experiments and corresponds to [188]

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E[\,\mathrm{GeV}]}}\right)^2 + \left(\frac{12\%}{E[\,\mathrm{GeV}]}\right)^2 + (0.3\%)^2. \tag{3.8}$$

The energy resolution is optimized for the energies of around 50 GeV to maximize the sensitivity for the measurements of Higgs boson decays into two photons [189]. The energy scale resolution for the electrons from Z boson decays ranges from 2–5% depending on the electron η [190].

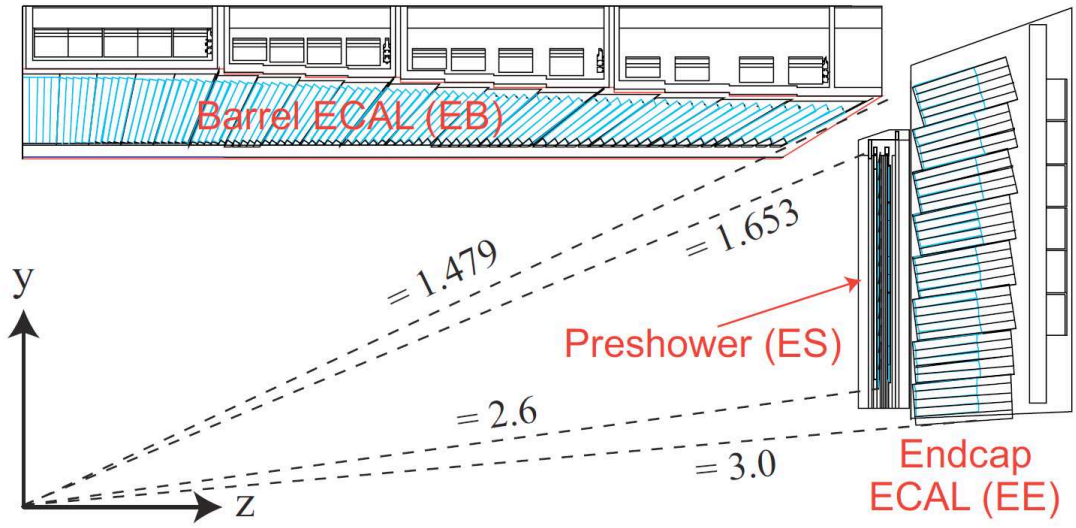

Figure 3.7: A sketch of one quadrant of the CMS detector for the ECAL [191].

The hadronic calorimeter

The hadronic calorimeter (HCAL) is designed to measure the energy of charged and neutral hadrons and consists also of barrel (HB) and endcap parts (HE) that are installed between the ECAL and the magnet coil. A schematic view of the HCAL is shown in Figure 3.8. The HCAL barrel is a sampling calorimeter, built of alternating layers of brass absorbers and plastic scintillator material. Hadronic showers are created through interaction of particles with the absorber material, which further leads to scintillation light emission, collected in photodiodes. An additional outer barrel part with lower granularity (HO) extends the HCAL outside the solenoid magnet, which is utilized by the HO as an additional absorber. In addition, the hadron forward detector (HF) is installed to cover high pseudorapidity ranges. Pseudorapidities in the range of $|\eta| < 1.4$ are covered by the barrel part, while the HCAL endcap together with the outer HCAL spans over the range of $1.3 < |\eta| < 3$. The HF, covering pseudorapidity ranges of $3 < |\eta| < 5.2$, needs to be more radiation hard in comparison to the barrel parts of the HCAL, because the energy deposit in the forward region is much larger. It therefore uses layers of steel and quartz fibres.

Using the data of dedicated test-beam experiments, the combined energy resolution of ECAL and HCAL is determined to be [192]

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{85\%}{\sqrt{E[\,\mathrm{GeV}]}}\right)^2 + (7.0\%)^2\,. \tag{3.9}$$

For the pp collisions in 2016 data taking period, a good agreement of the HCAL energy measurement between data and simulation, between 3–5%, is observed [193], as well as for the energy response [194].

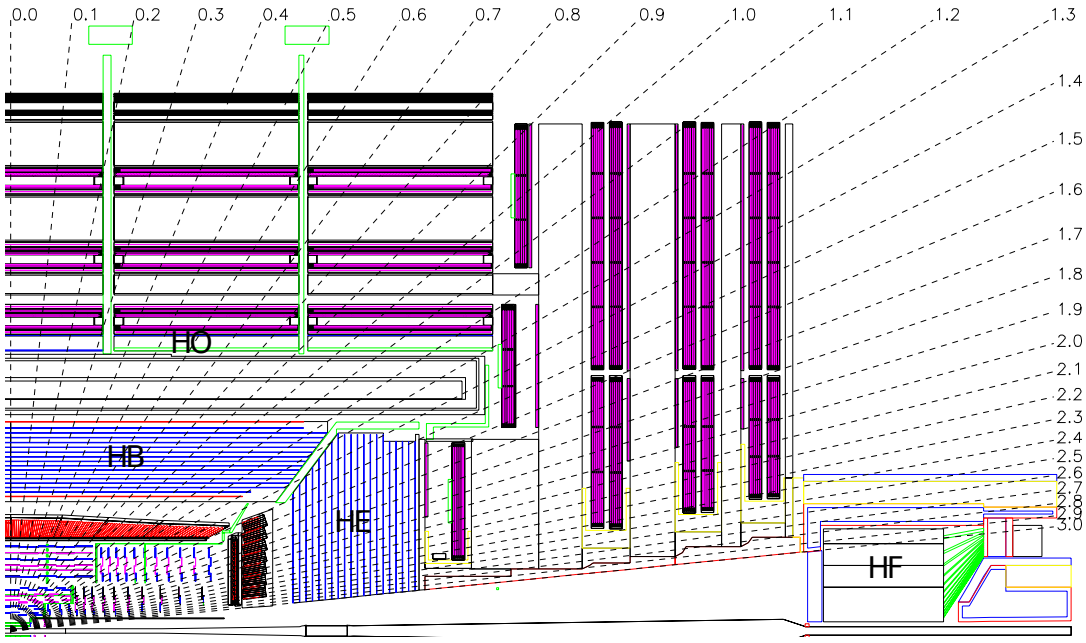

Figure 3.8: A sketch of one quadrant of the CMS detector [174]. The subparts of the HCAL, the HE, HB, HO, and HF are indicated separately.

The muon system

The outermost component of the CMS detector is the muon system (see Figure 3.9). It is located outside the solenoid and is only traversed by muons (and neutrinos, which are however not detectable with CMS), and allows for precise identification and measurement of muon trajectories by combining its signal with information of the tracker. Outside the solenoid, the magnetic field is weaker (approximately 2 T) than within the solenoid. The muon system is composed of three types of gaseous detectors that are interlaced with the flux return yoke. The barrel part ($|\eta| < 1.2$) is covered by four stations of drift tubes (DT) while cathode strip chambers (CSC) are located in the endcap regions for pseudorapidities in the range $0.9 < |\eta| < 2.4$. CSCs, in contrast to DTs, are characterized by a faster response time and larger radiation hardness, which makes them especially suitable for measurements in the forward region where a higher muon rate is expected. Additionally, six (three) layers of resistive plate chambers (RPC) are installed in the pseudorapidity range of $|\eta| < 1.2$ (1.6) in the barrel and endcaps. The RPCs show both good time resolution and fast response time of 1 ns and are thus mainly used for triggering purposes.

The muon reconstruction, identification, and isolation efficiency is found to be better than 95%, with a misidentification rate less than 1% [196]. The muon momentum resolution varies

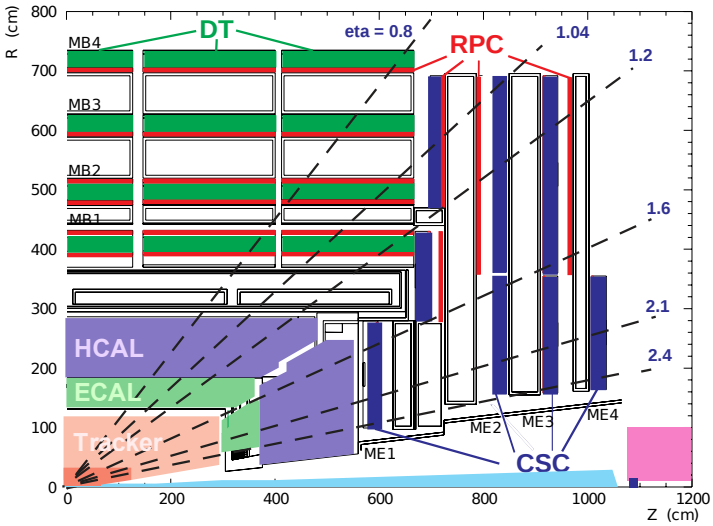

Figure 3.9: A sketch of one quadrant of the CMS detector with the main detector components [195], especially the different muon detectors, is shown.

between 10% in the barrel region to around 20% in the endcap region for muons of $p_T > 15\,\text{GeV}$, originating from Z boson decays, reconstructed only in the muon system [195]. In a combined reconstruction using also the tracker information, the muon resolution is on the order of 1–3% for muons with $p_T < 100\,\text{GeV}$ and about 10% for muons with p_T in the TeV range [196]. The energy resolution is optimized for muons from leptonic decays of the Higgs boson via ZZ or ZZ* [188].

The trigger system

With an instantaneous luminosity of about $L = 10^{34}\,\text{cm}^{-1}\,\text{s}^{-1}$ and a bunch spacing of 25 ns, i.e., a frequency of 40 MHz, the LHC provides pp collisions with a rate of 10^9 per second [180]. Considering the storage size of 1.5 MB per event, this results in a data output rate of about 950 TB per second [197]. Only a small fraction of these events is of potential interest for the LHC physics program, and the data acquisition (DAQ) system is capable of processing and saving only about 100 events per second for permanent storage. Thus, a significant reduction in the event rate, approximately by a factor of 10^6, has to be achieved by filtering or "triggering" only such events that feature signatures of particular interest. The CMS trigger system comprises two separate stages [198], the hardware based level 1 (L1) trigger system and the software based high level trigger (HLT) system. The L1 trigger system uses only the combined information from the calorimeters and the muon system, because 3.2 μs after recording of the data it has to be transferred to the readout system. Thus, the L1 trigger decisions are determined in less than 1 μs [180]. As a consequence, the maximum output rate is reduced by the L1 trigger system down to 100 kHz. A more complex reconstruction with increased precision is performed by the HLT trigger system. It has full access to the complete readout of all detector components and the reconstruction is thus close to the one used for offline analysis, cf. Section 4. The event reconstruction is further adapted to match the timing constraints of the software and data storage system. In total, the HLT system reduces the event rate to $\mathcal{O}(1\,\text{kHz})$ [199]. Using the decisions of the HLT, events are categorized into different subsets, based on event kinematics and the trigger objects, called primary data sets. A fixed set of L1 and HLT selection criteria is referred to as "trigger path". In contrast to the offline analysis, the events discarded at L1 or HLT level cannot be recovered.

Measurement of the integrated luminosity

The precise knowledge of the luminosity is an essential input to all measurements performed at the LHC. In CMS, five different subdetectors (luminometers) are employed in order to determine the luminosity based on event rates [174, 200]. The most precise method is based on pixel cluster counting (PCC) using the number of recorded pixel clusters in the pixel detector. Since the probability that one of the roughly 70 million pixels is hit by two different charged particles in the same bunch crossing is negligible, the number of reconstructed pixel clusters is linear dependent on the number of interactions per bunch crossing, and thus the instantaneous luminosity. The relation between the instantaneous luminosity L and the number of recorded pixel clusters N_{clusters} can be written as [200]

$$L = \frac{f_{\text{rev}}}{\sigma_{\text{vis}}^{\text{PCC}}} \cdot N_{\text{clusters}}, \tag{3.10}$$

with $\sigma_{\text{vis}}^{\text{PCC}}$ being the visible PCC cross section and f_{rev} the LHC orbit frequency. The unknown parameter $\sigma_{\text{vis}}^{\text{PCC}}$ is determined using so-called "Van der Meer" (VdM) scans [201, 202].

The idea behind VdM scans is to infer the visible cross section based on the rates at different beam displacements Δ_x and Δ_y in transverse direction. The benefit is that the determination becomes solely dependent on the experimental setup and calibrations, and is thus independent of modeling. The instantaneous luminosity can be written as

$$L = \frac{f_{\text{rev}} N_1 N_2}{A_{\text{eff}}}, \tag{3.11}$$

where N_1 and N_2 are the number of particles per bunch and A_{eff} denotes the effective overlap area between the bunches. By replacing $1/A_{\text{eff}}$ with the normalized density distributions of the beams which are assumed to be Gaussian and of the same width, the relation reduces to

$$L = \frac{f_{\text{rev}} N_1 N_2}{4\pi \sigma_x \sigma_y}, \tag{3.12}$$

where σ_x and σ_y are the root-mean-square (RMS) widths of the horizontal and vertical bunch profiles, respectively. Assuming Gaussian-distributed bunches with $\Sigma_i = \sqrt{2}\sigma_i$, $\sigma_{\text{vis}}^{\text{PCC}}$ can be calculated via

$$\sigma_{\text{vis}}^{\text{PCC}} = \frac{2\pi \Sigma_x \Sigma_y}{N_1 N_2} \mu_{\text{vis}}, \tag{3.13}$$

with μ_{vis} being the average interaction rate at $\Delta x = \Delta y = 0$, leading directly to the instantaneous luminosity via Equation 3.10. The quantities Σ_i are measured using VdM scans. The corresponding integrated luminosity is obtained by integrating L over time.

The total integrated luminosity collected by the CMS experiment during the 2016 data taking is estimated to be $36.3\,\text{fb}^{-1}$ with a total uncertainty of 1.2%, which is one of the most precise determinations of the luminosity at a hadron collider to date [200]. The total integrated luminosity for the full Run 2 data taking (2016-2018) is measured to be $138\,\text{fb}^{-1}$, with an uncertainty of 1.6% [200, 203, 204]. The integrated luminosity versus time as recorded by CMS and delivered by the LHC is shown in Figure 3.10, together with the average number of pp collisions.

3.3 The High Luminosity LHC upgrade

Both the LHC and the CMS experiment will undergo upgrades during the long shutdown 3 for the HL-LHC physics project starting in 2027 [48, 206–210], referred to as Phase 2, cf. Section 3.1.

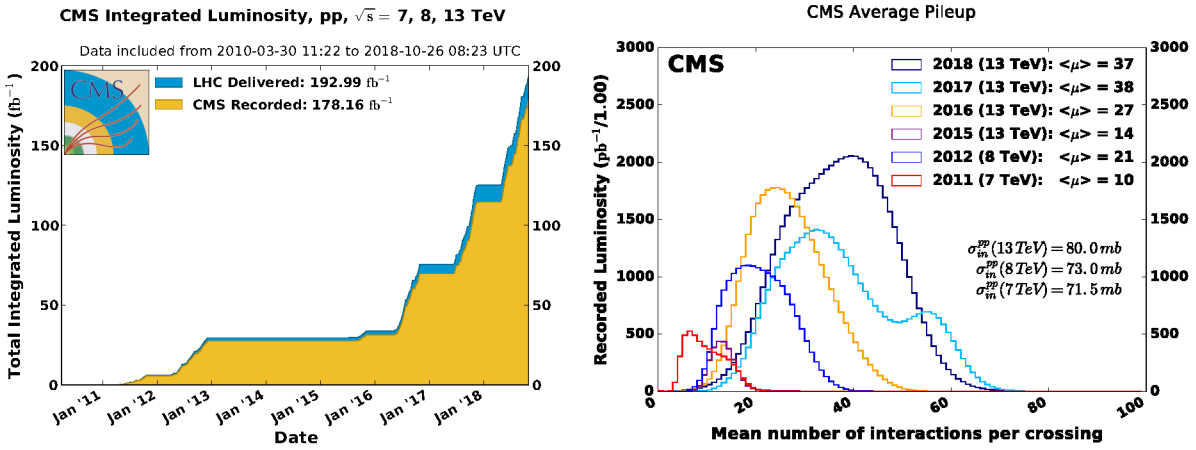

Figure 3.10: Integrated luminosity recorded by the CMS experiment (left) and average number of simultaneous pp interactions per bunch crossing for all years if data taking with the average $< \mu >$ and the inelastic minimum bias cross section for each center of mass energy (right) [205].

As a substantial increase in the instantaneous luminosity is necessary to result in a significant improved statistical precision after the LHC Run 3, new methods such as luminosity leveling have to be employed to achieve a target luminosity of $7.6 \cdot 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ at the HL-LHC. This is further realized by using improved superconducting crab cavities, which ensure head-on collisions for small beam widths, new radiation hard dipole and quadrupole magnets, and a revision of the LHC injector chain [206]. The HL-LHC target luminosity is a factor of seven larger than for the Run 2 LHC and leads to an average pileup of 200 interactions per bunch crossing. It will result in a collection of 300–350 fb^{-1} per year, and in turn in an integrated luminosity of about 4000 fb^{-1}. A development margin of 50% is considered for all HL-LHC developments, leading to a second running scenario of 140 pileup interactions with a reduced instantaneous luminosity. The physics program for the HL-LHC comprises beyond-the-SM searches up to high energy scales and unprecedented precision for SM measurements with significantly improved understanding of the Higgs boson sector. For example, it is expected to be sensitive to the Higgs boson self-coupling and Di-Higgs boson production.

Not only the collider, but also the experiments will receive upgrades to cope with the high radiation and high number of pileup interactions, leading to nearly a complete upgrade of the CMS detector. The upgraded tracker will cover up to $|\eta| = 4$ with the information of the particle momentum already used at L1 [207]. The current HCAL will be replaced with a new high granularity endcap calorimeter (HGCAL) to allow for a better coverage in the forward region motivated by planned investigations in the Higgs measurements [208]. This will be complemented by new muon detectors that cover the regions in $|\eta| < 2.8$ [209]. A new feature of CMS will be a timing detector, that will help to identify tracks and vertices belonging to the main interaction [210].

Consequently, also the DAQ and HLT systems have to be upgraded in order to deal with the new conditions. The new target rate for the L1 trigger is 5–7 times higher than for Run 2, and about 500–750 kHz. The output rate of the HLT will be approximately 5–7.5 kHz depending on the scenario under scrutiny (140 *vs.* 200 pileup interactions).

4 Event reconstruction

In this chapter, the details of the reconstruction employed in CMS are explained, with a focus on the objects relevant for this thesis. Section 4.1 introduces the particle flow (PF) algorithm [182], which is the core of the reconstruction of physics objects in CMS.

4.1 The particle-flow algorithm

The PF algorithm aims to perform a global identification of physics objects by exploiting the performance of the CMS tracking system, the high granularity of the ECAL, and the highly efficient muon tracking system, by combining the information from all subdetectors [182]. The concept is illustrated in Figure 4.1. The PF output consists of a collection of particle candidates such as leptons of different flavors, photons, jets, and derived observables such as $p_{\mathrm{T}}^{\mathrm{miss}}$, resulting in the complete event description with a precision better than using individual reconstruction methods for physics objects. It further allows for the identification of the primary vertex and associated tracks and objects, facilitating various pileup mitigation techniques. Moreover, for example the misidentification rate for PF jets is much lower than for jets solely reconstructed using calorimeter information. The PF reconstruction and identification procedure can be divided in three major steps, of which the first one is to identify tracks and clusters in each subdetector individually.

Vertices and tracks

The CMS track reconstruction combines the information from the pixel and strip trackers, aiming to maximize the track identification and reconstruction efficiency while retaining a low rate of misreconstructed "fake" tracks from incoherent or artificial hits.

Hits are reconstructed the following way. In a first step, referred to as "local reconstruction", signals above a defined threshold of the deposited charge in the pixel and strip detectors are collected and clustered locally in each sensor. Both the pixel and strip hit finding efficiency for tracks with $p_{\mathrm{T}} > 1\,\mathrm{GeV}$ are found to be better than 99.4% and 99.8%, respectively, when using only modules without defects [185]. The hit resolution is determined to vary between

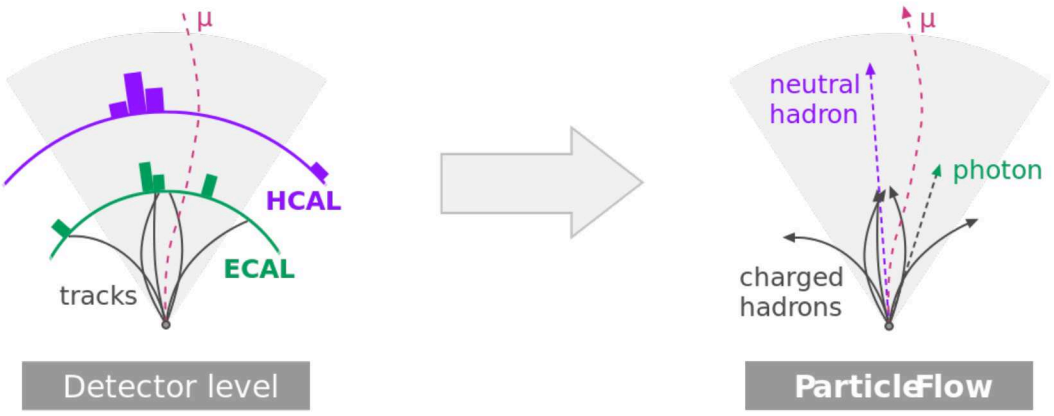

Figure 4.1: Illustration of the particle-flow concept. The algorithm combines information from different subdetectors to provide a global description of the event and identifies electrons, photons, muons, charged hadrons, and neutral hadrons [211].

10–50 μm for the pixel, and 20–50 μm for the strip tracker for tracks with $p_T > 12\,(10)$ GeV, respectively [185].

In the second step, hits are used as seeds in an iterative track finding algorithm, after translating their local position into a global coordinate system. In this step, possible differences between the actual and expected position of the sensor have to be taken into account. The procedure is referred to as alignment and uses LHC collision and cosmic data, yielding an accuracy of 10 μm on the position of the final aligned tracker modules [212].

The CMS tracking approach is based on the so-called combinatorial track finder (CTF) [185], which is an extension of the combinatorial Kalman filter [213–216]. The CTF sequence is processed multiple times iteratively in order to maximize the reconstruction efficiency, and to keep the fake rate at a minimum. The first few iterations aim to reconstruct tracks with higher p_T or closer distance to the interaction point, while later iterations target low p_T or significantly displaced tracks. In total, a maximum of six iterations are executed. In each iteration, hits are combined to seeds, and track properties are determined using an extrapolation of the estimated flight path based on the Kalman filter. With increasing number of iterations, the requirements to form seeds become less stringent by, e.g., allowing more displaced seed positions. At the end of each iteration, tracks not fulfilling selection requirements are discarded, and used hits are masked for the next iteration. For later iterations, these requirements are tightened to compensate for random track building and noise.

The tracking reconstruction efficiency is found to be better than 99% (80–99%) for isolated muons (pions) depending on the pseudorapidity range, and the fake rate is between 5 and 15% [185]. Excellent agreement between data and simulation is found, and the resolution of the track p_T is determined to be between 1 and 10% depending on the track p_T and η for muon tracks ranging from 1 to 100 GeV [185].

The reconstruction of primary, secondary, and pileup vertices in CMS is realized in three steps [185, 217]. First, tracks with high fit quality that are consistent with an origin close to the interaction point are selected. Then, these are clustered into groups that originate from the same interaction vertex using a deterministic annealing algorithm [218]. The clustering is performed considering the distance of the tracks to the center of the luminous region (the beam spot). Next, the vertex candidates with at least two associated tracks are fitted using an adaptive vertex finder [219] to determine the vertex parameters. In the fit each track is assigned a

weight ranging between 0 and 1, which represents the likelihood to originate from a particular vertex. For each vertex, the sum-of-track-weights is calculated and vertices with a weight sum below given thresholds are discarded. CMS also employs an independent track and vertex reconstruction approach using only pixel detector information yielding an extremely fast solution needed for the HLT. As the pixel-based approach has a worse efficiency and mistagging rate, it is not used as the default reconstruction method.

Tracking for electron/positron candidates

For electrons, an alternative track seeding is realized, based on reconstructed ECAL clusters used as seeds [220]. So-called "superclusters" are built out of ECAL clusters and accommodate for the electromagnetic radiation of electrons and collect all electron and Bremsstrahlung photon energy deposits in the ECAL. The superclusters are then used to draw conclusions about the position of hits expected in the innermost tracker layer under the electron or positron hypothesis.

To improve the reconstruction efficiency for electrons that are missed by the ECAL-based approach, seeds as determined by the iterative tracking approach are considered as inputs to a tracker-based electron reconstruction. However, the strong Bremsstrahlung of electrons can lead to non-Gaussian and abrupt energy losses and hence an inaccurate description of the tracks. Therefore, additionally a Gaussian-sum filter (GSF) [221] is employed to fit the same tracks again for which a large presence of Bremsstrahlung is assumed.

The seeds obtained by the two approaches are selected using a multivariate-analysis (MVA) based technique and combined as electron seeds in the following. In a first track building step, again a combinatorial Kalman Filter is used to collect all hits along the electron trajectory. To infer the track parameters, a GSF fit is performed assuming a Gaussian energy loss in each tracker layer.

For electrons from leptonic Z boson decays, the tracking efficiency is found to be about 95% and in good agreement between data and simulation for electrons with $25 \leq p_{\mathrm{T}} \leq 500\,\mathrm{GeV}$ and $|\eta| < 2.2$ [222].

Tracking for muon candidates

Muon tracking in CMS employs three different techniques, with the results combined in one final collection. Information of the tracker and the muon spectrometer is combined, yielding a high purity and identification efficiency. Using the inner tracker, the momenta of muon tracks can be determined with high precision [180, 220, 223].

"Standalone muons" are reconstructed using muon spectrometer information only. These "standalone muon tracks" are obtained by combining DT, SCS, and RSC hits, and applying pattern recognition. If standalone muon tracks after extrapolation can be matched to tracks in the inner tracker, they are denoted as "inner tracks". In a global fit, "global muon tracks" are obtained using hits from the inner tracks and the standalone muon tracks as inputs. Especially for high momentum muons that penetrate multiple layers, this fit improves the track reconstruction significantly [220]. "Tracker muon tracks" are built from the inner tracks with large enough momentum, extrapolated to the muon system. If a match to hits in the muon system can be found, the object are identified as a tracker muon.

For muons with p_{T} smaller than 10 GeV, the tracker muon reconstruction yields better results since only one muon segment is needed. Often, muons are reconstructed both as global and tracker muon candidates. If they share the inner track, these can be merged, leading to a total reconstruction efficiency of 99% [220].

Calorimeter clusters

Clustering in the ECAL and HCAL serves multiple purposes. The momenta and energy of neutral particles such as photons and hadrons are measured, but also have to be separated from deposits of charged particles to avoid double-counting in the energy measurement. Thus, electrons and Bremsstrahlung photons have to be identified using combined cluster and track information, which also leads to an improved energy measurement of charged hadrons.

Clustering is performed separately in each subdetector. Starting from cluster seeds that are identified as cells with largest energy deposit, topological clusters are built. This is achieved by combining cells that show energy deposits larger than the noise level and close to the seed. Based on Gaussian mixture models, the energy and position of the cluster are fitted using an expectation-maximization algorithm [220]. This model characterizes the energy deposit in each cell of the topological cluster as the combination of several Gaussian distributed energy deposits of each seed. The energy determination of clusters is subject to a precise calibration of the hadronic and electromagnetic energy response of the calorimeters to maximize the identification efficiency and minimize the misidentification rate. Electromagnetic (hadronic) calibrations for the ECAL (ECAL+HCAL) are carried out separately using test-beam and actual collision data [220].

Link algorithm and particle identification

The "PF link algorithm" is used to combine the obtained tracks and clusters. Links between tracks and a calorimeter cluster are determined in the η–ϕ plane by extrapolating the tracks to the corresponding cluster. Cluster-to-cluster links are built if the cluster envelope in the less granular calorimeter overlaps with the cluster position in the more granular calorimeter. For each link found, a distance is defined for the pair that quantifies the quality of the link. Distances are defined in the η–ϕ (x–y) plane for track–calorimeter and ECAL–HCAL (preshower–ECAL) links. If multiple links are found, only the one with the smallest distance is kept. Links between tracks and muon spectrometer information are determined as a part of the muon reconstruction.

After all links have been established, "PF blocks" are constructed by combining PF elements with a link. Consequently, in each PF block all particles are identified and reconstructed, and their tracks, clusters, and links are removed from the block. Muons are identified first, due to high reconstruction efficiency and low misidentification rate using the muon system. Electrons and prompt photons are identified next, since they produce very similar signatures in the ECAL. Finally, neutral and charged hadrons, and non-prompt photons, e.g., originating from meson decays, are identified. Based on the PF reconstruction, an analysis-specific physical object selection can be performed on top of the PF objects, which is discussed in the following.

4.2 Physics objects

The primary vertex and event filters

The primary vertex (PV) of an event is defined by the vertex with the largest sum of associated transverse momenta squared determined by vertex reconstruction algorithms [185, 217], and needs to be reconstructed within a distance of 24 cm in z direction, and 2 cm in the x–y plane. Events not retaining such a vertex candidate are discarded for the measurement. The resolution of the primary vertex position ranges between 10 and 100 μm, while the efficiency is found to be better than 98% (99.8%) for vertices with two (more than two) associated tracks [185]. Also, several event quality filters are applied to the event selection in order to reject contaminated events [224]. The events with faulty detector signals caused by noise or anomalies are rejected,

e.g., those with spurious energy deposits in ECAL endcap crystals, cosmic muons, muons being misidentified as charged hadrons, or muons produced in scattering processes of the beam halo.

Muons

For the measurement presented in this thesis, muon candidates are required to have a transverse momentum larger than 15 GeV. Because the muon chambers extend only to a pseudorapidity range of $|\eta| < 2.4$, muons need to be reconstructed within this region. In addition, muons have to pass *tight* quality requirements as defined in Ref. [196]. These requirements allow suppressing the presence of fake muons or muons from in-flight decays and ensure a selection of prompt muons with high purity. An additional isolation criterion is applied to remove contributions from leptons originating from QCD multijet events. Isolated muon candidates are selected if they fulfill the condition $I_{\text{rel}} < 0.15$, where I_{rel} is defined as the sum of transverse energy deposits from charged and neutral hadrons and photons, relative to the p_{T} of the lepton, inside a cone in the η-ϕ space of $\Delta R < 0.4$ around the muon. Pileup corrections are also taken into account in the isolation calculation [225]. Additionally, muon momentum scale and resolution are corrected [226] based on calibration determined in $\text{Z}/\gamma^* \to \mu^+\mu^-$ events [227, 228].

Using *tight* identification criteria as defined above, for muons with $p_{\text{T}} > 20\,\text{GeV}$ from $\text{Z}/\gamma^* \to \mu^+\mu^-$ decays, the efficiency to reconstruct and identify muons is determined to be better than 96–97% for 2016 collision data with an excellent agreement between data and simulation [196]. The efficiency of the isolation requirement is found to be better than 95–96% [196].

Electrons

For electron candidates, the same conditions on the pseudorapidity ($|\eta| < 2.4$) and transverse momentum ($p_{\text{T}} > 15\,\text{GeV}$) are imposed as for the muon candidates. Additionally, the gap between the barrel and endcap region of the ECAL ($1.4442 < |\eta_{\text{sc}}| < 1.5660$) is excluded, where η_{sc} is the pseudorapidity of the electron supercluster. To ensure a high purity sample of prompt electrons, a set of *tight* identification criteria is applied as defined in Ref. [190]. Electron candidates originating from photon conversions are rejected. A similar isolation criterion as for muons is applied, where all particle momenta within a distance of $\Delta R < 0.4$ are considered. Depending on the electron p_{T}, the maximum allowed value for I_{rel} varies from 0.05 to 0.1. Additional corrections are made to remove contributions from pileup [190]. The momentum scale and resolution corrections [190, 222, 229] are applied.

For *tight* electron identification criteria, the efficiency in data is measured to be about 60% for $p_{\text{T}} > 20\,\text{GeV}$ and 80% for $p_{\text{T}} > 60\,\text{GeV}$ [190, 222].

Jets

Jets, produced as a result of the fragmentation and hadronization of partons, are used to probe the properties and kinematics of initial quarks or gluons. Jets are identified by using jet clustering algorithms, that are insensitive to soft emissions and collinear gluon splitting. The mentioned requirements are commonly referred to as infrared and collinear safety, and popular implementations are the k_{T} [161, 230, 231] algorithms. In CMS, the anti-k_{T} algorithm [161] interfaced to the FastJet package [232] is used to cluster PF candidates into jets.

The clustering is performed as follows. For each particle i in the event, the distance d_{ij} to all other particles j defined as

$$d_{ij} = \min\left(\frac{1}{p_{\text{T},i}^2}, \frac{1}{p_{\text{T},j}^2}\right)\frac{\Delta R_{ij}^2}{R^2}, \tag{4.1}$$

and the distance to the beam B as $d_{iB} = 1/p_{\mathrm{T},i}^2$ are computed. Here, R is a parameter steering the radius of the cone size of the jet. If $d_{ij} < d_{iB}$, the objects i and j are clustered into a jet by summing their four-momenta. Otherwise, i is identified as a jet. The algorithm proceeds iteratively until no particles are left, and all jets are identified. The final jet momentum is consequently defined as the sum of the constituents' momentum.

The distance parameter used in the measurement presented in this thesis is $R = 0.4$. In order to mitigate the effect of overlapping pileup contributions, particles not associated to the PV are excluded from the clustering procedure. This technique is referred to as charged-hadron-subtraction (CHS) [233].

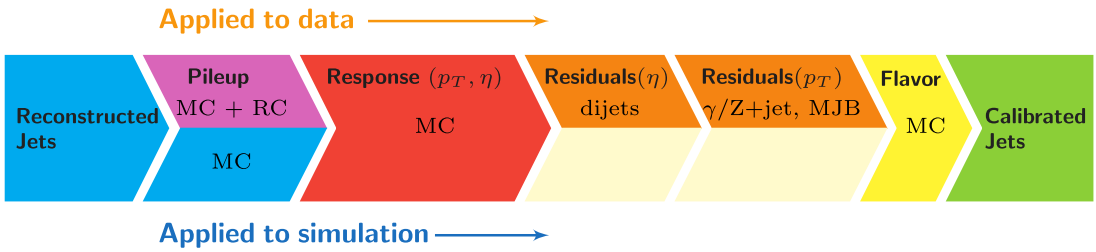

Figure 4.2: Illustration of the jet energy calibration procedure for data and simulation [234].

The calibration of the jet energy is subject to a set of corrections, that account for a manifold of different effects [234, 235], as illustrated in Figure 4.2. In a first step, pileup contributions from neutral particles are removed in offset corrections. The offset is estimated in a fit to account for the p_{T} and η dependence from the simulation of pileup events and is determined using the effective jet area and average energy density [232, 234–237]. The author of this thesis has contributed to improvements in the parameterization of the offset corrections, in order to simplify the fit model while keeping the same closure after subsequent response corrections. A dedicated correction is derived from data using the random cone method [235]. Subsequently, dedicated jet energy response corrections that correct for differences between the reconstructed and generated jet momenta are applied on jets in data and simulation as a function of the jet η and p_{T}. Finally, residual corrections are applied to jets in data as determined from QCD dijet, multijet, and Z+jets events [234, 238] based on momentum conservation, to correct for residual differences in the jet response corrections between data and simulation. In a similar manner, the jet p_{T} resolution is corrected in the simulation to match the one measured in data [234], as determined in Z+jets and γ+jets events.

A set of "tight" identification criteria are applied to all jet candidates to suppress contributions of jets emerging from noise or reconstruction failures [239], while also reducing the fraction of jets reconstructed from misreconstructed leptons. The efficiency for the selection is found to be better than 99% with a tag-and-probe technique using 2015 CMS pp collision events.

In order to further reject jets arising from pileup interactions, a dedicated pileup jet identification algorithm [240] is employed for jets with p_{T} of 30–50 GeV. It is based on an MVA approach and a working point with a misidentification rate of 40% for a fixed efficiency of 99% for jets with $|\eta| < 2.5$ is chosen [241].

The energy measurement of b jets is improved by about 6–12% using a deep-neural-network estimator [242], aiming to recover the energy carried away by neutrinos in leptonic b-hadron decays.

Jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and $|\eta| < 2.4$ are selected, and the η-ϕ distance ΔR to selected lepton candidates must be greater than 0.4.

Missing transverse momentum

Missing transverse momentum is the key observable to infer information about particles that leave the experiment undetected, such as neutrinos and potential BSM particles. Assuming that the transverse momenta of all partons taking part in the hard scattering process is negligible, the transverse momenta vector sum of all final state particles is assumed to be zero. Thus, the missing transverse momentum $p_{\mathrm{T}}^{\mathrm{miss}}$ can be defined [243] as the magnitude of the projection on the plane perpendicular to the beams of the negative vector sum of the momenta of all reconstructed particles in an event using the particle-flow algorithm:

$$p_{\mathrm{T}}^{\mathrm{miss}} = \left|\vec{p}_{\mathrm{T}}^{\mathrm{miss}}\right| = \left| - \sum_{i}^{N_{\mathrm{PF\ obj.}}} \vec{p}_{\mathrm{T,i}} \right|. \tag{4.2}$$

A similar relation cannot be made for longitudinal momentum conservation, because the initial momenta of the partons are unknown. The performance of $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction is sensitive to all reconstruction efficiencies and resolutions of other particles, and hence the whole PF algorithm. Additional effects, such as jet and lepton energy corrections, are propagated to the missing transverse momentum vector to avoid a bias in events with large hadronic activity.

As the calculation of the missing transverse momentum $p_{\mathrm{T}}^{\mathrm{miss}}$ is based on PF objects, it is also sensitive to pileup contributions. To reduce that dependence, the pileup-per-particle-identification (PUPPI) [244] algorithm is applied. PUPPI is an extension of the CHS algorithm where PF candidates are weighted by their probability to originate from the primary interaction vertex [224]. A comparison between the CHS and PUPPI approaches is shown in Figure 4.3.

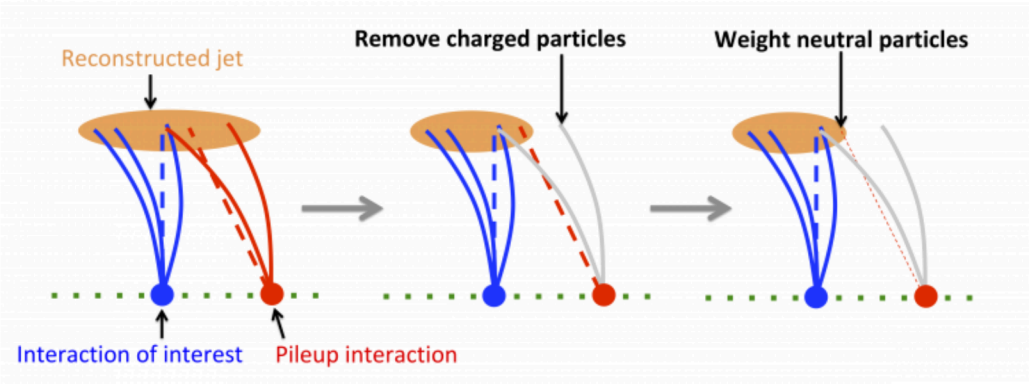

Figure 4.3: Illustration of pileup mitigation in jet reconstruction [245]. Pileup contributions are removed from jets by subtracting charged hadron energies (CHS, middle) or weighting the particle four momenta with their probability to originate from the primary vertex (PUPPI, right).

Using 2017 pp collision data, the performance of the $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction is assessed using Z/γ^{*} production with the Z boson or photon decaying leptonically [246]. Such events allow for a precise study as they are produced mostly without much additional hadronic activity and the leptons can be reconstructed with high accuracy and precision. Thus, the measured $p_{\mathrm{T}}^{\mathrm{miss}}$ originates solely from misreconstruction or miscalibration and can be studied by comparing $p_{\mathrm{T}}^{\mathrm{miss}}$ to the recoil of all (hadronic) objects besides the reconstructed Z boson candidate. The resolution is found to be about 25% for a boson p_{T} of 100 GeV and changes with increasing momentum to 50% for a p_{T} of 1 TeV [246]. The dependence on the number of pileup interactions is found to be more flat than for $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstructed using the CHS algorithm [224, 246, 247].

Heavy flavor jet identification

The identification of jets originating from c and b quarks (heavy-flavor jets) and light-flavor jets, originating from light-flavor quarks or gluons, is a very useful and powerful tool in high energy physics. Applications of b quark jet identification (b jet tagging) are often applied in analyses involving top quarks or Higgs bosons.

The idea of b jet tagging is based on the large mass of the b quark and, in turn, of the b-quark flavored hadrons, with masses of about 5 GeV and lifetimes of 1.5 ps [248]. The long lifetime leads to relatively long flight distances of these hadrons before decay, typically on the order of a few mm, resulting in a prominent secondary vertex (SV) in the detector and a significant displacement of associated tracks with respect to the jet axis. This is illustrated in Figure 4.4. Such displacement is investigated in heavy-flavor tagging algorithms. Further, by using leptonic observables in addition, a significant fraction of leptonic decays of b hadrons can be explored.

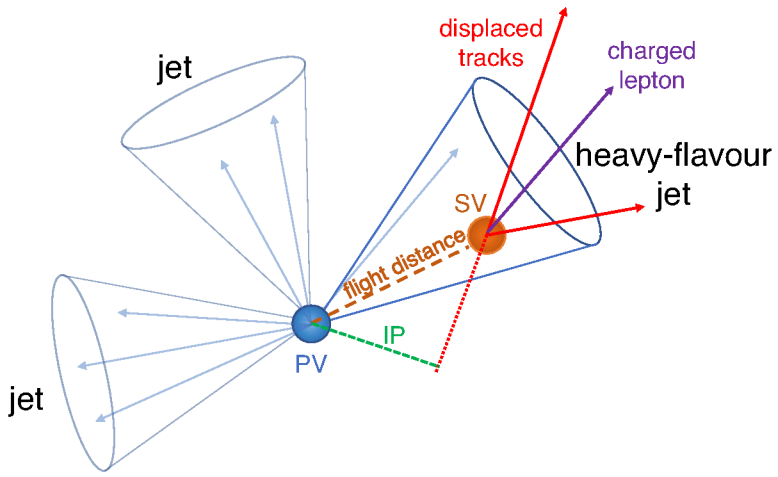

Figure 4.4: Illustration of a heavy-flavor jet including a secondary vertex (SV) from a decay of a b or c hadron that results in displaced tracks with respect to the primary vertex (PV) and a large impact parameter (IP) value [249].

The displacement of tracks is usually quantified by its impact parameter (IP), which is defined as the distance between the SV and PV at the point of closest approach, cf. Figure 4.4. It can be defined in two (2D) or three dimensions (3D). The sign of the impact parameter is defined such that it is positive if the angle between the jet axis and the IP is $< \pi/2$, or negative otherwise. For tracks originating from SVs from b hadron decays, large positive IP values are expected, while for light-flavor jet tracks the IP is distributed around zero according to the resolution of the detector. The significance defined as IP$/\sigma$ considers the uncertainty σ in the IP measurement. The 3D IP distribution for jets of different flavor is shown in Figure 4.5 (left).

In CMS, SVs are reconstructed using the inclusive vertex finder (IVF) [250]. Using all tracks with $p_{\mathrm{T}} > 0.8$ GeV and IP < 0.3 cm as inputs, the IVF algorithm builds track seeds according to 2D and 3D IP requirements. Next, clusters are built by testing the point of closest approach and angle between pairs of tracks, and the clusters are fitted for SVs using an adaptive vertex fitter (AVF) algorithm [251]. After the initial fit, SVs are removed from the collection of fitted vertices if they are found to be compatible with the PV, or if two or more SVs share tracks with each other. Additionally, tracks that are assigned both to the PV and a SV are removed from the SV to resolve any ambiguities. A second fit to determine the SV properties is performed using the same algorithm. Again, duplicates are removed from the obtained SV collection if they share, e.g., at least 20% of associated tracks. Finding a SV does not always succeed. For example, a jet

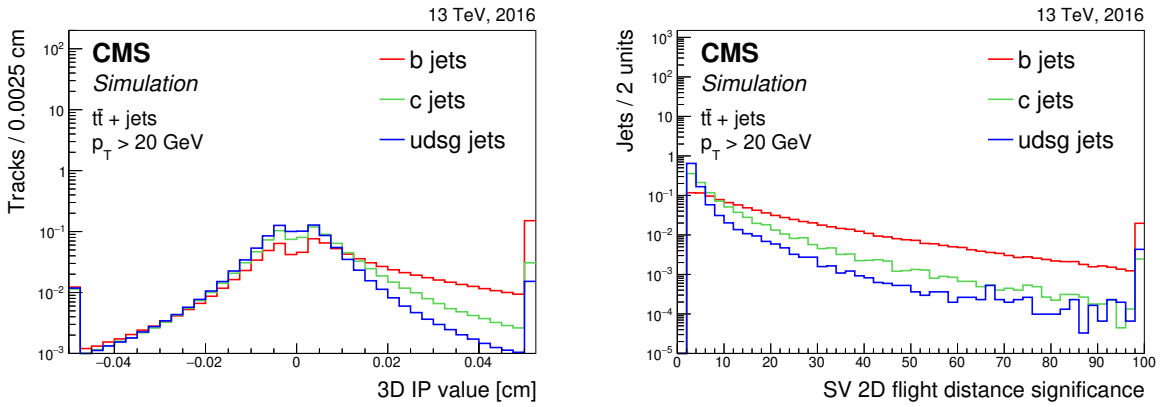

Figure 4.5: Distributions of the 3D impact parameter value (left) and secondary vertex flight distance (right) for tracks and vertices associated to light, c, and b jets in simulated $\mathrm{t\bar{t}}$ events [249]. The distributions are normalized to unit area.

can be too close to the PV, not enough tracks could be reconstructed, or they do not fulfill all selection requirements. The efficiency to find and reconstruct SVs is found to be 75% (37%) for b (c) jets in $\mathrm{t\bar{t}}$ events for jets with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ [249]. Properties or variables of SVs that help to discriminate between heavy and light flavor jets are for example the secondary vertex mass or its flight distance. The 2D (3D) flight distance is defined as the 2D (3D) distance between the PV and the SV. Analogously to the IP, a flight distance significance can be defined. Distributions for jets of different flavor for the 2D flight distance significance are shown in Figure 4.5 right.

Many of the variables described above could be directly used to infer the probability for a jet to originate from a b quark, but more sophisticated algorithms have been developed over time in CMS to maximize the tagging efficiency while keeping the mistagging rate at a minimum.

Tagging algorithms

Different tagging algorithms, often referred to as "taggers", are developed in CMS. In jet probability taggers (JP or JBP) [252] used during the LHC Run 1, the signed IP is used to construct a likelihood to originate from the PV. Although being less efficient as taggers developed later using MVA approaches, these still find application since they only rely on basic resolution curves for track parameters and are more robust compared to more complicated models and can be used as references.

The combined secondary vertex (CSV) [252] and CSVv2 [249] tagging algorithms combine information from tracks and SVs. In the following, only the CSVv2 algorithm is described, as it supersedes the CSV version. They are based on a NN, in particular a feed-forward multilayer perceptron with one hidden layer, cf. Section 5.2. The training data is split into three categories depending on the quality of the reconstructed SV, and the discriminator output of these trained algorithms is afterwards combined using a likelihood ratio. One tagger is trained to discriminate b jets with respect to c jets, and one with respect to light jets in a two-step training procedure. Their respective discriminator output is afterwards combined using a linear combination based on the flavor assumption in $\mathrm{t\bar{t}}$ events, yielding a binary classifier. For jets where input tracks do not satisfy quality criteria, a fixed output of -1 is assigned. The CSVv2 tagger was successfully used during LHC Run 2, especially for the pp collision data taken in 2016. Working points for taggers are commonly defined for a fixed light jet misidentification rate and are determined using simulation. In CMS, three working points are used, referred to as *loose*, *medium*, and *tight*, corresponding to light jet misidentification rates of 10%, 1%, and 0.1%. The b jet efficiencies for

the three working points are measured to be 81%, 63%, and 41% [249] for jets with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ in simulated $\mathrm{t\bar{t}}$ events. For c jet identification, the efficiencies are determined to be 37%, 12%, and 2.2%. Motivated by the exploration of deep machine learning algorithms for jet tagging [253],

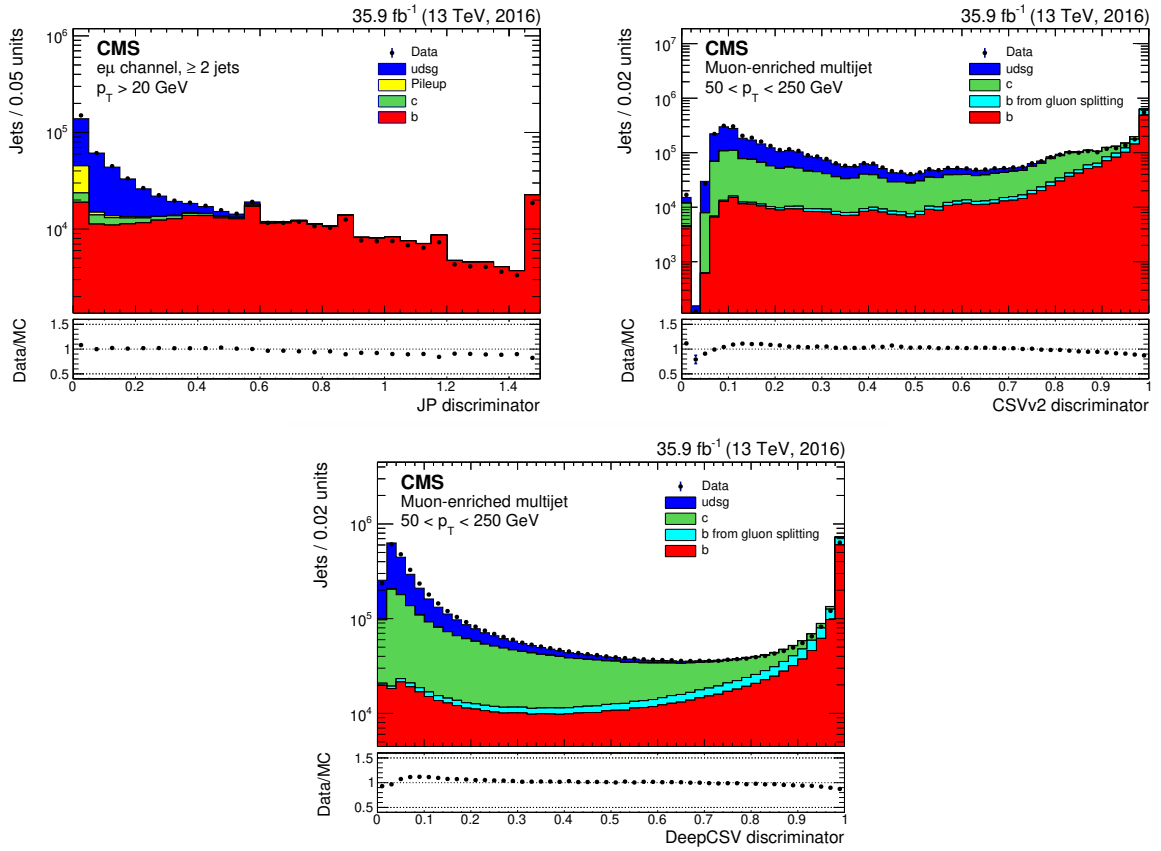

Figure 4.6: Distributions of the JP (upper left), CSVv2 (upper right), and DeepCSV (bottom) tagger outputs comparing data to simulation [249]. The JP distributions is obtained using $\mathrm{t\bar{t}}$ dileptonic events, while the CSVv2 and DeepCSV distributions are obtained using QCD multijet data and simulation.

the DeepCSV tagger is developed. It is based on CSVv2 but uses a deep neural network with more hidden layers and nodes per layer, and is trained simultaneously on all vertex categories and jet flavors. The same input variables are used for DeepCSV as for CSVv2 with the only difference being that only up to six tracks are considered. The output of DeepCSV consists of five values, corresponding to the output of five nodes in the last layer, that can be interpreted in terms of a probability for a jet to belong to a certain jet flavor category. The independent categories are defined for jets that contain exactly one b hadron, $P(\mathrm{b})$, at least two b hadrons, $P(\mathrm{bb})$, exactly one c hadron and no b hadron, at least two c hadrons and no b hadrons, or none of the mentioned possibilities. To discriminate heavy flavor jets from all other jets, i.e., c and light jets, the sum of the $P(\mathrm{b})$ and $P(\mathrm{bb})$ outputs has the best separation power. The output discriminant for data and simulation of the same taggers is shown in Figure 4.6 as obtained using $\mathrm{t\bar{t}}$ or QCD multijet events, respectively. Working points are defined according to a fixed light jet misidentification of 10%, 1%, and 0.1%. The b tagging efficiencies are measured to be 84%, 68%, and 50% using jets with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ from simulated $\mathrm{t\bar{t}}$ events [249]. For c jets, the tagging efficiencies are 41%, 12%, and 2.4%. The light flavor identification probability as a function of b tagging efficiency for the JP, CSVv2, and DeepCSV taggers are compared in Figure 4.7 (left).

Further improvement in tagging performance is achieved by the newly-developed DeepJet tagger [254]. In the DeepJet approach, two main changes with respect to previous taggers are combined. On the one hand, input variables are not subject to any kind of preselection and are combined with global event variables, which results in a more thorough description of the tagging problem. On the other hand, based on developments in machine learning, DeepJet uses a more advanced architecture.

The input variables to DeepJet include about 650 inputs, that can be associated to four distinct categories: the first one including global event variables such as jet kinematics or the primary vertex multiplicity, the second one based on 16 features of up to 25 charged PF candidates, the third one combining 6 inputs of up to 25 neutral PF candidates, and the last one based on 12 variables for up to 4 secondary vertices. PF candidate information includes, for example, track kinematics or displacement information, while SV variables include similar parameters as used for DeepCSV.

DeepJet uses a combination of recurrent layers in form of long-short-term-memory (LSTM) [255] designs and convolutional layers [256] individually for the charged and neutral PF, and SV input features. Using a feed-forward deep NN structure, the information is combined and fed to six output nodes. The motivation for this architecture is to reduce the dimensionality and complexity of the input variable space before feeding the information into the feed-forward part of the NN and to deal with variable input dimensions. The outputs represent probabilities for the jet to originate from a specific flavor category, allowing for b, c, and quark/gluon tagging. In detail, the orthogonal categories are defined for jets containing exactly one b hadron, at least two b hadrons, a leptonic decaying b hadron, at least one c hadron, coming from a gluon, or a light flavor (uds) quark. Using different combinations of outputs, classification of, e.g., b or c jets can be achieved. For b tagging, working points corresponding to fixed light jet misidentification rates of 10%, 1%, and 0.1% are defined, and the efficiencies are measured to be roughly 92%, 82%, and 65% [257].

The higher the b tagging efficiency for all misidentification efficiencies, the better the performance of the tagger. Thus, the light flavor identification probability as a function of the b tagging efficiency is the main performance indicator for the tagger. These are compared for DeepCSV and DeepJet in Figure 4.7.

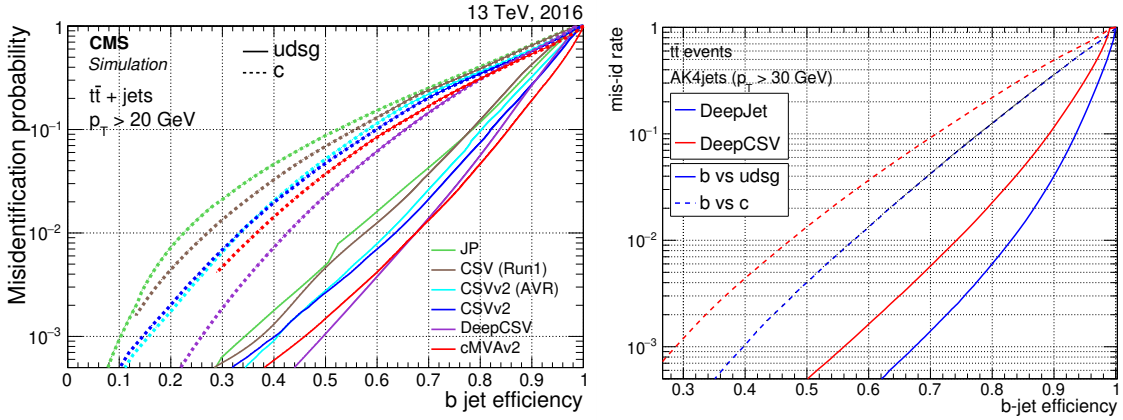

Figure 4.7: Misidentification probability for c and light jets versus b jet identification efficiency for various tagging algorithms used during the 2016 data taking (left) [249] and for DeepCSV and DeepJet (right) [254]. Both plots are obtained using $t\bar{t}$ simulated events.

4.3 Heavy flavor jet identification at the HLT

For measurements analyzing events with high b jet multiplicity, it is important to select events at the HLT with high efficiency. Both the CSVv2 and DeepCSV taggers are deployed at trigger level. Commonly, analyses with leptons and b jets in the final state, like the one presented in this thesis, use leptonic triggers to select signal events because of higher efficiency. However, measurements of processes expecting high b jet multiplicity, e.g., $HH \to 4b$ or $t\bar{t}H \to 4b + \text{jets}$, rely on hadronic triggers that make use of b tagging algorithms. During the Run 2 data taking, the CSVv2 tagger has been used at the HLT in 2016 and 2017, while the DeepCSV tagger has been fully deployed during the 2018 data taking.

For timing reasons, the HLT is not able to reconstruct all tracks and vertices (PVs and SVs), as done using the offline reconstruction. Thus, simplifications are made to the track, vertex reconstruction, and b identification. Two approaches for b tagging at HLT-level are employed. One uses jets that are clustered without the usage of PF, but using the calorimeter clusters only (called "calo jets & calo b tagging"), and another approach is closer to the offline reconstruction using PF jets. Calo b tagging is used to make fast decisions if a b jet could be contained in an event, and the selection can be refined using more accurate PF b tagging. In fact, for many trigger paths only calo b tagging is employed to assure maximal speed. Since both techniques make use of more restrictive selection criteria and simplifications compared to the methods used for offline analysis, lower tagging performance for highly displaced tracks and SVs is achieved.

Up to eight calo jets are considered in each event. In the first iteration, only pixel hit information are used, while in subsequent iterations all tracker hits are considered. After each iteration, the position of the PV is refitted, maximizing the resolution. Using the obtained track collection, SVs are constructed using the IVF algorithm.

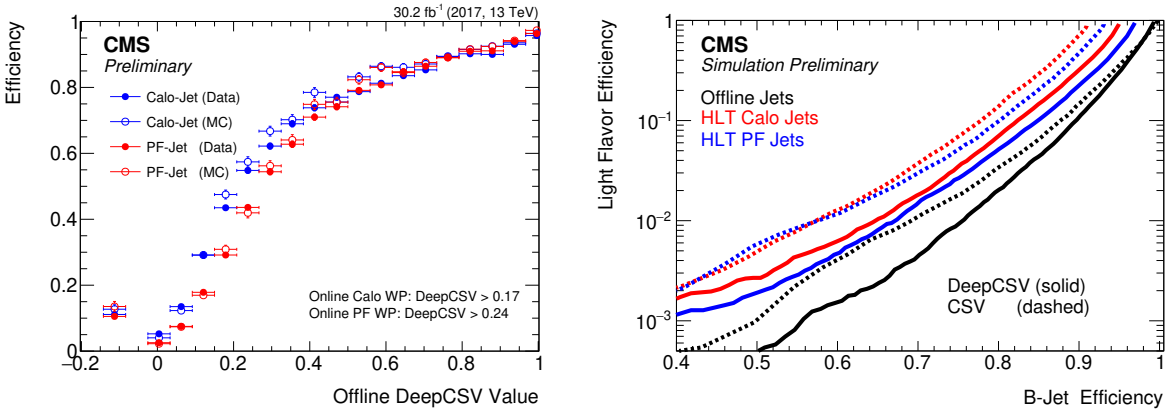

Figure 4.8: Efficiency to pass the online DeepCSV working points as given in the plot as a function of the offline DeepCSV value (left) for calo and PF jets [258]. Light flavor misidentification efficiency versus b tagging efficiency for CSV and DeepCSV and calo and PF jets (right) [258].

For PF b tagging, the procedure is very similar to the offline approach, but a simplified tracking is utilized to decrease the reconstruction time. The tagging performance of the calo and PF HLT b tagging algorithms is shown in Figure 4.8 in comparison to the offline approach. The left panel of Figure 4.8 shows the efficiency for a jet that passes the HLT DeepCSV working points as a function of the discriminator value. The studies are performed using collision data collected in 2018, and $t\bar{t}$ MC simulation. The calo b tagging approach has a similar efficiency compared to PF b tagging while not reaching the identical tagging performance by 1–10%. Both HLT algorithms do not reach the offline efficiency due to tracking inefficiencies at the HLT. The

right panel of Figure 4.8 shows the tagging performance of the CSVv2 and DeepCSV for calo and PF HLT and offline PF jets in terms of light flavor efficiency as a function of b jet efficiency. Similar observations as in the left panel of Figure 4.8 can be made when comparing CSVv2 and DeepCSV, but also the disadvantage of the calo b tagging is visible.

4.4 Developments for the Run 3 data taking and beyond

As preparation for the Run 3 data taking period of the LHC, started in 2022, at a center of mass energy of 13.6 and 14 TeV, multiple upgrades of the hardware and reconstruction software of the CMS experiment have been implemented [259]. Run 3 is expected to double the integrated luminosity in a similar amount of time with respect to Run 2. As a consequence, the instantaneous luminosity, and thus the pileup, increases. The average machine and beam conditions are similar to the one of the last runs in 2018. The tracking system, the hadronic calorimeter, and the muon system are upgraded to cope with the increased radiation and pileup. Additionally, the HLT computing farm is extended to use graphics processing units (GPUs) instead of central processing units (CPUs) to speed up the online event reconstruction. The ECAL, HCAL, and pixel tracker reconstruction are modified to make use of GPUs. In particular, the pixel-based tracking as part of the "patatrack" approach [260, 261] uses a heterogeneous architecture to efficiently reconstruct vertices and tracks. In the patatrack design, the complete procedure to obtain tracks and vertices from the raw data is constructed to be fully contained in modules that run on a GPU. This approach requires developments of a novel track and vertex reconstruction for Run 3, since the pixel tracks and pixel vertices can be reconstructed globally instead of using a regional approach based on jets as in Run 2.

In the new procedure, the globally reconstructed pixel tracks and vertices are directly used as seeds for the iterative tracking. Similarly, these collections can be used as a standalone track and vertex collection for PF reconstruction [262]. This concept is employed in the Run 3 scouting effort [262–264], where events are not further analyzed using the offline reconstruction but only the HLT information is kept. This significantly enhances the bandwidth available to store collected events. As a consequence of these developments, the calo b tagging approach had to be redefined and validated for the Run 3 data taking to exploit the new tracking efficiently. Similarly, studies are conducted to replace the default PF jet reconstruction and b tagging with a faster regional approach, and for the first time a dedicated recalibration of DeepCSV and DeepJet for the HLT is performed. The author of this thesis contributed significantly to the coordination and development of those efforts.

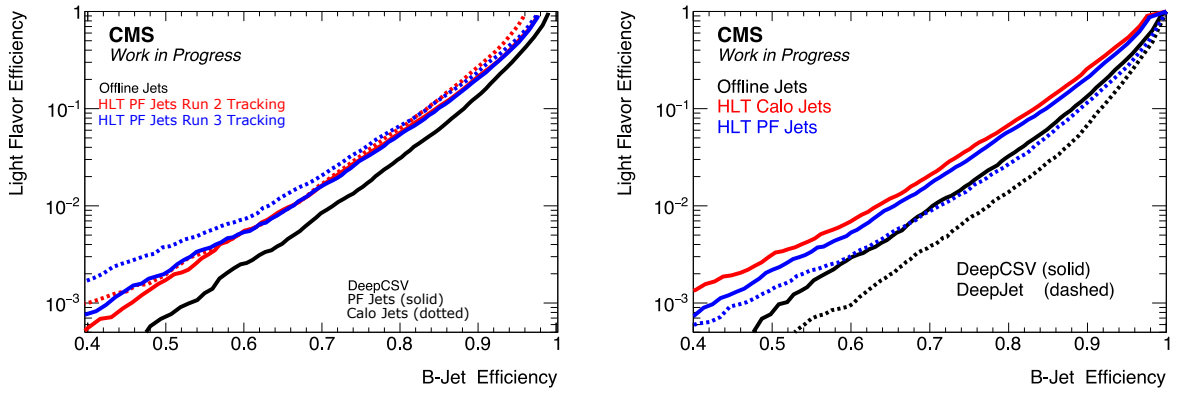

Figure 4.9: Light flavor misidentification efficiency versus b tagging efficiency for CSV and DeepCSV and calo and PF jets (right).

The new calo b tagging is based on the same approach that tracks, resulting from the iterative tracking, are geometrically matched to calorimetric jets. The difference for the Run 3 reconstruction is that all globally reconstructed pixel tracks are used as seeds for a single-iteration tracking procedure if they are in the vicinity of the eight most energetic calo jets. Afterwards, the geometric matching and the association of tracks to the jets is performed, giving rise to the necessary inputs to perform b tagging using DeepCSV. The standard iterative tracking used for PF reconstruction is also replaced centrally with a single-iteration tracking procedure seeded by pixel tracks. DeepJet cannot be employed as an alternative tagger using calo jets, since the PF algorithm is not used and thus no charged and neutral PF candidates are reconstructed. In Figure 4.9, the performance of the calo b tagging for the Run 2 and Run 3 scenarios is shown together with the performance of the offline and HLT PF reconstruction. Events from $t\bar{t}$ simulation are used, and jets are required to have $p_T > 25\,\text{GeV}$ and $|\eta| < 2.5$. The Run 3 performance for PF jets and calo jets using DeepCSV is found to be comparable, and a better efficiency for the calo b tagging is observed for larger light flavor efficiencies. Somewhat worse performance of the Run 3 tagging for very small b jet efficiencies can be explained by a poorer tracking performance of the single iteration approach. Additionally, statistical fluctuations due to small sample size can lead to distortions for small efficiencies.

For the first time, DeepJet at the HLT-level reconstruction is employed, using the offline calibration and the HLT inputs. A comprehensive performance comparison is shown in Figure 4.9 right for HLT calo jets and PF jets using DeepCSV, or alternatively, DeepJet. The same simulated samples and jet selection criteria as before are used. DeepJet yields a similar performance at the HLT-level as DeepCSV offline, although it is affected by significantly altered input distributions due to the HLT specific reconstruction. Since DeepJet relies on a large set of basic (low-level) inputs as track kinematic properties, it is particularly sensitive to changes in the reconstruction. Nevertheless, it yields a significant improvement in performance with respect to the previously used DeepCSV tagger. For a fixed light flavor mistagging rate of 10%, the b tagging efficiency improves by about 5% from 84% to 89%. Considering a light flavor mistagging rate of 0.1%, the improvement is on the order of 3% from 43% to 46%. The performance is expected to improve once DeepJet is retrained using the HLT input variables. Nevertheless, the potential of the DeepJet tagger to improve the b tagging at the HLT level is demonstrated.

In order to test the capabilities and limitations of the Run 3 global pixel tracking approach, additional tests are performed. The PF reconstruction is modified such that it uses non-globally reconstructed tracks based on the single-iteration tracking, but takes the outputs of the reconstruction used for calo b tagging, i.e., based on the eight highest energetic jets as inputs. This new regional approach is found to reduce the timing significantly, while a similar performance can be retained. In the following, this approach is referred to as "region-of-interest" (ROI). In Figure 4.10, the comparison of b tagging performance in terms of misidentification rate as a function of b tagging efficiency is shown for DeepCSV and DeepJet. Again, event and jet selection are identical to those in the previous studies. For DeepCSV, the ROI PF reconstruction leads to the same performance as when using the global PF reconstruction, while for DeepJet the performance is improved. The reason is a better description of the input distributions for PF candidates when using the ROI approach. Also, the key observables like the IP or the second vertex multiplicity, are found to be in better agreement with the offline reconstruction for the ROI than for the global-tracking approach. When employing a dedicated recalibration, both for the global and ROI reconstruction approaches, the same performance improvement is expected for both reconstruction approaches. For Run 3, the ROI approach is not yet employed, since detailed follow-up studies are needed to ensure the same efficiencies as in Run 2 for a variety of processes. However, the presented studies hint that ROI-based b tagging can be a beneficial

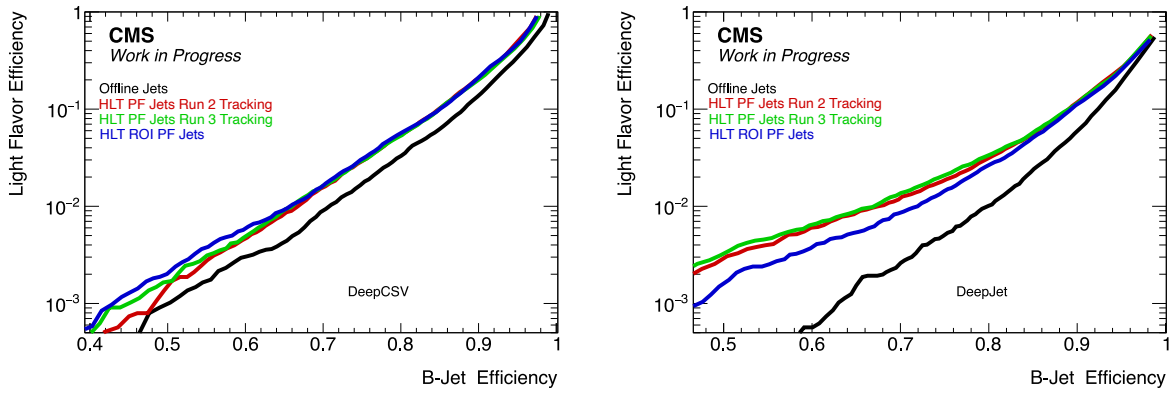

Figure 4.10: Light flavor misidentification efficiency versus b tagging efficiency for CSV and DeepCSV and calo and PF jets (right).

extension of the CMS scouting campaign, or can eventually be used for the HLT PF reconstruction to decrease the timing, while retaining identical performance. Future investigations targeting the definition of the geometrical regions based on calo jet seeds are expected to further improve its performance. The deployment of the ROI-based tracking and PF reconstruction will consequently lead to a significant acceleration of the HLT reconstruction while retaining similar efficiencies.

Since the offline calibration of DeepCSV is observed to lead to reduced tagging performance when evaluated using HLT inputs, for the first time, a dedicated recalibration is performed [47]. The identical training procedure as used for the offline DeepCSV and DeepJet calibration is employed [249, 254]. The training data is composed of simulated MC samples for $\mathrm{t\bar{t}}$ production, QCD multijet production in different jet p_T ranges, and Di-Higgs-boson (HH) production via the gluon-gluon-fusion and vector-boson-fusion processes. Jets with $p_\mathrm{T} > 30$ and $|\eta| < 2.5$ are selected to derive independent training, validation, and testing samples, cf. Section 5.2. The training set is augmented to achieve the same p_T and η distribution for all jet flavor categories to prevent the taggers from learning any flavor-dependence using these variables. Otherwise, the training could fail to generalize using events kinematically different from the training set. In total, about 40 million jets are used for the training. The performance is studied using jets satisfying the same p_T and η requirements as given above in $\mathrm{t\bar{t}}$ events, and is compared to the offline and HLT performance using the offline calibration. The light flavor misidentification rate as a function of b tagging efficiency is shown in Figure 4.11 (left). Using the recalibration, a decrease in light-flavor mistagging rate can be achieved for all b tagging efficiencies for both DeepCSV and DeepJet. The increase is more significant for the DeepJet tagger because it is more sensitive to basic input features and their correlations as compared to DeepCSV. However, both taggers are not able to reach the offline tagging performance due to HLT reconstruction inefficiencies. The retrained DeepJet tagger is able to surpass the offline DeepCSV performance. For a fixed working point of a light-flavor mistagging rate of 10%, the retrained DeepCSV (DeepJet) tagger increases the absolute b tagging efficiency by about 1.5 (2)%. For a tight working point with a light-flavor mistagging rate of 0.1%, the b tagging efficiency improves by about 4 (12)%.

Further, the charm jet rejection is studied. In Figure 4.11 (right) the c jet misidendification is shown as a function of b jet identification efficiency comparing the performance of online (HLT) and offline DeepCSV and DeepJet reconstruction for Run 2 and (retrained) Run 3 scenarios. In case of DeepCSV, the retrained algorithm shows a worse charm jet rejection as when using the offline training. This can be expected as the number of b and light flavor jets in the training

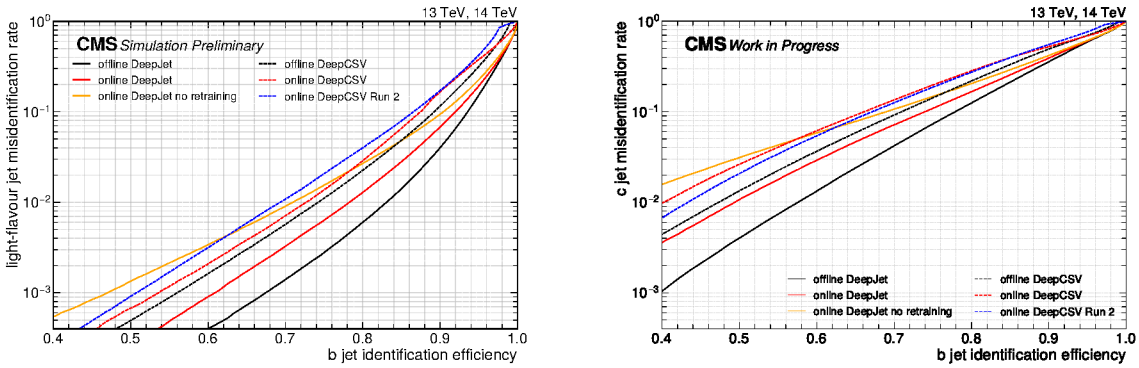

Figure 4.11: Light flavor (left) and c jet (right) misidentification efficiency as a function of the b tagging efficiency for CSV and DeepCSV and calo and PF jets [47].

set is much larger than the charm jet abundance. Given the training setup and the weighting of the categories, this effect can be mitigated by increasing the c jet sample in the training. In case of DeepJet, the charm rejection improves by almost an order of magnitude for all b tagging efficiencies.

Developments for the High Luminosity LHC

A detailed study is performed by CMS to develop a simplified HLT menu with new reconstruction algorithms and trigger paths for the HL-LHC upgrade, cf. Sec 3.3, to demonstrate that the timing and rate requirements can be met [48]. The author of this thesis has significantly contributed to the presented effort and led the studies for b tagging algorithms at the HLT presented in the Technical Design Report [48].

A new HLT tracking configuration is developed based on the iterative tracking approach, extending the track reconstruction to $\eta = 4.0$ and aiming to reconstruct tracks from all vertices. This approach is referred to in the following as "baseline". To further reduce the reconstruction time by about 30–40%, a "trimmed" approach is used, where constraints on the association of track candidates to the PV are imposed. The two approaches achieve the same tracking efficiency for the *pileup 140* scenario of more than 80%, while for a *pileup 200* scenario the baseline approach is a few percent better. Misidentification rates are also observed to be similar [48].

For jet object building, the PF algorithm is used. It is necessary to apply pileup mitigation techniques to reach optimal performance for jets at the HL-LHC, and PUPPI is currently the mitigation technique that works well for *pileup 140* and *200* scenarios [48].

The HLT b jet identification is based on the DeepCSV algorithm as its performance has been observed to be robust during the Run 2 data taking. Studies for DeepJet are also presented. The input definition and selection is the same as described in Section 4.2 with an additional $p_T > 1.4\,\text{GeV}$ requirement for the track selection. The reason for this raise is a speed increase of about 60% while keeping the same b tagging efficiency at only a cost of 5% increased light flavor mistagging rate. Also, SVs are reconstructed using the IVF algorithm as described in Section 4.2. Important observables, i.e., the 3D IP significance and SV multiplicity, for b tagging are shown in Figures 4.12 and 4.13 for both the "baseline" and "trimmed" tracking configurations compared to using the default offline tracking and PF reconstruction. The observables are also shown separately for light flavor and b jets. The "baseline" configuration shows a performance similar to the offline one, while the "trimmed" version shows inefficiencies in the negative and positive tails of the IP3D distribution. With the current b tagging algorithms, this leads to significant

performance loss when using the "trimmed" setup. Similar observations can be made for the SV finding probability.

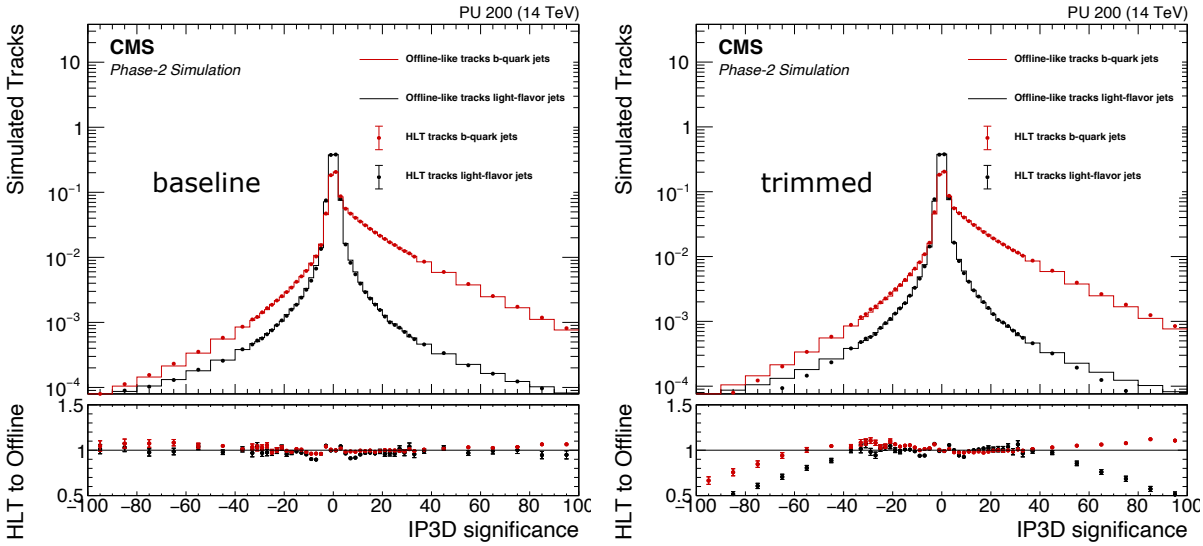

Figure 4.12: Left: comparison for the 3D IP distribution for tracks reconstructed with the "baseline" and offline tracking configurations. Right: the same distribution is shown, but for the "trimmed" setup. Tracks of light flavor (b quark) jets are shown in black (red). The figures are taken from Ref. [48] (modified).

The tagging performance of DeepCSV and DeepJet is measured using simulated $\mathrm{t\bar{t}}$ events for jets with $p_\mathrm{T} > 30\,\mathrm{GeV}$ and $|\eta| < 2.4$ in the *pileup 200* scenario as shown in Figure 4.14 (left). Additionally, the different tracking approaches are compared. In the case of DeepJet, the tagger suffers significantly from the fact that it heavily relies on usage of lower-level inputs as PF candidates and the calibration of the tagger is derived using the offline tracking configuration. Due to the tracking inefficiencies for the "trimmed" approach, the properties of charged and neutral PF candidates are fundamentally different compared to those in the offline reconstruction. In the following discussion on the trigger paths, only DeepCSV is considered as a tagger for the Phase-2 HLT.

The success of the heavy flavor tagging performance at the Phase-2 CMS experiment for jets with $|\eta| < 4$ is shown in Figure 4.14 (right). The performance for DeepCSV is shown separately for jets reconstructed in the barrel ($|\eta| < 1.5$), the HGCAL ($1.5 < |\eta| < 3$), and forward region ($|\eta| > 3$).

To demonstrate the impact of the reconstruction performance on physics analyses, several trigger paths are derived and tested. One critical application for b jet triggers will be the study of $\mathrm{HH} \to 4\mathrm{b}$ events. Candidate events are defined at the generator level by selecting four b jets with $p_\mathrm{T} > 30\,\mathrm{GeV}$ and $|\eta| < 2.4$. At the reconstruction level, events are selected at L1 using hadronic activity ($H_\mathrm{T} > 450\,\mathrm{GeV}$) and jet multiplicity ($p_\mathrm{T} > 70, 55, 40, 40\,\mathrm{GeV}$) requirements. The hadronic activity H_T is defined as the scalar sum of the transverse momenta of all selected jets with $p_\mathrm{T} > 30\,\mathrm{GeV}$ and $|\eta| < 2.4$ in an event. The output rate of this L1 trigger is 10 kHz. Two different HLT triggers are considered. The first trigger version implies conditions used in Run 2, with requirements of $H_\mathrm{T} > 330\,\mathrm{GeV}$ and $p_\mathrm{T} > 75, 60, 45, 40\,\mathrm{GeV}$. At least three jets are required to be b tagged. The second trigger version reduces all kinematic cuts to $H_\mathrm{T} > 200\,\mathrm{GeV}$ and $p_\mathrm{T} > 70, 40, 30, 30\,\mathrm{GeV}$ while relying mostly on b tagging of three jets ("b-tag only"). The working points for DeepCSV are adjusted such that the output rate of the trigger paths

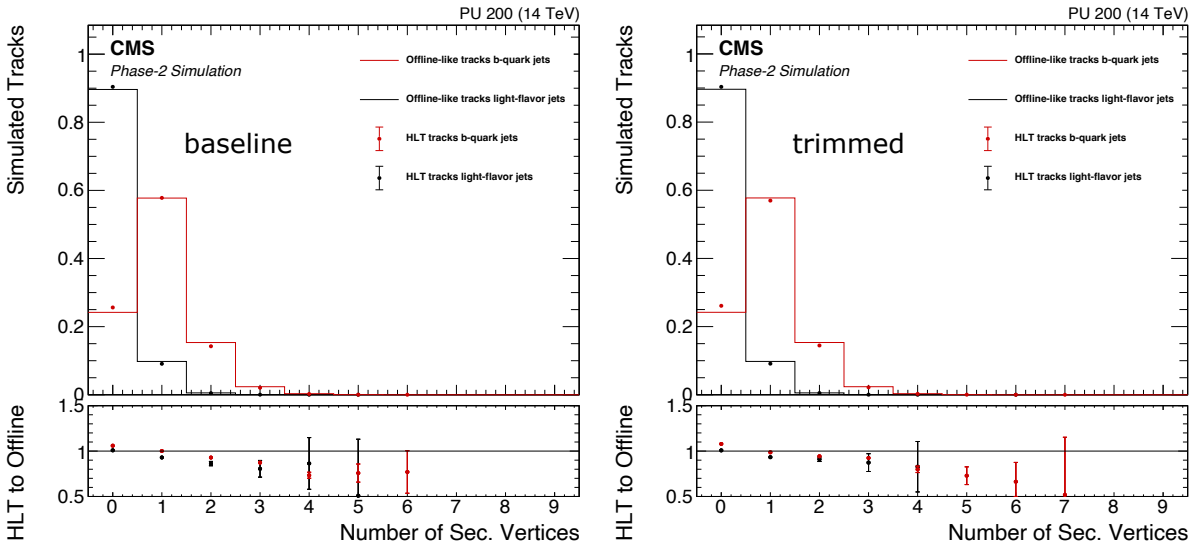

Figure 4.13: Left: comparison for the number of reconstructed SVs with the "baseline" and offline tracking configurations. Right: the same distribution is shown, but for the "trimmed" setup. SVs of light flavor (b quark) jets are shown in black (red). The figures are taken from Ref. [48] (modified).

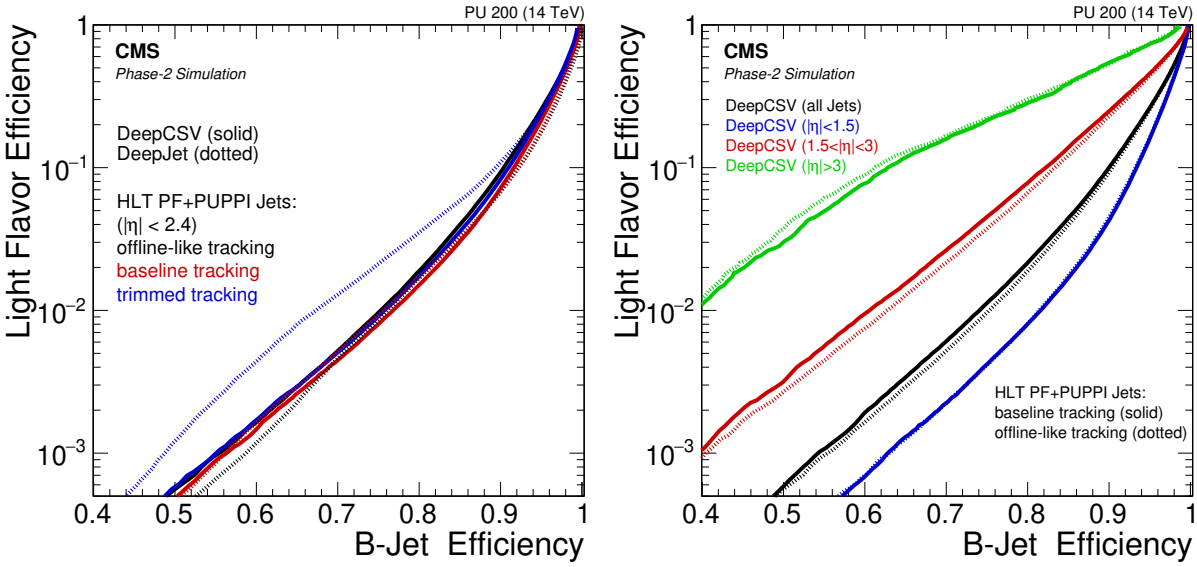

Figure 4.14: Left: b tagging performance of DeepCSV and DeepJet for different HLT tracking scenarios for jets with $|\eta| < 2.4$. Right: performance in different pseudorapidity regions. For both plots, jets are required to have $p_{\mathrm{T}} > 30\,\mathrm{GeV}$. The figures are taken from Ref. [48].

corresponds to 50 or 75 Hz, respectively. The efficiencies of the paths are shown in Figure 4.15. In the left panel of Figure 4.15, the invariant mass of the double-Higgs-boson system (m_{HH}) is shown for all generated events, selected by the L1, and the L1+HLT trigger paths. Similarly, the right panel of Figure 4.15 shows the L1+HLT selection efficiency with respect to all generated events as a function of m_{HH}. The conclusions is that kinematic cuts can be relaxed to efficiently select more HH events when using the "b-tag only" path. The trade-off is that b tagging, and thus also tracking, has to be performed at a higher rate (3.2 vs. 10 kHz) leading to an increased timing. The Run-2-like path is thus a compromise between timing and efficiency, and there is a clear dependence of b tagging on the available CPU or GPU capacity.

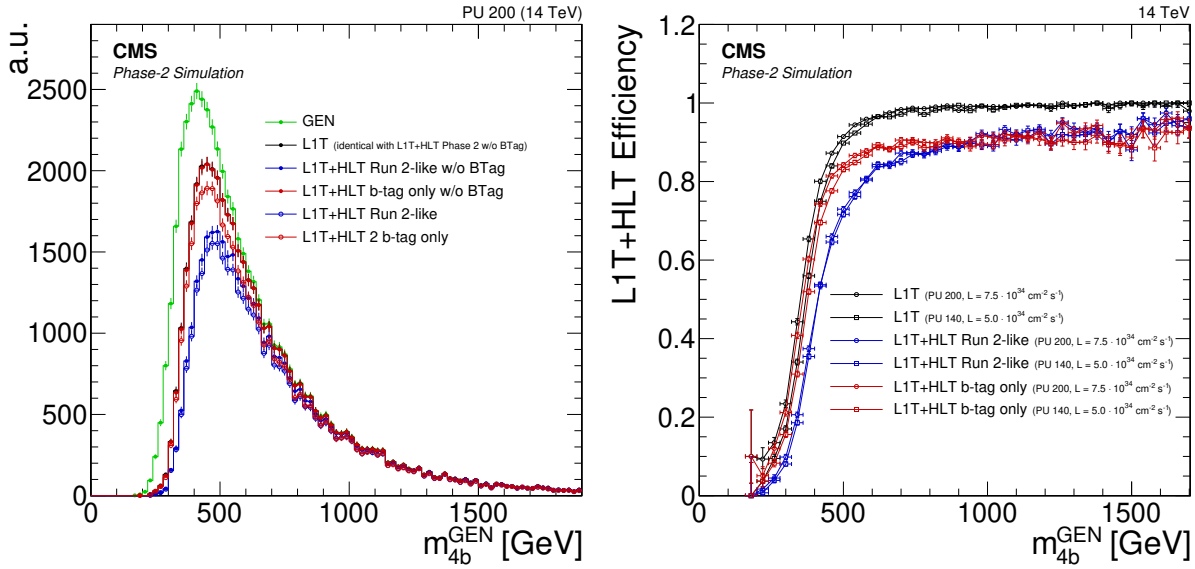

Figure 4.15: Left: invariant mass distribution for all generated, selected at L1, and HLT events for two trigger scenarios as explained in the text. Right: corresponding selection efficiencies are shown with respect to the generator level. The figures are taken from Ref. [48].

5 Analysis methodologies

In this chapter, the basics of the statistical methods, used in the presented analysis, are described. The first section introduces the main statistical methods needed for the $t\bar{t}$+jet cross section measurement and the extraction of the top quark mass. The techniques of parameter estimation, and the concept of goodness-of-fit tests relevant for the validation of input variables used for the neural network training are discussed. In the second section, the concept of machine learning and neural networks is presented. It comprises mathematical concepts, design strategies, and training techniques. The ideas of classification and regression approaches are also discussed.

5.1 Statistical methods

In this section, the most relevant statistical ingredients of the analysis, such as parameter estimation, often referred to as "fitting", and the concept of goodness-of-fit tests, are explained. In particular, the work of this thesis makes use of binned maximum likelihood and a χ^2 fit. The maximum likelihood fit is used to perform the binned cross section measurement, while the χ^2 fit is used to extract a value for the top quark pole mass from the measured differential cross sections.

Parameter estimation

The measurement of a parameter with unknown value based on a set of a number of experimental observations is often called parameter estimation [265]. For a set $\vec{x} = (x_0, \ldots, x_n)$ of N independent observations that share the same underlying probability density function depending on the parameter θ, the estimator of θ is denoted $\hat{\theta}(\vec{x})$. The numerical value of $\hat{\theta}(\vec{x})$ is called the estimate. The estimator $\hat{\theta}$ is itself a random variable with the sampling distribution $g(\hat{\theta}, \theta)$ with expectation value $E(\hat{\theta})$ and variance $V(\hat{\theta})$. For example, the data $\vec{x}$ could be measured events at the LHC as a function of some variable, and θ could be the cross section of a process. The goal of the analysis would then be to estimate the most likely value for θ with a corresponding uncertainty using an estimation method $\hat{\theta}(\vec{x})$. Important properties of estimators are discussed the following.

Consistency An estimator can be called consistent if its estimate $E(\hat{\theta})$ converges towards the true value θ once the number of measurements N increases.

Bias The bias b of an estimator is defined as the deviation of its expectation value $E(\hat{\theta})$ from the true value θ_0 of θ:

$$b_N = E(\hat{\theta}) - \theta_0 = E(\hat{\theta} - \theta_0). \tag{5.1}$$

Bias indicates the systematic deviation of an estimator and an estimator is called "unbiased" if $b = 0$ for any N or θ_0. Note that a consistent estimator can also be biased, or vice versa. Small or negligible biases are preferred, while constant or known biases could be removed from the estimate.

Variance The variance $V(\hat{\theta})$ can be understood as the "statistical" uncertainty of an estimator. Similar to the bias, smaller variances are preferred when choosing an appropriate estimator. Variance also corresponds to the "efficiency" of an estimator and increases the confidence that $\hat{\theta}$ is a good estimate of θ for specific N.

Commonly, estimators that have both zero bias and minimum variance are considered optimal, while usually there is a trade-off between the mentioned properties. For example, an estimator which is biased with $b > 0$ for all N but $b \to 0$ for $N \to \infty$ and has a variance $V \to 0$ for $N \to \infty$ is consistent. On the other hand, an unbiased estimator for which the variance is finite for $N \to \infty$, is inconsistent. Two of the most usual estimation methods used in high energy physics are the method of least-squares and maximum likelihood, of which the latter is described in the following section.

The minimum χ^2 method

One prominent estimator is the minimum χ^2 method. It shares many properties with the maximum likelihood method discussed below and is an extension of the χ^2 goodness-of-fit test [266–268]. The test statistic χ^2 for a set of N independent observations $\vec{x_i} = x_1, x_2, \ldots, x_N$ that are distributed according to a Gaussian distribution with variance σ_i and expected values $\vec{y} = y_1, y_2, \ldots, y_N$ depending on θ can be written as

$$\chi^2 = \sum_{i=1}^{N} \frac{(x_i - y_i(\theta))^2}{\sigma_i^2}. \tag{5.2}$$

When compared to the χ^2 goodness-of-fit test, essentially a test for the null hypothesis that the observation X_i equals Y_i is performed. The resulting sampling distribution of the test statistic follows the χ^2 distribution

$$p(x) = \frac{x^{\frac{k}{2}-1} \mathrm{e}^{-\frac{x}{2}}}{2^{\frac{k}{2}} \Gamma(\frac{k}{2})} \tag{5.3}$$

with $\Gamma(\frac{k}{2})$ being the Γ function and k the number of degrees of freedom (dof). Thus, when varying $\vec{y}(\theta)$, a best fit value for θ can be found where the corresponding χ^2 is minimal. Since the mean of the χ^2 distributions equals k and its variance $2k$, a good best fit value is usually found when $\chi^2_{\mathrm{min}} \approx k$ or also $\chi^2_{\mathrm{min}}/k \approx 1$. For example, an appropriate 68% confidence level (CL) can be found at $\chi^2 = \chi^2_{\mathrm{min}} + 1$ [269] in case of a one-parameter fit. In case of correlated observations, the definition of χ^2 can be rewritten as

$$\chi^2 = \mathbf{\Delta}^T \mathbf{V}^{-1} \mathbf{\Delta} \tag{5.4}$$

with $\mathbf{\Delta} = (x_0 - y_0, \ldots, x_N - y_N)^{\mathrm{T}}$ and $\mathbf{V}$ being the covariance matrix of the correlated data points.

The maximum likelihood method

For a set of N independent observations $\vec{x}$, i.e., $x_1, \ldots, x_N$, with a probability density function $f(\vec{x}|\theta)$ that depends on θ, which can be a single parameter or a set of parameters, the joint probability density function can be written as

$$\mathcal{L}(\vec{x}|\theta) = P(\vec{x}|\theta) = P(x_1, \ldots, x_N|\theta) = \prod_{i=1}^{N} f(x_i|\theta). \tag{5.5}$$

For a given set of observed data x_0, $\mathcal{L}(x_0|\theta) = \mathcal{L}(\theta)$ is called the likelihood function which now only depends on θ.

The values $\hat{\theta}$ for θ where $\mathcal{L}$ reaches its maximum are called maximum likelihood estimators of θ, given the observed data x_0. The condition for the existence of a maximum of $\mathcal{L}(x_0, \theta)$ is called likelihood equation:

$$\frac{\partial \mathcal{L}}{\partial \theta}(x_0, \theta) = 0. \tag{5.6}$$

Compared to the minimum χ^2 method, the advantage of the maximum likelihood method is that it works for data distributed according to any probability density function. Often, instead of maximizing the likelihood, the negative log likelihood is minimized equivalently:

$$-\ln \mathcal{L}(\vec{x}, \theta) = \sum_{i=1}^{N} \ln f(x_i, \theta). \tag{5.7}$$

Since for random variables distributed according to a Gaussian probability density function $f(\vec{x}|\theta)$ the negative log-likelihood function agrees with the χ^2 function with a constant offset and a factor 2 in difference, commonly $-2 \ln \mathcal{L}$ is minimized instead.

Applying the likelihood equation to Equation 5.7 yields:

$$\frac{\partial}{\partial \theta} \sum_{i=1}^{N} \ln f(x_i, \theta) = \frac{\partial}{\partial \theta} \ln \mathcal{L}(x_0, \theta) = 0. \tag{5.8}$$

The maximum likelihood estimator itself is a consistent estimator, but not necessarily unbiased. However, it is asymptotically unbiased for a sample size $N \to \infty$.

In presence of large data sets and complex probability density functions, the computation of the log-likelihood function and its derivative can be very complicated. One solution is to perform the measurement on binned data. The number of events per bin follows a Poisson distribution and $\mathcal{L}$ can be rewritten [265, 270] as

$$\mathcal{L} = \prod_{i=1}^{N} \frac{\mathrm{e}^{-v_i} v_i^{n_i}}{n_i!}, \tag{5.9}$$

where i represents the number of the bin, and v_i and n_i denote the expected and observed number of events for each counting experiment. The index i in this case can describe one bin of a single distribution, or even one bin of a multidimensional distribution in various categories or measurements of multiple experiments. In the context of the analysis addressed in this thesis, the information entering v_i contains details about production, decay and background composition, and experimental features as integrated luminosity, acceptance and efficiencies,

as well as calibrations. A common way to implement this information is based on templates determined from a MC simulation.

The expectation model is associated with a number of uncertainties, both systematic and statistical ones, which can be incorporated in the likelihood function as nuisance parameters (NP) [271]. The prior knowledge on these is taken into account by the prior probability density functions that are based on experimental measurements or theoretical rationale. Such, the likelihood function can be extended as

$$\mathcal{L} = \prod_{i=1}^{N} \frac{\mathrm{e}^{-v_i} v_i^{n_i}}{n_i!} \prod_j \pi(\lambda_j), \tag{5.10}$$

where λ_j are the NPs and $\pi(\lambda_j)$ are the prior probability functions modeling the dependence of the expectation on the NPs. Depending on the origin and effect of the associated uncertainty, different bins and processes can be affected with varying intensity. This leads to different implementations and probability models, where for the analysis presented in this thesis three main classes are important: rate-changing NPs, shape- and rate-changing NPs, and statistical uncertainties of simulated events.

Rate-changing nuisance parameters Many uncertainties, such as the uncertainty in the total integrated luminosity or theoretical prediction for the signal cross section, affect all bins of one or more processes identically, affecting only normalization but not the shapes. It is natural to model such nuisance parameters with Gaussian probability density functions,

$$p(\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\theta-\tilde{\theta})^2}{2\sigma^2}\right), \tag{5.11}$$

with σ being the width corresponding to a relative uncertainty at 68% CL. The disadvantage is that the Gaussian model can lead to negative values for positive defined observables like event counts or cross sections once the uncertainties are large. One solution would be to truncate the Gaussian distribution at zero. A better approach is to use log-normal probability density functions, defined as

$$p(\theta) = \frac{1}{\sqrt{2\pi}\ln(\kappa)} \exp\left(-\frac{(\ln(\theta/\tilde{\theta}))^2}{2(\ln(\kappa))^2}\right) \frac{1}{\theta}, \tag{5.12}$$

with κ being the width of the distribution. In Figure 5.1, the Gaussian and the log-normal probability density functions for different parameter sets are compared. The log-normal function does not allow for negative values of x.

Shape- and rate-changing nuisance parameters In contrast to the type of uncertainties described above, a variety of uncertainties affect correlated bins with different strength, changing the rate and the shape of a distributions. These uncertainties are usually taken into account by deriving additional templates according to the variation of the corresponding uncertainty up and down by one standard deviation. The implementation in the likelihood function however requires a continuous function rather than a discrete function at two points. To solve this, usually a template "morphing" procedure is employed where the interpolation as a function of the nuisance parameter follows a quadratic-linear approach [271]. Hence, the dependence is modeled quadratically (linearly) within (beyond) one standard deviation in a continuous and differentiable way. To retain the correct normalization under template morphing, templates are first normalized before morphing, and are rescaled before and afterwards by the contribution of the content of each bin to the integral of the template. The normalization is modeled by a log-normal probability density function. Special attention has to be paid to the population of bins, because due to a finite number of MC-generated events the morphing algorithm can

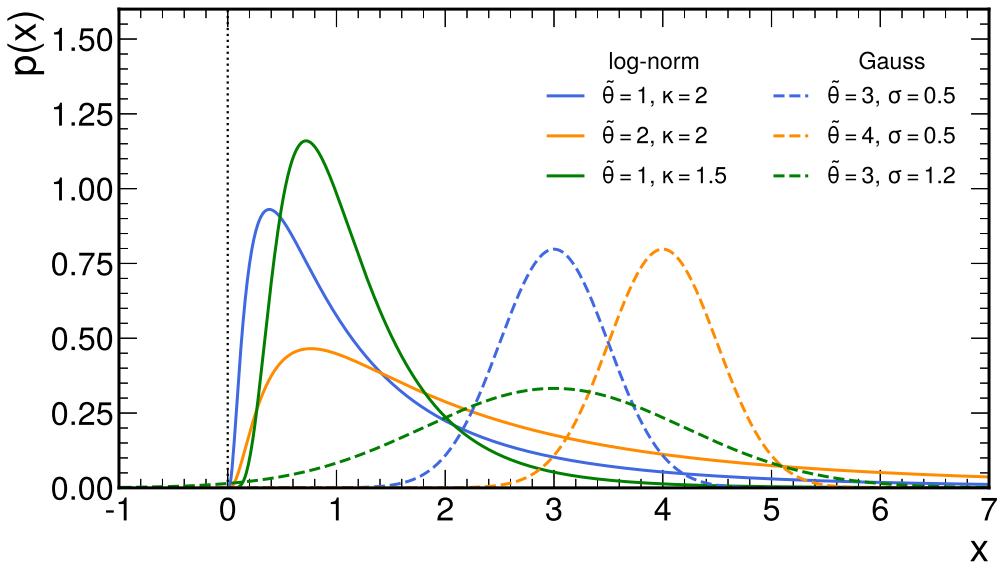

Figure 5.1: Examples for the log-norm (solid) and Gauss (dashed) probability density functions. For each class of functions, three different curves for varying parameter choices are shown, indicating the effect on the shape of the distribution.

predict negative yields for specific values of the nuisance parameter. Thus, it is sensible to choose adequate bin widths.

Statistical uncertainties of simulated events The templates are usually obtained using the MC simulation, consequently leading to two effects. Firstly, in case of limited MC statistics, each bin in a template is associated with a statistical uncertainty based on the number of simulated events in each bin. The proper way to treat this additional uncertainty in the likelihood function is to implement the corresponding single nuisance parameter modeled with a Gaussian or Poisson prior per process and bin. As a result, the number of such nuisance parameters increases significantly and complicates the minimization process. This is partially related to the strong (anti-) correlations of all parameters within a bin. A solution for sufficiently populated bins is to approximate all contributions to each bin with a single Poisson or Gaussian distribution. This is often referred to as the "Barlow-Beeston"-lite approach [272, 273], leading to N additional nuisance parameters in the likelihood function. Secondly, the template morphing itself is subject to statistical bin-by-bin fluctuations. Using toy experiments, the impact of the statistical uncertainty in the fit results can be fully evaluated [42, 274].

The profiled maximum likelihood fit

In order to determine the estimates for the parameters of interest, usually a minimization of the negative log-likelihood is performed. The estimates are consequently known as best fit values. Since the negative log-likelihood function (NLL), $-2\ln\mathcal{L}(\mu_i, \theta_j)$, typically depends on the parameters of interest (POIs) μ_i and the nuisance parameters θ_j, the minimization can be complex and computationally very intensive. Usually, the nuisance parameters θ_j are profiled, leading to the concept of the profiled maximum likelihood fit. In a profiled maximum likelihood fit, the function $-2\mathcal{L}(\mu_i, \theta_j)$ is minimized for fixed μ_i, so that the problem depends only on θ_j. The values of θ_j for which the minimum of $-2\mathcal{L}(\mu_i, \theta_j)$ is reached are called profiled values. To quote an (asymmetric) uncertainty for the estimate $\hat{\mu}_i$, a common approach is to scan the

profiled NLL curve as a function of μ_i and take the points at which μ_i $-2\Delta \ln \mathcal{L} = 1$ with respect to the minimum. The reasoning is the relation of the likelihood function to the χ^2 function for Gaussian distributed variables such that the extracted uncertainties can correspond to a 68% CL.

The same procedure can be followed to determine the best fit values for all NPs θ_j, resulting in the "post-fit" description of their probability distributions. When comparing those post-fit values to the corresponding "pre-fit" parameters, observables with a larger impact on the fit result can be identified. The pull of a NP is defined as the deviation of its post-fit central value $\hat{\theta}$ from the pre-fit value θ_0 with respect to the pre-fit uncertainty $\Delta\theta$:

$$\text{pull} = \frac{\hat{\theta} - \theta_0}{\Delta\theta}. \tag{5.13}$$

The constraint is defined as the ratio between post-fit and pre-fit uncertainty:

$$\text{constraint} = \frac{\Delta\theta_{\text{post}}}{\Delta\theta}. \tag{5.14}$$

The impacts of a NP $\Delta\mu_i$ are computed as the difference of the best fit value $\hat{\mu}_i^{\text{up/down}}$ to the best fit value $\hat{\mu}_i$ when fixing a particular NP to its post-fit value varied by one standard deviation of the post-fit uncertainty up and down, while all other NPs are fixed to their best-fit estimate.

Saturated goodness-of-fit tests

The agreement between the prediction and the observation is quantified by the goodness-of-fit (GOF) test. It can be described as a hypothesis test when there is no alternative hypothesis present. Since, according to the Neyman–Pearson lemma [275], the best possible test statistic depends on the presence of an alternative hypothesis, in this case no optimal solution exists. The test model used in the presented analysis is a generalization of a χ^2 GoF and is applicable for the data in a form of binned histrograms [276].

Assuming data distributed according to a Gaussian probability density function, the likelihood function can be written as

$$\mathcal{L} = \prod_i^N \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(x_i - \mu_i)^2}{2\sigma_i^2}\right) \tag{5.15}$$

for per-bin observations x_i with an expected mean of μ_i and variance σ_i. In the absence of an alternative hypothesis, an artificial model can be developed to mimic the expectation. Such a model is called "saturated" model [277]. The likelihood for a saturated model reads

$$\mathcal{L}_{\text{saturated}} = \prod_i^N \frac{1}{\sqrt{2\pi\sigma_i^2}}. \tag{5.16}$$

This leads to ratio of the two likelihoods

$$\lambda = \frac{\mathcal{L}}{\mathcal{L}_{\text{saturated}}} = \prod_i^N \exp\left(-\frac{(x_i - \mu_i)^2}{2\sigma_i^2}\right) \tag{5.17}$$

and the conclusion that $-2\ln\lambda$ follows a χ^2 distribution and hence can be used to define a p-value for a test, observing a value of the test statistic as extreme as the observation. It is

defined as the probability of obtaining test results at least as extreme as the result actually observed, under the assumption that the null hypothesis is correct.

The Kolmogorov–Smirnov test

The Kolmogorov–Smirnov (KS) test [278, 279] is a common tool to test if a sample originates from a known distribution. Although it is originally defined for continuous distributions, it can also be used as a GoF test to test if two samples originate from the same distribution. Given the random samples $\vec{x} = x_1, \ldots, x_N$ and $\vec{y} = y_1, \ldots, y_N$ from two independent groups, the empirical distribution function for both $\vec{x}$ and $\vec{y}$ is defined as

$$\begin{aligned} F_N(\vec{x}) &= \tfrac{n(i)}{N} \text{ and} \\ G_N(\vec{y}) &= \tfrac{n(i)}{N}. \end{aligned} \tag{5.18}$$

Here, $n(i)$ is the number of points less than x_i (y_i) and $\vec{x} = x_1, \ldots, x_N$ ($\vec{y} = y_1, \ldots, y_N$) are ordered from the smallest to largest value.

The test statistic D of the KS test is defined using the empirical distribution function as

$$D = \max\left(|F_i(\vec{x}) - G_i(\vec{y})|\right). \tag{5.19}$$

It is also known as the maximum KS-distance. Using the value of D, a p-value can be determined for the null hypothesis that the two samples originate from the same distribution, given the sample size N and a certain CL.

For the measurement presented in this thesis, the KS test is performed as implemented in the ROOT [280] software package for binned data.

5.2 Machine learning and neural networks

In the following, only the machine-learning aspects relevant for the analysis in this thesis are discussed, while an extensive review of machine learning and artificial intelligence can be found in Refs. [281, 282].

Mathematical foundations

A neural network (NN) in its simplest representation can be described as a non-linear mapping of a set of inputs $\vec{x}$ to a corresponding set of outputs $\vec{y}$. Both $\vec{x}$ and $\vec{y}$ can have arbitrary dimensions and are commonly denoted as "input features" and "prediction", respectively. The most important variant of neural networks is called feedforward neural network, or multilayer perceptron. Its objective is to approximate an arbitrary, known or unknown, function f^* by determining the mapping function $\vec{y} = f(\vec{x}, \theta)$, where θ can describe a considerable set of parameters including simple or complex mathematical functions. The procedure of finding the set of θ that yields the best approximation of f^* is called "training" and is realized iteratively, based on an objective function that describes quantitatively the disagreement between expected and predicted outputs. This function strongly depends on the architecture of the NN and its specific task and application. It is usually described as "loss" or "loss function".

A general depiction of a feedforward NN is given in Figure 5.2. Because of the layerwise structure of the network and composition of f, these individual components or functions are called "layers". Hence, the set of inputs $\vec{x}$ (outputs $\vec{y}$) is called "input layer" ("output layer"), and the number of "hidden" layers in between determines the depth d of a NN. Further, every layer consists of

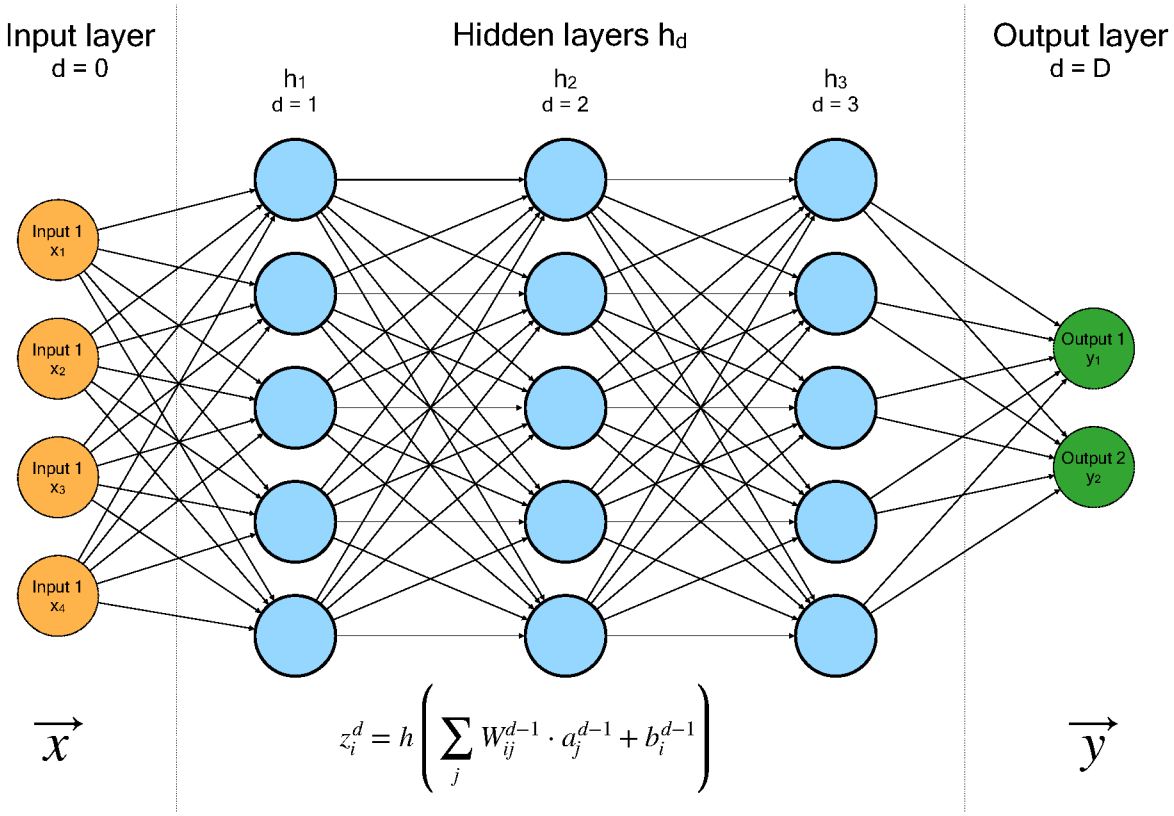

Figure 5.2: An illustration of a feedforward fully connected neural network. The orange (green) blobs indicate nodes of the input (output) layer with depth $d = 0$ ($d = D$), corresponding each to one input (output) variable. The blue blobs stand for nodes in the hidden layers with depth $0 < d < D$.

a fixed number of "nodes", which represent mathematical operations that receive values from the previous layer and combine them to one single output value to be"fed" to the nodes of the respective next layer. The number of layers and nodes per layer, and thus the dimension of the network, depend on the specific application and need to be determined externally. Parameters of that kind are called "hyperparameters". Connections between the nodes are denoted by arrows, representing multiplicative weight tensors W_{ij}^d between node i in layer $d-1$, a_i^{d-1}, and node j in layer d, a_j^d. Additional additive bias terms are denoted as b_j^d. At each node, an operation

$$\tilde{z} = \widetilde{a_j^d} = \sum_i W_{ij}^d \cdot a_i^{d-1} + b_j^d \tag{5.20}$$

is computed, formulating all computations in the network as matrix multiplications and vector additions. Since the operation in Eq. 5.20 is linear, a combination of an arbitrary number of such computations would still be a linear function of the input features. This would prevent the network to learn any nonlinear behavior to model non-trivial non-linear functions. Thus, non-linear "activation" functions $h(\tilde{z})$ are used to apply transformations to the output vector $\tilde{z}$ for each element. The final output of a node reads

$$z = a_j^d = h\left(\sum_i W_{ij}^d \cdot a_i^{d-1} + b_j^d\right). \tag{5.21}$$

Overall, individual applications strongly favor a specific activation function, especially for the output layer while the type of activation function to be applied per hidden layer is a subject to hyperparameter optimization. Typical activation functions are displayed in Figure 5.3, where the scaled exponential linear unit function, SELU(z) [283], is defined as

$$\mathrm{SELU}(z) = \lambda \begin{cases} z & \text{if } z > 0 \\ \alpha \mathrm{e}^z - \alpha & \text{if } z \leq 0. \end{cases} \tag{5.22}$$

For regression problems, a natural choice of the output activation is linear, while for multi-class classification problems the softmax activation function is used:

$$\mathrm{softmax}(z) = \frac{\mathrm{e}^z}{\sum_i \mathrm{e}^{z_i}}, \tag{5.23}$$

which transforms all outputs in such a way that their sum is equal to unity and each output has properties of a probability.

Altogether, a NN can be expressed as a composition of its nodes and mathematical operations as

$$\vec{y} = f(\vec{x}, \theta) = f(\vec{x}, \underbrace{W, b}_{\theta}) = \underbrace{(a^0 \cdot a^1 \ldots \cdot a^d)}_{\text{network}}(\vec{x}) \tag{5.24}$$

with the set of free parameters θ.

Neural network training and hyperparameter optimization

The training of a NN goes hand in hand with the optimization, i.e., the minimization of the loss function that compares a set of predicted outputs $\hat{\vec{y}}$ to a set of expected or truth outputs $\vec{y}$. For

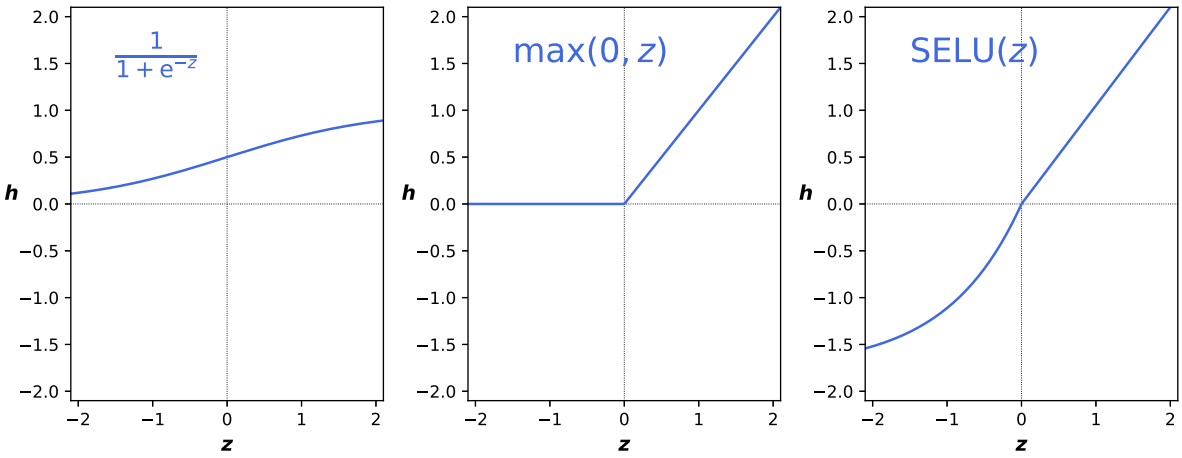

Figure 5.3: Examples for commonly used activation functions. In the left (center) figure the sigmoid (rectified linear unit) activation function is shown. The right plot shows the SELU activation function defined in the text.

a regression application, the mean squared error (MSE) or the mean absolute percentage error (MAPE) are adequate loss functions:

$$L_{\text{MSE}}(y, \hat{y}) = \frac{1}{N} \sum_{i=0}^{N} (y_i - \hat{y}_i)^2, \tag{5.25}$$

$$L_{\text{MAPE}}(y, \hat{y}) = \frac{1}{N} \sum_{i=0}^{N} \frac{|y_i - \hat{y}_i|}{y_i}. \tag{5.26}$$

For multiclass classification tasks, the cross-entropy (CE) is a better choice for the loss function:

$$L_{\text{CE}}(y, \hat{y}) = \frac{1}{N} \sum_{i=0}^{N} -y_i \log \hat{y}_i. \tag{5.27}$$

Here, the vectors $\vec{y}$ and $\widehat{\vec{y}}$ have the dimension N, where N denotes the number of output classes.

The loss is usually expressed as the arithmetic mean over a distinct number of loss evaluations, yielding a statistically reliable representation of the total data set and is referred to as "batch". The batch size is also subject to hyperparameter optimization.

The task is to find the set of θ for which L reaches a global minimum. The minimization procedure is performed using a stochastic process, called gradient descent. It derives rules of how individual θ_i have to be updated in each iteration t to go along the gradient towards the minimum. The update rule can be written as

$$\theta^{\text{t+1}} = \theta^t + \Delta\theta = \sum_i \theta_i^t - \alpha \frac{\partial L(\theta)}{\partial \theta_i}, \tag{5.28}$$

where α is an additional hyperparameter that is known as "learning rate" or "step-size". The calculation of the derivatives for each weight tensor and bias vector can be, depending on the size of the NN, computationally very intensive. A simpler solution is found by applying the so-called "backpropagation" algorithm [284]. By making use of the layer-wise structure of a NN,

it can be shown by applying the chain rule [281, 282], that using δ^d as a "local error" for each layer d, the derivatives can be written as

$$\frac{\partial L}{\partial \theta} = \begin{cases} a_j^{d-1} \cdot \delta_i^d, & \text{if } \theta \text{ is a weight tensor,} \\ \delta_i^d, & \text{if } \theta \text{ is a bias vector.} \end{cases} \tag{5.29}$$

Here, δ^d is defined as

$$\delta_j^d = \begin{cases} \frac{\partial L}{\partial y} \cdot h'(z), & \text{for } d = D, \\ \left(\sum_\ell W_{j\ell}^{d+1} \delta_\ell^{d+1} \right) \cdot h'(z), & \text{for } d < D. \end{cases} \tag{5.30}$$

The described procedure of updating all free parameters of the NN is repeated until a minimum of the loss function is found. An iteration where the weights and biases are updated multiple times for all batches of the data, is referred to as "epoch". Several more involved update rules, based on various "optimizers" exist, and offer faster convergence in less epochs. One of them is the Adam optimizer [285] which accounts for the weights of previous gradients when updating the parameters.

Because the data used for training a NN is usually limited, special attention has to be paid on the network's ability to generalize the learned behavior. Hence, the data is split into orthogonal "training", "validation", and "testing" sets. During the training procedure, the training set is used to calculate all gradients and update free parameters, while the loss is also calculated using the statistically independent validation set. In case the model begins to be sensitive to properties of the training set and fails to generalize also on the validation set, the training is stopped to prevent "overtraining". Special measures are to be taken to prevent overtraining, as discussed in the following. The validation or the testing set can further be used to optimize hyperparameters, e.g., the learning rate or the batch size.

Methods to optimize hyperparameters are manifold [281]. The simplest one is the so-called grid-search approach where a distinct set of parameters of the possible parameter space is defined a-priori, and all permutations are tested. Often, a randomized approach, where only some of all defined parameters are employed, shows faster convergence. Another approach is based on Bayesian probability and keeps, in contrast to the approaches above, track of past iterations. Based on the previous trials, a prior distribution is updated towards a posterior distribution to calculate an approximation of the objective function. This posterior function is then used to derive the next set of hyperparameters for the following iteration [286, 287].

The vanishing gradient problem and overtraining

The analysis of this thesis utilizes additional machine learning concepts when deriving NNs used for regression and multiclass classification applications.

One common problem when optimizing the weights of the NN in an iterative procedure is the dependence of gradients of all activation functions h' on the layer input z. Depending on the loss function chosen, a certain range of input values for the activation function can be preferred, as for input values with a value very different from zero, the gradients can be close to zero. This, however, leads to negligible weight updates and prevents the network to learn any additional information with growing training epochs. This problem is known as the "vanishing gradient problem" and can be addressed by several measures [281, 282].

For all activation functions, the gradients are significantly different from zero if the input value is close to zero. Hence, one straightforward way to ensure non-vanishing gradients is to set

upper limits on the parameter range of weights and biases to prevent extreme output values. Another approach is to transform each input z_i such that its distribution over the whole batch has a mean of zero and a standard deviation of unity. Such a transformation reads

$$z \to z' = \frac{z - \mu}{\sigma}, \tag{5.31}$$

where μ (σ) is the mean (standard deviation) of the original distribution. Such a behavior is, for example, obtained when using batch normalization [288]. Similar results can also be achieved when using the SELU activation function, as it shows a self-normalizing behavior [283]. Further optimization can be accomplished when deriving the initial values for all weights and bias terms of the nodes from specific probability density distributions, such as the uniform or normal distribution. This is implemented, for example, in the Glorot initialization process [289], where biases are initialized to zero and weights are derived from a uniform distribution from an interval respecting the dimension of the respective input layer.

In general, the performance of a NN is strongly correlated to its "capacity", i.e., the ability to approximate or learn complex models and is related to, e.g., the number of free parameters and depth of a NN. Varying capacity goes along with two types of errors that have to be considered when training a network, bias and variance. Bias is defined as the error between the average model response and the truth, and can be inferred from the performance using the training set. Variance is associated to the capability or flexibility of the model to generalize on different related data sets. High variance (bias) usually corresponds to a large (low) capacity. As illustrated in Figure 5.4, there is a connection between bias and variance, where the task is to find the compromise between the two, and thus, the optimal capacity for a given task and model.

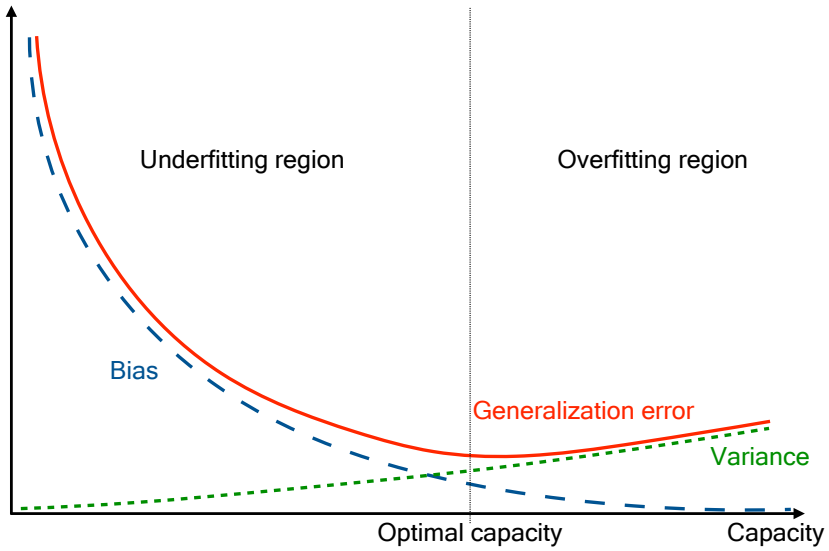

Figure 5.4: Sketch for the dependence of a NN model on its capacity, explaining typical effects such as over- and underfitting and properties as variance and bias. The optimal capacity is marked with a dotted line, yielding small bias and variance.

Another representation of the same problem is depicted in Figure 5.5. Complex problems rely on the model to learn multidimensional correlations of the input data and require an adequate number of free parameters. If, on the other hand, the capacity is too low or the training data set is too small, the consequence often is underfitting the data. While an appropriate capacity seems reasonable to describe the training data, it might lead to generalization issues when

evaluating on independent data sets because no appropriate complex representation of the data has been learned. Although a larger capacity can lead to overfitting, deeper connections in the data are learned, therefore often larger capacities are preferred. In order to limit the variance, regularization has to be applied.

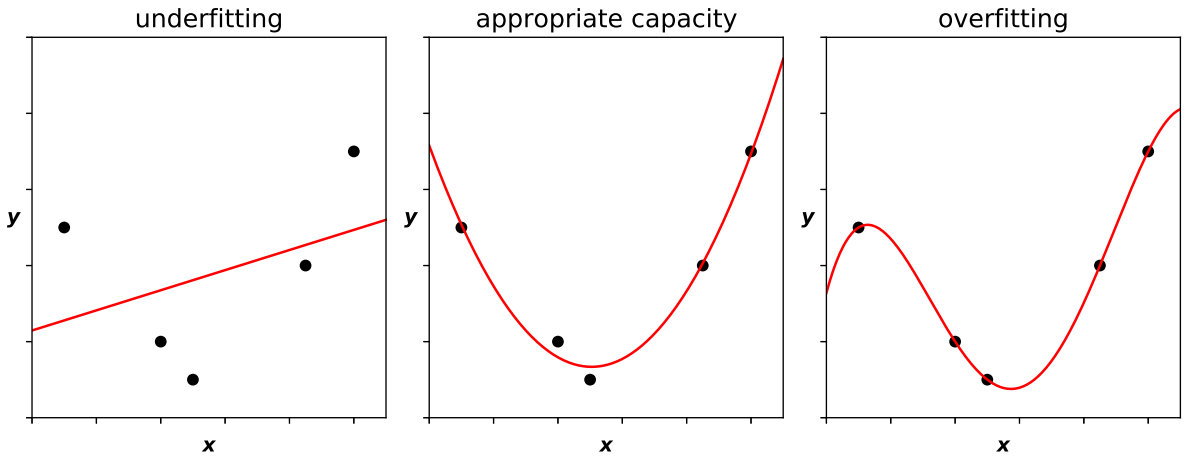

Figure 5.5: Illustration of typical states for the capacity of a NN model, leading to underfitting (left), overfitting (right), or an optimal description (center).

One common regularization method is called "weight decay" or L_2 regularization. Here, an additional penalty term

$$\lambda \sum_{i,j} W_{i,j}^2 \tag{5.32}$$

is added to the loss function. The hyperparameter λ regulates the strength of the regularization and $W_{i,j}$ are the weights of the weight tensors. The penalty term leads to favoring smaller weights and a following reduction of the bias towards specific inputs or nodes during the training.

Another method of regularization is to implement so-called "dropout" units, that randomly, based on a flat prior function, set the output of a node to zero during the training [290]. That ensures that the network becomes less reliant on the output of a particular node but rather allocates the information from a larger set of nodes. It can be understood as inserting the noise into the data set or the training of an ensemble of networks at once.

Domain adaption by backpropagation

When training a NN with a particular task on a given set of data sets with distinct properties, one assumes that both the training and the test data set are derived from a joint probability distribution and share identical features. This constraint is important for a model to generalize but can be violated easily. Solutions to adapt this problem of diverging response in different domains are called domain adaption [291]. Several methods have been proposed, but the one used in the analysis of this thesis is known as "domain adaption by backpropagation" [292]. It can be used not only to reduce the discrepancy between the performance on specific domains, but also to make the response of a model invariant with respect to a particular variable. Domain adaption by backpropagation can be implemented by incorporating additional outputs with the target of classifying the domain into the network. Similarly, a regression-NN part can be implemented, achieving decorrelation with respect to a certain variable. Figure 5.6 shows a possible architecture. Three parts can be identified: the network G_f performing the feature extraction from the inputs and learning different representations of the data, a second network G_y

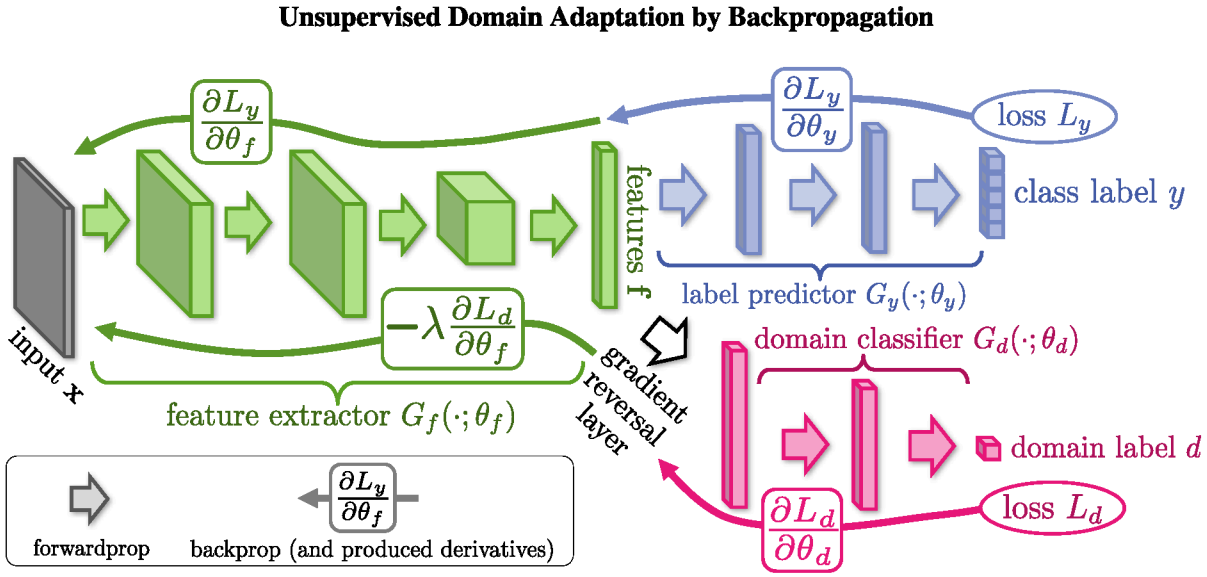

Figure 5.6: Sketch of a neural network architecture employing domain adaption by back-propagation [292]. Standard parts of the NN model are shown in green for the feature extractor G_f and the classifier G_y in blue. The additional layers G_d responsible for the domain adaption are shown in pink.

performing the original target task, and the extension G_d which distinguishes between different domains. In front of the latter a special layer is implemented that switches the sign of the gradient by multiplying it with a constant factor $-\beta$. The consequence is that during training the model will learn to perform the target task while keeping the performance identical for both domains. Inverting the sign of the gradient ensures that every information needed to distinguish between the domains is forgotten during the training and does not impact the output of G_d. Instead of classifying between multiple domains, a regression task can be implemented to decorrelate the output of G_d from a particular variable.

6 Measurement of the $\mathrm{t\bar{t}}$+jet cross section and the top quark pole mass

In this chapter, the measurement of the differential $\mathrm{t\bar{t}}$+jet production cross section as a function of the ρ observable is discussed. The measured cross section is used to extract the top quark pole mass as defined in the so-called on-shell scheme. The decays of the top quarks into two leptons are used. The pp collision data at the LHC at a center of mass energy of 13 TeV are recorded with the CMS experiment in 2016 and correspond to the integrated luminosity of $36.3\,\mathrm{fb}^{-1}$. The chapter is structured as follows. In Section 6.1, the general measurement strategy is motivated and explained. The data sets and event selection are discussed in Section 6.2, where also the main background contributions to the measurement are outlined. Also, the employed MC generated samples are introduced. Section 6.3 discusses various approaches of how to reconstruct $\mathrm{t\bar{t}}$ event kinematic properties, and a new approach using NNs that is developed for this measurement is presented. An event classification procedure, also based on NNs, is introduced in Section 6.4. The statistical procedure used to extract the differential production cross section is presented in Section 6.5. The relevant uncertainties are described in Section 6.6, and the results of the measurement are presented in Section 6.7. The extraction of the top quark pole mass is also discussed therein, together with a description of the theoretical framework employed. The described analysis is developed solely for the purpose of this thesis, and is explained in a public Physics Analysis Summary (PAS) [293], which is superseded by a document submitted to the Journal of High Energy Physics (JHEP) [49].

6.1 Analysis strategy

The normalized differential cross section of the $\mathrm{t\bar{t}}$+jet production as a function of the ρ observable is sensitive to the value of the top quark pole mass, as explained in Section 2.3. This fact is exploited in the extraction of the top quark pole mass in this work. The idea for such an analysis was introduced in [44] and realized first by the ATLAS collaboration using the LHC data at

center-of-mass energies of 7 TeV [35] and 8 TeV [46], with an accuracy of the results of 1.31% and 0.71%, respectively. The work in this thesis considers events where top quarks decay into two leptons, accounting for combinations of e^+e^-, $\mu^+\mu^-$, and $e^\pm\mu^\mp$. These final states exhibit most clear experimental signature, since it is less affected by the QCD background with related jet production due to imposed requirements of high-energy isolated leptons. However, the branching fraction for $t\bar{t} \to \ell\ell$ is only about 9%. By restricting the lepton selection to combinations of electrons and muons, also via τ-lepton decays, this contribution gets further reduced to less than 5%, yielding an upper limit on the total acceptance of the measurement.

In the measurement, the geometric acceptance of the detector and the event selection requirements on the final-state objects define the visible phase space. For the measurement presented in this thesis, the acceptance is defined by the p_T and $|\eta|$ requirements on the lepton and jet candidates. The differential cross section is measured in the visible phase space, and is extrapolated to the full phase space as defined at the parton level.

The neutrinos originating from the W boson decays cannot be measured directly and instead are identified by a missing transverse momentum, p_T^{miss}, in an event. The ambiguities related to the reconstruction of the two neutrinos in the $t\bar{t}$ decays challenge the reconstruction of the event kinematic properties and deteriorate the resolution of the final state observables. Additionally, p_T^{miss} can arise from neutrinos originating from the leptonic decay of τ leptons.

Further challenges originate from the presence of the jet associated to the production of the $t\bar{t}$ pair in the event, due to related uncertainties of the jet energy scale. Moreover, the parton-level definition of an additional jet implies applying kinematic cuts on the transverse momentum p_T and pseudorapidity η of the jet. These leads to the irreducible background formed by events with 0 jets ($t\bar{t}$+0 jet), originating from phase space migrations. In particular, at the parton-level the additional jet does not pass the kinematic requirements while it does at the detector level. This background contribution is dominant and of the same amount as the signal in the $\rho \sim 1$ range.

The strategy of the analysis presented in this thesis is based on the optimization of the event reconstruction and background mitigation, and a significant improvement in the resolution of the ρ observable.

Experimentally, the final state is characterized by the presence of a top quark pair followed by top quark decays $t \to W^+b$ and $\bar{t} \to W^-\bar{b}$, and subsequent decays of either $W \to \mu\nu$ or $W \to e\nu$, together with one additional jet. Furthermore, events with leptonic W boson decays into τ leptons are considered as signal events, if all τ leptons in the event decay further to electrons or muons. Possible additional extra jets could arise from QCD radiation. However, to keep a maximum acceptance, events are selected requiring the presence of two leptons.

The event selection is based on the definition of physics objects introduced in Section 4, considering the following background contributions for the analysis: top quark-antiquark pair production with no additional energetic jet ($t\bar{t}$+0 jet), single top quark production in association with a W boson (tW), Z/γ^* bosons produced in association with additional jets, and smaller background contributions such as WW, WZ, ZZ diboson production, $t\bar{t}V$ associated production, or W+jets production.

Both $t\bar{t}$+0 jet and tW, with additional QCD radiation, result in the same final state and signature and are therefore irreducible backgrounds. The Z+jets backgrounds is present predominantly in the same-flavor channels e^+e^- and $\mu^+\mu^-$, while in the $e^\pm\mu^\mp$ channel it originates mainly from $Z \to \tau^+\tau^-$ events. Due to lepton misidentification, semileptonic $t\bar{t}$ and W+jets are additional sources of background, while diboson and $t\bar{t}V$ events lead to backgrounds due

to the leptonic decays of the W and Z bosons or top quarks with a non-negligible amount of $p_{\mathrm{T}}^{\mathrm{miss}}$ present coming from leptonic W decays. The contribution of QCD multijet production is negligible.

6.2 Data set, Monte-Carlo simulation, and event selection

The total integrated luminosity of the complete data sample used for the analysis presented in this thesis is $36.3\,\mathrm{fb}^{-1}$ $\pm 1.2\%$ [200]. The data is collected by using combinations of dilepton and single lepton trigger paths. This is done in order to recover interesting dilepton events, which failed to pass the dilepton trigger requirements. Usage of the single lepton triggers in this analysis allows a recovery of approximately 10% of the dilepton events discarded by dilepton triggers due to inefficiencies. Minimum requirements on the event selection imposed by the trigger paths are at least either one of the following:

- one muon with $p_{\mathrm{T}} > 24\,\mathrm{GeV}$,
- one electron with $p_{\mathrm{T}} > 27\,\mathrm{GeV}$,
- one muon with $p_{\mathrm{T}} > 17\,\mathrm{GeV}$ and one muon with $p_{\mathrm{T}} > 8\,\mathrm{GeV}$,
- one muon with $p_{\mathrm{T}} > 23\,\mathrm{GeV}$ and one electron with $p_{\mathrm{T}} > 12\,\mathrm{GeV}$,
- one electron with $p_{\mathrm{T}} > 23\,\mathrm{GeV}$ and one muon with $p_{\mathrm{T}} > 8\,\mathrm{GeV}$,
- or one electron with $p_{\mathrm{T}} > 23\,\mathrm{GeV}$ and one electron with $p_{\mathrm{T}} > 12\,\mathrm{GeV}$.

Leptons are further required to be reconstructed within the tracker acceptance.

MC simulation

The signal and background $\mathrm{t\bar{t}}$ processes are simulated using the POWHEG (version 2) [130, 294, 295] MC event generator at NLO in QCD for the ME calculations. The top quark mass in the simulation, $m_{\mathrm{t}}^{\mathrm{MC}}$, is set to 172.5 GeV. To describe the proton structure, the NNPDF 3.1 PDF at NNLO [169, 296] is used. The parton showering is performed using PYTHIA (version 8.230) [139] employing the CP5 tune [132]. The h_{damp} parameter in POWHEG which effectively regulates the matching scale between the ME and the PS, is set to $1.379\,m_{\mathrm{t}}^{\mathrm{MC}}$ [132]. For this thesis, the $\mathrm{t\bar{t}}$ simulation is divided into few different components, modeling the $\mathrm{t\bar{t}}$+jet signal and the background contributions as described later in the text.

In addition, simulated samples of the $\mathrm{t\bar{t}}$ process are generated using MG5_AMC@NLO [297] with up to two extra partons at the NLO ME level. These samples are used to derive calibrations of NNs, as discussed in Sections 6.3 and 6.4. Events are matched to the PYTHIA PS using the FxFx [298] prescription , while the decays of the top quark and antiquark at LO are modeled using MADSPIN [299].

For all the background samples, PYTHIA is used to model the parton showering, hadronization, and multiparton interactions. The tW process is simulated using POWHEG [300, 301] at NLO. The identical tune and modeling parameter settings as for the $\mathrm{t\bar{t}}$ simulation are used, while the CUETP8M1 [302–304] tune is used for the remaining background simulation. The Z+jets process is generated using MG5_AMC@NLO at leading order (LO) with up to four additional partons at ME level, and the matching to the PYTHIA PS is done using the MLM [305, 306] prescription. Similarly, W+jets events are simulated using MG5_AMC@NLO at LO with up to four additional partons at ME level. To derive the NN calibrations, an independent Z+jets simulation is produced using MG5_AMC@NLO at NLO with the FxFx merging prescription. Diboson production is simulated at LO using PYTHIA.

All predictions are normalized to their theoretical cross section and the corresponding integrated luminosity of the data. The cross section for the $t\bar{t}$ signal is calculated using the TOP++ (version 2.0) program [153] at NNLO, including resummation of next-to-next-to-leading logarithmic soft-gluon terms [150, 307–311], and assuming $m_t = 172.5\,\text{GeV}$. The NNPDF3.1 PDF set at NNLO is used, and the resulting cross section value is $830.91\,^{+20.39}_{-29.96}\,(\text{scale})\,^{+3.92}_{-3.83}\,(\text{PDF}+\alpha_S)\,\text{pb}$. Here, "scale" ($\alpha_S$) refers to the uncertainty assessed by varying the renormalization and factorization scales (the value for the strong coupling constant). All other cross sections are calculated at approximate NNLO for tW production [312], NLO for Z+jets, NNLO for W+jets [313], and NLO for diboson production [314].

For all simulated samples, the interaction of the final state particles with the CMS detector volume is simulated with the GEANT4 [315] program. To model the effect of pileup, additional minimum bias interactions are added to the simulated events.

Signal definition

An event is considered as signal, if it contains a $t\bar{t}$ pair in the dileptonic decay channel and at least one additional jet with $p_T > 30\,\text{GeV}$ and $|\eta| < 2.4$ at the parton level. The parton level is defined after the QCD and electromagnetic radiation with the on-shell top quarks, while the additional jets are clustered using the anti-k_T algorithm with a distance parameter of 0.4, using all particles before hadronization that do not originate from the top quark decays.

Moreover, the nominal $t\bar{t}$ simulation is split into multiple contributions, modeling the $t\bar{t}$+jet signal in different ρ bins and $t\bar{t}$ background processes on the basis of the aforementioned jet definition. The $t\bar{t}$+0 jet process is defined, in which no jet at the parton level is retained after the signal-definition requirements. The individual $t\bar{t}$+jet signal processes are defined based on the value of ρ at the parton level in the following binning: 0–0.3, 0.3–0.45, 0.45–0.7, and 0.7–1. The binning choice is further discussed in Section 6.5. Since $t\bar{t}$ events with two leptons, also originating from the decays of τ leptons, are included in the signal definition, the remaining events are treated as background. The unfolded results are extrapolated to the full phase space using the $t\bar{t}$ MC simulation, as described in Section 6.5.

Event selection

The event selection is optimized in order to maximize the phase space of the analyzed process and to suppress the background contributions. Selection criteria are defined such that only events are selected that have well understood properties, i.e., the experimental data are well described by the MC simulation. This includes sufficient reconstruction and identification efficiencies and implies further corrections to the simulated events and, if needed, to the data. Typically, this is true for high-momentum objects reconstructed in the central part of the detector.

Events selected by the trigger paths are then reprocessed and further used for offline analysis. Further selection criteria are applied: for each event, at least one PV has to be reconstructed, and additional event filters are applied. All lepton candidates are required to pass the identification and selection as defined in Section 4.2. The kinematic requirements on the lepton p_T arise in the context of trigger efficiency measurements, that are discussed briefly in the following. Minimum requirements of 25 GeV (20 GeV) for the leading (subleading) lepton are imposed. A presence of an opposite-sign lepton pair (e^+e^-, $\mu^+\mu^-$, $e^\pm\mu^\mp$) is required. If more than two leptons are reconstructed in the event, the two leptons with the highest p_T are chosen. The event is then unambiguously classified as e^+e^-, $e^\pm\mu^\mp$, or $\mu^+\mu^-$ depending on the type of the selected lepton pair. Furthermore, the invariant mass of the selected lepton pair is required to be larger than 20 GeV to suppress events from decays of heavy-flavor resonances and low-mass Drell–Yan processes.

Jets and $p_{\mathrm{T}}^{\mathrm{miss}}$ are reconstructed and identified as defined in Section 4.2. No requirement on the jet multiplicity is imposed. Instead, the number of reconstructed jets is used to define suitable categories in the analysis. Jets arising from b quarks are identified using the DeepCSV tagger, cf. Section 4.2. A high efficiency, i.e., "loose", working point is used with a fixed light flavor mistagging rate of 10% and a b (c) tagging efficiency of 80–90% (40%) as determined in $\mathrm{t\bar{t}}$ events [249]. Further, the events have to contain at least one b-tagged jet.

Background contributions of Z+jets production are reduced by discarding events with $p_{\mathrm{T}}^{\mathrm{miss}} < 40\,\mathrm{GeV}$ and an invariant dilepton mass, $m_{\ell\ell}$, in the Z mass window of $76 < m_{\ell\ell} < 106\,\mathrm{GeV}$ for the same-flavor lepton channels.

The described event selection without the b jet requirement is defined as "inclusive" selection, or preselection. Events with at least three jets are considered to be in the signal region, and NNs are trained to enhance the sensitivity to the $\mathrm{t\bar{t}}$+jet signal, as described in Sections 6.3 and 6.4.

Typical purities for the $\mathrm{t\bar{t}}$+jet signal obtained after preselection without requirements on the jet or b jet multiplicity range from about 15% in the $\mathrm{e^+e^-}$ and $\mu^+\mu^-$ channels to about 30% in the $\mathrm{e^{\pm}}\mu^{\mp}$ channel. The events selected with at least three reconstructed jets, of which one is b tagged, have the purity of about 75 %.

Corrections to simulated events

To reduce the discrepancies between the simulation and the data, caused by deficits in the event generation/simulation or detector inefficiences, a number of corrections are applied to the simulation by using the aforementioned scale factors (SFs). Usually, weights are derived as a function of observables correlated to the effect under scrutiny, as, e.g., the lepton p_{T} and η for the identification efficiency. Efficiencies are measured in simulation and data, and the corresponding weights are determined. Similar corrections can arise also for the energy scale or resolution.

To reproduce the pileup distribution in the data, weights are applied to the simulated events, determined from the bunch-to-bunch instantaneous luminosity and assuming a total inelastic pp cross section of 69.2 mb [316]. Dedicated trigger SFs are derived as described below directly for the measurable phase space. Differences in the lepton reconstruction, identification, and isolation are accounted for by the SFs as determined in Refs. [190, 196, 222, 317, 318]. The different effects are considered uncorrelated and individual efficiencies for muons and electrons, i.e., also SFs, are derived. The SFs are thus applied multiplicatively to the full event and are determined individually per object. The lepton energy-scale and resolution corrections are determined as described in Refs. [190, 222, 227–229]. In a similar fashion, pileup jet identification efficiency corrections are measured [240] and applied to the simulation. Jet energy scale and resolution corrections are determined and adopted as described in Section 4.2.

In 2016, a gradual timing shift of the electromagnetic calorimeter (ECAL) was not properly propagated to the Level-1 (L1) trigger primitives (TP), leading to an incorrect association of a significant fraction of high-$|\eta|$ TP to the previous bunch crossing. Due to this effect and as L1 forbids firing of two consecutive bunch crossings, events can self veto if a significant amount of ECAL energy is found in the region of $2 < |\eta| < 3$ (known as L1 ECAL prefiring). The effect is not described by the simulation and is taken into account using event weights [319, 320].

To correct for the different b-jet-tagging efficiency in both data and simulation, SFs are measured for the true jet flavor as a function of jet p_{T}. The light flavor misidentification SFs are determined using inclusive QCD multijet events for light flavor jets [249]. Similarly, the b tagging SFs for b and c jets are derived using muon-enriched QCD multijet events. The per-jet SFs are transferred to per-event corrections as follows. The probabilities to identify all b jets in an event in data

(P_{data}) and in the simulation (P_{MC}) for a fixed chosen working point are calculated, and their ratio is applied as a weight to the simulated events. The probabilities P_{data} and P_{MC} are calculated via

$$P_{\text{data}} = \prod_{f}^{\text{flavors}} \prod_{i}^{\text{tagged}} \epsilon_i^f\left(p_{\text{T}}\right) \text{SF}_i^f\left(p_{\text{T}}\right) \prod_{j}^{\text{untagged}} \left[1 - \text{SF}_j^f\left(p_{\text{T}}\right) \epsilon_j^f\left(p_{\text{T}}\right)\right], \tag{6.1}$$

and

$$P_{\text{MC}} = \prod_{f}^{\text{flavors}} \prod_{i}^{\text{tagged}} \epsilon_i^f\left(p_{\text{T}}\right) \prod_{j}^{\text{untagged}} \left[1 - \epsilon_j^f\left(p_{\text{T}}\right)\right]. \tag{6.2}$$

Here, ϵ_i^f and SF_i^f denote the b tagging efficiency and SF in MC for the jet i with a flavor f, dependent on its p_{T}. The final event weight applied to the simulation is defined as $P_{\text{data}}/P_{\text{MC}}$.

Measurement of the trigger efficiency

The trigger selection aims to yield maximum efficiency in order to collect as many $\mathrm{t\bar{t}}$ events as possible. Its efficiency is measured with respect to the offline lepton selection. Triggers used for this measurement have asymmetric requirements on the lepton momenta. Several dilepton and single lepton triggers are combined by using a logical "OR" in the trigger selection.

The trigger efficiency is measured using an independent data set recorded using an orthogonal set of $p_{\text{T}}^{\text{miss}}$ triggers. This procedure allows for an unbiased measurement of the trigger efficiency in data and simulation under the condition that the $p_{\text{T}}^{\text{miss}}$ baseline triggers are uncorrelated to the signal lepton triggers. The minimal thresholds on $p_{\text{T}}^{\text{miss}}$ vary between 75–250 GeV for the set of baseline triggers. An advantage of using $p_{\text{T}}^{\text{miss}}$ triggers is that from dileptonic $\mathrm{t\bar{t}}$ events a non-negligible amount of $p_{\text{T}}^{\text{miss}}$ is expected for the final state, so that the efficiency can be determined in a data sample containing $\mathrm{t\bar{t}}$ events. The trigger efficiency ϵ_{trig} can be obtained as

$$\epsilon_{\text{trig}} = \frac{N_{\text{Events}}\left(\text{pass selection} + p_{\text{T}}^{\text{miss}}\text{ trigger} + \text{lepton trigger}\right)}{N_{\text{Events}}\left(\text{pass selection} + p_{\text{T}}^{\text{miss}}\text{ trigger}\right)}, \tag{6.3}$$

where N_{Events} is the number of events passing the selection given in the brackets. For the numerator this comprises the offline selection, the signal and baseline triggers, and for the denominator this includes the offline selection and the baseline triggers. Statistical uncertainties are estimated using Clopper–Pearson intervals [321, 322] that describe exact confidence intervals for Bernoulli variables distributed according to a binomial probability density function. The trigger efficiencies as a function of the lepton p_{T} are shown in Figure 6.1 for all three lepton channels. For the optimal p_{T} cuts of 25 (20) GeV for the leading (subleading) lepton, the trigger efficiencies are observed to be close to unity over the whole p_{T} spectrum. The region where the triggers reach their maximum efficiency is denoted as "plateau", and the transition region, where the efficiency steeply rises, is commonly called "turn-on". For the dielectron case, the turn-on region is much more pronounced as for the $\mathrm{e^{\pm}\mu^{\mp}}$ or dimuon ones. A good agreement between data and simulation is observed for all three channels, within statistical uncertainties. Additionally, the dependence of the trigger efficiency on, e.g., the number of reconstructed jets or pileup vertices is also studied, and no dependence is found.

To correct for differences in the trigger efficiency between data and simulation, SFs are derived as a function of the lepton p_{T}. They are defined as the ratio of the trigger efficiency in data with respect to the efficiency determined for the simulation. The results are shown in Figure 6.2 for

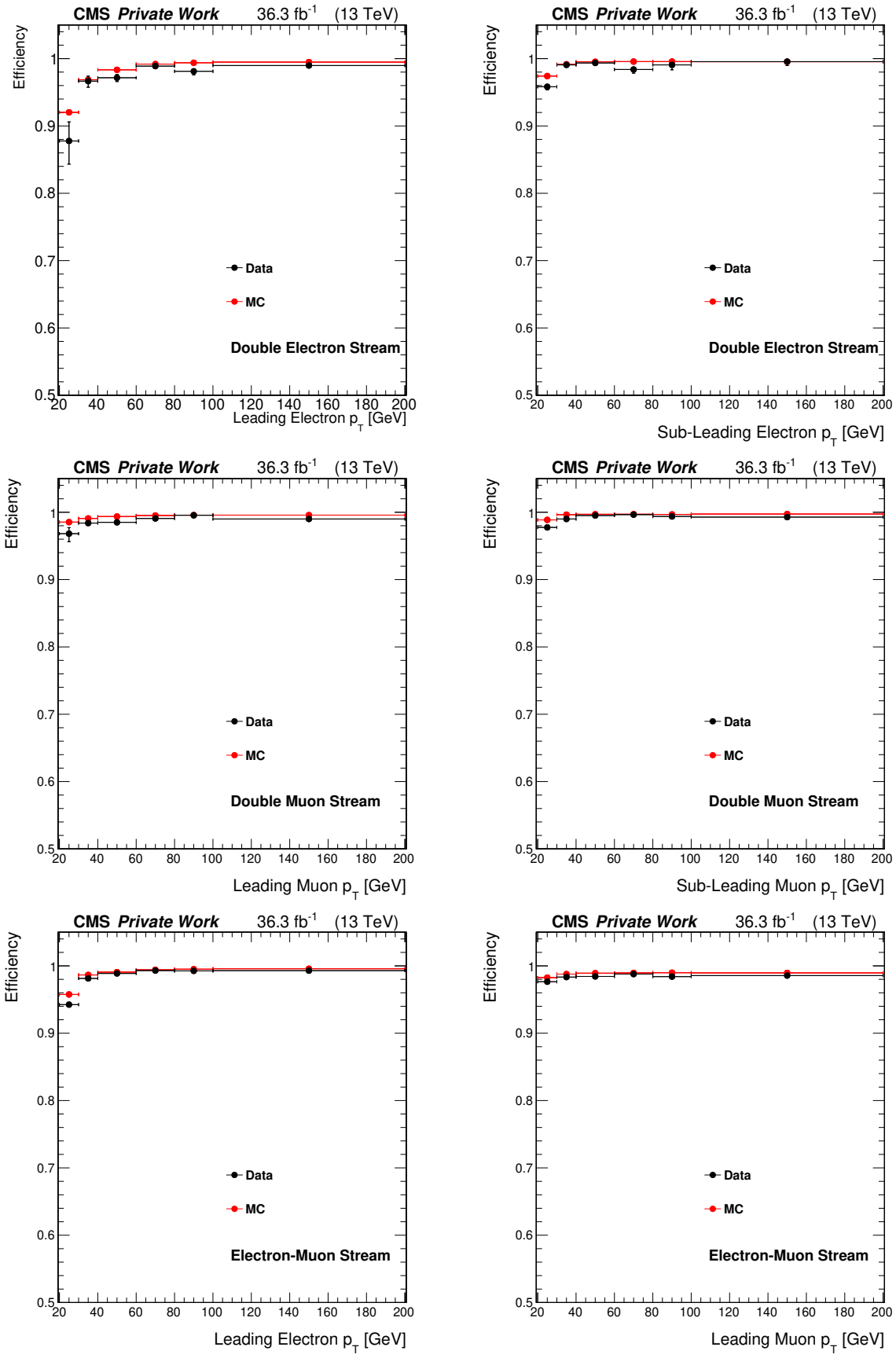

Figure 6.1: Trigger efficiencies for the dielectron (upper row), dimuon (middle row) and electron-muon channel (bottom row) as measured in data (black) and MC simulated samples (red), shown as a function of the electron and muon p_{T}.

the three lepton channels. To estimate the uncertainties originating from detector inefficiencies and related phase-space dependence of the measurement in different kinematic regions, the SFs are determined by varying kinematic requirements in the selection. The following regions are investigated:

- number of reconstructed jets, N_{jet}, < 3; ≥ 3,
- number of primary vertices, NPV, < 30; ≥ 30,
- and missing transverse momentum, $p_{\text{T}}^{\text{miss}}$, $< 150\,\text{GeV}$: $\geq 150\,\text{GeV}$.

Additionally, the run-range dependence and the correlation between baseline and signal triggers are considered as systematic uncertainties. Effects to the total uncertainty are as follows: the correlation between baseline and signal triggers (0.1–0.5%); the run-range dependency (0.02–0.3%); of statistical origin (0.1–0.2%); and from the measurement in different kinematic regions (0.2–2.9%).

Data and MC compatibility

Figures 6.3 and 6.4 show a comparison between data and simulation for basic kinematic observables for all three lepton channels combined. The simulation is observed to describe the data well. A small discrepancy between the normalization of the data and simulation is present, which does not affect the measurement presented in this thesis. The reason is that the signal normalization is directly fitted in the likelihood fit as described in Section 6.5, and the normalized cross section is used to extract the value for $m_{\text{t}}^{\text{pole}}$. Furthermore, some trends can be observed in the jet and lepton momentum distributions, as the spectrum predicted by the simulation is harder than the one observed in data. This effect can be associated to the modeling of the top quark p_{T} in the POWHEG+PYTHIA8 $t\bar{t}$ simulation, and was observed already in previous measurements [323–326] when performing comparisons of the data to detector-level or particle-level predictions. The jet and b jet multiplicity distributions are well described within the total uncertainties. In particular, the shapes of these particular distributions are used to discriminate the signal and the background processes through introduction of event categories.

6.3 Reconstruction of the top quark kinematic properties

The top quarks are investigated experimentally by measuring their decay products and their kinematic properties. In the fully-hadronic decay channel, all decay products are reconstructed. In the dileptonic channel however, the two neutrinos from the W decay are not measured, and hence lead to ambiguities in the reconstruction problem.

The full kinematic reconstruction

Several methods of kinematic reconstruction of $t\bar{t}$ pairs have been developed for former measurements [38, 42, 323–327]. The approach presented in the following is based on the algebraic approach suggested in Ref. [327] and was used in $t\bar{t}$ cross section measurements at $\sqrt{s} = 8$ and 13 TeV [323, 324]. Further improvements have been implemented in Ref. [328]. It is denoted further in this thesis as "full kinematic reconstruction" and is described in the following. Based on the four-momenta of the six decay products $(\ell, \bar{\ell}, \text{b}, \bar{\text{b}}, \nu_\ell, \overline{\nu_\ell})$, the following relations can be derived, for the energy and four-momenta of the two neutrinos:

$$\begin{aligned} E_\nu^2 &= m_\nu^2 + p_{x,\nu}^2 + p_{y,\nu}^2 + p_{z,\nu}^2, \\ E_{\overline{\nu}}^2 &= m_{\overline{\nu}}^2 + p_{x,\overline{\nu}}^2 + p_{y,\overline{\nu}}^2 + p_{z,\overline{\nu}}^2, \end{aligned} \tag{6.4}$$

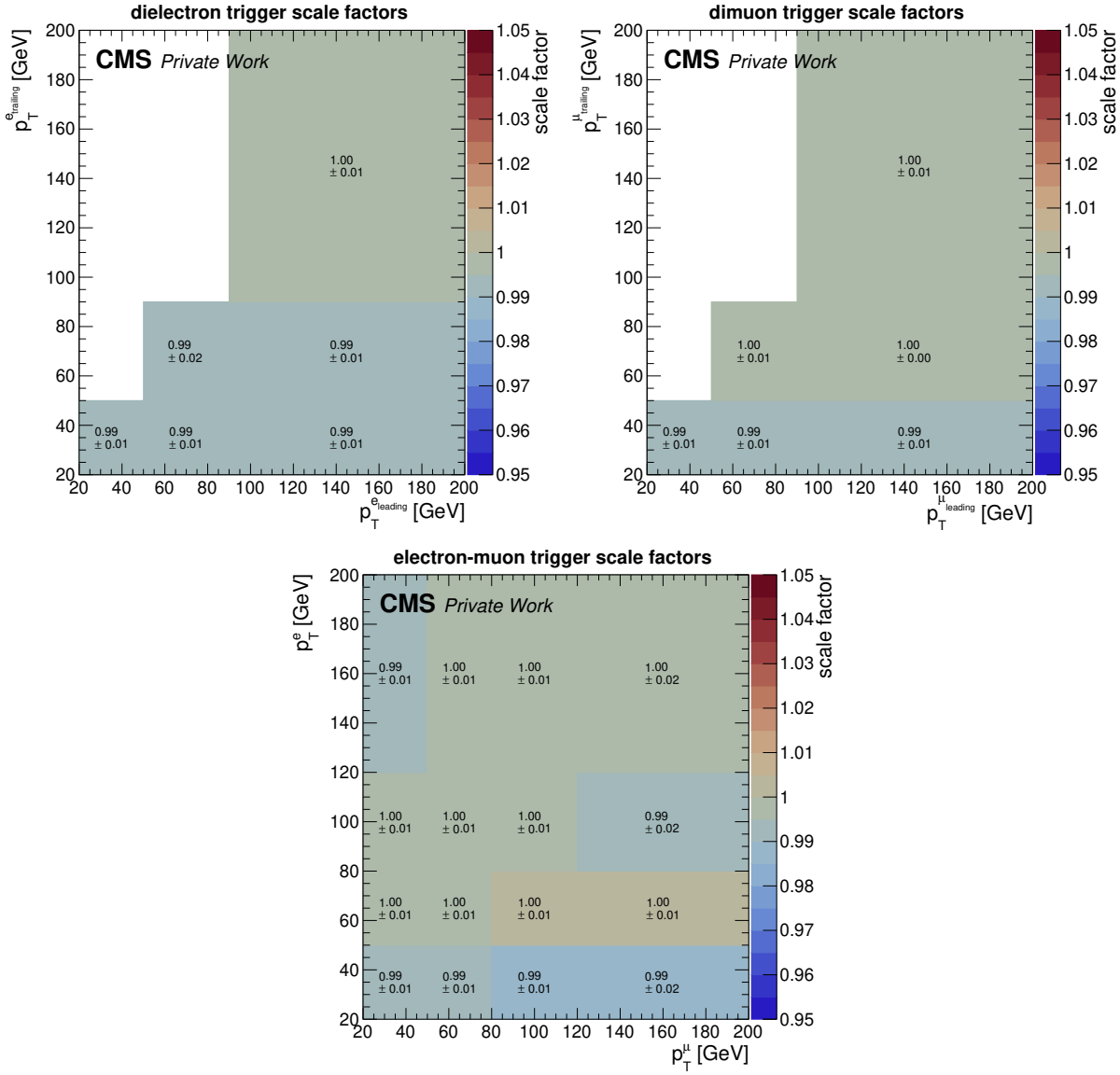

Figure 6.2: Trigger scale factors for the e^+e^- (upper left), $\mu^+\mu^-$ (upper right), and $e^\pm\mu^\mp$ (bottom) lepton channels. The uncertainties denote the quadratic sum of statistical and systematic uncertainties.

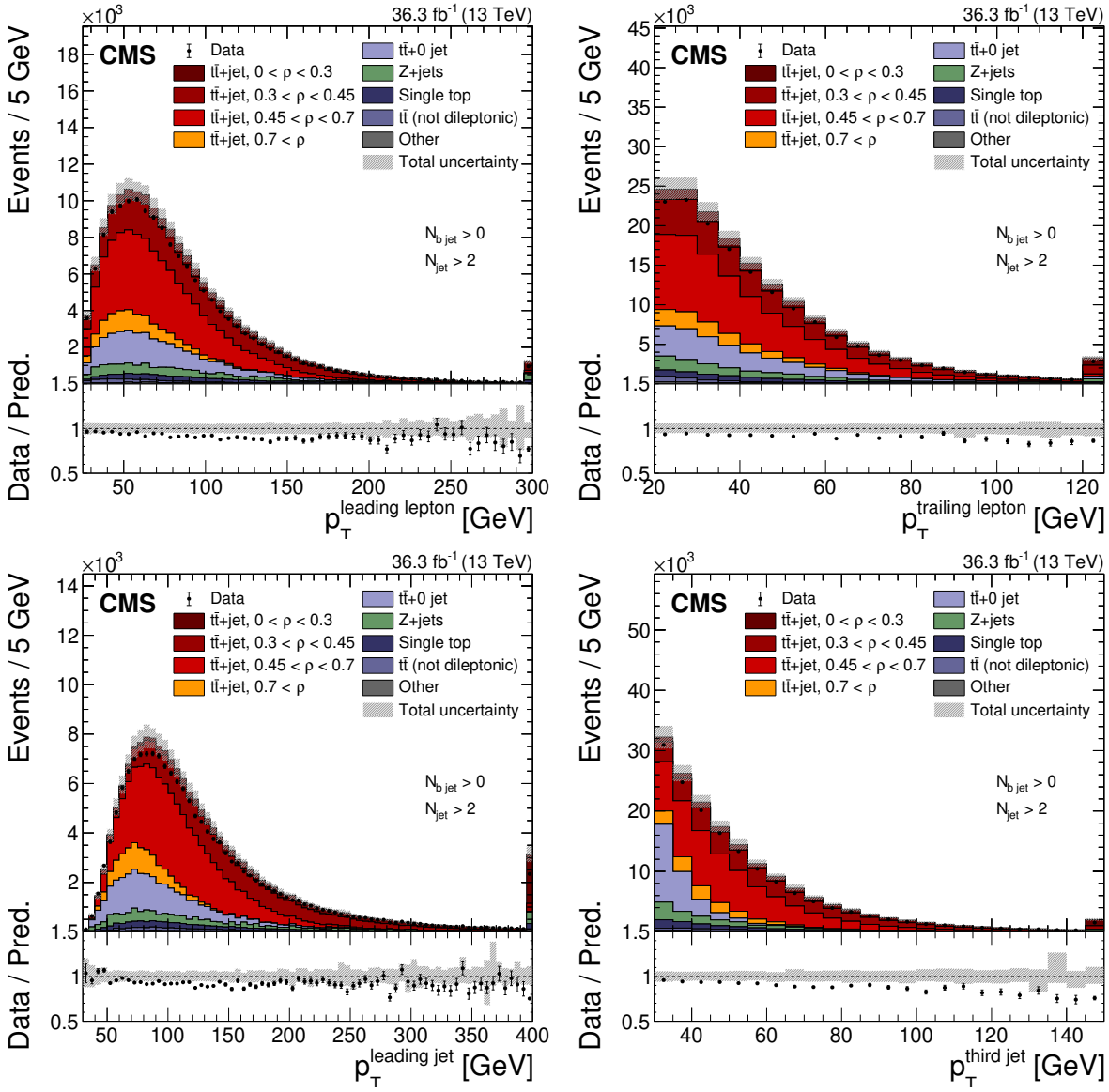

Figure 6.3: The observed (points) and predicted (stacked histograms) signal and background yields as a function of the leading (upper left) and subleading (upper right) lepton p_{T}, and the leading (lower left) and third-highest (lower right) jet p_{T}, after applying the preselection. The vertical bars on the points represent the statistical uncertainty in the data. The hatched band represents the total uncertainty in the sum of the simulated signal and background predictions. The lower panels show the ratio of the data to the sum of the signal and background predictions [49].

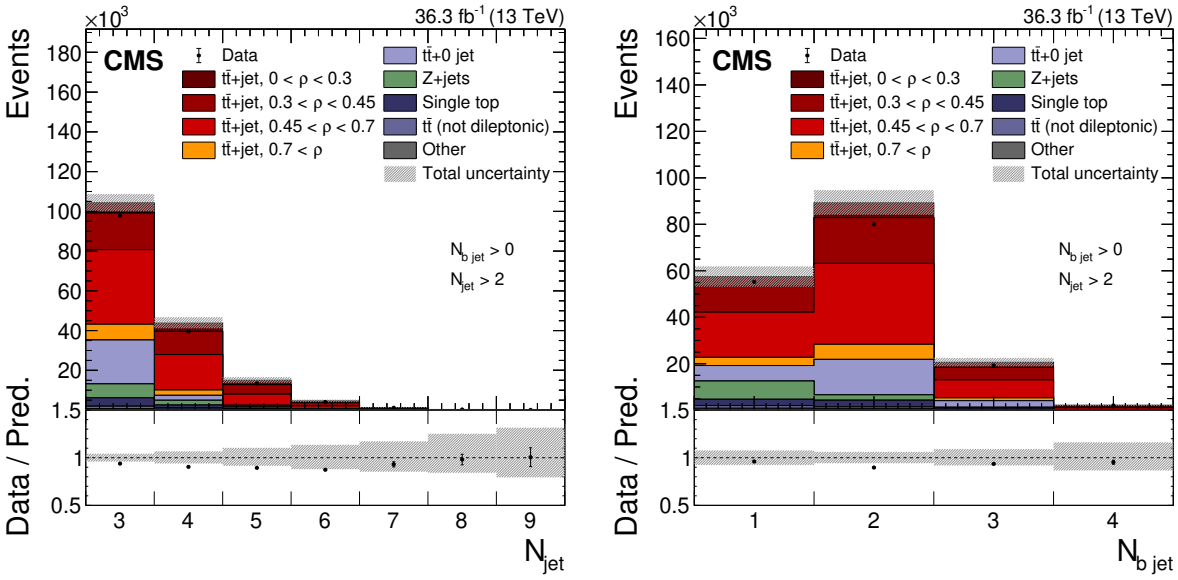

Figure 6.4: The observed (points) and predicted (stacked histograms) signal and background yields as a function of the jet (left), and b jet (right) multiplicities, after applying the preselection. The vertical bars on the points represent the statistical uncertainty in the data. The hatched band represents the total uncertainty in the sum of the simulated signal and background predictions. The lower panels show the ratio of the data to the sum of the signal and background predictions [49].

for the mass of the W bosons based on their decay products:

$$m^2_{\mathrm{W}^+} = (E_{\ell^+} + E_\nu)^2 - (p_{x,\ell^+} + p_{x,\nu})^2 - (p_{y,\ell^+} + p_{y,\nu})^2 - (p_{z,\ell^+} + p_{z,\nu})^2, \tag{6.5}$$

$$m^2_{\mathrm{W}^-} = (E_{\ell^-} + E_{\overline{\nu}})^2 - (p_{x,\ell^-} + p_{x,\overline{\nu}})^2 - (p_{y,\ell^-} + p_{y,\overline{\nu}})^2 - (p_{z,\ell^-} + p_{z,\overline{\nu}})^2, \tag{6.6}$$

and finally, for the top quark and antiquark, the following relations:

$$\begin{aligned} m^2_{\mathrm{t}} &= (E_{\ell^+} + E_\nu + E_{\mathrm{b}})^2 - (p_{x,\ell^+} + p_{x,\nu} + p_{x,\mathrm{b}})^2 \\ &- (p_{y,\ell^+} + p_{y,\nu} + p_{y,\mathrm{b}})^2 - (p_{z,\ell^+} + p_{z,\nu} + p_{z,\mathrm{b}})^2, \end{aligned} \tag{6.7}$$

$$\begin{aligned} m^2_{\overline{\mathrm{t}}} &= (E_{\ell^-} + E_{\overline{\nu}} + E_{\overline{\mathrm{b}}})^2 - (p_{x,\ell^-} + p_{x,\overline{\nu}} + p_{x,\overline{\mathrm{b}}})^2 \\ &- (p_{y,\ell^-} + p_{y,\overline{\nu}} + p_{y,\overline{\mathrm{b}}})^2 - (p_{z,\ell^-} + p_{z,\overline{\nu}} + p_{z,\overline{\mathrm{b}}})^2. \end{aligned} \tag{6.8}$$

Primarily, the subject of the reconstruction is to obtain solutions for the unknown neutrino momenta, and reconstruct the $\mathrm{t\bar{t}}$ system thereafter.

The set of the six nonlinear Equations 6.4 to 6.8 cannot be solved directly, but by assuming that the measured missing transverse $E_{\mathrm{T}}^{\mathrm{miss}}$ originates from the two neutrinos, the following additional relation can be inserted:

$$\begin{aligned} E_x^{\mathrm{miss}} &= p_{x,\nu} + p_{x,\overline{\nu}}, \\ E_y^{\mathrm{miss}} &= p_{y,\nu} + p_{y,\overline{\nu}}. \end{aligned} \tag{6.9}$$

Neglecting neutrino masses, the energies E_ν can be obtained by using Equation 6.4. The masses of the b quarks are set to the values used in the simulation, while lepton masses are assumed to

be negligible. The remaining set of equations has six unknown parameters $p_{x,\nu}$, $p_{y,\nu}$, $p_{z,\nu}$, $p_{x,\bar{\nu}}$, $p_{y,\bar{\nu}}$, and $p_{z,\bar{\nu}}$, which can be obtained using the following constraints:

- Two unknown parameters are removed by using the momentum balance conditions in Equation 6.9.
- By fixing the masses of the two W bosons to 80.4 GeV [248], two more unknowns can be removed.
- Similarly, the top quark and antiquark mass of 172.5 GeV can be assumed, eliminating the last two unknowns.

Without loss of generality, the set of equations can be transformed into a single equation in the form of a fourth order polynomial depending only on the remaining unknown $p_{x,\bar{\nu}}$:

$$0 = \sum_{i=0}^{4} c_i(m_{\mathrm{t}}, p_{\ell^+}, p_{\ell^-}, p_{\mathrm{b}}, p_{\bar{\mathrm{b}}})\, p_{x,\bar{\nu}}^{i}. \tag{6.10}$$

The remaining coefficients c_i depend only on the final state four momenta, and the equation can be solved analytically with a maximum four-fold ambiguity. For this thesis, the solution with the smallest invariant mass of the $t\bar{t}$ pair is selected in case of multiple solutions. This choice has been motivated in Ref. [328] and yields the best solution in a wide kinematic range.

Depending on the jet multiplicity, the number of permutations grows as $\frac{N_{\mathrm{jet}}!}{(N_{\mathrm{jet}}-2)!}$. Here, the preference is always given to the permutation with the highest b jet multiplicity, if a physical solution exists. In case multiple solutions are found for the same b jet multiplicity, weights are derived based on the mass of the reconstructed lepton and b jet pairs. The extraction of the weights is discussed in the following.

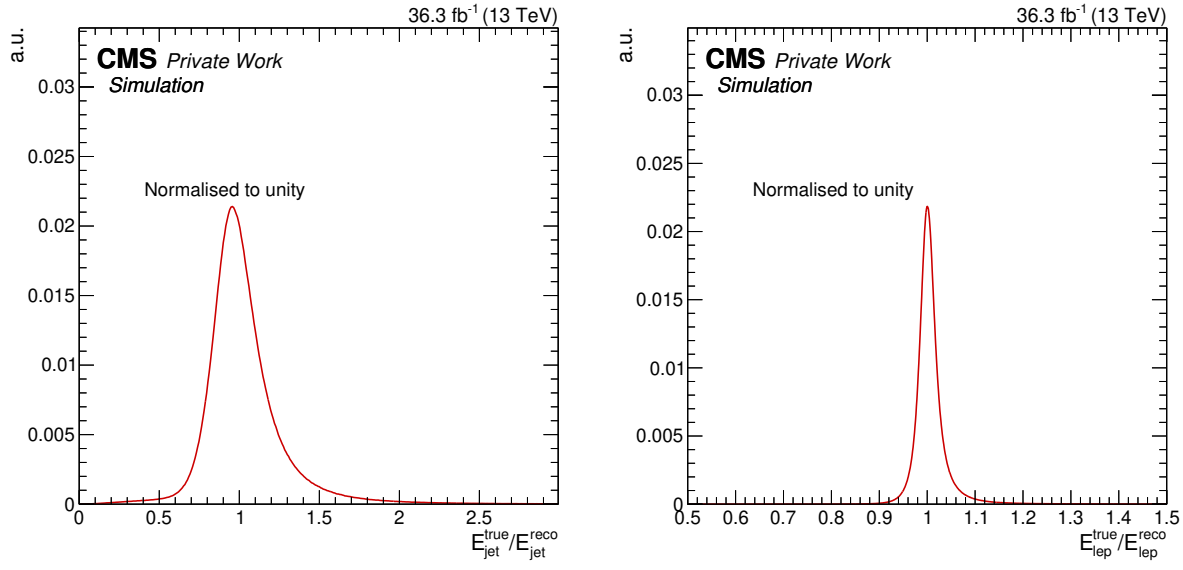

Figure 6.5: Distributions of the energy factors used for the energy smearing in the full kinematic reconstruction of the top quark kinematic properties. The factors are shown for the b quarks (left) and for the leptons (right) .

Given the detector response and the resolution effects, for a sizable amount of events no solution can be determined. To mitigate this issue, in each event the energies and the directions of the two jets and leptons are smeared. The smearing functions are derived using simulation

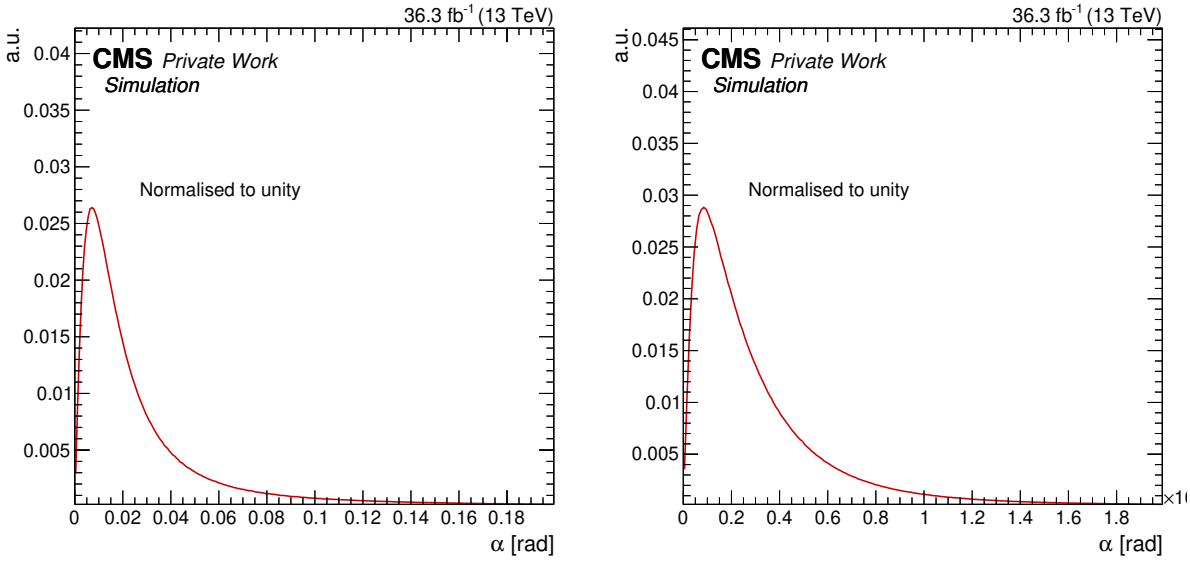

Figure 6.6: Distributions of the angle between the true and the reconstructed directions for jets (left) and leptons (right) used in the smearing for the full kinematic reconstruction.

and are based on the measured and true momentum of the leptons and b jets stemming from the top quarks. The energy smearing is performed using a correction factor $E^{\text{true}}/E^{\text{reco}}$, and corresponding curves are shown in Figure 6.5. For the angular smearing, a random Gaussian smearing is performed, where the width of the Gaussian is set to the measured angular resolution α. Distributions of α are displayed in Figure 6.6. Further, the mass of the W boson is varied according to a relativistic Breit–Wigner function, estimated using the generator W boson mass. This is illustrated in Figure 6.7 (left). The smearing procedure is conducted 100 times for each event, and for each solution a weight $w = w_{m_{\bar{\ell}\text{b}}} \times w_{m_{\ell\bar{\text{b}}}}$ is derived. The values of $w_{m_{\bar{\ell}\text{b}}}$ and $w_{m_{\ell\bar{\text{b}}}}$ are taken from true distributions at the particle level as expected from the simulation. The curve for $m_{\ell\text{b}}$ is shown in Figure 6.7 (right). The final solution for a given lepton-jet permutation is calculated as the weighted average $\langle \vec{p}_{\text{t}} \rangle$ over all solutions:

$$\langle \vec{p}_{\text{t}} \rangle = \frac{1}{w_s} \sum_{i=1}^{100} w_i \vec{p}_{\text{t},\,i}, \quad \text{with} \quad w_s = \sum_{i=1}^{100} w_i. \tag{6.11}$$

Here w_i denotes the weight and $\vec{p}_{\text{t},\,i}$ is the reconstructed top quark three momentum obtained for the i-th smearing of the event. If for a given smearing no kinematic solution is found, the values of w_i and $\vec{p}_{\text{t},\,i}$ are set to zero. The four momentum vector of the top quark is determined by its energy, which is calculated from $\langle \vec{p}_{\text{t}} \rangle$, and the top quark mass of 172.5 GeV. The kinematic properties of the top antiquark is analogously determined.

The loose kinematic reconstruction

Alternatively, another algorithm is used to reconstruct the $\text{t}\bar{\text{t}}$ kinematic properties as a whole without using the m_{t} constraint, which makes this method especially useful for the determination of m_{t}. It was originally developed in the context of the measurement described in Ref. [38] and is also investigated in the analysis of this thesis. This approach is referred to in the following as "loose kinematic reconstruction". Inputs for this method are also the two leptons, the missing transverse momentum, and the two b jets. Because the $\text{t}\bar{\text{t}}$ pair is directly reconstructed, the

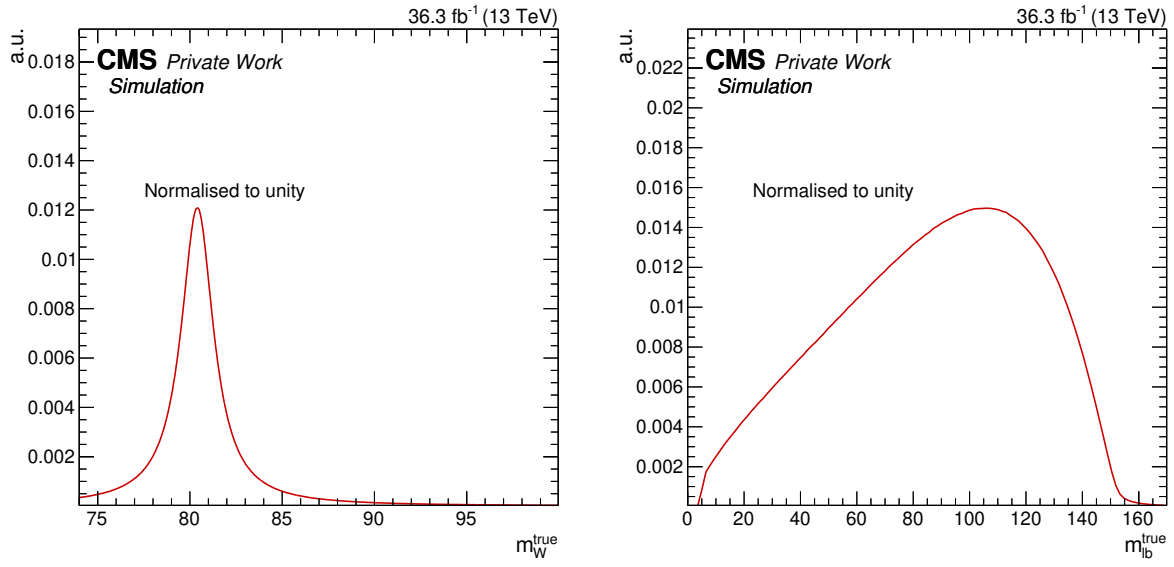

Figure 6.7: Distributions of the true W boson mass (left) and m_{lb} distribution (right).

lepton pair is used in contrast to individual leptons. This reduces the amount of possible permutations drastically. Two jets are selected via the following ranking and selection procedure:

- the invariant mass with one of the leptons $m_{\ell \mathrm{b}}$ must be smaller than 180 GeV,
- b-tagged jets are preferred over non b-tagged jets,
- and for the same b jet multiplicity, jets are ranked in p_{T}.

The $m_{\ell \mathrm{b}}$ variable is determined for each pair of jets individually, while the maximum is chosen for the top quark and antiquark decays, and the minimum over the two jet-lepton assignments is considered.

The transverse momentum of the neutrinos, which are assumed to be collinear to the charged leptons of the W decay, is solved as follows:

1. Set the transverse components of the dineutrino four-vector to:
   $p_{x,\nu\bar{\nu}} = E^{\mathrm{miss}}_{x,\nu\bar{\nu}}$ and $p_{y,\nu\bar{\nu}} = E^{\mathrm{miss}}_{y,\nu\bar{\nu}}$.
2. Determine the longitudinal momentum of the dineutrino system as:
   If $p_{\mathrm{T},\nu\bar{\nu}} < E_{\ell\bar{\ell}}$, set $p_{z,\nu\bar{\nu}} = p_{z,\ell\bar{\ell}}$, or $p_{z,\nu\bar{\nu}} = 0$ otherwise.
3. Determine the energy of the dineutrino system via:
   If $p_{\nu\bar{\nu}} < E_{\ell\bar{\ell}}$, set $E_{\nu\bar{\nu}} = E_{\ell\bar{\ell}}$, or $E_{\nu\bar{\nu}} = p_{\ell\bar{\ell}}$ otherwise.
4. Calculate the four momentum sum of of $\ell\bar{\ell}\nu\bar{\nu}$.
5. In case $m_{\ell\bar{\ell}\nu\bar{\nu}} < 2m_{\mathrm{W}} = 2 \times 80.4\,\mathrm{GeV}$, the four momentum of $\ell\bar{\ell}\bar{\nu}$ is parameterized with p_x, p_y, y, m, and m is set to $2 \times m_{\mathrm{W}}$.
6. The four momentum sum of $t\bar{t}$ is calculated by using the four-momenta of the $\ell\bar{\ell}\nu\bar{\nu}$ system, as determined in step 3., and of the two b jets b and $\bar{\mathrm{b}}$ via $\ell\bar{\ell}\nu\bar{\nu} + \mathrm{b}\bar{\mathrm{b}}$.

The reconstruction efficiency $\epsilon_{\text{kin. reco.}} = \frac{N_{\text{solved}}}{N_{\text{total}}}$ for both algorithms is shown in Figure 6.8 as a function of the number of reconstructed jets and the leading jet p_{T}. It is calculated by deriving the ratio between the number of events with a valid solution and the number of events

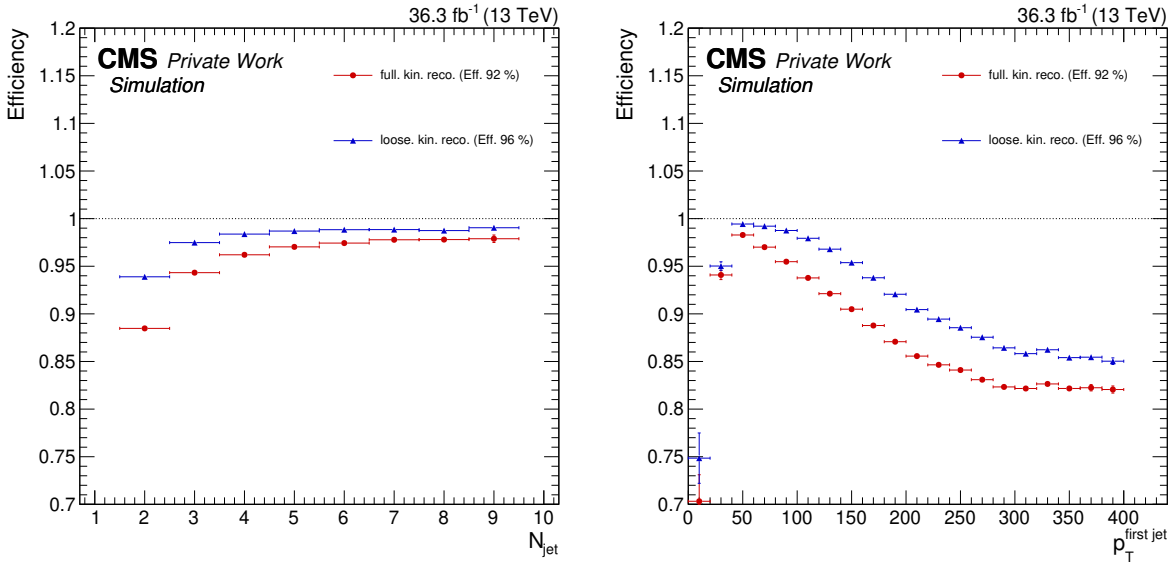

Figure 6.8: Efficiencies for the full (red) and loose kinematic reconstruction (blue) for $t\bar{t}$ dileptonic events in bins of the jet multiplicity (left) and the leading jet p_T (right). The uncertainties shown are of statistical origin.

used as inputs for the algorithm. The study is performed using the $t\bar{t}$ signal simulation. The integrated efficiency of the full (loose) kinematic reconstruction is 92% (96%). A significant trend with respect to the jet p_T and the number of jets can be observed, with the efficiency in both kinematic extremes going down to about 85%. The low p_T range of 30–50 GeV has a 5% inefficiency, which is especially relevant for the presented measurement.

Multivariate analysis using neural networks

Since both, the full and the loose, kinematic reconstructions are developed for $t\bar{t}$ production without additional jet radiation, an MVA approach is developed in this thesis based on a regression NN in order to optimize the resolution in the ρ observable. Furthermore, the efficiency and in turn the acceptance is maximized by employing an MVA method. Using the TensorFlow [329] package and the Keras [330] backend, a fully connected feed-forward NN is trained.

The NN is trained using preselected events, as described in Section 6.2 with the additional requirement of at least three reconstructed jets. The training sample, which is statistically independent of the analysis sample, is obtained using a $t\bar{t}$ MC events sample simulated with the MG5_aMC@NLO[FXFX] event generator interfaced with Pythia8. The final performance of the NN is cross-checked using alternatively the POWHEG+Pythia8 simulation. A combined training for all lepton channels (e^+e^-, $e^\pm\mu^\mp$, $\mu^+\mu^-$) is performed, because no significant improvement is observed deriving separate calibrations.

The full set of input variables considered for the training of the NN is listed in Tables 6.1 to 6.4. These include low-level variables as basic four momenta of selected final-state particles, but also high-level variables as solutions of the kinematic reconstruction algorithms used. In total, 102 variables are considered. They can be grouped into the following categories:

- **Individual kinematic quantities** include basic four-momenta or tag information of jets and leptons. Only the properties of the first three leading jets are used. These are considered low-level variables.

- **Event variables** contain general event kinematic properties, like the jet or b jet multiplicity in the event. These are considered low-level and high-level variables, depending on the variable.
- **Geometric variables** are computed for all combinations of objects, or object systems, in an event. These are considered high-level variables.
- **Mass variables** are also derived for different object- or object-system combinations. These are considered high-level variables.
- **Solutions of the kinematic reconstruction** methods employed give rise to $t\bar{t}$, t, $\bar{t}$, and additional jet information. These are considered high-level variables.

Table 6.1: Total set of input variables considered for the NN training, first part. The checkmarks indicate if the variable is used for the final regression or classification NN.

Name	Symbol	Description	Reg.	Class.
`jet1Pt`	$p_{\mathrm{T}}^{\text{leading jet}}$	p_{T} of the leading jet		✓
`jet1Eta`	$\eta^{\text{leading jet}}$	η of the leading jet		
`jet1Phi`	$\phi^{\text{leading jet}}$	ϕ of the leading jet		
`jet1M`	$m^{\text{leading jet}}$	mass of the leading jet		
`jet1BTag`	b tag$^{\text{leading jet}}$	boolean b tagged status of the leading jet		
`jet2Pt`	$p_{\mathrm{T}}^{\text{second jet}}$	p_{T} of the second-hardest jet		
`jet2Eta`	$\eta^{\text{second jet}}$	η of the second-hardest jet		
`jet2Phi`	$\phi^{\text{second jet}}$	ϕ of the second-hardest jet		
`jet2M`	$m^{\text{second jet}}$	mass of the second-hardest jet		
`jet2BTag`	b tag$^{\text{second jet}}$	boolean b tagged status of the second-hardest jet		
`jet3Pt`	$p_{\mathrm{T}}^{\text{third jet}}$	p_{T} of the third-hardest jet		✓
`jet3Eta`	$\eta^{\text{third jet}}$	η of the third-hardest jet		
`jet3Phi`	$\phi^{\text{third jet}}$	ϕ of the third-hardest jet		
`jet3M`	$m^{\text{third jet}}$	mass of the third-hardest jet		
`jet3BTag`	b tag$^{\text{third jet}}$	boolean b tagged status of the third-hardest jet		
`lep1Pt`	$p_{\mathrm{T}}^{\text{leading lepton}}$	p_{T} of the leading lepton		✓
`lep1Eta`	$\eta^{\text{leading lepton}}$	p_{T} of the leading lepton		
`lep1Phi`	$\eta^{\text{leading lepton}}$	η of the leading lepton		
`lep1M`	$m^{\text{leading lepton}}$	mass of the leading lepton		✓
`lep2Pt`	$p_{\mathrm{T}}^{\text{trailing lepton}}$	p_{T} of the trailing lepton	✓	
`lep2Eta`	$\eta^{\text{trailing lepton}}$	η of the trailing lepton		
`lep2Phi`	$\phi^{\text{trailing lepton}}$	ϕ of the trailing lepton		
`lep2M`	$m^{\text{trailing lepton}}$	mass of the trailing lepton		✓
`dilepPt`	$p_{\mathrm{T}}^{\ell\ell}$	p_{T} of the dilepton system		✓
`dilepEta`	$\eta^{\ell\ell}$	η of the dilepton system		
`dilepPhi`	$\phi^{\ell\ell}$	ϕ of the dilepton system		
`dilepM`	$m_{\ell\ell}$	mass of the dilepton system	✓	✓
`metPt`	$p_{\mathrm{T}}^{\text{miss}}$	missing transverse momentum	✓	✓
`metPhi`	ϕ^{miss}	ϕ component of $p_{\mathrm{T}}^{\text{miss}}$		

Table 6.2: Total set of input variables considered for the NN training, second part. The checkmarks indicate if the variable is used for the regression or classification NN.

Name	Symbol	Description	Reg.	Class.
`llj1Pt`	$p_{\mathrm{T}}^{\ell\ell j_1}$	p_{T} of the dilepton plus leading jet system		
`llj1Eta`	$\eta^{\ell\ell j_1}$	η of the dilepton plus leading jet system		
`llj1Phi`	$\phi^{\ell\ell j_1}$	ϕ of the dilepton plus leading jet system		
`llj1M`	$m^{\ell\ell j_1}$	mass of the dilepton plus leading jet system	✓	
`llj2Pt`	$p_{\mathrm{T}}^{\ell\ell j_2}$	p_{T} of the dilepton plus second-hardest jet system		
`llj2Eta`	$\eta^{\ell\ell j_2}$	η of the dilepton plus second-hardest jet system		
`llj2Phi`	$\phi^{\ell\ell j_2}$	ϕ of the dilepton plus second-hardest jet system		
`llj2M`	$m^{\ell\ell j_2}$	mass of the dilepton plus second-hardest jet system	✓	
`llj3Pt`	$p_{\mathrm{T}}^{\ell\ell j_3}$	p_{T} of the dilepton plus third-hardest jet system		
`llj3Eta`	$\eta^{\ell\ell j_3}$	η of the dilepton plus third-hardest jet system		
`llj3Phi`	$\phi^{\ell\ell j_3}$	ϕ of the dilepton plus third-hardest jet system		
`llj3M`	$m^{\ell\ell j_3}$	mass of the dilepton plus third-hardest jet system		
`l1j1Pt`	$p_{\mathrm{T}}^{\ell_1 j_1}$	p_{T} of the leading lepton plus leading jet system		
`l1j1Eta`	$\eta^{\ell_1 j_1}$	η of the leading lepton plus leading jet system		
`l1j1Phi`	$\phi^{\ell_1 j_1}$	ϕ of the leading lepton plus leading jet system		
`l1j1M`	$m^{\ell_1 j_1}$	mass of the leading lepton plus leading jet system		
`l1j2Pt`	$p_{\mathrm{T}}^{\ell_1 j_2}$	p_{T} of the leading lepton plus second-hardest jet system		
`l1j2Eta`	$\eta^{\ell_1 j_2}$	η of the leading lepton plus second-hardest jet system		
`l1j2Phi`	$\phi^{\ell_1 j_2}$	ϕ of the leading lepton plus second-hardest jet system		
`l1j2M`	$m^{\ell_1 j_2}$	mass of the leading lepton plus second-hardest jet system	✓	
`l1j3Pt`	$m^{\ell_1 j_3}$	p_{T} of the leading lepton plus third-hardest jet system		
`l1j3Eta`	$\eta^{\ell_1 j_3}$	η of the leading lepton plus third-hardest jet system		
`l1j3Phi`	$\phi^{\ell_1 j_3}$	ϕ of the leading lepton plus third-hardest jet system		
`l1j3M`	$m^{\ell_1 j_3}$	mass of the leading lepton plus third-hardest jet system		
`l2j1Pt`	$p_{\mathrm{T}}^{\ell_2 j_1}$	p_{T} of the trailing lepton plus leading jet system		
`l2j1Eta`	$\eta^{\ell_2 j_1}$	η of the trailing lepton plus leading jet system		
`l2j1Phi`	$\phi^{\ell_2 j_1}$	ϕ of the trailing lepton plus leading jet system		
`l2j1M`	$m^{\ell_2 j_1}$	mass of the trailing lepton plus leading jet system	✓	
`l2j2Pt`	$p_{\mathrm{T}}^{\ell_2 j_2}$	p_{T} of the trailing lepton plus second-hardest jet system		
`l2j2Eta`	$\eta^{\ell_2 j_2}$	η of the trailing lepton plus second-hardest jet system		
`l2j2Phi`	$\phi^{\ell_2 j_2}$	ϕ of the trailing lepton plus second-hardest jet system		
`l2j2M`	$m^{\ell_2 j_2}$	mass of the trailing lepton plus second-hardest jet system	✓	
`l2j3Pt`	$p_{\mathrm{T}}^{\ell_2 j_3}$	p_{T} of the trailing lepton plus third-hardest jet system		
`l2j3Eta`	$\eta^{\ell_2 j_3}$	η of the trailing lepton plus third-hardest jet system		
`l2j3Phi`	$\phi^{\ell_2 j_3}$	ϕ of the trailing lepton plus third-hardest jet system		
`l2j3M`	$m^{\ell_2 j_3}$	mass of the trailing lepton plus third-hardest jet system		

Table 6.3: Total set of input variables considered for the NN training, third part. The checkmarks indicate if the variable is used for the final regression or classification NN.

Name	Symbol	Description	Reg.	Class.
`ht`	H_T	the transverse sum of all hadronic momentum		
`nBJets`	$N_{\mathrm{b\,jet}}$	number of b-tagged jets		
`nJets`	N_jet	number of jets		✓
`mlbmin`	$m_{\ell\mathrm{b}}^{\mathrm{min}}$	minimal mass of lepton and b jet combinations		
`l1j1DR`	$\Delta R(\ell_1, j_1)$	ΔR between leading lepton and hardest jet		
`l2j1DR`	$\Delta R(\ell_2, j_1)$	ΔR between trailing lepton and hardest jet		
`l1j2DR`	$\Delta R(\ell_1, j_1)$	ΔR between leading lepton and second-hardest jet		
`l2j2DR`	$\Delta R(\ell_2, j_1)$	ΔR between trailing lepton and second-hardest jet		
`l1j3DR`	$\Delta R(\ell_1, j_1)$	ΔR between leading lepton and third-hardest jet		
`l2j3DR`	$\Delta R(\ell 2 j_1)$	ΔR between trailing lepton and third-hardest jet		
`j1j2DR`	$\Delta R(j_1, j_2)$	ΔR between hardest and second-hardest jet		
`j1j3DR`	$\Delta R(j_1, j_3)$	ΔR between hardest and third-hardest jet		
`j2j3DR`	$\Delta R(j_2, j_3)$	ΔR between second-hardest and third-hardest jet		
`kr_rho`	$\rho_{\mathrm{kin.\,reco.}}$	ρ from the full kin. reco.	✓	
`kr_ttbarPt`	$p_\mathrm{T}^{\mathrm{t\bar{t}_{kin.\,reco.}}}$	p_T of the $\mathrm{t\bar{t}}$ system from the full kin. reco.		
`kr_ttbarEta`	$\eta_{\mathrm{t\bar{t}}}^{\mathrm{kin.\,reco.}}$	η of the $\mathrm{t\bar{t}}$ system from the full kin. reco.		
`kr_ttbarPhi`	$\phi_{\mathrm{t\bar{t}}}^{\mathrm{kin.\,reco.}}$	ϕ of the $\mathrm{t\bar{t}}$ system from the full kin. reco.		
`kr_ttbarM`	$m_{\mathrm{t\bar{t}}}^{\mathrm{kin.\,reco.}}$	mass of the $\mathrm{t\bar{t}}$ system from the full kin. reco.		
`kr_topPt`	$p_\mathrm{T}^{\mathrm{t_{kin.\,reco.}}}$	p_T of the t from the full kin. reco.		
`kr_topEta`	$\eta_{\mathrm{t}}^{\mathrm{kin.\,reco.}}$	η of the t from the full kin. reco.		
`kr_topPhi`	$\phi_{\mathrm{t}}^{\mathrm{kin.\,reco.}}$	ϕ of the t from the full kin. reco.		
`kr_atopPt`	$p_\mathrm{T}^{\mathrm{\bar{t}_{kin.\,reco.}}}$	p_T of the $\bar{\mathrm{t}}$ from the full kin. reco.		
`kr_atopEta`	$\eta_{\bar{\mathrm{t}}}^{\mathrm{kin.\,reco.}}$	η of the $\bar{\mathrm{t}}$ from the full kin. reco.		
`kr_atopPhi`	$\phi_{\bar{\mathrm{t}}}^{\mathrm{kin.\,reco.}}$	ϕ of the $\bar{\mathrm{t}}$ from the full kin. reco.		
`kr_jetPt`	$p_\mathrm{T}^{\mathrm{add.\ jet_{kin.\,reco.}}}$	p_T of the add. jet from the full kin. reco.		✓
`kr_jetEta`	$\eta_{\mathrm{add.\ jet}}^{\mathrm{kin.\,reco.}}$	η of the add. jet from of the full kin. reco.		
`kr_jetPhi`	$\phi_{\mathrm{add.\ jet}}^{\mathrm{kin.\,reco.}}$	ϕ of the add. jet from of the full kin. reco.		
`kr_jetM`	$m_{\mathrm{add.\ jet}}^{\mathrm{kin.\,reco.}}$	mass of the add. jet from the full kin. reco.		

Starting from the full set of variables, using as as first step Bayesian hyperparameter optimization [286, 287, 331], cf. Section 5.2, a subset containing the ten most relevant input variables is selected. This selection procedure is performed using a dedicated software package [332]. It is based on the concept of Shapley values as used in models of cooperative game theory. Each input variable is considered a "player", whereas the prediction is considered the "payout" [333]. The contribution of each player to the outcome, i.e., the prediction, is computed, while some approximations to calculate gradients of the NN output with respect to its input in order to determine the Shapley values are made. Various techniques as proposed in the literature [334, 335] are combined in this approach.

Table 6.4: Total set of input variables considered for the NN training, fourth part. The checkmarks indicate if the variable is used for the final regression or classification NN.

Name	Symbol	Description	Reg.	Class.
`lkr_rho`	$\rho_{\text{loose kin. reco.}}$	ρ from the loose kin. reco.	✓	
`lkr_ttbarPt`	$p_{\text{T}}^{\text{t}\bar{\text{t}}_{\text{loose kin. reco.}}}$	p_{T} of the $\text{t}\bar{\text{t}}$ system from the loose kin. reco.		
`lkr_ttbarEta`	$\eta_{\text{t}\bar{\text{t}}}^{\text{loose kin. reco.}}$	η of the $\text{t}\bar{\text{t}}$ system from the loose kin. reco.		
`lkr_ttbarPhi`	$\phi_{\text{t}\bar{\text{t}}}^{\text{loose kin. reco.}}$	ϕ of the $\text{t}\bar{\text{t}}$ system from the loose kin. reco.		
`lkr_ttbarM`	$m_{\text{t}\bar{\text{t}}}^{\text{loose kin. reco.}}$	mass of the $\text{t}\bar{\text{t}}$ system from the loose kin. reco.		
`lkr_jetPt`	$p_{\text{T}}^{\text{add. jet}_{\text{loose kin. reco.}}}$	p_{T} of the add. jet from the loose kin. reco.		
`lkr_jetEta`	$\eta_{\text{add. jet}}^{\text{loose kin. reco.}}$	η of the add. jet from the loose kin. reco.		
`lkr_jetPhi`	$\phi_{\text{add. jet}}^{\text{loose kin. reco.}}$	ϕ of the add. jet from the loose kin. reco.		
`lkr_jetM`	$m_{\text{add. jet}}^{\text{loose kin. reco.}}$	mass of the add. jet from the loose kin. reco.		

The relevance and the ranking of the input variables are shown in Figure 6.9 for the final set of ten inputs. The major information in the reconstruction is given by the two classical kinematic

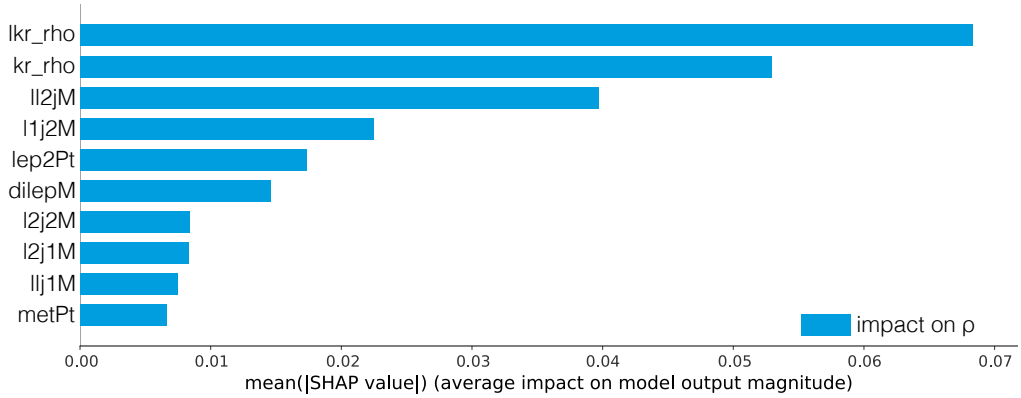

Figure 6.9: Input variable ranking of the ten most relevant input variables for the regression NN.

reconstruction solutions, and is supported by kinematic variables characterizing the energy and mass of various lepton–jet systems.

Ordered by their impact, the relevant input variables are

- the solution for ρ using the loose kinematic reconstruction,
- the solution for ρ using the full kinematic reconstruction,
- the invariant mass of the dilepton and subleading jet system,
- the invariant mass of the leading lepton and subleading jet system,
- the p_{T} of the subleading lepton,
- the invariant mass of the dilepton,
- the invariant mass of the subleading lepton and subleading jet system,
- the invariant mass of the subleading lepton and leading jet system,

- the invariant mass of the dilepton and leading jet system,
- and $p_{\mathrm{T}}^{\mathrm{miss}}$.

If the input variables are not defined for an event like in the case of the kinematic reconstruction, default values of zero are used.

Since NNs are able to learn complex correlations and representations of the input variable space, these are sensitive to imperfect modeling using the MC simulation. That can lead to discrepancies in the NN response when evaluating it using data and simulated samples. In order to ensure a good description of the variables using the simulation to be compared to data, an input validation is performed for the chosen set of variables. The test procedure and the result are independent of the true cross section in the data and are not sensitive to m_{t}. Binned GoF tests including the full uncertainty model (i.e. statistical and systematic uncertainties as described in Section 6.6) using a saturated model, cf. Section 5.1 are performed as follows. Simultaneous fits for all lepton channels ($\mathrm{e^+e^-}$, $\mu^+\mu^-$, $\mathrm{e^\pm}\mu^\mp$) are carried out, making use of all parameter correlations. A threshold on the p-value measured on the marginal distribution that is returned by the GoF test is set to 5%. Input variables with a p-value below 5% are discarded. An observed test-statistics t_0 is computed by fitting the model to the data. The expected test-statistic distribution $f(t)$ is calculated by performing the same fit multiple times, using toy data sampled from the expected distributions. In total, 500 toys are considered. Consequently, the final p-value is calculated using $p = \int_{t_0}^{\infty} f(t)dt$. An exemplary GoF test result for the $\rho_{\text{kin. reco.}}$ variable is shown in Figure 6.10. The distribution for the toy test statistics is compared to the observed test statistics t_0. A p-value of 1.00 is computed, suggesting very good agreement between the simulation and the data. None of the ten considered variables is observed to have a p-value below 8%. All p-values are given in Table 6.5.

The event sample obtained using the MG5_aMC@NLO[FXFX] simulation is split into independent train, validation, and testing sets in the ratio of 0.5–0.2–0.3. All corrections used as in the further measurement are applied. Additional event weights are derived based on the true ρ distribution in 200 equally sized bins to remove any possible bias coming from the kinematic shape of the ρ distribution and ensure the same importance in the loss functions for all events despite its ρ value.

The NN architecture used consists of one input layer, two hidden layers with 512 nodes each, and one output layer. Details of the used methods can be found in Section 5.2. SELU activation functions [283] are used for all hidden layers, while a linear activation function is used for the one dimensional output layer. All layers, except for the output layer, are complemented with batch normalization layers [288] to improve the minimization process and result in a more generalizable model. To reduce potential overtraining, randomized dropout units [290] with a "drop-probability" of 0.3601 are complemented by applying L2 regularization with a weight of 7.9857×10^{-5}. In addition, the Glorot normal method [289] is used to initialize the parameters of the weight tensors. The weights of each hidden unit are required to have a norm ≤ 5, as it is observed to stabilize the training process. A batch size of 25723 is used, and the learning rate is set to 1×10^{-4}. In total, about 1.3 million events are used for the training itself. All hyperparameters, i.e., the batch size, the L2 regularization strength, the dropout rate, the number of layers, the number of neurons per layer, and the type of activation function, are optimized in a Bayesian optimization procedure. The target function to be optimized during the Bayesian optimization procedure is the Jensen–Shannon distance [336] between the true, ρ_{true}, and predicted, ρ_{reco}, values of the test data set. Since one of the properties of the Jensen–Shannon distance is that it yields a measure of the similarity between the distributions, directly this similarity can be maximized. During the NN training, the mean absolute percentage error

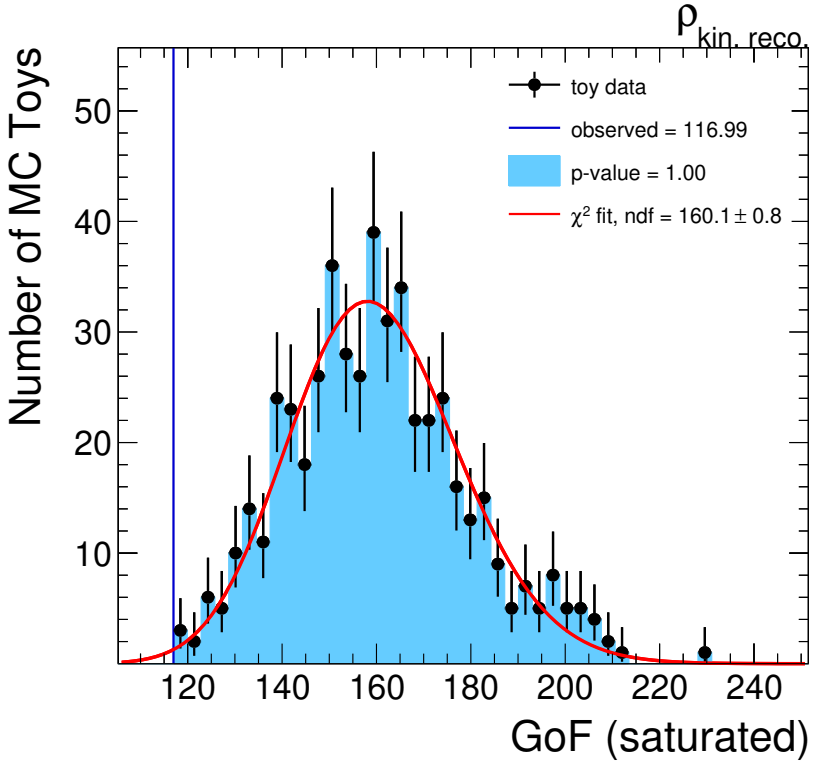

Figure 6.10: Example Goodness-of-Fit test for the ρ observable as obtained from the kinematic reconstruction. The distribution $f(t)$ of the test statistics for the generated pseudo data is shown as black dots, the values of test statistic when the observed data is evaluated is indicated by the vertical blue line, and a comparison to a χ^2 distribution is shown in red. The p-value is indicated by the shaded blue area.

Table 6.5: Input variables used for the regression and classification NNs. The p-value as obtained using the saturated GoF test model as explained in the text are also given.

Name	Symbol	p-value	Reg.	Class.
`jet1Pt`	$p_{\mathrm{T}}^{\text{leading jet}}$	0.95		✓
`jet3Pt`	$p_{\mathrm{T}}^{\text{third jet}}$	0.80		✓
`nJets`	N_{jet}	0.88		✓
`dilepPt`	$p_{\mathrm{T}}^{\ell\ell}$	0.87		✓
`dilepM`	$m_{\ell\ell}$	0.71	✓	✓
`lep1Pt`	$p_{\mathrm{T}}^{\text{leading lepton}}$	0.98		
`lep1M`	$m^{\text{leading lepton}}$	0.08		✓
`lep2Pt`	$p_{\mathrm{T}}^{\text{trailing lepton}}$	0.70	✓	
`lep2M`	$m^{\text{trailing lepton}}$	0.17		✓
`metPt`	$p_{\mathrm{T}}^{\text{miss}}$	0.85	✓	✓
`llj1M`	$m^{\ell\ell j_1}$	0.51	✓	
`llj2M`	$m^{\ell\ell j_2}$	0.85	✓	
`l1j2M`	$m^{\ell_1 j_2}$	1.00	✓	
`l2j1M`	$m^{\ell_2 j_1}$	1.00	✓	
`l2j2M`	$m^{\ell_2 j_2}$	0.92	✓	
`lkr_rho`	$\rho_{\text{loose kin. reco.}}$	0.90	✓	
`kr_rho`	$\rho_{\text{kin. reco.}}$	1.00	✓	
`kr_jetPt`	$p_{\mathrm{T}}^{\text{add. jet}_{\text{kin. reco.}}}$	0.82		✓

loss function is minimized with the ADAM optimizer [285] based on the validation data. To assess for a possible overtraining, a KS test, cf. Section 5.1, between the training and the test response distributions is performed. A p-value of 0.75 is obtained, reflecting a good agreement between the two distributions and hinting to no overtraining.

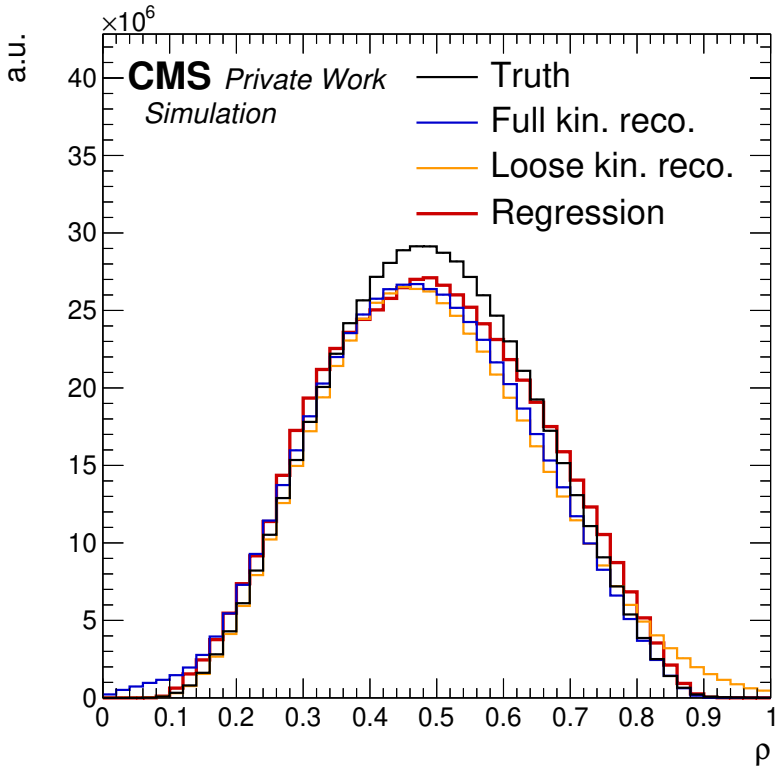

Figure 6.11: Comparisons between the ρ_{true} distribution (black) and ρ_{reco} for the kinematic reconstruction (blue), the loose kinematic reconstruction (orange), and the regression NN (red).

The response of the NN reconstruction is shown in Figure 6.11, compared to the true parton-level distribution for events that are reconstructed and pass the preselection, together with the loose and the full kinematic reconstruction. The histograms are not normalized to unity but show the MC weighted events for the ρ observable at the parton level, the predictions obtained using the kinematic reconstruction, loose kinematic reconstruction, and the NN regression. The loose kinematic reconstruction exhibits a trend towards larger values of ρ, while for the NN regression a shift towards the bulk of the distribution is present. In case of the full kinematic reconstruction, the truth and reconstructed shapes reveal a good agreement.

In Figure 6.12, the correlation between the parton-level true value (ρ_{true}) and the reconstructed value (ρ_{reco}) is shown for the three reconstruction techniques. The regression NN shows the largest correlation factor between ρ_{true} and ρ_{reco} with a value of 0.87, while the loose (full) kinematic reconstruction has only a value of 0.84 (0.78).

The information of Figure 6.12 can also be used to study the bias and the resolution in ρ. In Figure 6.13, the bias (left) and the resolution (right) are displayed as a function of ρ_{true}. The bias is defined as the mean of $\rho_{\text{true}} - \rho_{\text{reco}}$ per bin of ρ_{true}, and the ρ_{reco} resolution in each bin of ρ_{true} is defined as the root-mean-square of the difference between ρ_{reco} and ρ_{true}, divided by $1 + \langle \rho_{\text{true}} - \rho_{\text{reco}} \rangle$. All three approaches show a similar trend for the bias. For small values of ρ_{true}, i.e., higher energies, a slightly negative bias is observed. With increasing

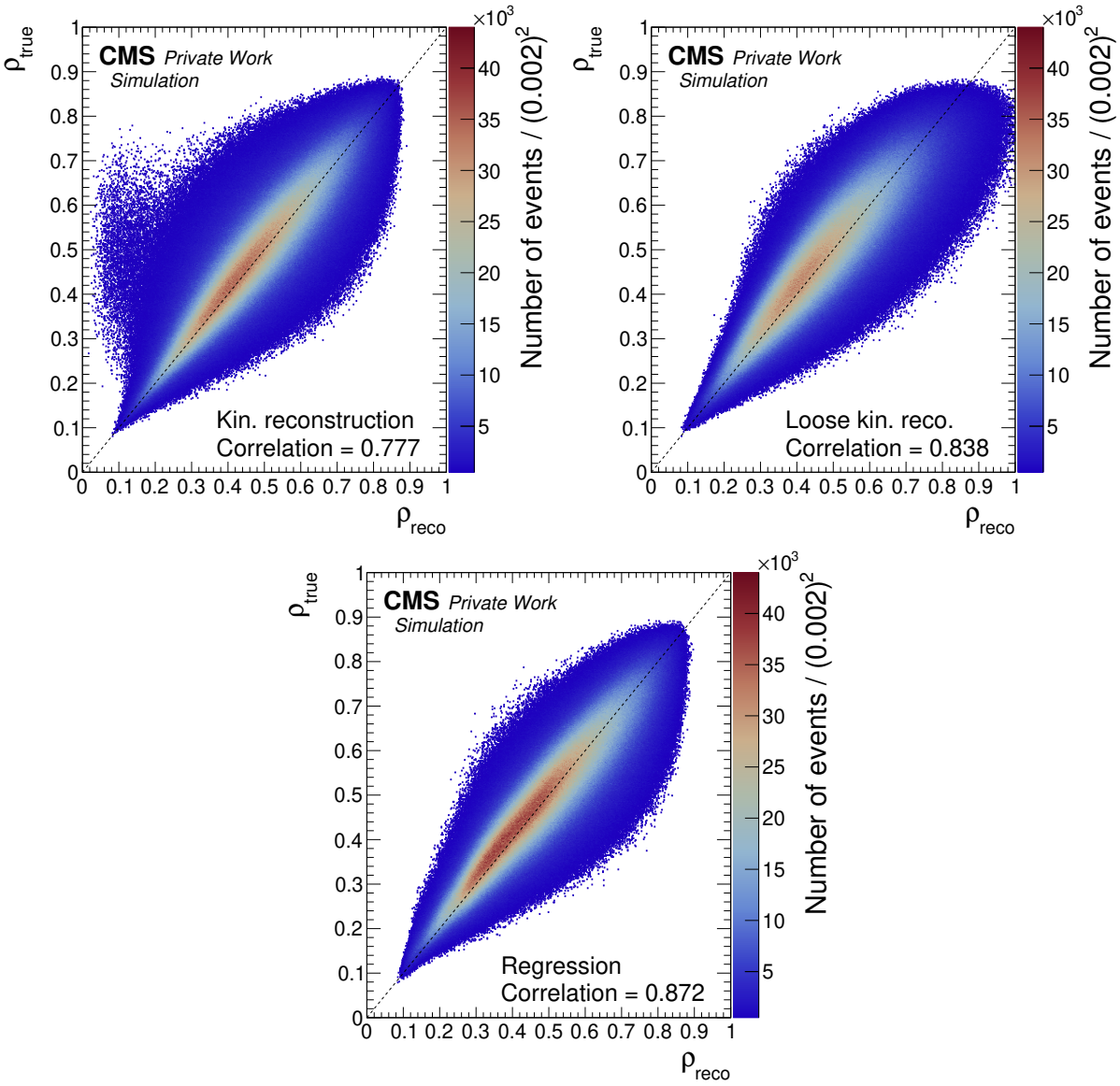

Figure 6.12: The correlation between ρ_{true} and ρ_{reco} is shown for the kinematic reconstruction (upper left), the loose kinematic reconstruction (upper right), and the regression NN reconstruction method (bottom).

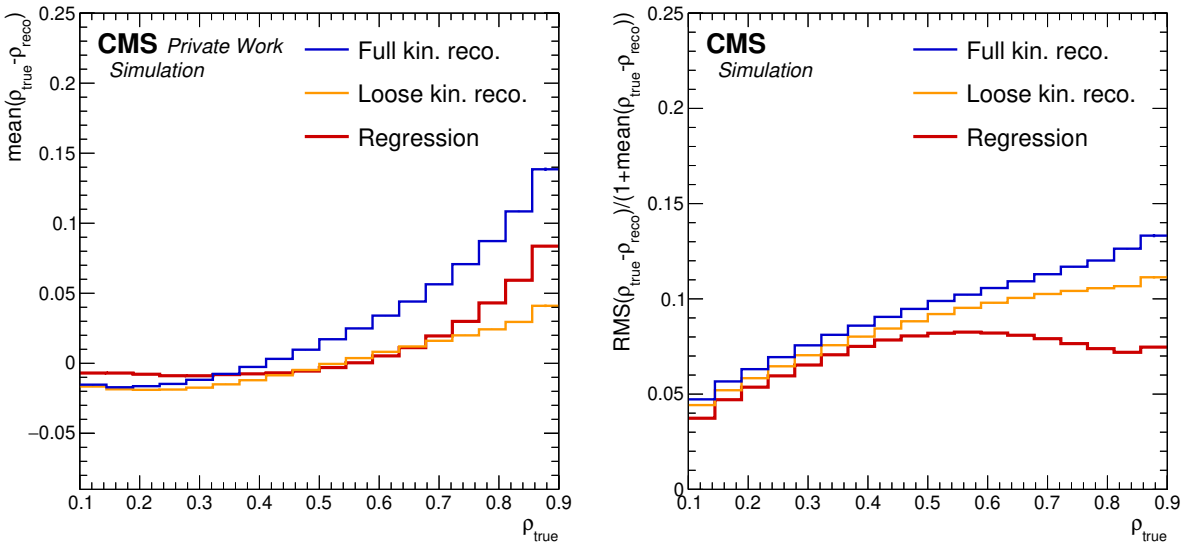

Figure 6.13: The distribution of the mean (left) and response-corrected RMS (right [49] of $\rho_{\text{true}} - \rho_{\text{reco}}$ versus ρ_{true} is shown, for the kinematic reconstruction (blue), the loose kinematic reconstruction (orange), and the regression NN (red).

ρ_{true}, the values for the bias become increasingly more positive, reaching a maximum for large ρ_{true}. The regression NN features the smallest bias for small ρ_{true} and performs better than the full kinematic reconstruction. Compared to the loose kinematic reconstruction, the bias of the regression NN is always comparable or better, while for large ρ_{true} the bias is slightly larger.

A big advantage of the NN reconstruction is that a resolution between 0.05–0.08 is obtained in the whole spectrum, which is an improvement by a factor of two as compared to the two analytical reconstruction methods. This improvement brings benefit to the unfolding applied in the analysis presented in this thesis. Furthermore, unlike the kinematic reconstruction methods, which are affected by reconstruction inefficiencies, the NN approach yields a measurement of ρ_{reco} for every event. This improves the reconstruction efficiency by 5–10% with respect to the loose and full kinematic reconstruction, cf. Figure 6.8. Based on these considerations, in the analysis presented in this thesis the solution of the MVA regression is used, and is indicated with ρ_{reco}. Distributions of ρ_{reco} for the data and simulation are shown in Figure 6.14 for the $e^{\pm}\mu^{\mp}$ (left) and same-flavor dilepton channels (right). A good agreement for the NN response between the data and the simulation within the total systematic uncertainties is observed.

6.4 Event classification

In this thesis, MVA methods in form of NN are also employed to significantly improve the discriminating power between the $t\bar{t}$+jet signal and the background. Discriminating variables are based on the shape differences of the same distribution for different processes.

From low-level variables as, e.g., p_{T} of the studied objects, high-level variables are built, like the total hadronic activity in the event, H_{T}, or the invariant mass of multiple investigated objects to increase the discrimination capability. Using MVA methods, the discrimination power of multiple low- and high-level variables are combined and are of advantage. Figure 6.15 shows two basic discriminating variables for the $t\bar{t}$+jet signal versus all background contributions, the p_{T} of the hardest and of the third-hardest jet. To illustrate the discrimination power, the distributions are normalized to unity and display only the shapes. The two p_{T} distributions

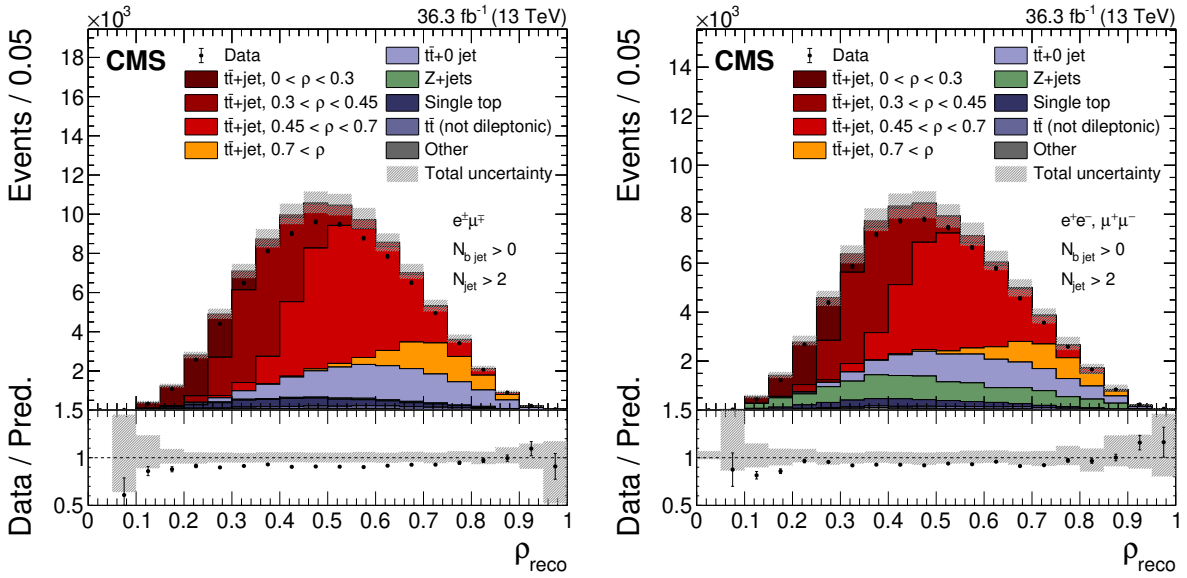

Figure 6.14: The observed (points) and MC predicted (stacked histograms) signal and background yields as a function of ρ_{reco} as determined by the NN reconstruction method for the $e^{\pm}\mu^{\mp}$ (left) and same-flavor dilepton channels (right). The vertical bars on the points represent the statistical uncertainty in the data. The hatched band represents the total uncertainty in the sum of the simulated signal and background predictions. The lower panels show the ratio of the data to the sum of the signal and background predictions. The plots are taken from Ref. [49].

offer certain discrimination potential, but especially at large ρ, the shapes of the $t\bar{t}$+jet signal and background become similar.

The conceptual design of the NN classifier used in this thesis is based on multi-class classification. From a distinct set of input variables on an event-level basis, the NN classifier predicts the probabilities p_i for an event to originate from a certain group of processes (classes). They are further denoted as output scores s_i. Although the main task is to separate signal from background, it is observed that the separation into multiple processes provides a better discrimination of the signal versus background events as compared to a binary-classifier approach. The classes considered are the $t\bar{t}$+jet signal class and the $t\bar{t}$+0 jet and Z+jets background classes. The contribution of all other minor backgrounds is found to be negligible.

Using the same interface as for the regression NN, a fully connected feed-forward event-classification NN is defined, with the same strategies for the input-variable selection and optimization. The NN is trained using preselected events as described in Section 6.2, requiring at least three reconstructed jets. Independent sets of training samples are obtained by using MG5_aMC@NLO[FXFX]+PYTHIA 8 simulation at NLO for the $t\bar{t}$ and Z+jets events, which are not used in the likelihood fit. A combined training for all lepton channels (e^+e^-, $e^{\pm}\mu^{\mp}$, $\mu^+\mu^-$) is performed, since no significant improvement is observed while deriving separate calibrations.

Furthermore, in order to ensure a well-understood behavior of the NN classifier, all used inputs are validated and tested for their consistent modeling in the simulation with respect to the data, as it is also performed for the regression NN. The corresponding p-values are given in Table 6.5.

The same set of input variables are considered for the training of the NN, as are listed in Tables 6.1 to 6.4. Starting from the full set of variables, also here, using a first step Bayesian

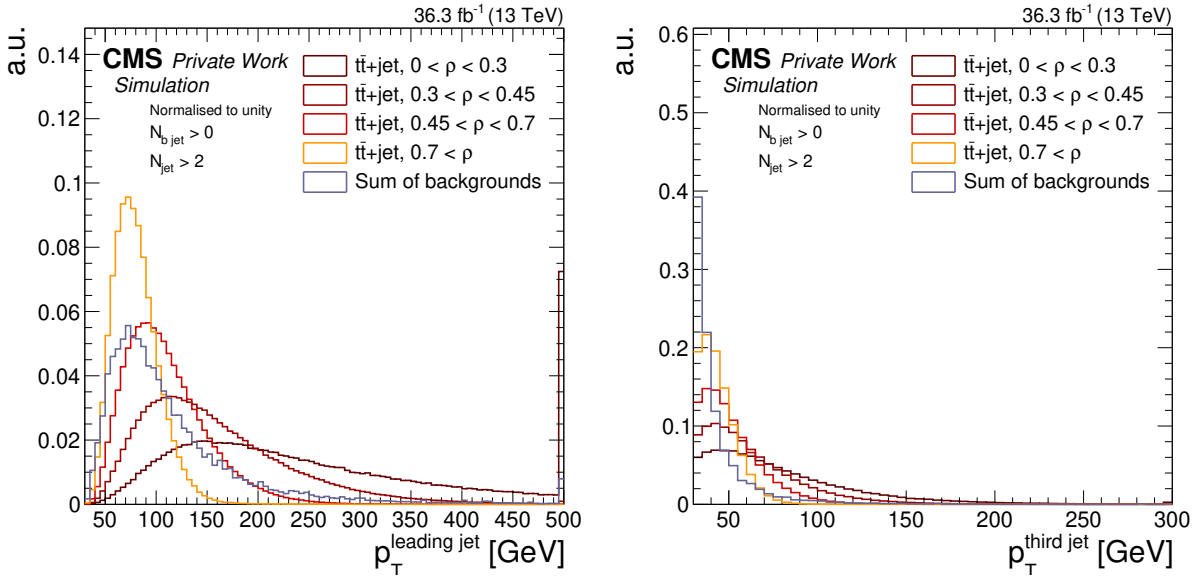

Figure 6.15: The hardest jet p_T distributions for the $t\bar{t}$+jet signal and background processes are shown (left). The distributions are normalized to unity. The right plot shows the same information, but for the third-hardest-jet p_T.

hyperparameter optimization, a subset of the ten most relevant input variables is selected. This procedure is identical to the one described above for the regression NN. The input variable relevance and the ranking are shown in Figure 6.16. The input variables chosen for this NN, ranked by the impact, are

1. the p_T of the additional third jet beyond the two jets from the top quark decays, as determined in the solution of the full kinematic reconstruction,
2. the p_T of the third-highest-p_T jet,
3. the invariant mass of the dilepton system,
4. p_T^{miss},
5. the mass of the leading lepton,
6. the mass of the subleading lepton,
7. the p_T of the leading jet,
8. the p_T of the dilepton system,
9. the number of reconstructed jets,
10. the p_T of the leading lepton.

Variables that are related to the energy of the additional or third jet are the most useful when discriminating between the $t\bar{t}$+jet signal and the $t\bar{t}$+0 jet and Z+jets backgrounds. Additionally, leptonic information helps to differentiate between the kinematic characteristics for different lepton configurations. Global event variables such as jet multiplicity and p_T^{miss} are also useful to infer the origin process for an event.

Because the processes with the highest contribution are the signal ($t\bar{t}$+jet), the $t\bar{t}$+0 jet, and the Z+jets backgrounds, only these are considered in the training of the NN classifier to simplify the training procedure. The signal is composed of multiple individual signal processes, corresponding

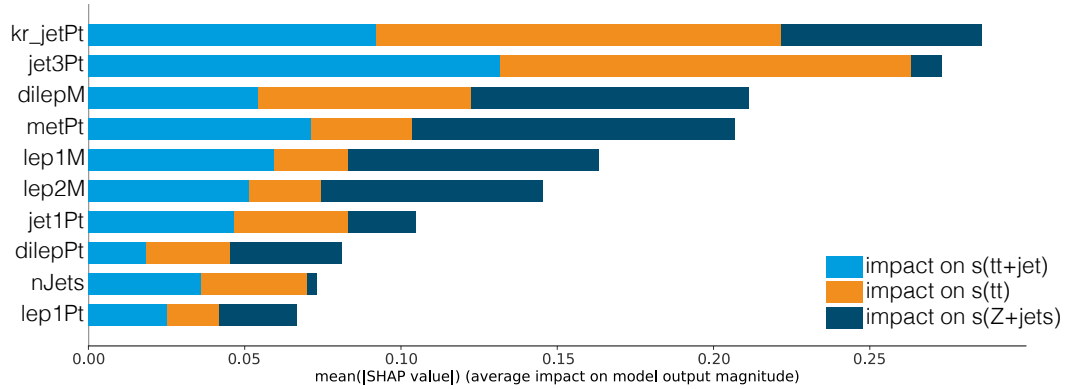

Figure 6.16: Input variable ranking of the ten most relevant input features for the classification NN. The impact on the three different output nodes is indicated by the different colors in the bar plot.

to the bins at parton level. The various signal contributions are thus weighted equally in the training, which ensures the same performance for the signal process independent of ρ_{reco}.

An overall train–validation–test splitting of the available data is done in a ratio of 0.5–0.2–0.3. To remove the imbalance of the number of training events for the three classes considered, class weights are introduced so that the overall sum of weights is identical for the three output classes.

The main part of the NN architecture consists of one input layer, four hidden layers with 204 nodes each, and one output layer with three output nodes, corresponding to the three process classes ($t\bar{t}$+jet, Z+jets, and $t\bar{t}$+0 jet). It is further denoted as G_{class}.

Details about the methods listed in the following are given in Section 5.2. The outputs of each hidden layer are inputs to the SELU activation function [283], while a sigmoid activation function is applied to the final layer. After each but the last layer, batch normalization [288] is applied to improve the minimization process. Dropout layers [290] with a dropout rate of 0.288 are used in addition to L2 regularization with a regularization-strength of 7.835×10^{-5} to mitigate overtraining. With the usage of Glorot normal initialization [289] the training process is stabilized. It initializes the parameters of the weight tensors and constraints on the weights incident to each hidden unit to have a norm less than or equal to ≤ 5. A batch size of 15563 is used and the learning rate is set to 1×10^{-3}. The categorical cross-entropy loss function is minimized with the Adam optimizer [285] based on the validation data.

Because the response of the output nodes is used at a later step in the unfolding procedure (Section 6.5), any bias in the obtained response with respect to the ρ observable is reduced by making use of unsupervised domain adaptation by back propagation, cf. Section 5.2.

In this method, additional layers are added to the NN and to G_{class}. The optimization target of this additional part G_{reg} is the regression of ρ_{reco} as an ancillary output node whereas the sign of the gradient in the back-propagation algorithm is inverted such that the NN does not learn to use input information sensitive to ρ_{reco}. The loss function minimized in this task is the mean-squared-error loss. A weight of 0.01 for this part of the loss in the overall loss function computation is added. A relative strength parameter of 0.3 is added to the flipped gradient in the gradient reversal layer in order to balance the training of the overall NN. The cost in terms of performance is estimated to be on the order of maximum 0.05 in the receiver operating characteristic (ROC) curve integral.

The hyperparameters (batch size, L2 regularization strength, dropout rate, number of layers, number of neurons per layer, activation function) are optimized in a Bayesian optimization procedure. Here, a target function, the integral of the ROC curve averaged over the three output nodes is optimized as a measure of the performance, in order to separate the signal from the background processes. Beforehand, the relative strength parameter in the gradient reversal layer and the weight of the loss function in the regression task is optimized using a 20 point grid search.

For further studies, the ancillary regression part G_{reg} of the full classification NN is abandoned, and only the classification layers and their corresponding weights G_{class} are used. The final NN classification performance is then evaluated using an independent testing data set. The result is shown in Figure 6.17 for all three output nodes. To check for overtraining, KS-tests between the training and test response distributions are performed. The resulting p-values are given in the plots, and they indicate a good agreement between the training and testing distributions.

The same plots also illustrate how well the classification NN can separate the three classes, as for every output node the normalized shapes are given. Especially the Z+jets background class is well discriminated from the other two classes.

The performance of the classification NN is further studied by determining the ROC curves for all three output scores. The distributions are displayed in Figure 6.18 (left). The ROC is obtained using the false (FPR) and true positive rates (TPR) evaluated at various thresholds of the classifier output scores, where the false (true) positive rate is defined as the fraction of events that have an output score s_i larger than the threshold and belong to the class i (or not). These represent the classification efficiency and misidentification rate, respectively. The area-under-curve value, i.e., the integral, yields a powerful indicator of the classifier's performance: the larger the value, the better the discrimination. Its maximum is unity. Additionally, the micro- and macro-averaged ROC curves are displayed to give a global overview of the classification NN's performance, where all three classes are taken into account. For the calculation of the macro-averaged ROC curve, the individual FPR and TPR metrics are evaluated independently for each class and averaged afterwards, whereas for the micro-averaged ROC, the average metrics are computed. They characterize different aspects of the classification capability. The micro-averaged ROC is much more sensitive to the class imbalances and different number of events for the three classes, while the macro-averaged ROC gives a more general average performance number.

The areas under the ROC curves demonstrate that the Z+jets discrimination performs best of the three separation problems and that the $\mathrm{t\bar{t}}$+jet and $\mathrm{t\bar{t}}$+0 jet separation is much more challenging. Due to the physics similarities between the processes, this behavior is expected. However, with all the values larger than 0.83, the classification strength is also large in these two cases. Another characteristic measure for the performance of a classification NN is the confusion matrix. Each row of the matrix corresponds to the true classes while each column represents predicted classes. The entries are calculated using the events of the test set and an event is assigned to the class where the corresponding output node is maximal. The truth-normalized confusion matrix shown in Figure 6.18 (right) displays that in more than 82% of the cases Z+jets events can be correctly identified, while for the $\mathrm{t\bar{t}}$+jet ($\mathrm{t\bar{t}}$+0 jet) events this fraction is 60% (76%). The confusion matrix is evaluated by considering the maximum output score for each event.

The background rejection of the event-classification NN is shown in Figures 6.19 and 6.20, as characterized by the distributions of the NN output scores for the $\mathrm{t\bar{t}}$+jet signal $s(\mathrm{t\bar{t}}$+jet), $\mathrm{t\bar{t}}$+0 jet background $s(\mathrm{t\bar{t}})$, and Z+jets background s(Z+jets) events from data and simulation.

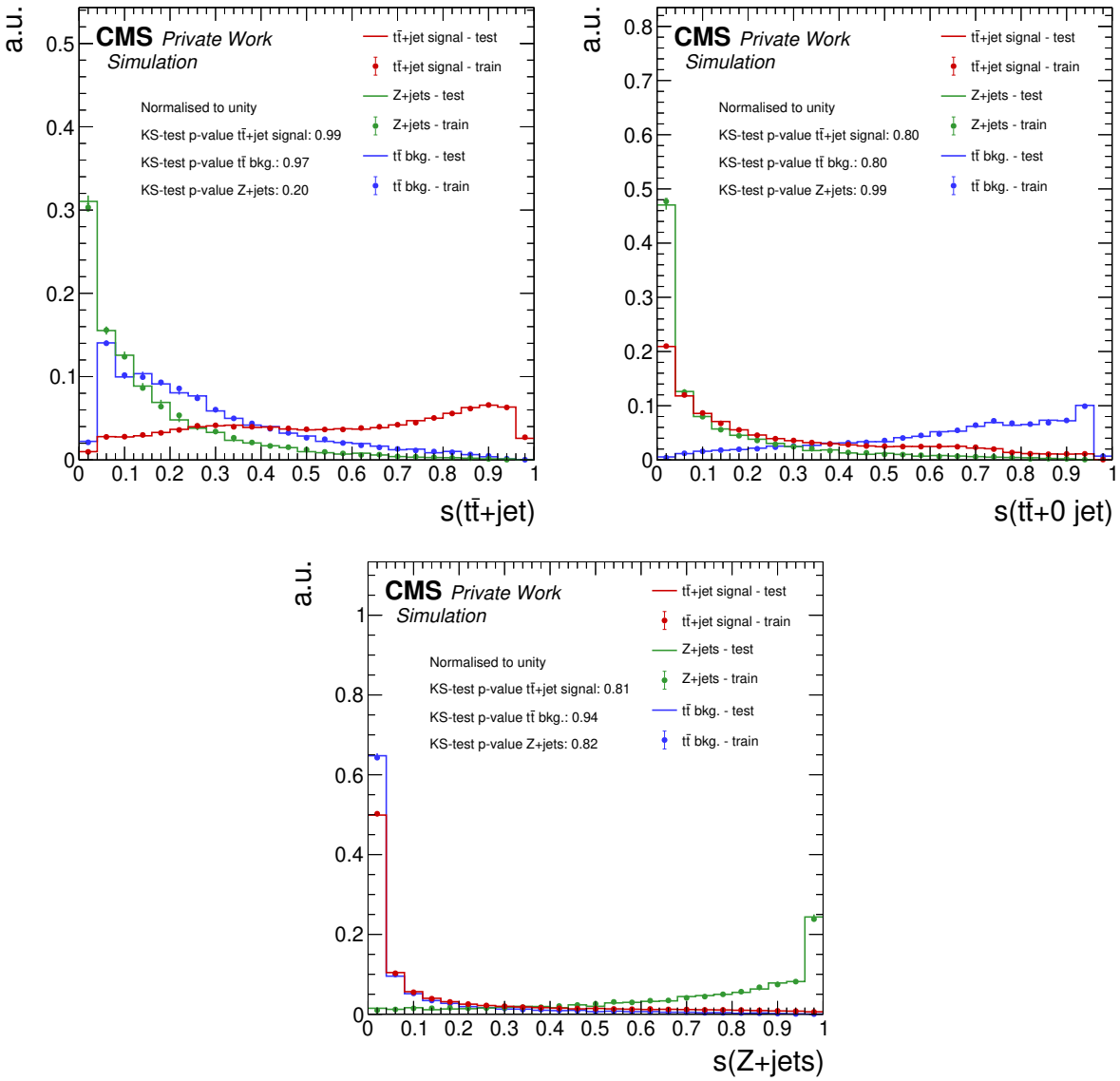

Figure 6.17: Distributions for the signal and background predictions as a function of the NN $t\bar{t}$+jet (top left), the $t\bar{t}$+0 jet (top right), and the Z+jets output scores (bottom), calculated for the training and testing set are shown. The agreement between the training and test distributions is evaluated using KS-tests, and the corresponding p-values are given in the plots.

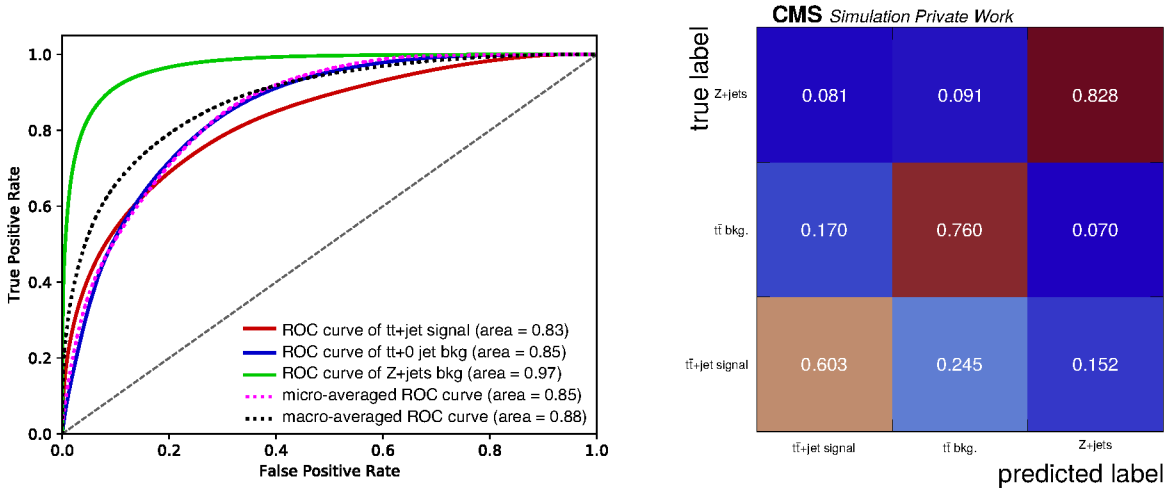

Figure 6.18: The receiver operating characteristic (ROC) curve is shown in the left. For the exact definitions refer to the text. In the right, the row-normalized confusion matrix is given.

For $e^{\pm}\mu^{\mp}$ and same-flavor dilepton channels, the distributions are shown in the left and right column, respectively.

Because $t\bar{t}$+jet is the dominant background process, the signal response of the NN classifier with respect to the $t\bar{t}$+jet background rejection (R_{NN}) is used as the discriminating observable in all signal event categories. It is defined as $R_{\mathrm{NN}} = s(t\bar{t}\text{+jet})/[s(t\bar{t}\text{+jet}) + s(t\bar{t}\text{+0 jet})]$. The data to simulation agreement for this variable is shown in Figure 6.21 for the $e^{\pm}\mu^{\mp}$ and same-flavor dilepton channels.

6.5 Unfolding and cross section measurement

For extraction of m_{t} in a certain renormalization scheme, the measured observables are compared with the theoretical calculations at fixed perturbative order in QCD.

Since the latter cannot be directly compared to the measured observables, either unfolding of the detector-level distributions to the parton level or folding of the theory prediction mimicking detector response, have to be performed. Both approaches rely on the same mapping between the binned parton-level and the detector-level data in form of a so-called response matrix $\mathcal{R}$. Unfolding is preferred in many cases over folding, because it allows for easier reinterpretation of the data by the theoretical community. The response matrix is usually built based on the information available in the MC simulation, and can be defined for a distribution with n bins as

$$\mathcal{R}_{ij} = \frac{N_j^{rec}}{N_i^{gen}}, 1 \leq i \leq n, 1 \leq j \leq n, \tag{6.12}$$

where each entry denotes the transition probability for the events generated in bin i to be reconstructed in bin j. The diagonal elements indicate the probabilities for the events to be reconstructed in the correct bins in the detector-level spectrum. The relation between the true distribution $\vec{y}$ and the reconstructed distribution $\vec{x}$, can be written as

$$\vec{x} = \mathcal{R} \cdot \vec{y} + \vec{b} \tag{6.13}$$

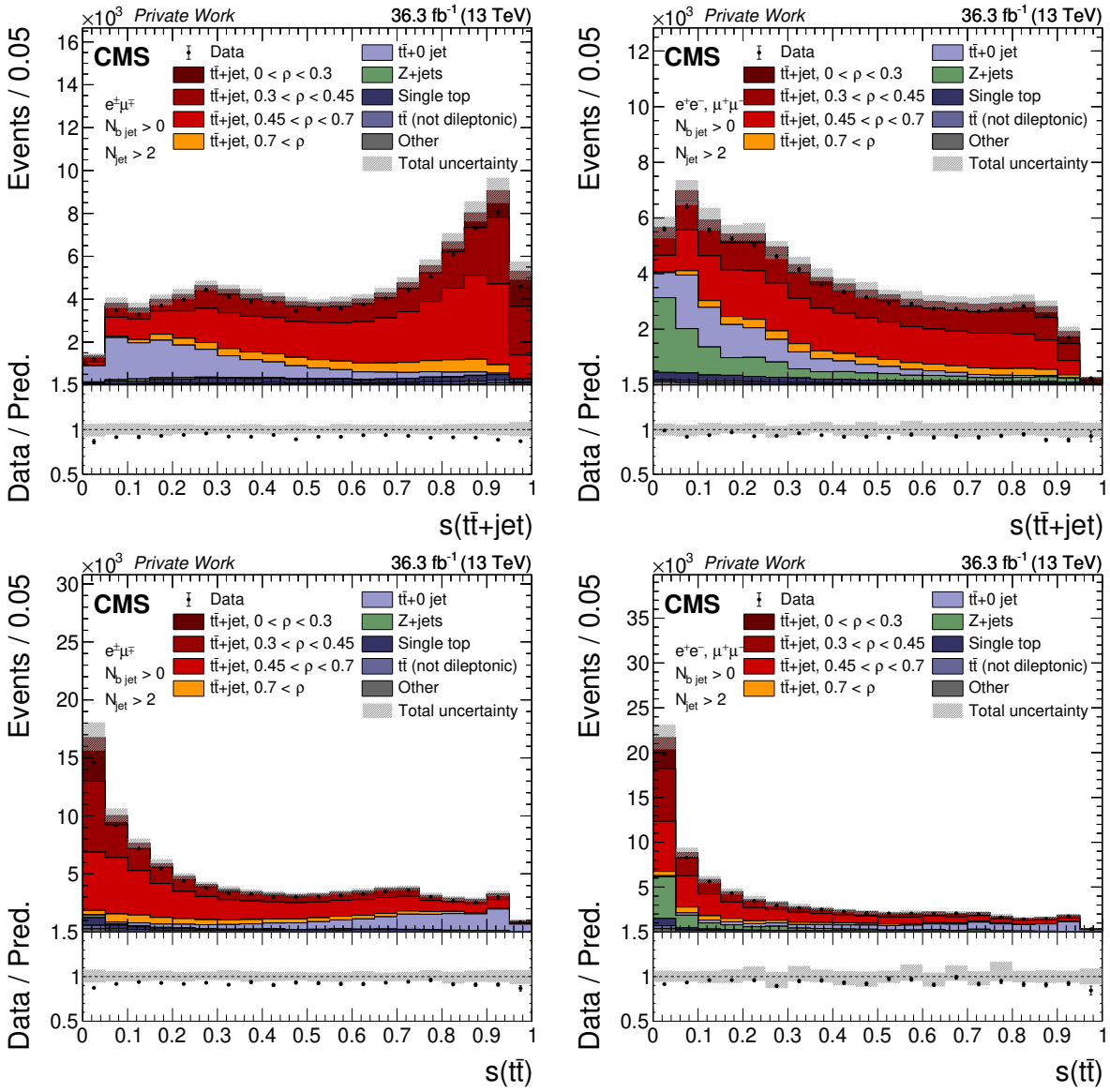

Figure 6.19: The observed (points) and MC predicted (stacked histograms) signal and background yields as a function of the signal (upper row) and $t\bar{t}$+0 jet (lower row) output node scores of the classifier NN. The left (right) column shows the $e^{\pm}\mu^{\mp}$ (same-flavor) dilepton channel. The vertical bars on the points represent the statistical uncertainty in the data. The hatched band represents the total uncertainty in the sum of the simulated signal and background predictions. The lower panels show the ratio of the data to the sum of the signal and background predictions.

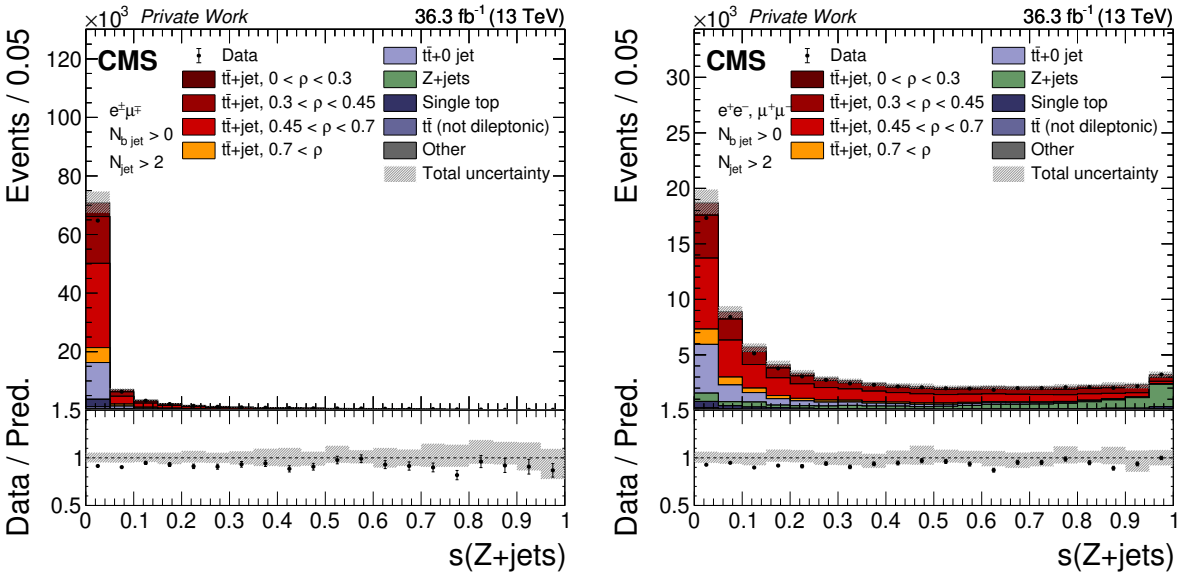

Figure 6.20: The observed (points) and MC predicted (stacked histograms) signal and background yields as a function of the Z+jets background output node. The left (right) plot shows the $e^{\pm}\mu^{\mp}$ (same-flavor) dilepton channel. The vertical bars on the points represent the statistical uncertainty in the data. The hatched band represents the total uncertainty in the sum of the simulated signal and background predictions. The lower panels show the ratio of the data to the sum of the signal and background predictions.

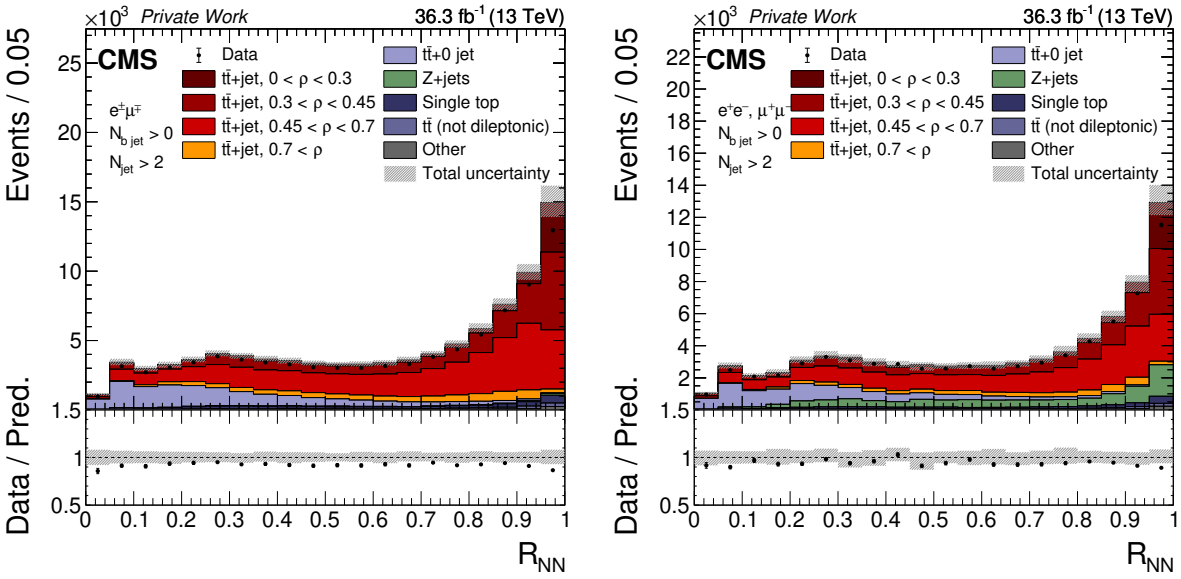

Figure 6.21: The observed (points) and MC predicted (stacked histograms) signal and background yields as a function of the R_{NN} variable. The left (right) plot shows the $e^{\pm}\mu^{\mp}$ (same-flavor) dilepton channel. The vertical bars on the points represent the statistical uncertainty in the data. The hatched band represents the total uncertainty in the sum of the simulated signal and background predictions. The lower panels show the ratio of the data to the sum of the signal and background predictions.

using the response matrix $\mathcal{R}$ and considering $\vec{b}$ as the number of background events in the reconstructed spectrum. The unfolding problem is now formulated as the inverse relation, i.e., inferring the true distribution $\vec{y}$ from $\vec{x}$, $\vec{b}$, and $\mathcal{R}$. It is often known to be an ill-posed problem, being very sensitive with respect to small variations in the input distributions [337]. The simplest approach to solve the problem is based on the inversion of the response matrix. But even for cases where this yields one unique solution, usually large bin-to-bin anticorrelations arise, originating from amplified contributions of these high-frequency variations [338].

Different statistical methods are available of how to solve such unfolding problems. Several methods use matrix inversion techniques, while others rely on Bayesian inference [339] or are based on χ^2 minimization [338].

In this thesis, a different approach is employed, solving the unfolding problem using a profiled maximum likelihood fit, originally developed and applied for the measurement of the running of the top quark mass [42]. In this thesis, this unfolding technique is extended and applied to the cross sections corresponding to bins in the parton-level differential distribution. The signal extraction implies a template fit to the observed final state distributions. Binning of the distributions is accounted for in the fit by building a model describing ensembles of counting experiments, where the event yield is described by a Poissonian distribution. The relevant method is described in detail in Section 5.1. A key property of the method is the fact that NPs are fitted to the data, leading to significantly improved description of the measured distributions and at the same time constraining the systematic uncertainties. This is a direct consequence of the method, as any discrepancy between the prior assumptions and the data is compensated by adjusting the parameters of the NP central values (pulls). If the associated uncertainty for a NP is a-posteriori found to be smaller than assumed a priori, the width of its probability function is correspondingly reduced (constrained). The choice of the event categorization and of the final state observables used in the fit is driven by their potential to constrain relevant NPs.

The additional benefits of the used unfolding technique are the following. Auxiliary information can be used to enhance the discrimination power among different contributions of the parton level bins in the final state distributions, helping to reduce migration effects, or to constrain NPs. Further, correlations of the systematic uncertainties are fully taken into account through the NPs and the background subtraction is directly realized in the fit. Therefore, the method used for the measurement of this thesis leads to superior precision as compared to other techniques.

The expected signal and background distributions and the effect of the systematic uncertainties are estimated using the MC simulations. The total cross section in each bin k of the ρ distribution, $\sigma^k_{\mathrm{t\bar{t}+jet}}$, is measured at the parton level. The values are related to the differential cross section $\mathrm{d}\sigma_{\mathrm{t\bar{t}+jet}}/\mathrm{d}\rho$ as

$$\sigma^k_{\mathrm{t\bar{t}+jet}} = \int_{\rho^k_{\mathrm{low}}}^{\rho^k_{\mathrm{high}}} \frac{\mathrm{d}\sigma_{\mathrm{t\bar{t}+jet}}}{\mathrm{d}\rho}\,\mathrm{d}\rho, \tag{6.14}$$

where ρ^k_{low} and ρ^k_{high} are the lower and the upper bounds of the k-th generator-level bin in ρ, respectively.

The likelihood function $\mathcal{L}$, cf. Section 5.1, assumes that the number of observed events in each bin follows a Poisson distribution and can be written as:

$$\mathcal{L} = \prod_i \frac{\mathrm{e}^{-v_i} v_i^{n_i}}{n_i!} \prod_j \pi(w_j) \prod_m \pi(\lambda_m). \tag{6.15}$$

Here, i denotes the index of the bin in the final-state distribution, and v_i and n_i are the expected and observed number of events in bin i, respectively. The additional terms $\pi(w_j)$ and $\pi(\lambda_m)$ stand for the prior probability density functions (pdfs) for the normalization of each background process w_j and NP λ_m, respectively.

The expected number of events v_i can be written as:

$$v_i = \sum_k s_i^k \left(\sigma^k_{\mathrm{t\bar{t}+jet}}, \vec{\lambda}, m_\mathrm{t}^\mathrm{MC}\right) + \sum_j b_i^j \left(w_j, \vec{\lambda}, m_\mathrm{t}^\mathrm{MC}\right), \tag{6.16}$$

where s_i^k is the number of expected $\mathrm{t\bar{t}}$+jet signal events from the k-th generator-level bin for a reconstructed ρ value in bin i. The value of s_i^k depends on $\sigma^k_{\mathrm{t\bar{t}+jet}}$, the NPs $\vec{\lambda}$, and the top quark mass used in the simulation m_t^MC. Similarly, b_i^j is the number of expected background events from the process j, and depends on its normalization w_j, the NPs $\vec{\lambda}$, and further on m_t^MC in the case of the $\mathrm{t\bar{t}}$ and tW backgrounds.

Since v_i implicitly incorporates the dependence of the parton-level cross section $\sigma^k_{\mathrm{t\bar{t}+jet}}$ on the experimental acceptance and detector response, the minimization of the negative log-likelihood function yields the results that are directly unfolded to the parton level.

To each cross section bin k, a signal strength parameter r_k is assigned, which is defined as

$$r_k = \frac{\sigma^k_{\mathrm{t\bar{t}+jet}}}{\sigma^k_{\mathrm{t\bar{t}+jet}}\,(\mathrm{MC})}. \tag{6.17}$$

Here, the denominator is the cross section corresponding to the normalization of the MC prediction. In the likelihood fit, all signal strength parameters are fitted simultaneously. The $\sigma^k_{\mathrm{t\bar{t}+jet}}(\mathrm{MC})$ values are determined using the nominal NLO POWHEG+PYTHIA8 MC simulation, and the NPs are constrained simultaneously with the values of the cross section in each kinematic bin.

To mitigate the correlation between the fitted signal strength parameters r_k and m_t^MC, the latter is added as an additional free parameter to the fit. The relevant predictions are obtained from two dedicated MC simulations where the value of m_t^MC is set to 169.5 and 175.5 GeV, alternatively. The dependence of the measured $\sigma^k_{\mathrm{t\bar{t}+jet}}$ on m_t^MC is fully taken into account in the fit. Thus, the resulting differential cross section $\mathrm{d}\sigma_{\mathrm{t\bar{t}+jet}}/\mathrm{d}\rho$ can be compared directly to fixed-order predictions without assumptions on the relation between m_t^MC and $m_\mathrm{t}^\mathrm{pole}$ [340]. This method was proposed in Refs. [34, 340] and was used in previous CMS measurements [31, 42]. The likelihood construction and the minimization of $-2\ln(\mathcal{L})$ using MINUIT [341] follows the procedures described in Refs. [342–344]. The final uncertainties are estimated using MINOS [341], as discussed in Refs. [342–344]. The NP modeling and the template morphing follow the descriptions in Section 5.1.

Event categorization and binning

In order to maximize the visible phase space by keeping as many events as possible, multiple event categories are introduced. Signal and background events populate these categories differently, which is used to select the signal and constrain backgrounds in a more efficient way. This way, the normalization of the main background process $\mathrm{t\bar{t}}$+0 jet can be fitted simultaneously with the signal strength parameters, while all correlations are fully taken into account. This argument also applies to migrations of the signal events into background-dominated phase space regions or event categories.

As a consequence of maximizing the acceptance, a significant contribution of background events are retained in the final event selection. The kinematic differences between the signal and background are used to group events with similar properties in categories. The objective is to create event categories with significantly different compositions of signal and background events. The basic categorization criteria are based on the multiplicity of the final state objects, such as reconstructed leptons of certain flavor, the jet multiplicity, or the number of measured b-tagged jets. By separating events in three different channels as e^+e^-, $\mu^+\mu^-$, $e^\pm\mu^\mp$, effects of contributions from uncertainties in the muon or electron reconstruction and identification efficiencies can be distinguished, and the Z+jets background can be separated between the $e^\pm\mu^\mp$ and the same-flavor lepton channels.

By using the b jet multiplicity, the tW single top process, treated as a background here, can be isolated, as its final state exhibits one b jet, while for $t\bar{t}$ decays exactly two b jets are expected. Categorization in the number of reconstructed jets leads to clear separation of the signal $t\bar{t}$+jet events from the $t\bar{t}$+0 jet background. Another benefit is the better control over the uncertainties in the modeling and a possibility to constrain the normalization of the backgrounds from the background-enriched kinematic regions.

Binning definition To choose a suitable binning for ρ both at the parton and reconstructed level, several aspects are considered. First, each bin needs to be well populated by enough observed and simulated events in all dilepton channels. In addition, a diagonal response matrix is desired for the selected signal events to account for the migrations with respect to the parton level spectrum. Quality criteria such as "purity" and "stability" are accessed to determine the binning for the whole ρ distribution. Purity (stability) is sensitive to migrations into (out of) the bin and is defined as the number of signal events that are generated and reconstructed in the same bin, divided by the total number of signal events generated (reconstructed) in that bin:

$$purity = \frac{N_i^{reco\&gen}}{N_i^{gen}}; \qquad stability = \frac{N_i^{reco\&gen}}{N_i^{reco}}. \tag{6.18}$$

Ideally, with the absence of bin-to-bin migrations, both quantities would be unity. A good solution is found with four equal bins both at parton and reconstruction level. The resulting response matrix is shown in Figure 6.22. Corresponding values for the purity and stability for each bin are indicated. Since three reconstructed jets enter the determination of ρ_{reco}, values for the purity and stability on the order of 50% are not unexpected due to the limited jet energy resolution and $p_{\text{T}}^{\text{miss}}$. The condition number for the inversion of the response matrix is determined to be ~ 8. Combined with the observed values for the studied quality criteria, the determined binning shows appropriate properties for an unfolding procedure [345, 346]. Furthermore, the low condition number indicates that the unregularized maximum likelihood estimation is well suited to solve the present unfolding problem.

In Figure 6.23, the reconstructed ρ distribution is shown for the chosen binning in the e^+e^-, $\mu^+\mu^-$, and $e^\pm\mu^\mp$ lepton channels for all three years. A good agreement between data and prediction is observed within statistical and systematic uncertainties.

Regions included in the fit As introduced in Section 6.1, the signal region is defined by events with at least three reconstructed jets and at least one b jet, for which ρ_{reco} is evaluated using the regression NN, and are fed through the classification NN. The signal region is further divided into categories based on the reconstructed ρ_{reco} bin, i.e., introducing four additional subcategories, which are labeled as "Reconstructed ρ" categories. Each of these subcategories is divided into event categories corresponding to different b jet multiplicities ($N_{\text{b jet}} = 1$ and $N_{\text{b jet}} \geq 2$) to enhance the sensitivity to the signal by creating regions with different $t\bar{t}$+jet

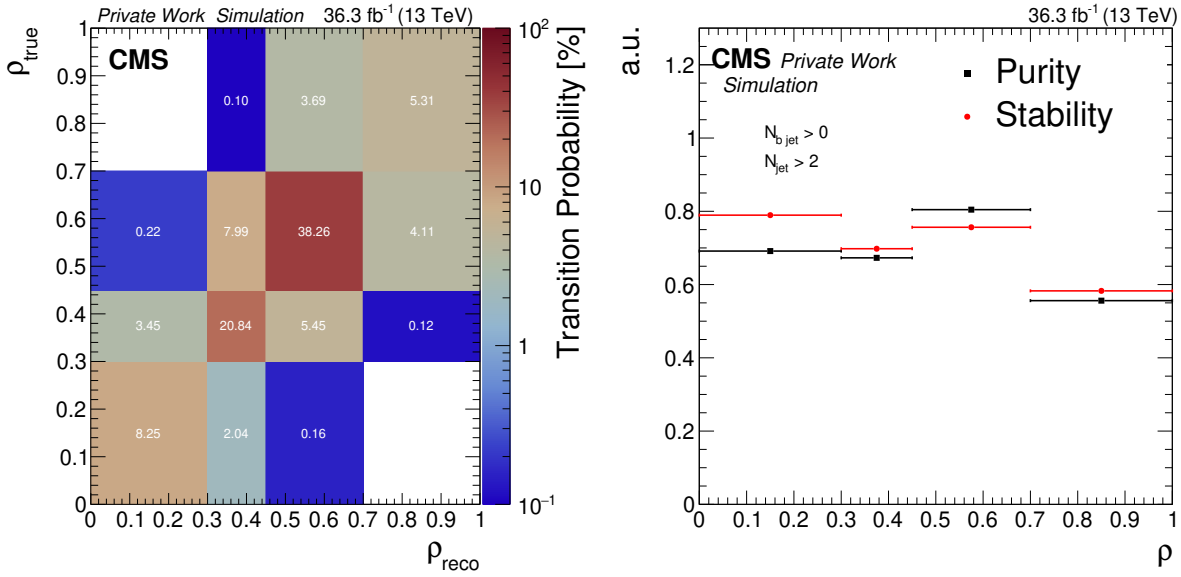

Figure 6.22: The response matrix showing the transition probability and bin-to-bin migrations as determined from the MC simulation (left). A plot showing purity and stability determined from the nominal simulated samples, including the statistical uncertainty for the number of simulated events (right).

purity. To constrain the various background contributions and minimize their extrapolation into the signal region, events with less than three jets are assigned to a dedicated category (in the following denoted as "No reconstructed ρ" event categories). They are additionally separated according to the same b jet multiplicity requirements, but also in different jet multiplicity event categories ($N_{\text{jet}} = 1$, $N_{\text{jet}} = 2$, and $N_{\text{jet}} \geq 3$).

Final-state distributions for each category are chosen such that they maximize the signal sensitivity. The relative signal response R_{NN} of the NN classifier is used as the observable for event categories with $N_{\text{jet}} > 2$. To increase the fit sensitivity to m_{t}^{MC}, the minimum invariant mass $m_{\ell\text{b}}^{\text{min}}$ of the lepton and b jet combinations is used as the observable for the events with $N_{\text{jet}} = 2$ and $N_{\text{b jet}} \geq 2$. In the remaining event categories, the p_{T} of the lowest-p_{T} jet is fitted to constrain JES uncertainties. All three dileptonic channels are kept separate in the fit and are considered as independent categories. This allows to disentangle the effects of systematic uncertainties that differ for the electrons and muons as, e.g., lepton identification efficiencies. An overview of the event categories and the chosen observables are given in Table 6.6.

Table 6.6: A list of the event categories and distributions used in the likelihood fit.

	Reconstructed ρ				No reconstructed ρ	
	$N_{\text{jet}} \geq 3$				$N_{\text{jet}} = 1$	$N_{\text{jet}} = 2$
	$\rho < 0.3$	$0.3 < \rho < 0.45$	$0.45 < \rho < 0.7$	$\rho > 0.7$		
$N_{\text{b jet}} = 1$	R_{NN}	R_{NN}	R_{NN}	R_{NN}	$p_{\text{T}}^{\text{leading jet}}$	$p_{\text{T}}^{\text{subleading jet}}$
$N_{\text{b jet}} \geq 2$	R_{NN}	R_{NN}	R_{NN}	R_{NN}	—	$m_{\ell\text{b}}^{\text{min}}$

In the binning of each observable, the statistical uncertainty for all samples used is taken into account to reduce the impact of limited statistics on the extracted values of r_k, w_j, and $\vec{\lambda}$ after the fit [42]. In particular, the observables given in Table 6.6 are reduced to the total event yield (N_{Events}) if the number of simulated events is insufficient to model the systematic variations

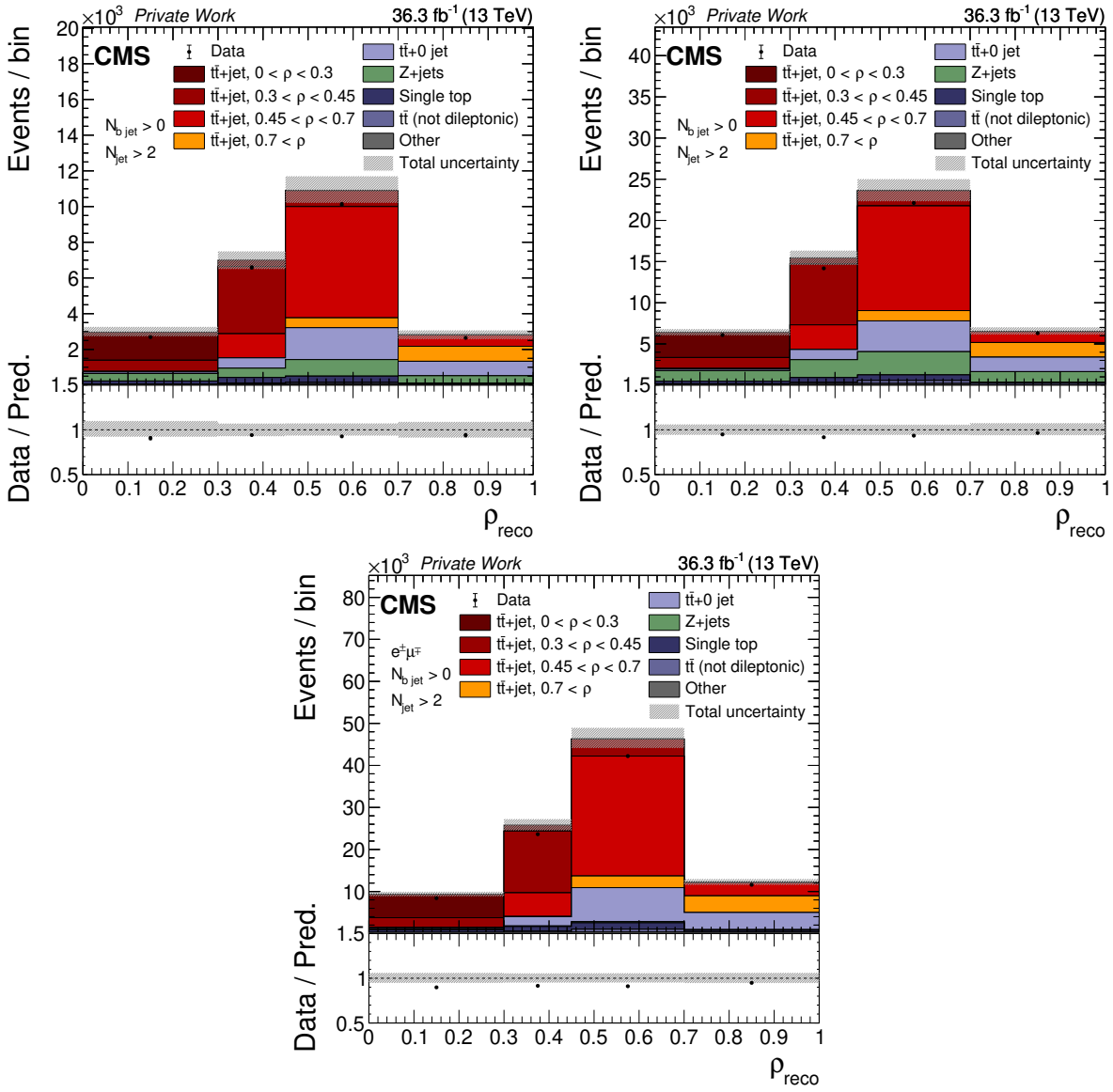

Figure 6.23: The observed (points) and predicted (stacked histograms) signal and background yields as a function of ρ_{reco} in the e^+e^- (left), $\mu^+\mu^-$ (right), and $e^{\pm}\mu^{\mp}$ (bottom) dilepton channels, after applying the signal selection. The vertical bars on the points represent the statistical uncertainty in the data. The hatched band represents the total uncertainty in the sum of the simulated signal and background predictions. The lower panels show the ratio of the data to the sum of the signal and background predictions.

properly. As a requirement, the statistical uncertainty in the number of simulated events for the signal sample with the lowest total number of generated events is not allowed to exceed 0.5% per bin. The final input distributions to the fit, in the event categories, are shown in Figure 6.24, where the data are compared to the simulated signal and pre-fit background distributions. Good agreement between data and simulation is found.

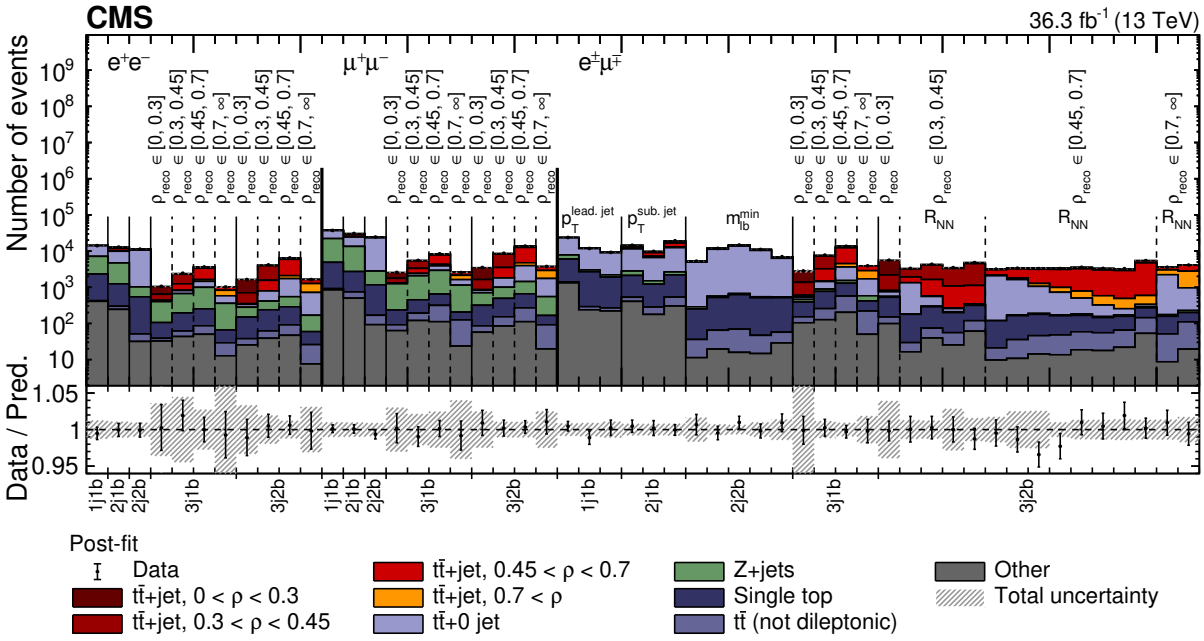

Figure 6.24: The pre-fit distributions from data (points) and simulated signal and background (colored histograms) used in the maximum likelihood fits. The distributions are shown for each dilepton type and each event category, where the x-axis label "mjnb" refers to events with m jets and n b jets. The vertical bars on the points show the statistical uncertainty in the data. The hatched band represents the total uncertainty in the sum of the simulated signal and background predictions. The lower panel gives the ratio of the data to the sum of the simulated predictions. The plot is taken from Ref. [49].

6.6 Systematic uncertainties

Contributions to the systematic uncertainties from various sources are modeled as NPs in the fit, as described in Section 5.1. For each variation, dedicated templates are obtained describing the effect in each background source or signal contribution.

Experimental uncertainties

Most of the experimental systematic uncertainties are estimated by varying the SFs used to correct any differences between the data and simulation. The following experimental uncertainties are considered.

Luminosity The integrated luminosity used to normalize the simulated samples has a relative uncertainty of 1.2% [200].

Pileup The PF algorithm is used to remove contributions from pileup events as good as possible. However, pileup events also affect many other kinematic variables such as $p_{\mathrm{T}}^{\mathrm{miss}}$. To

estimate the uncertainty in the amount of pileup, the simulated pileup distribution is varied by performing a ±4.6% variation of the minimum bias cross section [316].

Trigger efficiency To correct for differences in the trigger efficiencies between data and simulation, the derived SFs are varied within their uncertainties, which are composed of statistical and systematic components and found to be smaller than 3%.

Lepton selection Electron and muon identification and isolation efficiencies are measured using the "tag-and-probe" method with Z boson event samples in bins of the lepton p_T and η [190, 196]. Corresponding SFs are consistent with unity within 10 and 3% for electrons and muons, respectively. The uncertainties of the SFs are typically on the order of 2–5% for electrons, and 0.5–1.5% for muons. They are varied within their uncertainties in the simulation individually for every lepton flavor (electron/muon) and type (identification/isolation/reconstruction) to estimate the corresponding systematic uncertainty. On top of the systematic uncertainty for the muon isolation, a 0.5% additional uncertainty per muon is added, to account for the extrapolation from the Drell–Yan to $t\bar{t}$ phase space. In case of the electrons, a 1% extrapolation uncertainty is considered.

Uncertainties due to correcting the electron and muon energy scales and resolutions are estimated separately by varying them within their uncertainties in the simulation. The typical uncertainty in the energy resolution is on the order of 0.5% and 5% for electrons and muons, respectively [190, 196]. The energy variations are propagated to $\vec{p}_T^{\,miss}$.

Jet energy scale To determine the uncertainty due to the JES, 23 individual uncertainty contributions of the p_T- and η-dependent corrections [234] are assessed. The total uncertainty in the JES ranges from 1.0–3.5% depending on the jet kinematic properties [234]. The vectorial changes of the jet momenta are propagated to the missing transverse momentum vector.

Jet energy resolution uncertainty The uncertainty in the JER is determined by a variation of the JER SFs in the simulated samples by their uncertainties in two different η regions. The effect is on the order of 2–8%, depending on the η region [234]. The uncertainty in the SFs is on the order of 2–8%.

Jet pileup ID Uncertainties in the application of the pileup jet identification are estimated by varying the efficiency and mistag rate within their uncertainties [240, 241]. The agreement between the data and simulated samples is on the order of 2–10% [240], depending on the jet kinematic properties. The uncertainties range between 5–30%, depending on the jet η and p_T.

Unclustered p_T^{miss} In order to account for the unclustered missing energy, $\vec{p}_T^{\,miss}$ is recalculated varying the deposited energy from the charged and neutral hadrons and photons according to the corresponding energy resolutions. Its effect on the resolution is on the order of 5–30% [224].

L1 ECAL and muon prefiring During the 2016 data taking, a gradual shift in the timing of the inputs of the ECAL L1 trigger in the region at $|\eta| > 2.0$ caused a specific trigger inefficiency. For events containing an electron (a jet) with p_T larger than ≈50 GeV (≈100 GeV), in the region $2.5 < |\eta| < 3.0$ the efficiency loss is ≈10–20%, depending on p_T, η, and time [320]. A similar effect was present for L1-trigger muons because of the finite time resolution of the muon detectors. This lead to an effect on the order of 0.5–1.5% per muon [319]. Dedicated SFs are derived for correcting the simulation for these effects known as "prefiring" and they are varied within their uncertainties.

b tagging To correct for different b tagging efficiencies and mistagging rates of light-quark and gluon jets in the data and simulation, SFs are derived using simulated QCD multijet events and are applied as a function of the jet p_{T} [249]. To estimate the impact of that systematic uncertainty, the SFs are varied within their estimated uncertainties, which are split into individual subsources. The heavy flavor (b and c) jets are considered fully correlated, while light and gluon jets are considered uncorrelated to the b jets. For light-quark (heavy-quark) jets, the uncertainty ranges from 5–10% (1–5%). The uncertainties are split into four subsources, taking into account the effect from pileup, the JES, number of selected events, and effects coming from the SF derivation method and the modeling.

Theoretical uncertainties

The impact of theoretical assumptions in the modeling is assessed by performing appropriate variations of the model parameters of the nominal POWHEG+PYTHIA8 simulation by the usage of dedicated simulated samples with altered parameters, or by varying the reference simulation with a use of weights. Besides the PDF and h_{damp} variations, all the above-mentioned uncertainties are also assessed for tW production and are treated as correlated, if applicable. In the case of Z+jets production, μ_{R} and μ_{F} variations are also considered. In all cases, the total cross section per individual process contribution, background or signal, is kept fixed to the nominal value. The following theoretical uncertainties are considered.

Matrix element μ_R and μ_F scales The uncertainty in the modeling of the hard-production process is assessed through changes in the renormalization and factorization scales (μ_R and μ_F respectively) in the ME in the POWHEG sample. Two separate variations are computed, varying μ_R and μ_F independently by a factor of two up and down with respect to their nominal values.

Similarly, for Z+jets and single top processes, scale variations are assessed in a similar way, individually.

Variation of the parton shower scales Uncertainties in the modeling of the PS are determined by varying the corresponding scales for the initial- (ISR) and final-state radiation (FSR) individually by a factor of two up and down with respect to their nominal values. Uncertainties for tW production are considered uncorrelated to the uncertainties for $\mathrm{t\bar{t}}$ production.

Parton distribution function The uncertainty arising from the choice of the PDF is assessed by reweighting the $\mathrm{t\bar{t}}$ signal sample according to the 100 eigenvector variations in the NNPDF3.1 PDF set [169, 296]. Each is treated as an individual NP in the fit. Additionally, the value of $\alpha_S(m_Z)$, where m_Z is the Z boson mass, is varied within its uncertainty in the PDF set.

Matrix element – Parton shower matching The uncertainty in the matching of the ME to the PS models is evaluated by means of dedicated POWHEG+PYTHIA8 simulations with the up/down variation of the h_{damp} parameter with respect to its nominal value, $h_{\mathrm{damp}} = 1.379\,^{+0.926}_{-0.5052} m_{\mathrm{t}}^{\mathrm{MC}}$, determined in a dedicated tuning procedure [132] .

Underlying event tune The uncertainty due to the choice of the UE tune CP5 is estimated using dedicated POWHEG+PYTHIA8 simulations for $\mathrm{t\bar{t}}$ and single top with variations of the tuned parameters up and down according to their uncertainties that are determined in the tuning process [132].

Color reconnection For the default CR model as implemented in PYTHIA 8 and used in the reference $\mathrm{t\bar{t}}$ and single top simulation, certain assumptions are made. While early resonance decays (ERD) are switched off in the nominal setup, dedicated additional samples with ERD

enabled are generated, and the difference is considered as a systematic uncertainty. Two more CR schemes are studied, and the systematic uncertainty is estimated likewise using a gluon-move scheme [135] and a QCD-inspired scheme [136].

b quark fragmentation The fragmentation of a b quark into b hadrons is described by the Bowler–Lund function in the nominal CP5 tune of PYTHIA 8 used for $t\bar{t}$ and single top simulations, with the Bowler-Lund parameter set to 1.056 [142, 302]. The corresponding uncertainty is evaluated by reweighting the relevant transfer function in the reference $t\bar{t}$ and single top samples. Alternatively, the Peterson or the Bowler-Lund fragmentation functions with their default parameters in PYTHIA 8 are used, and the fragmentation function in the nominal $t\bar{t}$ and single top simulations are reweighted accordingly.

b semileptonic branching fraction The uncertainties originating from the semileptonic branching fractions of b hadrons is estimated by varying those within their uncertainties, as estimated in Ref. [12].

Top quark p_{T} reweighting Differential measurements of the $t\bar{t}$ cross section at $\sqrt{s} = 13\,\mathrm{TeV}$ have demonstrated that the p_{T} distribution of the top quark is softer than predicted by the POWHEG simulation [323–326]. An additional uncertainty, which is estimated by reweighting the simulation to data [323, 347–349] is considered.

Additional uncertainties

Additional uncertainty contributions arise from the normalization uncertainties in the background processes, the finite number of events in the MC simulation, and the corresponding effects on the templates from these sources.

Background normalization The uncertainty in the background normalization due to the uncertainties in the used cross section for the background coming from inclusive $t\bar{t}$ production in the semileptonic and hadronic decay channels is coming from two sources that are added quadratically: the scale uncertainty, assumed to be $^{+2.5\%}_{-3.6\%}$, and the combined PDF and α_s uncertainty, assumed to be $^{+0.47\%}_{-0.46\%}$. The uncertainties in the normalizations for single top quark production and smaller background contributions, such as diboson and W+jets production, are taken to be 30% and modeled with a log-normal prior pdf, following the prescriptions from previous analyses [31, 38, 326]. For the Z+jets background, separate uncertainties are assigned to each b jet category in order to remove the dependence of the fit result on the prediction of the b jet multiplicity distribution from the LO Z+jets simulation. Similarly, the Z+jets background is assigned an additional uncertainty of 5, 10, 30, and 50% for events with exactly 0, 1, 2, and 3 or more jets, respectively. The first three uncertainties are estimated by performing scale variations in W+jets predictions with NLO precision, whereas the last one is assigned conservatively [31]. For the $t\bar{t}$+0 jet background, no prior pdf is used because the normalization is left free in the fit to help constrain it with the background-dominated categories.

Limited size of the MC samples The limited size of the background and signal MC simulated samples results in statistical fluctuations of the templates. This is taken into account via bin-by-bin NPs with the Barlow–Beeston "light" method [272, 273]. In addition, the impact of the finite statistical precision on the predictions of the MC samples with dedicated systematic variations is estimated using toy experiments, following the approach described in Ref. [42], and is found to be negligible.

Extrapolation uncertainties Additional extrapolation uncertainties, which reflect the impact of modeling uncertainties in the acceptance, by evaluating their impact on the signal acceptance [31, 42] are taken into account. In this procedure possible post-fit constraints on these sources of uncertainties, obtained from events reconstructed in the visible phase space, are ignored.

The extrapolation uncertainty is determined for each relevant model systematic source j as described in the following. All NPs except the one under study are fixed to their post-fit values. The NP λ_j is set to values corresponding to one standard deviation up and down, and the variations of the acceptance are recorded. The resulting variations of $\sigma^{k}_{\mathrm{t\bar{t}+jet}}$ with respect to the nominal value, obtained with the post-fit value of λ_j, are taken as the additional extrapolation uncertainties.

6.7 Results and extraction of the top quark pole mass

The differential $\mathrm{t\bar{t}}$+jet cross section as a function of ρ is measured in four bins. The resulting distributions after the likelihood fit to the MC signal and background prediction are shown in Figure 6.25. Very good agreement between the data and the expectation is observed.

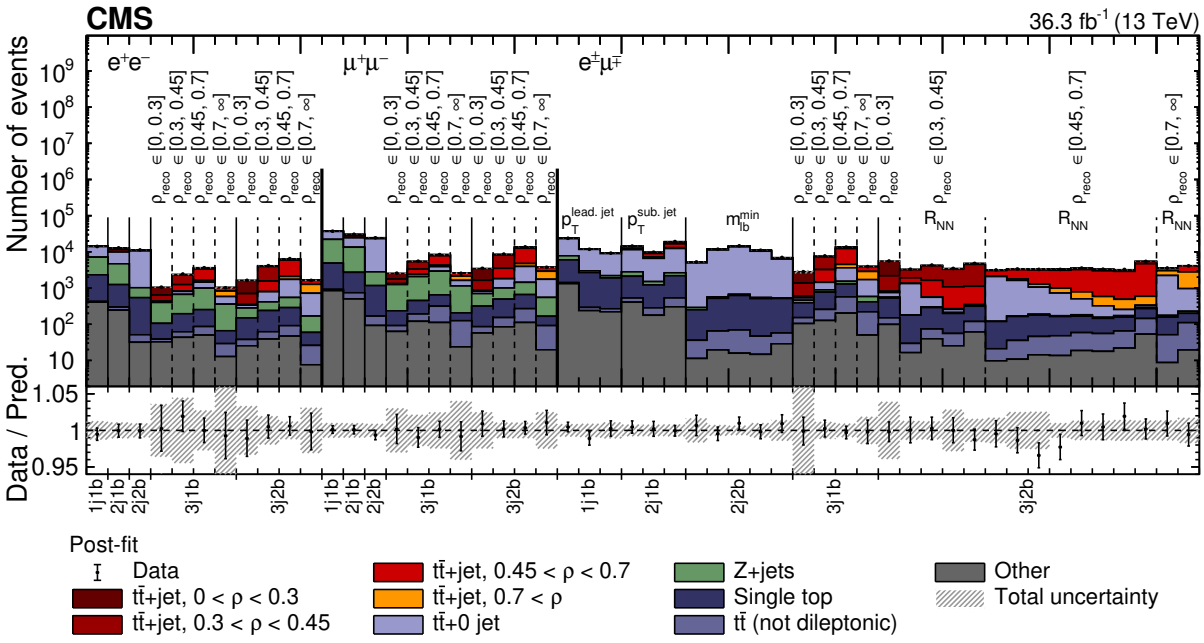

Figure 6.25: The distributions from data (points) and simulated signal and background (colored histograms) used in the maximum likelihood fits after the fit to the data. The distributions are shown for each dilepton type and each event category, where the x-axis label "mjnb" refers to events with m jets and n b jets. The vertical bars on the points show the statistical uncertainty in the data. The hatched band represents the total uncertainty in the sum of the simulated signal and background predictions. The lower panel gives the ratio of the data to the sum of the simulated predictions. The plot is taken from Ref. [49].

The measured absolute differential $\mathrm{t\bar{t}}$+jet cross section $\mathrm{d}\sigma_{\mathrm{t\bar{t}+jet}}/\mathrm{d}\rho$ is shown in Figure 6.26 (left), in comparison to NLO QCD predictions. The predictions are obtained using the $\mathrm{t\bar{t}}$+jet process implemented in POWHEG-BOX [164], using the ABMP16NLO [167] PDF set, and assuming

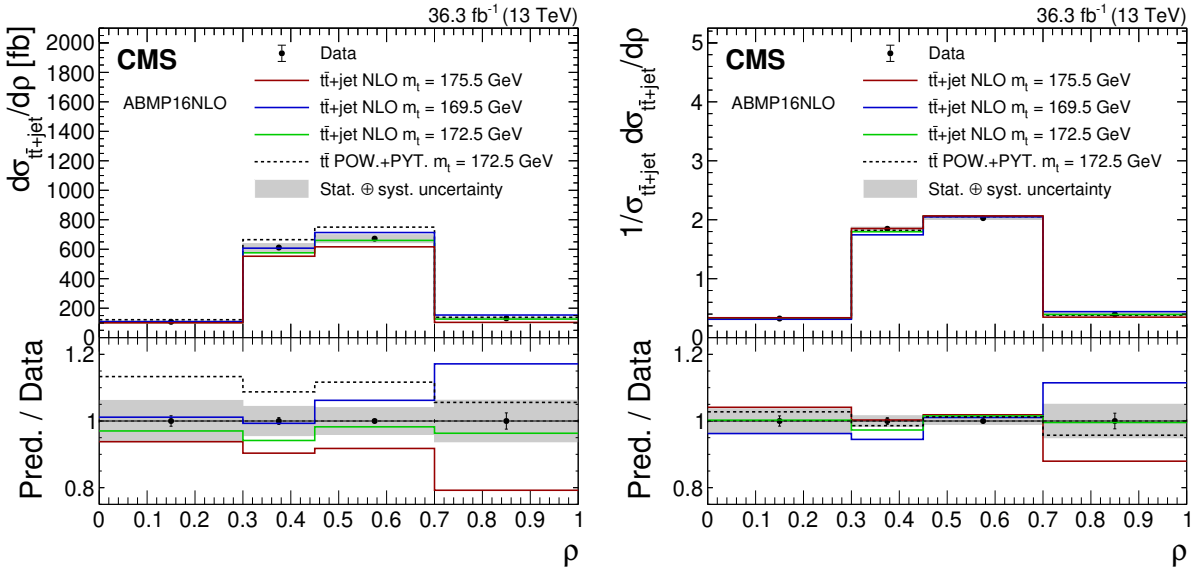

Figure 6.26: The absolute (left) and normalized (right) $t\bar{t}$+jet differential cross section as a function of ρ for the data (points) and theoretical predictions described in the text using the AMBP16NLO PDF set from the NLO MC with three different m_t values and from the POWHEG (POW) + PYTHIA8 (PYT) calculations (lines). The vertical bars on the points show the statistical uncertainty in the data and the shaded region represents the total uncertainty in the measurement. The lower panels give the ratio of the predictions to the data. The plots are taken from Ref. [49].

m_t^{pole} values of 169.5, 172.5, and 175.5 GeV. Additionally, the CT18NLO PDF set [166] is also considered for the mass extraction.

The QCD scale in the theoretical prediction is dynamically set to $H_T^B/2$, as suggested in Ref. [165] and discussed in Ref. [160], where H_T^B is defined as the scalar sum of the top quark and antiquark transverse masses and the p_T of the additional jet.

The statistical uncertainty is evaluated by keeping all systematic NPs fixed to their post-fit values. By subtracting the statistical uncertainty quadratically from the total uncertainty, the systematic uncertainty is obtained.

The relative uncertainties $\Delta\sigma^k_{t\bar{t}+jet}$ in the parton-level cross section values $\sigma^k_{t\bar{t}+jet}$ and contributions of each individual source in each bin k of the ρ distribution are given in Table 6.7. The contributions to the uncertainty are obtained by repeating the fit after fixing the NPs related to the uncertainty under scrutiny to their post-fit values. Each partial uncertainty is determined by subtracting the uncertainty obtained with this procedure from the total uncertainty. The quadratic sum of all contributions differs from the total uncertainty because of correlations between the NPs accounted for in the fit and ignored in externalizing individual sources.

In total, a relative precision of 6% for the measured cross section is achieved. The largest contributions to the total uncertainty originate from the modeling, affecting the acceptance and extrapolation to the parton level. Experimental uncertainties are dominated by uncertainties in the JES, JER, and pileup jet identification, which is expected for measurements involving jets.

The fitted parameter values, expected and observed constraints, and impacts on the signal strength parameters r_k for the 30 most relevant NPs after the fit to data are shown in Figure 6.27. Pulls, constraints, and impacts are defined in Section 5.1. For the NPs associated with the

Table 6.7: The relative uncertainties $\Delta\sigma^{k}_{\mathrm{t\bar{t}+jet}}$ in the parton-level cross section values $\sigma^{k}_{\mathrm{t\bar{t}+jet}}$ and their sources in each bin k of the ρ distribution.

Uncertainty Source	$\Delta\sigma^{1}_{\mathrm{t\bar{t}+jet}}$ [%]	$\Delta\sigma^{2}_{\mathrm{t\bar{t}+jet}}$ [%]	$\Delta\sigma^{3}_{\mathrm{t\bar{t}+jet}}$ [%]	$\Delta\sigma^{4}_{\mathrm{t\bar{t}+jet}}$ [%]
Experimental				
Muon identification	1.8	1.5	1.5	1.4
Muon energy scale and resolution	0.7	0.2	0.3	0.5
Electron identification	2.0	1.7	1.7	2.1
Electron energy scale and resolution	0.9	1.0	0.9	1.5
Jet energy scale	2.6	2.0	2.2	3.6
Jet energy resolution	0.6	0.5	0.5	0.4
Jet identification	1.1	0.8	0.8	1.3
$p_{\mathrm{T}}^{\mathrm{miss}}$	0.2	0.3	0.4	0.8
b jet identification	1.0	0.7	0.6	1.2
Trigger efficiency	1.8	1.2	1.1	1.8
Total	4.0	3.1	3.1	4.7
Background normalization				
$\mathrm{t\bar{t}}$+0 jet	2.2	2.0	1.7	0.7
Z+jets	2.4	1.9	1.7	2.6
Single top quark	0.9	0.8	0.7	0.1
Total	3.1	2.5	2.4	2.7
Modeling				
Z+jets ME scale	0.7	0.4	0.2	0.3
Single top quark ME/FSR/ISR scales	1.2	0.6	0.4	0.1
$\mathrm{t\bar{t}}$ PDF	0.1	0.1	0.1	0.6
$\mathrm{t\bar{t}}$ ME scale	1.0	0.5	0.6	0.4
$\mathrm{t\bar{t}}$ ISR scale	1.2	0.8	0.6	1.6
$\mathrm{t\bar{t}}$ FSR scale	1.3	0.8	0.6	1.7
$\mathrm{t\bar{t}}$ top quark p_{T}	2.0	1.3	0.1	1.2
b fragmentation	0.9	0.7	0.8	0.8
Color reconnection	0.5	0.6	0.2	0.7
$\mathrm{t\bar{t}}$ matching scale	0.6	0.5	0.6	$\leq$0.1
Underlying-event tune	0.2	0.5	0.2	0.5
Total	3.2	1.9	1.8	3.1
Integrated luminosity	1.2	1.4	1.3	1.2
$m_{\mathrm{t}}^{\mathrm{MC}}$	1.7	1.0	0.4	2.3
Finite size of simulated samples	2.0	1.4	1.3	2.2
Total systematic	5.0	3.4	3.2	5.6
Statistical	1.6	1.0	0.8	2.4
Total	5.2	3.6	3.3	6.1

$t\bar{t}$+0 jet normalization and the value of m_t^{MC}, the post-fit values are given in the figure instead of showing $(\hat{\theta} - \theta_0)/\Delta\theta$ because no prior pdf is assigned to them. The post-fit values are compatible with the corresponding pre-fit values within their prior uncertainties, reflecting the good agreement between data and simulation before the fit. Some pulls for NPs describing the background normalization for the single top and Z+jets production, considered as background, are larger. This feature is expected, because the background processes are not well modeled in the signal phase space. The NP associated with the modeling of the top quark p_T spectra also shows a larger post-fit value and a pull close to one standard deviation. This is because of the differences between the data and simulation discussed in Section 6.2.

The observed constraints and impacts agree with the expectations derived from fitting Asimov pseudo-data built using the nominal simulation. Larger constraints are obtained on the normalizations of the Z+jets and single top quark production. This is explained by large contributions of these events in the background-dominated categories that helps to constrain the large pre-fit normalization uncertainty. The NP associated with the electron identification efficiency shows a tight constraint from combining decay channels with different electron and muon multiplicities in the fit. This effect was seen in previous CMS measurements [31, 42] and is expected.

The constraints on NPs associated with the JES, JER, and jet identification demonstrate the power of the fit method, since the analysis phase space region and fitted kinematic distributions are sensitive to these effects. Also, because the analyzed phase space region is different from the one used to derive the corresponding calibrations, deviations from the pre-fit values and significant constraints are expected. In the fit, the value for m_t^{MC} is determined to be 171.93 ± 0.65 GeV, where the uncertainty includes both the statistical and systematic components.

Normalized differential cross section By calculating the normalized cross section, normalization uncertainties largely cancel out and the overall uncertainty is reduced. Thus, a higher precision can be reached extracting the top quark mass by comparing normalized differential cross sections. The values and the uncertainty in the normalized differential cross section are derived from the fit results for the absolute cross section by taking into account the correlation and covariance matrices as obtained in the fit to data. By dividing the value of the absolute differential cross section for each bin by the sum of the values of all the bins, the normalized differential cross section is determined. The uncertainties are symmetrized prior to the normalization procedure as provided by HESSE [341]. The propagation of the uncertainties to the normalized cross sections is performed via Gaussian error propagation individually for the statistical and total uncertainties:

$$\sigma_f^2 = |\frac{df}{dx}|^2\sigma_x^2 + |\frac{df}{dy}|^2\sigma_y^2 + 2\frac{df}{dy}\frac{df}{dy}\sigma_{xy}, \tag{6.19}$$

where $\sigma_{xy} = \sigma_x\sigma_y\rho_{xy}$ is the covariance between x and y and ρ_{xy} their correlation, and x and y refer to the per-bin cross section or total cross section, respectively. This approach is valid as the number of events is large enough to assume the Gaussian limit of the likelihood function. The normalization procedure is validated using a toy approach, as described in Ref. [42].

The correlation matrices for the pure statistical and total uncertainties are given in Figure 6.28 for the absolute differential cross section, and in Figure 6.29 for the normalized differential cross section. The resulting normalized cross section is shown in Figure 6.26 (right) compared to the theoretical predictions. While the absolute cross section shows a stronger dependence on α_S, this dependence is small for the normalized cross section [160]. This benefits the mass extraction, as the measurement becomes generally independent of α_S.

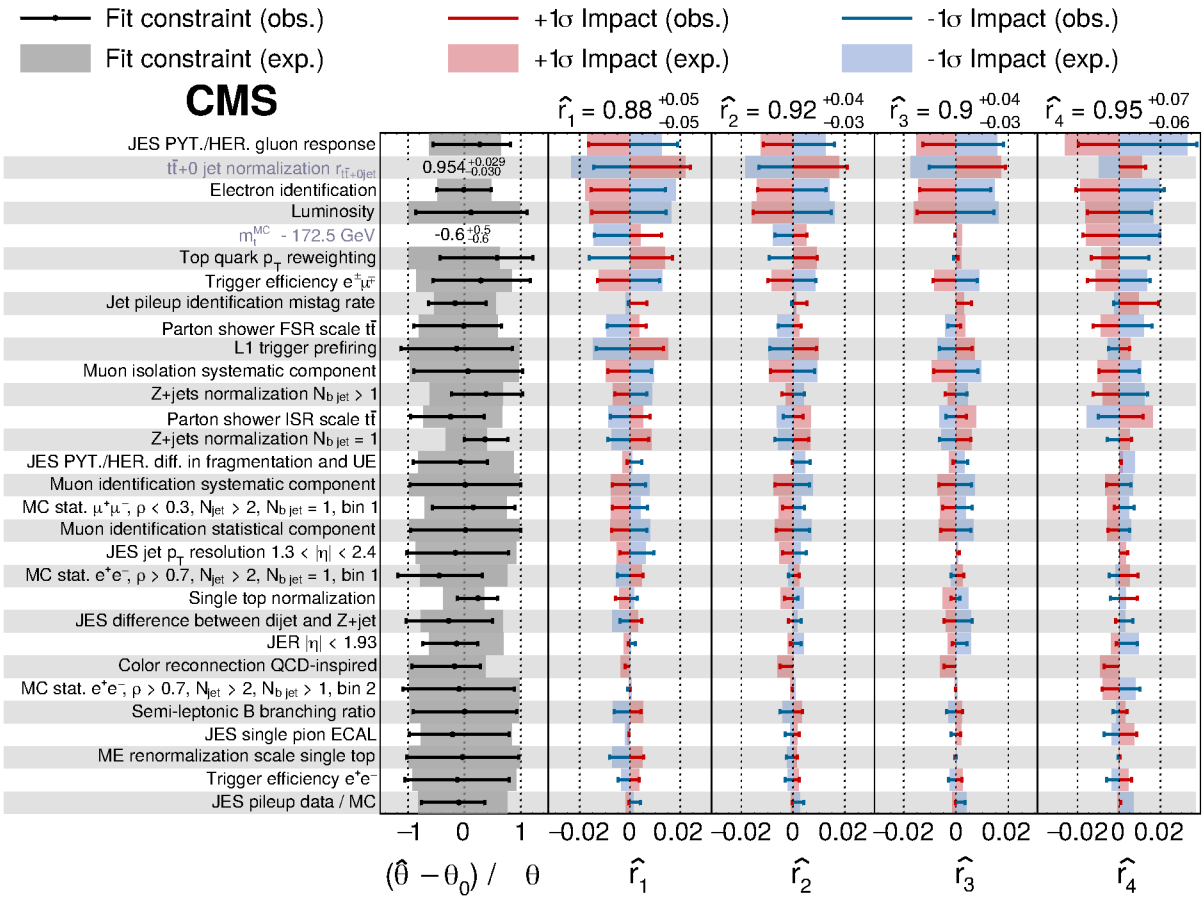

Figure 6.27: The fitted nuisance-parameter values and their impacts $\Delta\hat{r}_k$ on the signal strengths $\hat{r}_k$ from the fit to the data, ordered by their relative summed impact. Only the 30 highest ranked parameters are shown. The resulting fitted values of $\hat{r}_k$ and their total uncertainties are also given. The nuisance-parameter values ($\hat{\theta}$, black lines) are shown in comparison to their input values θ_0 before the fit and relative to their uncertainty $\Delta\theta$. The impact $\Delta\hat{r}_k$ for each nuisance parameter is the difference between the nominal best fit value of r_k and the best fit value when only that nuisance parameter is set to its best fit value $\hat{\theta}$ while all others are left free. The red and blue lines correspond to the variation in $\Delta\hat{r}_k$ when the nuisance parameter is varied up and down by its fitted uncertainty ($\Delta\theta$), respectively. The corresponding gray, red, and blue regions show the expected values from fits to pseudo-data. For the nuisance parameters associated with the $t\bar{t}$+0 jet normalization and m_t^{MC}, the values after the fit to the data are given, because no prior pdf is assigned. The plot is taken from Ref. [49].

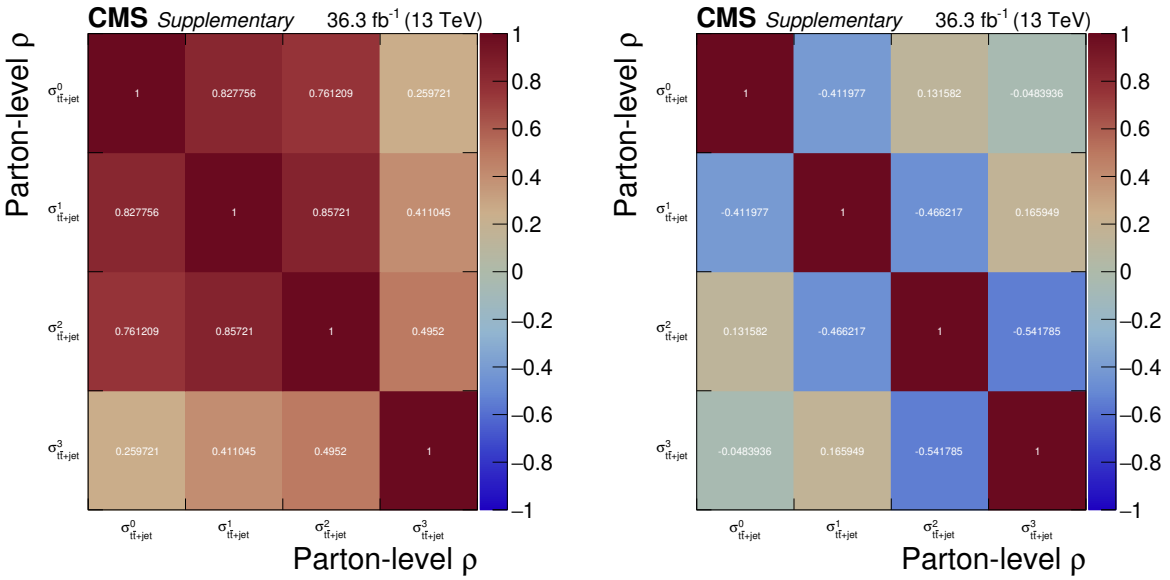

Figure 6.28: Correlation matrices for the total uncertainty (left) and statistical uncertainty (right) between the bins of the absolute differential cross section.

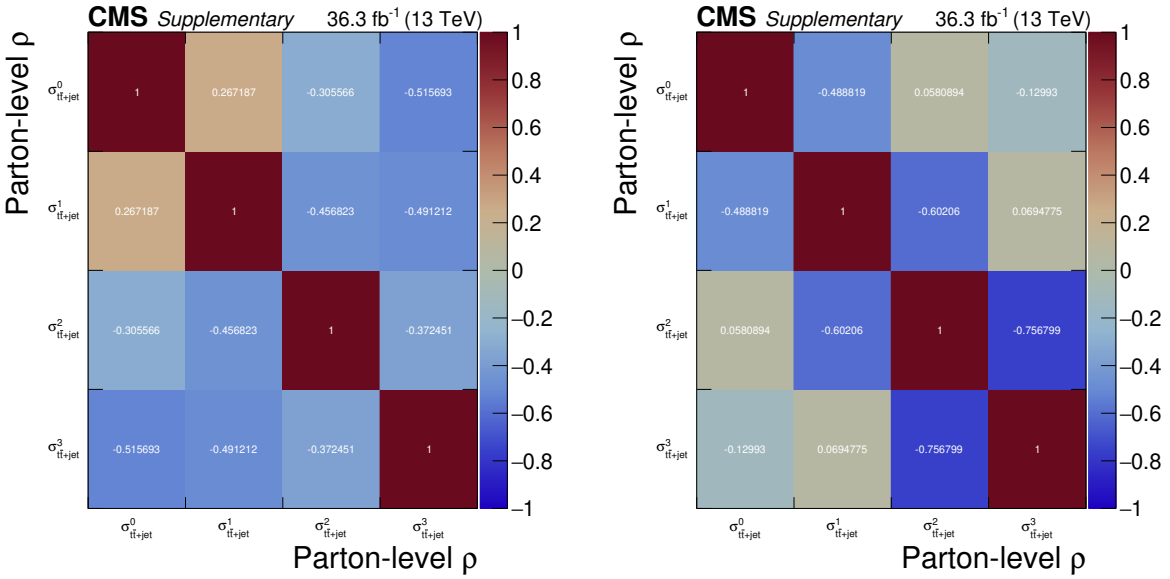

Figure 6.29: Correlation matrices for the total uncertainty (left) and statistical uncertainty (right) between the bins of the normalized differential cross section.

Mass extraction

The value of m_t^{pole} is extracted from a χ^2 fit of the NLO theoretical predictions to the normalized differential cross section as explained in Section 5.1, where the value of m_t^{pole} is varied from 168.5 to 176.5 GeV. Since in a fit to a normalized distribution, the value of the measurement in one bin is correlated to the measurements in the others, the measurement in the first bin is removed from the χ^2 fit. The χ^2 values and the fit result are independent of the choice of the removed bin. PDF uncertainties are evaluated in each bin for each eigenvector of the PDF uncertainties, and are included in the total covariance matrix $\mathbf{V}$ used for the χ^2 fit. They are assumed to be fully correlated among all bins. For CT18NLO, the uncertainties evaluated at 90% CL are rescaled to the 68% CL for consistency and symmetrized by taking the maximum of the up and down variations. Additionally, the extrapolation uncertainties for all relevant theoretical uncertainties are symmetrized by taking the average of the positive and negative impacts and are included in the calculation of the χ^2 by adding them to $\mathbf{V}$. They are further assumed to be fully correlated among all bins and the sign of the variation is considered. The resulting χ^2 curves for both the ABMP16NLO and CT18NLO comparisons are shown in Figure 6.30 (left). The best fit values for m_t^{pole} are extracted at the χ^2 minimum, and the total fit uncertainty is obtained using the tolerance criterion of $\Delta\chi^2 = 1$.

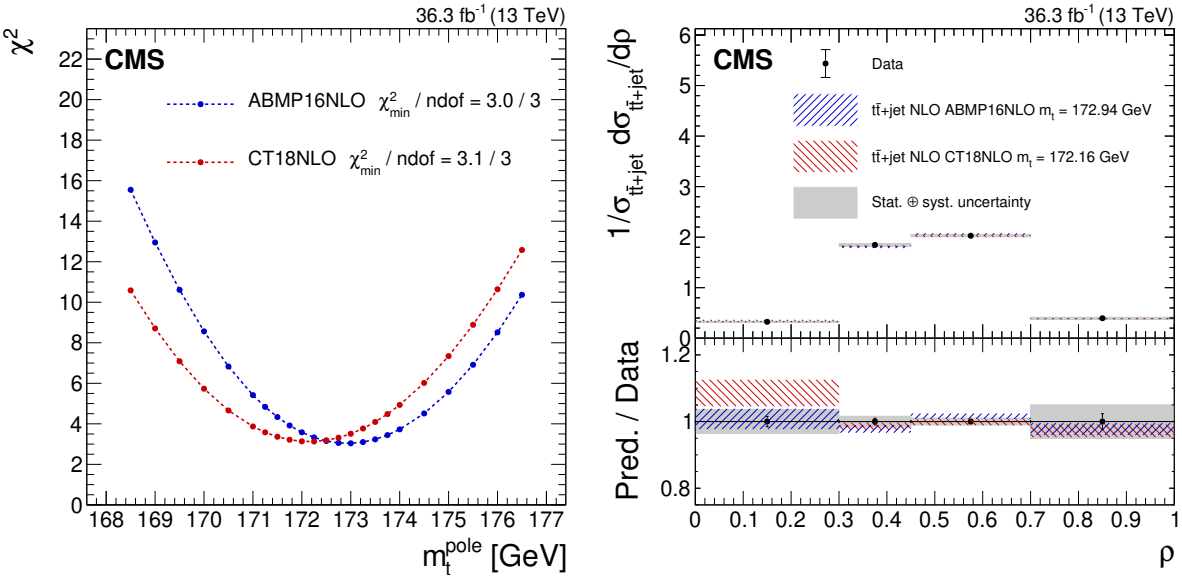

Figure 6.30: Left: the χ^2 values versus m_t^{pole} from the fit of the measured normalized $t\bar{t}$+jet differential cross sections to the theoretical predictions using the ABMP16NLO (blue points) and CT18NLO (red points) PDF sets. The minimum χ^2 value and the number of degrees of freedom (ndof) are given for each fit. Right: the measured normalized $t\bar{t}$+jet differential cross section (points) as a function of ρ, compared to the predictions using the two PDF sets and the corresponding best fit values for m_t^{pole} (hatched bands). The lower panel gives the ratio of the theoretical predictions to the measured values. For both panels, the vertical bars on the points show the statistical uncertainty in the data, the height of the hatched bands represent the theoretical uncertainties in the predictions, and the gray band gives the total uncertainty in the measured cross section. The plots are taken from Ref. [49].

To estimate the impact of the scale variation uncertainty, the χ^2 fits are repeated using the theoretical predictions with μ_R and μ_F varied independently by factors of 0.5 and 2, avoiding

cases where $\mu_F/\mu_R = 4$ or $1/4$. The total scale uncertainty is then estimated by taking the maximum difference in the m_t^{pole} results with respect to the nominal one. The resulting value of

$$m_t^{pole} = 172.93 \pm 1.26\,(\text{fit})\,^{+0.51}_{-0.43}\,(\text{scale})\,\text{GeV}$$

is determined using the ABMP16NLO PDF set. The fit uncertainty corresponds to the total statistical and systematic uncertainties from the χ^2 fit including the PDF and extrapolation uncertainties, and the scale uncertainty originates from the variation in the μ_R and μ_F scales. Using the CT18NLO PDF set instead, the measured value is

$$m_t^{pole} = 172.13 \pm 1.34\,(\text{fit})\,^{+0.50}_{-0.40}\,(\text{scale})\,\text{GeV}.$$

The total uncertainty in m_t^{pole} corresponds to 1.36 (1.43) GeV for the ABMP16NLO (CT18NLO) PDF set. Comparisons between the unfolded data at parton level and theoretical predictions for the determined values of m_t^{pole} for both PDF sets used are given in Figure 6.30 (right). Good agreement between the fitted prediction and the measured cross section for both PDF sets is observed, with the values for the χ^2_{min} per degree of freedom (ndof) of 3.0 (3.1) for the ABMP16NLO (CT18NLO) PDF set.

The final results for m_t^{pole} are in good agreement with previous measurements of m_t^{pole} using $t\bar{t}$+jet events [46] at a center-of-mass energy of 8 TeV by the ATLAS collaboration and the triple-differential cross sections for $t\bar{t}$ production at a center-of-mass energy of 13 TeV by CMS [38]. Compared to the ATLAS result, somewhat lower sensitivity to m_t^{pole} is expected in this analysis due to the higher center-of-mass energy, as discussed in Chapter 2. Furthermore, in this analysis the systematic uncertainty due to the PDFs is fully taken into account in the fit, and the $t\bar{t}$+0 jet background is considered a separate process with a dedicated normalization uncertainty in the signal extraction. This leads to an increase in the total uncertainty in m_t^{pole} with respect to previous measurements.

7 Future prospects

During the LHC long shutdown 2, the CMS Collaboration initiated a collaboration-wide effort to recalibrate the data and optimize the simulation and reconstruction to reprocess the data collected during Run 2. New MC samples are simulated based on improved detector description and tunes, and the data has been recalibrated and reanalyzed. Several improvements are identified for the detector performance [350] and object reconstruction [246, 351], so that the impact on the $m_{\mathrm{t}}^{\mathrm{pole}}$ measurement is expected to be larger due to, e.g., improved JES and JER calibrations, and from the increased MC sample size. In addition, background processes such as Z+jets production can be modeled using higher-order simulation based on POWHEG using MiNNLO$_{\mathrm{PS}}$ [352, 353]. This helps to reduce the modeling uncertainties of this background contribution significantly, allowing to include the $N_{\mathrm{b\,jet}} < 1$ event category additionally in the fit, leading to even further reduced extrapolation uncertainties and a larger acceptance. All these improvements for the data taken in 2017 and 2018 were not available at the time the presented analysis was performed.

The measurement described in Chapter 6 is based on the data recorded by CMS in 2016 only. In this section, the MC-based projection of the $m_{\mathrm{t}}^{\mathrm{pole}}$ extraction for the full Run 2 data of the LHC is presented. The fit is performed not to the actual data, but to an Asimov dataset [354], which is obtained from the nominal MC prediction, considering the 2016, 2017 and 2018 CMS data-taking configurations. By definition, all fitted parameters agree with their generated values, yielding the expected result from the model used. The increase in the total integrated luminosity by a factor of four leads to an improvement in the statistical precision of the measurement by a factor of two. As a result, tighter constraints on systematic uncertainties are expected, leading to a significant reduction in the total uncertainty.

Particular attention is put on the modeling of all systematic uncertainties and their correlations among the three different years. The expected results are evaluated using a likelihood fit using a toy setup, and the $m_{\mathrm{t}}^{\mathrm{pole}}$ extraction is performed based on theoretical predictions to evaluate the final sensitivity.

7.1 Setup of the study

This part of the analysis is carried out using identical MC simulated samples as for the measurement presented in Section 6.2, but adapted to the conditions and detector changes of the

2017 and 2018 data taking periods. Likewise, all MC simulated samples are corrected using the CMS centrally determined SFs and other corrections as for the 2016 analysis. The trigger efficiency is determined separately for 2017 and 2018 data, following the main analysis methodology. Although the exact details of some of the signal trigger paths vary, the requirements on the minimal p_{T} of the lepton candidates at HLT for the dilepton triggers are identical to the 2016 analysis. In case of the single lepton triggers in 2017, the requirement on the p_{T} of the electron (muon) candidates changes by $+5\,\mathrm{GeV}$ ($+3\,\mathrm{GeV}$). The signal efficiency remains unaffected, and the event selection is kept consistent with the main analysis.

The regression and classification NNs are not recalibrated for this study, but are used as for the main measurement since the response is not expected to differ significantly among the three data sets. All three data sets are included in the likelihood fit by treating them as individual event categories to extract the signal strength parameters r_k from a simultaneous fit.

All other event categories and distributions that are fitted are unchanged so that the 2016 part of the analysis is retained, and the 2017 and 2018 categories are as consistent as possible with 2016. For each year, the binning of the distributions and hence the templates are evaluated independently to reflect the specific properties of each data set. The data and pre-fit predictions for the signal and background are shown in Figures 7.1 and 7.2. In general, good agreement within the total systematic uncertainties is observed.

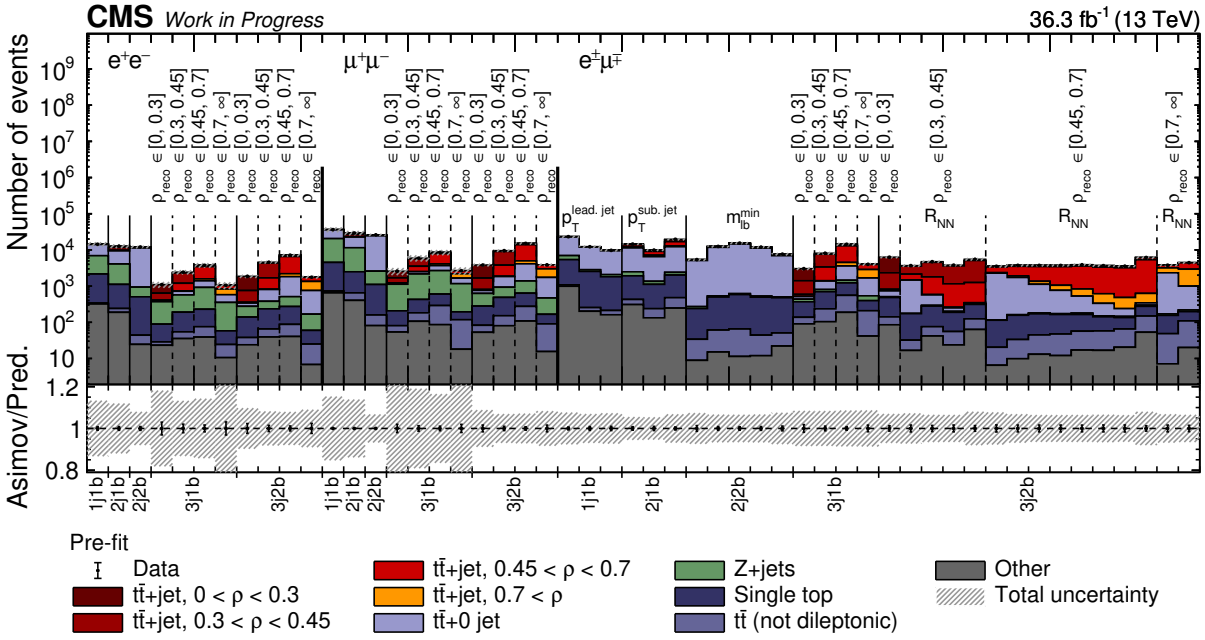

Figure 7.1: The distributions from 2016 Asimov data (points) and simulated signal and background (colored histograms) used in the maximum likelihood fit before the fit to the data. The distributions are shown for each dilepton type and each event category, where the x-axis label "mjnb" refers to events with m jets and n b jets. The vertical bars on the points show the statistical uncertainty in the data. The hatched band represents the total uncertainty in the sum of the simulated signal and background predictions. The lower panel gives the ratio of the data to the sum of the simulated predictions.

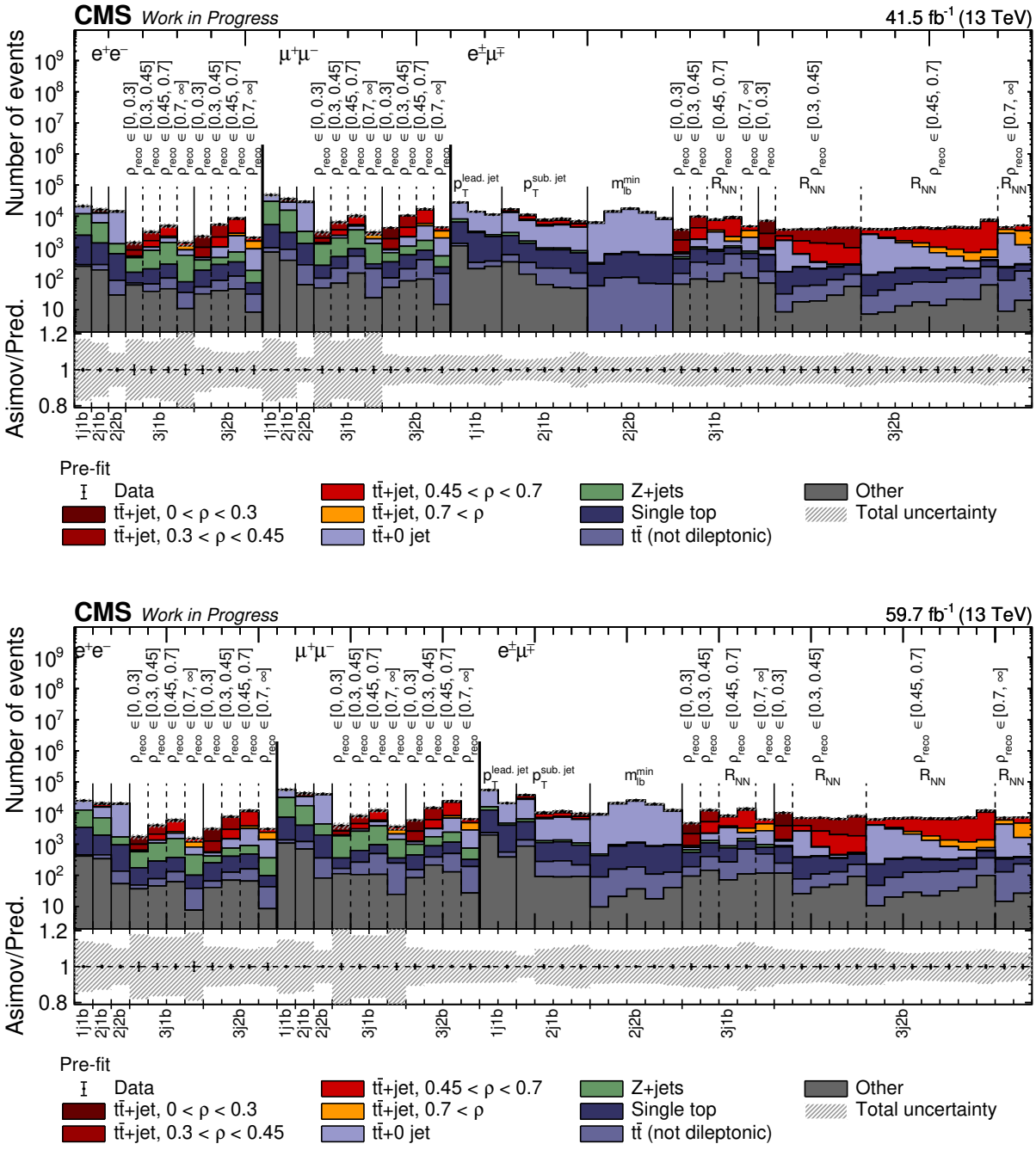

Figure 7.2: The distributions from 2017 (up) and 2018 (bottom) Asimov data (points) and simulated signal and background (colored histograms) used in the maximum likelihood fit before the fit to the data. The distributions are shown for each dilepton type and each event category, where the x-axis label "mjnb" refers to events with m jets and n b jets. The vertical bars on the points show the statistical uncertainty in the data. The hatched band represents the total uncertainty in the sum of the simulated signal and background predictions. The lower panel gives the ratio of the data to the sum of the simulated predictions.

Systematic uncertainties and correlation assumptions

Although the statistical uncertainty is not dominant, by using the three independent data sets of the LHC Run 2, the precision of the measurement can be improved significantly. Since the same MC simulated samples are used to model the signal and background, all modeling uncertainties are correlated among the three data sets and can be better constrained in the fit. Examples are the ME scales, the background normalizations, or the top quark MC mass $m_{\mathrm{t}}^{\mathrm{MC}}$. The same holds for many experimental calibrations that are limited by systematic uncertainties. Therefore, a careful evaluation of the correlations of all systematic uncertainties, and in turn the NPs, has to be performed. Since the differential cross section $\sigma^{k}_{\mathrm{t\bar{t}+jet}}(\mathrm{MC})$ is fully correlated among the years, there is a common signal strength parameter r_k for the full Run 2 measurement.

Where possible, the contribution of the experimental uncertainties based on calibrations is split into different subsources treated as fully correlated or uncorrelated. In cases where the statistical uncertainty of the data or simulation dominates, the correlation is assumed to be zero. The same holds for calibrations where the method is changed for different data taking periods, while for systematically limited sources where the same method is used, a full correlation is considered. An example for a fully decorrelated uncertainty is the trigger efficiency, since the trigger paths differ for all three years, or the jet energy resolution, which is limited by the number of data events used in the calibrations. The correlation assumptions for all the systematic uncertainties, together with the number of individual NPs, N_{nuis}, used in the likelihood fit, are given in Table 7.1.

In addition to the systematic uncertainties considered in the nominal measurement, an additional effect has to be considered for the 2018 data set. During the data taking, power issues of two modules of the HCAL detector lead to complications in the jet energy measurement in the regions of $-1.57 < \phi < 0.87$ & $-2.5 < \eta < -1.3$ and $-1.57 < \phi < 0.87$ & $-3.0 < \eta < -2.5$. It is further denoted as "HEM15/16" issue. To assess the impact of this effect, the jet energy in the two affected kinematic ranges is scaled down by a factor of 20 and 35%, respectively, and a one-sided uncertainty is constructed. The effect is fully propagated to $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$. In addition to the uncertainty considered, a dedicated treatment is determined for the analysis presented in this thesis to reduce the impact of the HEM issue. Events are removed from the selection if they contain at least either

- an electron with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and passing loose identification criteria,
- a muon with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and passing loose identification criteria,
- or a jet with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$,

within the HEM-affected region. The removal efficiency is validated in data and simulation by evaluating it as a function of the jet multiplicity. It is measured in two independent kinematic regions, where one is constructed orthogonal to the preselection by inverting the requirement of $m_{\ell\ell}$ inside the Z boson mass window in the $\mathrm{e^+e^-}$ and $\mu^+\mu^-$ channels. The other one corresponds to the preselection. Further, the efficiency is studied using data events from early 2018 runs that are unaffected by the HEM15/16 issue. Good agreement for the removal efficiency is observed for the full kinematic range in the data and simulation, so that no dedicated correction or uncertainty has to be considered.

7.2 Results

The post-fit predictions for the signal and background after the fit to the Asimov data are presented in Figures 7.3 and 7.4 for 2016, 2017, and 2018. Corresponding NP constraints and

Table 7.1: The correlations among the three years for all systematic uncertainties are given. Additionally, the number of resulting components and nuisance parameters is listed as N_{nuis}.

Systematic uncertainty	2016–2017	2016–2018	2017–2018	N_{nuis}
Trigger efficiency	0%	0%	0%	9
L1 ECAL and muon prefiring	100%	100%	100%	1
Electron identification efficiencies	100%	100%	100%	1
Electron reconstruction efficiencies	100%	100%	100%	1
Electron scale systematical — correlated	100%	100%	100%	2
Electron scale statistical — uncorrelated	0%	0%	0%	6
Electron resolution	100%	100%	100%	2
Muon identification efficiencies — statistical	0%	0%	0%	3
Muon identification efficiencies — systematical	100%	100%	100%	1
Muon isolation efficiencies — statistical	0%	0%	0%	3
Muon isolation efficiencies — systematical	100%	100%	100%	1
Muon scale — systematical	100%	100%	100%	4
Muon scale — statistical	0%	0%	0%	3
Jet pileup identification	100%	100%	100%	2
Background normalization	100%	100%	100%	11
m_{t}^{MC} dependence	100%	100%	100%	1
Pileup modeling	100%	100%	100%	1
b tagging heavy flavors — statistical	0%	0%	0%	3
b tagging heavy flavors — systematical	100%	100%	100%	3
b tagging light flavors — correlated	100%	100%	100%	1
b tagging light flavors — uncorrelated	0%	0%	0%	3
Unclustered $p_{\text{T}}^{\text{miss}}$	0%	0%	0%	3
μ_{R} and μ_{F} scales	100%	100%	100%	2
Parton shower α_S^{FSR}	100%	100%	100%	1
Parton shower α_S^{ISR}	100%	100%	100%	1
ME–PS matching (h_{damp})	100%	100%	100%	1
Underlying event tune CP5	100%	100%	100%	1
PDF eigenvectors	100%	100%	100%	100
PDF α_S	100%	100%	100%	1
Color reconnection	100%	100%	100%	3
b-fragmentation	100%	100%	100%	3
b semileptonic branching fraction	100%	100%	100%	1
Jet energy resolution	0%	0%	0%	6
Jet energy scale — correlated	100%	100%	100%	16
Jet energy scale — uncorrelated	100%	100%	100%	22
Luminosity — uncorrelated	0%	0%	0%	3
Luminosity — correlated	100%	100%	100%	1
Luminosity — correlated 2017/2018	0%	100%	100%	1

impacts are shown in Figure 7.5. As in the case of the 2016-only analysis, NPs associated to background process normalizations are tightly constrained. Similarly, constraints for NP of uncertainties that have a large impact on the fitted values for r_k and are correlated among the three years show significant constraints, as for example individual JES or modeling parameters. The value of $m_{\mathrm{t}}^{\mathrm{MC}}$ is expected to be determined with a precision of 0.35 GeV, while the $\mathrm{t\bar{t}}$+0 jet normalization is expected to be determined with a 2.2% precision. The results are significantly improved in comparison to the measurement of $m_{\mathrm{t}}^{\mathrm{MC}}$ and the $\mathrm{t\bar{t}}$+0 jet normalization using 2016 data, presented in Chapter 6, with respective uncertainties of 0.6 GeV and 3%.

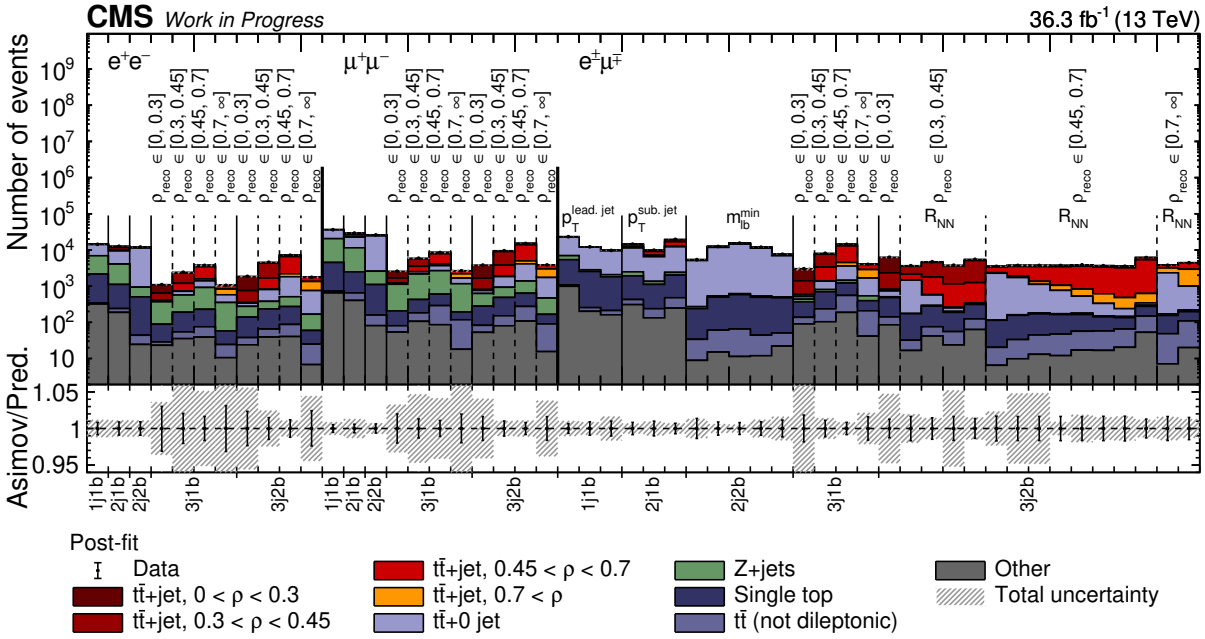

Figure 7.3: The distributions from 2016 data (points) and simulated signal and background (colored histograms) used in the maximum likelihood fits after the fit to the data. The distributions are shown for each dilepton type and each event category, where the x-axis label "mjnb" refers to events with m jets and n b jets. The vertical bars on the points show the statistical uncertainty in the data. The hatched band represents the total uncertainty in the sum of the simulated signal and background predictions. The lower panel gives the ratio of the data to the sum of the simulated predictions.

The individual uncertainty contributions are listed in Table 7.2. Compared to the results of the 2016 measurement, the statistical uncertainty improves by a factor of about two, while the total uncertainty improves by a factor of about 1.5–1.8, depending on the bin in ρ.

The $\mathrm{t\bar{t}}$+jet differential cross section at parton level, as expected from the Asimov fit, is given in Figure 7.6. With the reduced uncertainty, the sensitivity of the ρ distribution to $m_{\mathrm{t}}^{\mathrm{pole}}$ in a larger range can be explored.

Using the same NLO $\mathrm{t\bar{t}}$+jet theoretical predictions as for the nominal measurement, the χ^2 scan is repeated using the Asimov expectation. Here, the data is represented by the theoretical prediction itself with a nominal value of $m_{\mathrm{t}}^{\mathrm{pole}} = 172.5\,\mathrm{GeV}$ for each PDF set, while the relative expected uncertainties from the Asimov fit are applied. The resulting χ^2 distribution is shown

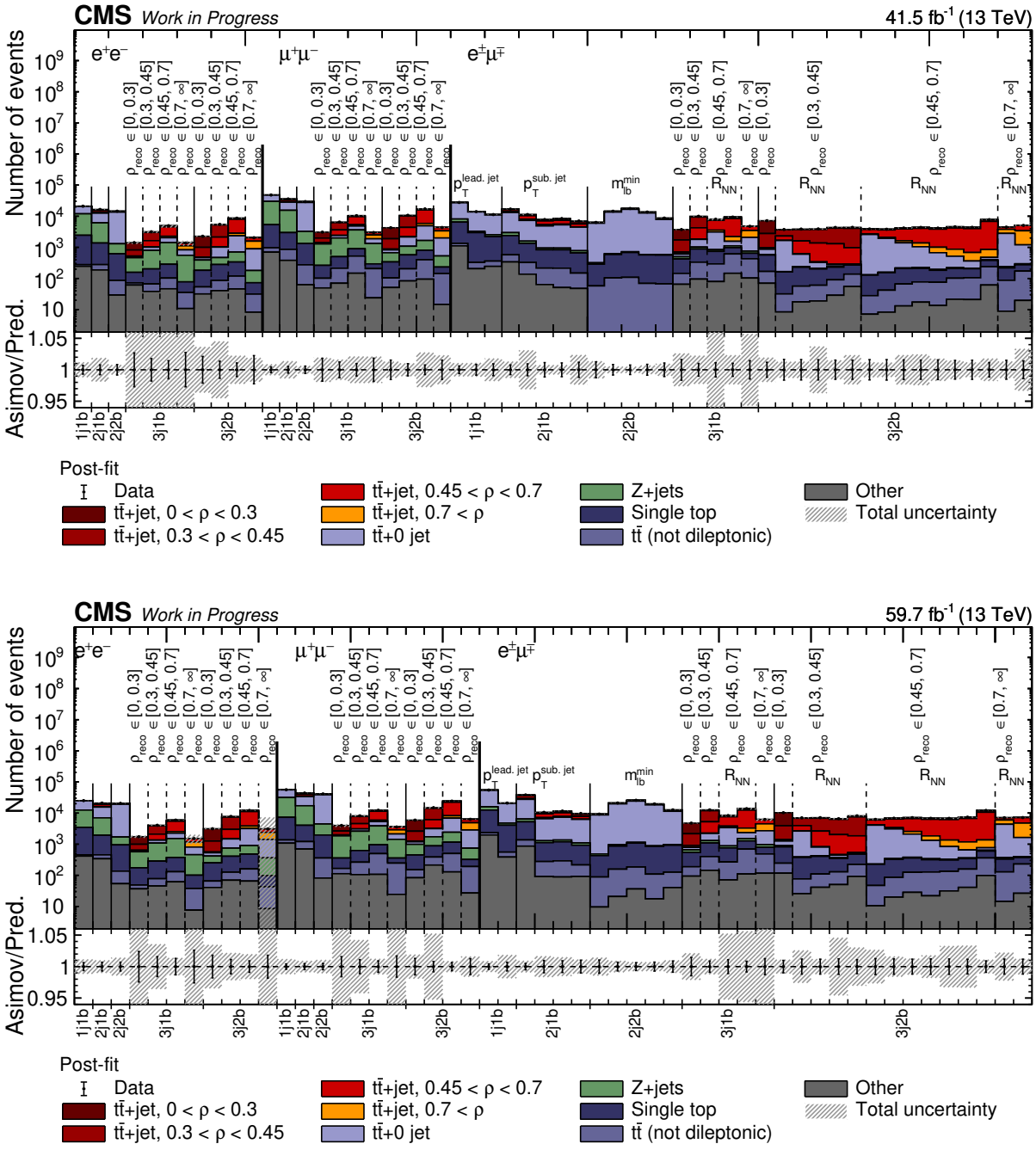

Figure 7.4: The distributions from 2017 (up) and 2018 (bottom) data (points) and simulated signal and background (colored histograms) used in the maximum likelihood fits after the fit to the data. The distributions are shown for each dilepton type and each event category, where the x-axis label "mjnb" refers to events with m jets and n b jets. The vertical bars on the points show the statistical uncertainty in the data. The hatched band represents the total uncertainty in the sum of the simulated signal and background predictions. The lower panel gives the ratio of the data to the sum of the simulated predictions.

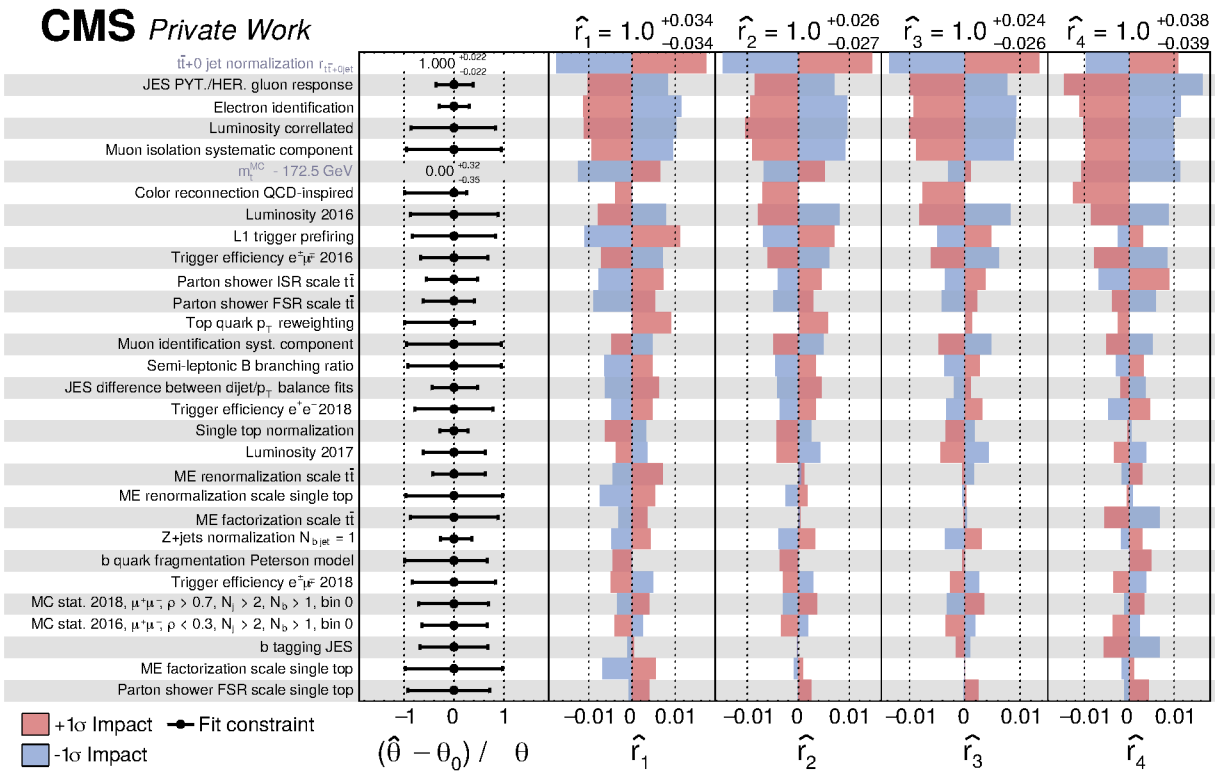

Figure 7.5: The fitted nuisance-parameter values and their impacts $\Delta\hat{r}_k$ on the signal strengths $\hat{r}_k$ from the fit to the data, ordered by their relative summed impact. Only the 30 highest ranked parameters are shown. The resulting fitted values of $\hat{r}_k$ and their total uncertainties are also given. The nuisance-parameter values ($\hat{\theta}$, black lines) are shown in comparison to their input values θ_0 before the fit and relative to their uncertainty $\Delta\theta$. The impact $\Delta\hat{r}_k$ for each nuisance parameter is the difference between the nominal best fit value of r_k and the best fit value when only that nuisance parameter is set to its best fit value $\hat{\theta}$ while all others are left free. The red and blue lines correspond to the variation in $\Delta\hat{r}_k$ when the nuisance parameter is varied up and down by its fitted uncertainty ($\Delta\theta$), respectively. The corresponding gray, red, and blue regions show the expected values from fits to pseudo-data. For the nuisance parameters associated with the $t\bar{t}$+0 jet normalization and m_t^{MC}, the values after the fit to the data are given, because no prior pdf is assigned.

Table 7.2: The relative uncertainties $\Delta\sigma^{k}_{\mathrm{t\bar{t}+jet}}$ in the parton-level cross section values $\sigma^{k}_{\mathrm{t\bar{t}+jet}}$ and their sources in each bin k of the ρ distribution.

Uncertainty Source	$\Delta\sigma^{1}_{\mathrm{t\bar{t}+jet}}$ [%]	$\Delta\sigma^{2}_{\mathrm{t\bar{t}+jet}}$ [%]	$\Delta\sigma^{3}_{\mathrm{t\bar{t}+jet}}$ [%]	$\Delta\sigma^{4}_{\mathrm{t\bar{t}+jet}}$ [%]
Experimental				
Muon identification	1.2	1.2	1.1	1.4
Muon energy scale and resolution	0.4	0.3	0.1	0.6
Electron identification	1.4	1.1	1.1	1.4
Electron energy scale and resolution	0.2	0.3	0.3	0.6
Jet energy scale	1.3	1.0	1.0	2.1
Jet energy resolution	0.1	0.3	0.2	0.7
Jet identification	0.4	0.4	0.1	0.6
$p_{\mathrm{T}}^{\mathrm{miss}}$	0.5	0.4	0.2	0.6
b jet identification	0.3	0.3	0.4	0.9
Trigger efficiency	0.7	0.4	0.7	1.1
Total	2.5	2.0	1.9	2.8
Background normalization				
$\mathrm{t\bar{t}}$+0 jet	1.8	1.6	1.5	1.4
Z+jets	1.4	1.0	1.0	1.2
Single top quark	0.6	0.2	0.2	0.7
Total	2.2	1.8	1.7	1.7
Modeling				
Z+jets ME scale	0.4	0.4	0.2	0.6
Single top quark ME/FSR/ISR scales	1.0	0.4	0.4	0.6
$\mathrm{t\bar{t}}$ PDF	0.5	0.4	0.2	0.6
$\mathrm{t\bar{t}}$ ME scale	0.8	0.4	0.3	0.9
$\mathrm{t\bar{t}}$ ISR scale	0.8	0.3	0.3	0.9
$\mathrm{t\bar{t}}$ FSR scale	0.7	0.1	0.2	0.6
$\mathrm{t\bar{t}}$ top quark p_{T}	0.7	0.3	0.1	0.6
b fragmentation	0.5	0.2	0.1	0.5
Color reconnection	0.2	0.6	0.6	1.3
$\mathrm{t\bar{t}}$ matching scale	0.3	0.4	0.2	0.5
Underlying-event tune	0.1	0.1	0.2	0.6
Total	1.9	1.1	1.0	1.9
Integrated luminosity	0.6	0.8	0.8	1.0
$m_{\mathrm{t}}^{\mathrm{MC}}$	0.9	0.5	0.1	1.3
Finite size of simulated samples	1.2	0.9	0.9	1.5
Total systematic	3.5	2.7	2.6	4.0
Statistical	0.7	0.5	0.4	1.2
Total	3.5	2.8	2.6	4.2

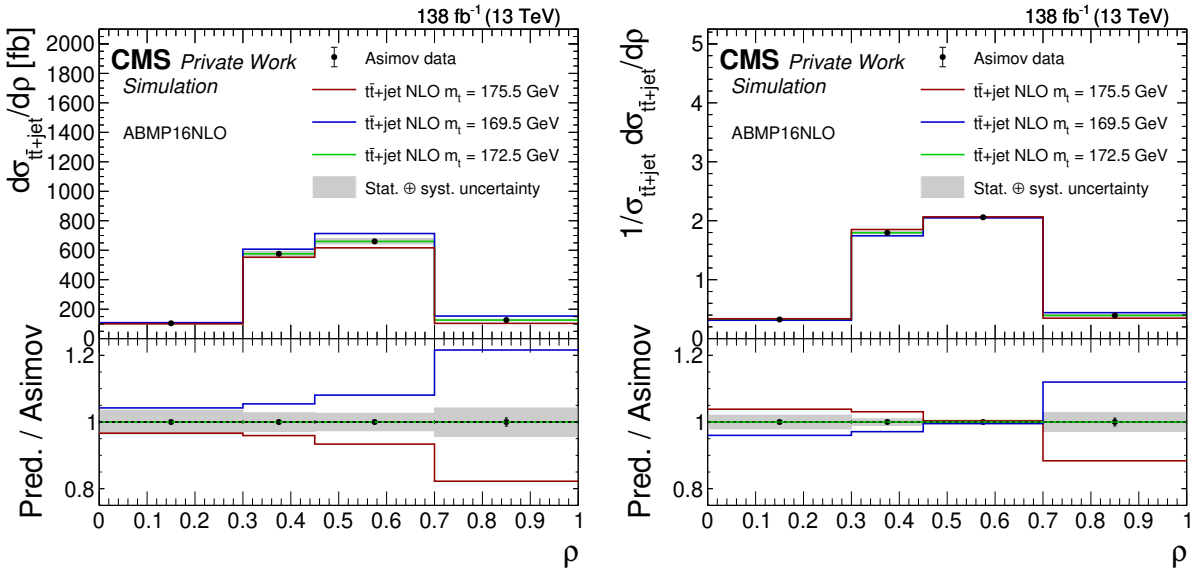

Figure 7.6: The absolute (left) and normalized (right) $\mathrm{t\bar{t}}$+jet differential cross section as a function of ρ for the Asimov data (points) and theoretical predictions described in the text using the AMBP16NLO PDF set from the NLO MC with three different m_t values and from the POWHEG (POW) + PYTHIA8 (PYT) calculations (lines). The vertical bars on the points show the statistical uncertainty in the data and the shaded region represents the total uncertainty in the measurement. The lower panels give the ratio of the predictions to the data.

in Figure 7.7, where by definition the minimum agrees with the input value ($\chi^2_\mathrm{min} = 0$). Values for $m_\mathrm{t}^\mathrm{pole}$ extracted using the ABMP16NLO PDF set yield

$$m_\mathrm{t}^\mathrm{pole} = 172.46 \pm 0.80\,(\mathrm{fit})\,^{+0.36}_{-0.34}\,(\mathrm{scale})\,\mathrm{GeV}.$$

For the CT18NLO PDF set, the resulting value is

$$m_\mathrm{t}^\mathrm{pole} = 172.50 \pm 0.80\,(\mathrm{fit})\,^{+0.38}_{-0.32}\,(\mathrm{scale})\,\mathrm{GeV}.$$

Compared to the 2016 results, the significant improvement of the "fit" uncertainty is expected, reducing the total uncertainty from 1.4 to 0.9 GeV, which would result in one of the most precise measurements of $m_\mathrm{t}^\mathrm{pole}$.

The Asimov $\mathrm{t\bar{t}}$+jet cross section compared to the theoretical predictions is shown in Figure 7.7. For all the ρ bins except the last one, the theoretical uncertainty dominates for the Run 2 measurement.

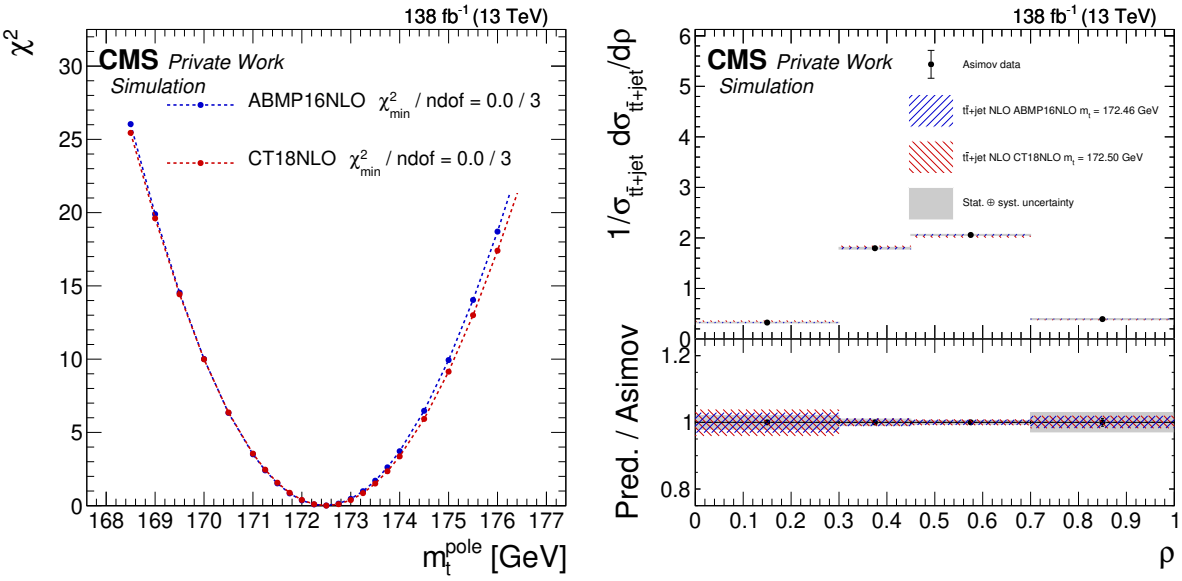

Figure 7.7: Left: the χ^2 values versus m_t^{pole} from the fit of the measured normalized $t\bar{t}$+jet differential cross sections to the theoretical predictions using the ABMP16NLO (blue points) and CT18NLO (red points) PDF sets. The minimum χ^2 value and the number of degrees of freedom (ndof) are given for each fit. Right: the expected measured normalized $t\bar{t}$+jet differential cross section (points) as a function of ρ, compared to the predictions using the two PDF sets and the corresponding best fit values for m_t^{pole} (hatched bands). The lower panel gives the ratio of the theoretical predictions to the measured values. For both panels, the vertical bars on the points show the statistical uncertainty in the Asimov data, the height of the hatched bands represent the theoretical uncertainties in the predictions, and the gray band gives the total uncertainty in the measured cross section.

8 Conclusions

This thesis presents the extraction of the top quark pole mass, $m_{\mathrm{t}}^{\mathrm{pole}}$, by using the events where the top quark-antiquark pair ($\mathrm{t\bar{t}}$) is produced in association with at least one energetic jet ($\mathrm{t\bar{t}}$+jet) in proton-proton collisions at the CERN LHC at a center-of-mass energy of 13 TeV. The presence of a jet shows significantly enhanced sensitivity to $m_{\mathrm{t}}^{\mathrm{pole}}$, compared to $\mathrm{t\bar{t}}$ production.

The analyzed data are collected by the CMS experiment in 2016 and correspond to the total integrated luminosity of $36.3\,\mathrm{fb}^{-1}$. The dileptonic decays of the top quarks are used, and the transverse momentum of the additional jet is required to be above 30 GeV. The $\mathrm{t\bar{t}}$+jet production cross section is measured as a function of the invariant mass of the $\mathrm{t\bar{t}}$+jet system, $m_{\mathrm{t\bar{t}+jet}}$, and is used to extract the value of $m_{\mathrm{t}}^{\mathrm{pole}}$ at next-to-leading-order (NLO) in quantum chromodynamics. In particular, the ρ observable defined as $\rho = 2m_0/m_{\mathrm{t\bar{t}+jet}}$, with the scaling constant $m_0 = 170\,\mathrm{GeV}$, is considered.

Candidate events are selected if they contain at least two opposite-sign leptons ($\mathrm{e^+e^-}$, $\mu^+\mu^-$, $\mathrm{e^\pm}\mu^\mp$) and one jet identified as originating from a b quark in the final state (b jet). In each event, the kinematic properties of the $\mathrm{t\bar{t}}$+jet event are inferred using a novel machine-learning-based reconstruction technique, developed in this thesis. This is necessary, as the two neutrinos in dileptonic $\mathrm{t\bar{t}}$ events cannot be reconstructed and result in a significant amount of missing transverse momentum. In particular, the value of ρ as defined at the parton level is reconstructed with a regression neural network (NN), using a combination of low-level and high-level variables. The performance of the regression NN is found to lead to a superior performance compared to analytical solutions employed in the previous publications. The reconstruction efficiency is maximized and the resolution in the ρ observable improves by to a factor of two with respect to the conventional reconstruction methods.

The sensitivity to $m_{\mathrm{t}}^{\mathrm{pole}}$ is further enhanced by applying a NN-based multiclassifier, separating the candidate $\mathrm{t\bar{t}}$+jet signal events from the background. Advanced techniques, such as domain adaption by backpropagation, are employed to assure identical classification performance with respect to the value of ρ, mitigating possible bias. The outputs of the NN classifier are used to define the R_{NN} variable, which maximizes the discrimination power against the dominant background arising from $\mathrm{t\bar{t}}$ production without additional jet radiation ($\mathrm{t\bar{t}}$+0 jet).

The $\mathrm{t\bar{t}}$+jet differential cross section is measured at the parton level using a multidimensional profiled-maximum-likelihood fit to the final-state distributions. Events are categorized based on the jet and b jet multiplicities in bins of the reconstructed value of ρ. This approach maximizes

the visible phase space of the measurement, defined by the geometric acceptance of the detector and the p_{T} and $|\eta|$ requirements on the final-state objects. The systematic uncertainties are treated as nuisance parameters in the fit, constraining the uncertainties and their correlations from the experimental data. In the fit, the signal and background processes are fitted simultaneously, while the normalization of $\mathrm{t\bar{t}}$+0 jet is left unconstrained in the fit. A similar approach is applied to mitigate the dependence on the value of m_{t} assumed in the simulation, $m_{\mathrm{t}}^{\mathrm{MC}}$, treating the latter as a free parameter in the fit. In particular, the value of $m_{\mathrm{t}}^{\mathrm{MC}}$ is constrained by including in the fit the minimum invariant mass of a lepton and a b jet, $m_{\ell\mathrm{b}}^{\mathrm{min}}$. Additionally, sensitivity to relevant systematic uncertainties arising from the limited knowledge about the jet energy scale is achieved by exploring the transverse momentum distributions of the jets with the lowest p_{T} in event categories with less than three jets. For events with at least three reconstructed jets, R_{NN} is used as the auxiliary variable in the fit in each ρ bin.

An extrapolation to the full phase space is performed using the acceptance determined by the Monte Carlo simulation, and additional extrapolation uncertainties are considered.

The measured normalized $\mathrm{t\bar{t}}$+jet cross section as a function of ρ, $1/\sigma_{\mathrm{t\bar{t}+jet}}\ \mathrm{d}\sigma_{\mathrm{t\bar{t}+jet}}/\mathrm{d}\rho$, is compared to the theoretical prediction at NLO [160], where the value of $m_{\mathrm{t}}^{\mathrm{pole}}$ is varied from 168.5 GeV to 176.5 GeV. Two alternative sets of parton distribution functions are used in the theory predictions. The value of $m_{\mathrm{t}}^{\mathrm{pole}}$ is extracted from the χ^2 scan comparing the measured and the predicted cross sections. In the fit, the correlations of the uncertainties are fully taken into account by using the covariance matrix as obtained from the likelihood fit. The value of $m_{\mathrm{t}}^{\mathrm{pole}}$ is obtained as

$$m_{\mathrm{t}}^{\mathrm{pole}} = 172.93 \pm 1.36\,\mathrm{GeV}$$

using the ABMP16NLO [167] and

$$m_{\mathrm{t}}^{\mathrm{pole}} = 172.13 \pm 1.43\,\mathrm{GeV}$$

using the CT18NLO [166] parton distribution functions in the theory predictions, respectively. The uncertainties include the total statistical, systematic, and extrapolation uncertainties, and the theoretical uncertainties due to uncertainties in the parton distribution functions and the choice of the renormalization and factorization scales. The results are in good agreement with previous measurements of $m_{\mathrm{t}}^{\mathrm{pole}}$. The precision of the measurement is limited by the uncertainties in the jet energy corrections, and the modeling of the $\mathrm{t\bar{t}}$/$\mathrm{t\bar{t}}$+jet process. The presented analysis, entirely developed and performed for the purpose of this thesis, is also documented in Ref. [293]. This is the first time that the sensitivity of $\mathrm{t\bar{t}}$+jet production is used to extract $m_{\mathrm{t}}^{\mathrm{pole}}$ in proton-proton collisions at $\sqrt{s} = 13\,\mathrm{TeV}$, using a machine learning based reconstruction for the $\mathrm{t\bar{t}}$ kinematic properties.

During the LHC long shutdown in 2018–2022, the CMS Collaboration has significantly improved the calibrations of the data collected in Run 2, employing the optimized simulation and a revised reconstruction. These improvements were yet not available at the time of performing the presented measurement, based on the data collected in 2016. A sensitivity study is performed to estimate the expected precision for $m_{\mathrm{t}}^{\mathrm{pole}}$, once the full LHC Run 2 data set is used, increasing the data sample by a factor of four as compared to the one used in the presented measurement. By using the simulated events, a detailed study of the correlations for all systematic uncertainties is realized. Based on the same theoretical predictions as used for the main measurement, the final precision for $m_{\mathrm{t}}^{\mathrm{pole}}$ is estimated to be 0.9 GeV, which would result in one of the most precise $m_{\mathrm{t}}^{\mathrm{pole}}$ measurements.

Furthermore, the author of the thesis has significantly contributed to the b-quark jet identification (b tagging) at the CMS High Level Trigger (HLT). For the first time, the identification algorithm based on advanced neural network architectures, called DeepJet, is deployed at the HLT for the Run 3 data taking. In the same effort, the first recalibration of b tagging algorithms dedicated for the HLT is performed. Additionally, feasibility studies for the High-Luminosity LHC are carried out, and the first HLT paths using b jets are developed. These results are documented in Refs. [47, 48].

Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen benutzt zu haben.

Die eingereichte schriftliche Fassung entspricht der auf dem elektronischen Speichermedium.

Die Dissertation wurde in der vorgelegten oder einer ähnlichen Form nicht schon einmal in einem früheren Promotionsverfahren angenommen oder als ungenügend beurteilt.

Hamburg, den 25. Oktober 2022 Sebastian Wuchterl

Bibliography

[1] P. W. Higgs, "Broken symmetries and the masses of gauge bosons", *Phys. Rev. Lett.* **13** (1964) 508–509, doi:10.1103/PhysRevLett.13.508.

[2] F. Englert and R. Brout, "Broken symmetry and the mass of gauge vector mesons", *Phys. Rev. Lett.* **13** (1964) 321–323, doi:10.1103/PhysRevLett.13.321.

[3] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett. B* **716** (2012) 30–61, doi:https://doi.org/10.1016/j.physletb.2012.08.021.

[4] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett. B* **716** (2012) 1–29, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.

[5] CDF Collaboration, "Observation of top quark production in $\bar{p}p$ collisions", *Phys. Rev. Lett.* **74** (1995) 2626–2631, doi:10.1103/PhysRevLett.74.2626, arXiv:hep-ex/9503002.

[6] D0 Collaboration, "Observation of the top quark", *Phys. Rev. Lett.* **74** (1995) 2632–2637, doi:10.1103/PhysRevLett.74.2632, arXiv:hep-ex/9503003.

[7] G. Degrassi et al., "Higgs mass and vacuum stability in the standard model at NNLO", *JHEP* **08** (2012) 098, doi:10.1007/JHEP08(2012)098, arXiv:1205.6497.

[8] S. Alekhin, A. Djouadi, and S. Moch, "The top quark and Higgs boson masses and the stability of the electroweak vacuum", *Phys. Lett. B* **716** (2012) 214, doi:10.1016/j.physletb.2012.08.024, arXiv:1207.0980.

[9] Gfitter Group Collaboration, "The global electroweak fit at NNLO and prospects for the LHC and ILC", *Eur. Phys. J. C* **74** (2014) 3046, doi:10.1140/epjc/s10052-014-3046-5, arXiv:1407.3792.

[10] J. de Blas et al., "Electroweak precision observables and Higgs-boson signal strengths in the standard model and beyond: present and future", *JHEP* **12** (2016) 135, doi:10.1007/JHEP12(2016)135, arXiv:1608.01509.

[11] S. Alekhin, J. Blümlein, S. Moch, and R. Placakyte, "Parton distribution functions, α_S, and heavy-quark masses for LHC Run II", *Phys. Rev. D* **96** (2017) 014011, doi:10.1103/PhysRevD.96.014011, arXiv:1701.05838.

[12] Particle Data Group, P. A. Zyla et al., "Review of particle physics", *Prog. Theor. Exp. Phys.* **2020** (2020) 083C01, doi:10.1093/ptep/ptaa104.

[13] M. Beneke and V. M. Braun, "Heavy quark effective theory beyond perturbation theory: renormalons, the pole mass and the residual mass term", *Nucl. Phys. B* **426** (1994) 301–343, doi:10.1016/0550-3213(94)90314-X, arXiv:hep-ph/9402364.

[14] I. I. Y. Bigi, M. A. Shifman, N. G. Uraltsev, and A. I. Vainshtein, "The Pole mass of the heavy quark. Perturbation theory and beyond", *Phys. Rev. D* **50** (1994) 2234–2246, doi:10.1103/PhysRevD.50.2234, arXiv:hep-ph/9402360.

[15] M. C. Smith and S. S. Willenbrock, "Top-quark pole mass", *Phys. Rev. Lett.* **79** (1997) 3825–3828, doi:10.1103/PhysRevLett.79.3825.

[16] ATLAS Collaboration, "Measurement of the top quark mass in the $t\bar{t} \to$ dilepton channel from $\sqrt{s} = 8\,\text{TeV}$ ATLAS data", *Phys. Lett. B* **761** (2016) 350, doi:10.1016/j.physletb.2016.08.042, arXiv:1606.02179.

[17] ATLAS Collaboration, "Top-quark mass measurement in the all-hadronic $t\bar{t}$ decay channel at $\sqrt{s} = 8\,\text{TeV}$ with the ATLAS detector", *JHEP* **09** (2017) 118, doi:10.1007/JHEP09(2017)118, arXiv:1702.07546.

[18] ATLAS Collaboration, "Measurement of the top quark mass in the $t\bar{t} \to$ lepton+jets channel from $\sqrt{s} = 8\,\text{TeV}$ ATLAS data and combination with previous results", *Eur. Phys. J. C* **79** (2019) 290, doi:10.1140/epjc/s10052-019-6757-9, arXiv:1810.01772.

[19] CMS Collaboration, "Measurement of the top quark mass using proton-proton data at $\sqrt{s} = 7$ and $8\,\text{TeV}$", *Phys. Rev. D* **93** (2016) 072004, doi:10.1103/PhysRevD.93.072004, arXiv:1509.04044.

[20] CMS Collaboration, "Measurement of the top quark mass in the dileptonic $t\bar{t}$ decay channel using the mass observables $m_{\text{b}\ell}$, m_{T2}, and $m_{\text{b}\ell\nu}$ in pp collisions at $\sqrt{s} = 8\,\text{TeV}$", *Phys. Rev. D* **96** (2017) 032002, doi:10.1103/PhysRevD.96.032002, arXiv:1704.06142.

[21] CMS Collaboration, "Measurement of the top quark mass with lepton+jets final states using pp collisions at $\sqrt{s} = 13\,\text{TeV}$", *Eur. Phys. J. C* **78** (2018) 891, doi:10.1140/epjc/s10052-018-6332-9, arXiv:1805.01428.

[22] CMS Collaboration, "Measurement of the top quark mass in the all-jets final state at $\sqrt{s} = 13\,\text{TeV}$ and combination with the lepton+jets channel", *Eur. Phys. J. C* **79** (2019) 313, doi:10.1140/epjc/s10052-019-6788-2, arXiv:1812.10534.

[23] CMS Collaboration, "A profile likelihood approach to measure the top quark mass in the lepton+jets channel at $\sqrt{s} = 13\,\text{TeV}$", CMS Physics Analysis Summary CMS-PAS-TOP-20-008, CERN, 2022.

[24] S. Ferrario Ravasio, T. Ježo, P. Nason, and C. Oleari, "A theoretical study of top-mass measurements at the LHC using NLO+PS generators of increasing accuracy", *Eur. Phys. J. C* **78** (2018) 458, doi:10.1140/epjc/s10052-018-5909-7, arXiv:1906.09166. [Addendum: doi:10.1140/epjc/s10052-019-7336-9].

[25] M. Butenschön et al., "Top quark mass calibration for Monte Carlo event generators", *Phys. Rev. Lett.* **117** (2016) 232001, doi:10.1103/PhysRevLett.117.232001, arXiv:1608.01318.

[26] ATLAS, CDF, CMS, and D0 Collaborations, "First combination of Tevatron and LHC measurements of the top-quark mass", (2014). arXiv:1403.4427.

[27] S. Moch et al., "High precision fundamental constants at the TeV scale", (2014). arXiv:1405.4781.

[28] A. Juste et al., "Determination of the top quark mass circa 2013: methods, subtleties, perspectives", *Eur. Phys. J. C* **74** (2014) 3119, doi:10.1140/epjc/s10052-014-3119-5, arXiv:1310.0799.

[29] A. H. Hoang, "What is the top quark mass?", *Ann. Rev. Nucl. Part. Sci.* **70** (2020) 225, doi:10.1146/annurev-nucl-101918-023530, arXiv:2004.12915.

[30] A. H. Hoang, "The top mass: interpretation and theoretical uncertainties", in *Proc. 7th Int. Workshop on Top Quark Physics (TOP2014): Cannes, France, September 28–October 3, 2014*. 2014. arXiv:1412.3649.

[31] CMS Collaboration, "Measurement of the $\mathrm{t\bar{t}}$ production cross section, the top quark mass, and the strong coupling constant using dilepton events in pp collisions at $\sqrt{s} = 13\,\text{TeV}$", *Eur. Phys. J. C* **79** (2019) 368, doi:10.1140/epjc/s10052-019-6863-8, arXiv:1812.10505.

[32] ATLAS Collaboration, "Measurement of the $\mathrm{t\bar{t}}$ production cross-section using $\mathrm{e}\mu$ events with b-tagged jets in pp collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS detector", *Eur. Phys. J. C* **74** (2014) 3109, doi:10.1140/epjc/s10052-014-3109-7, arXiv:1406.5375. [Addendum: doi:10.1140/epjc/s10052-016-4501-2].

[33] CMS Collaboration, "Measurement of the $\mathrm{t\bar{t}}$ production cross section in the $\mathrm{e}\mu$ channel in proton-proton collisions at $\sqrt{s} = 7$ and 8 TeV", *JHEP* **08** (2016) 029, doi:10.1007/JHEP08(2016)029, arXiv:1603.02303.

[34] CMS Collaboration, "Measurement of the $\mathrm{t\bar{t}}$ production cross section using events with one lepton and at least one jet in pp collisions at $\sqrt{s} = 13\,\text{TeV}$", *JHEP* **09** (2017) 051, doi:10.1007/JHEP09(2017)051, arXiv:1701.06228.

[35] ATLAS Collaboration, "Determination of the top-quark pole mass using $\mathrm{t\bar{t}}$ + 1-jet events collected with the ATLAS experiment in 7 TeV pp collisions", *JHEP* **10** (2015) 121, doi:10.1007/JHEP10(2015)121, arXiv:1507.01769.

[36] D0 Collaboration, "Measurement of the inclusive $\mathrm{t\bar{t}}$ production cross section in $\mathrm{p\bar{p}}$ collisions at $\sqrt{s} = 1.96\,\text{TeV}$ and determination of the top quark pole mass", *Phys. Rev. D* **94** (2016) 092004, doi:10.1103/PhysRevD.94.092004, arXiv:1605.06168.

[37] ATLAS Collaboration, "Measurement of lepton differential distributions and the top quark mass in $\mathrm{t\bar{t}}$ production in pp collisions at $\sqrt{s} = 8\,\text{TeV}$ with the ATLAS detector", *Eur. Phys. J. C* **77** (2017) 804, doi:10.1140/epjc/s10052-017-5349-9, arXiv:1709.09407.

[38] CMS Collaboration, "Measurement of $\mathrm{t\bar{t}}$ normalised multi-differential cross sections in pp collisions at $\sqrt{s} = 13\,\text{TeV}$, and simultaneous determination of the strong coupling strength, top quark pole mass, and parton distribution functions", *Eur. Phys. J. C* **80** (2020) 658, doi:10.1140/epjc/s10052-020-7917-7, arXiv:1904.05237.

[39] CMS Collaboration, "Measurement and QCD analysis of double-differential inclusive jet cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP* **02** (2022) 142, doi:10.1007/JHEP02(2022)142, arXiv:2111.10431.

[40] U. Langenfeld, S. Moch, and P. Uwer, "Measuring the running top-quark mass", *Phys. Rev. D* **80** (2009) 054009, doi:10.1103/PhysRevD.80.054009, arXiv:0906.5273.

[41] J. Fuster et al., "Extracting the top-quark running mass using $t\bar{t}$ + 1-jet events produced at the Large Hadron Collider", *Eur. Phys. J. C* **77** (2017) 794, doi:10.1140/epjc/s10052-017-5354-z, arXiv:1704.00540.

[42] CMS Collaboration, "Running of the top quark mass from proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Phys. Lett. B* **803** (2020) 135263, doi:10.1016/j.physletb.2020.135263, arXiv:1909.09193.

[43] M. M. Defranchis, J. Kieseler, K. Lipka, and J. Mazzitelli, "Running of the top quark mass at NNLO in QCD", arXiv:2208.11399.

[44] S. Alioli et al., "A new observable to measure the top-quark mass at hadron colliders", *Eur. Phys. J. C* **73** (2013) 2438, doi:10.1140/epjc/s10052-013-2438-2, arXiv:1303.6415.

[45] G. Bevilacqua et al., "Top quark mass studies with $t\bar{t}j$ at the LHC", *JHEP* **03** (2018) 169, doi:10.1007/JHEP03(2018)169, arXiv:1710.07515.

[46] ATLAS Collaboration, "Measurement of the top-quark mass in $t\bar{t}$ + 1-jet events collected with the ATLAS detector in pp collisions at $\sqrt{s} = 8\,\text{TeV}$", *JHEP* **11** (2019) 150, doi:10.1007/JHEP11(2019)150, arXiv:1905.02302.

[47] CMS Collaboration, "Expected Performance of Run-3 HLT b-quark jet identification", CMS Detector Performance Note CMS-DP-2020-030, CERN, 2022.

[48] CMS Collaboration, "The Phase-2 upgrade of the CMS data acquisition and High Level Trigger", CMS Technical Design Report CERN-LHCC-2021-007, CMS-TDR-022, CERN, 2021.

[49] CMS Collaboration, "Measurement of the top quark pole mass using $t\bar{t}$+jet events in the dilepton final state in proton-proton collisions at $\sqrt{s} = 13$ TeV", arXiv:2207.02270.

[50] M. Thomson, "Modern particle physics". Cambridge University Press, Cambridge, 2013. doi:10.1017/CBO9781139525367, ISBN 9781107034266.

[51] J. Campbell, J. Huston, and F. Krauss, "The black book of quantum chromodynamics: a primer for the LHC era". Oxford University Press, Oxford, 2017. doi:10.1093/oso/9780199652747.001.0001, ISBN 9780199652747.

[52] R. K. Ellis, W. J. Stirling, and B. R. Webber, "QCD and collider physics", volume 8 of *Cambridge monographs on particle physics, nuclear physics and cosmology*. Cambridge University Press, Cambridge, 1996. ISBN 0521581893.

[53] G. Sterman, "Renormalization". Cambridge University Press, 1993. doi:10.1017/CBO9780511622618.

[54] A. Einstein, "Die Grundlage der allgemeinen Relativitätstheorie", *Ann. Phys.* **354** (1916) 769–822, doi:https://doi.org/10.1002/andp.19163540702, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.19163540702.

[55] G. 't Hooft and M. J. G. Veltman, "Regularization and renormalization of gauge fields", *Nucl. Phys. B* **44** (1972) 189–213, doi:10.1016/0550-3213(72)90279-9.

[56] E. Noether, "Invariante Variationsprobleme", *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* **1918** (1918) 235–257.

[57] D. G. Carsten Burgard, "Standard model of physics.". Accessed: 2022-08-01. https://texample.net/tikz/examples/model-physics/.

[58] Particle Data Group Collaboration, "Review of Particle Physics (RPP)", *Phys.Rev.* **D86** (2012) 010001, doi:10.1103/PhysRevD.86.010001.

[59] G. Bachy et al., "The LEP collider: construction, project status and outlook", *Part. Accel.* **26** (1990).

[60] ALEPH, DELPHI, L3, OPAL, SLD, and LEP Electroweak Working Group, SLD Electroweak Group, SLD Heavy Flavour Group Collaboration, "Precision electroweak measurements on the Z resonance", *Phys. Rept.* **427** (2006) 257–454, doi:10.1016/j.physrep.2005.12.006, arXiv:hep-ex/0509008.

[61] S. L. Glashow, "Partial-symmetries of weak interactions", *Nucl. Phys.* **22** (1961) 579–588, doi:https://doi.org/10.1016/0029-5582(61)90469-2.

[62] S. Weinberg, "A model of leptons", *Phys. Rev. Lett.* **19** (1967) 1264–1266, doi:10.1103/PhysRevLett.19.1264.

[63] A. Salam, "Weak and Electromagnetic Interactions", volume 680519, pp. 367–377. 1968. doi:10.1142/9789812795915_0034.

[64] C. S. Wu et al., "Experimental test of parity conservation in beta decay", *Phys. Rev.* **105** (1957) 1413–1415, doi:10.1103/PhysRev.105.1413.

[65] T. Nakano and K. Nishijima, "Charge Independence for V-particles", *Prog. Theor. Phys.* **10** (1953) 581–582, doi:10.1143/PTP.10.581.

[66] M. Gell-Mann, "The interpretation of the new particles as displaced charge multiplets", *Nuovo Cim.* **4** (1956) 848–866, doi:10.1007/BF02748000.

[67] ATLAS Collaboration, "A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery", *Nature* **607** (2022) 52–59, doi:10.1038/s41586-022-04893-w, arXiv:2207.00092.

[68] CMS Collaboration, "A portrait of the Higgs boson by the CMS experiment ten years after the discovery", *Nature* **607** (2022) 60–68, doi:10.1038/s41586-022-04892-x, arXiv:2207.00043.

[69] ATLAS Collaboration, "Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector", *Phys. Lett. B* **784** (2018) 173–191, doi:10.1016/j.physletb.2018.07.035, arXiv:1806.00425.

[70] CMS Collaboration, "Observation of $\mathrm{t\bar{t}H}$ production", *Phys. Rev. Lett.* **120** (2018) 231801, doi:10.1103/PhysRevLett.120.231801, arXiv:1804.02610.

[71] ATLAS Collaboration, "Observation of $\mathrm{H} \to \mathrm{b\bar{b}}$ decays and VH production with the ATLAS detector", *Phys. Lett. B* **786** (2018) 59–86, doi:10.1016/j.physletb.2018.09.013, arXiv:1808.08238.

[72] CMS Collaboration, "Observation of Higgs boson decay to bottom quarks", *Phys. Rev. Lett.* **121** (2018) 121801, doi:10.1103/PhysRevLett.121.121801, arXiv:1808.08242.

[73] ATLAS Collaboration, "Cross-section measurements of the Higgs boson decaying into a pair of τ-leptons in proton-proton collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ with the ATLAS detector", *Phys. Rev. D* **99** (2019) 072001, doi:10.1103/PhysRevD.99.072001, arXiv:1811.08856.

[74] CMS Collaboration, "Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector", *Phys. Lett. B* **779** (2018) 283–316, doi:10.1016/j.physletb.2018.02.004, arXiv:1708.00373.

[75] CMS Collaboration, "Evidence for Higgs boson decay to a pair of muons", *JHEP* **01** (2021) 148, doi:10.1007/JHEP01(2021)148, arXiv:2009.04363.

[76] ATLAS Collaboration, "A search for the dimuon decay of the standard model higgs boson with the ATLAS detector", *Phys. Lett. B* **812** (2021) 135980, doi:https://doi.org/10.1016/j.physletb.2020.135980.

[77] CMS Collaboration, "Search for Higgs boson decay to a charm quark-antiquark pair in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", (2022). arXiv:2205.05550.

[78] ATLAS Collaboration, "Direct constraint on the Higgs-charm coupling from a search for Higgs boson decays into charm quarks with the ATLAS detector", (2022). arXiv:2201.11428.

[79] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, "Evidence for the 2π decay of the K_2^0 meson", *Phys. Rev. Lett.* **13** (1964) 138–140, doi:10.1103/PhysRevLett.13.138.

[80] N. Cabibbo, "Unitary symmetry and leptonic decays", *Phys. Rev. Lett.* **10** (1963) 531–533, doi:10.1103/PhysRevLett.10.531.

[81] M. Kobayashi and T. Maskawa, "CP violation in the renormalizable theory of weak interaction", *Prog. Theor. Phys.* **49** (1973) 652–657, doi:10.1143/PTP.49.652.

[82] C. Alpigiani et al., "Unitarity triangle analysis in the standard model and beyond", in *5th Large Hadron Collider Physics Conference*. 2017. arXiv:1710.09644.

[83] Super-Kamiokande Collaboration, "Evidence for oscillation of atmospheric neutrinos", *Phys. Rev. Lett.* **81** (1998) 1562–1567, doi:10.1103/PhysRevLett.81.1562, arXiv:hep-ex/9807003.

[84] SNO Collaboration, "Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory", *Phys. Rev. Lett.* **89** (2002) 011301, doi:10.1103/PhysRevLett.89.011301, arXiv:nucl-ex/0204008.

[85] OPERA Collaboration, "Final Results of the OPERA Experiment on ν_τ Appearance in the CNGS Neutrino Beam", *Phys. Rev. Lett.* **120** (2018) 211801, doi:10.1103/PhysRevLett.120.211801, arXiv:1804.04912. [Erratum: Phys.Rev.Lett. 121, 139901 (2018)].

[86] Z. Maki, M. Nakagawa, and S. Sakata, "Remarks on the unified model of elementary particles", *Prog. Theor. Phys.* **28** (1962) 870–880, doi:10.1143/PTP.28.870.

[87] B. Pontecorvo, "Inverse beta processes and nonconservation of lepton charge", *Zh. Eksp. Teor. Fiz.* **34** (1957) 247.

[88] S. Bilenky, J. Hošek, and S. Petcov, "On the oscillations of neutrinos with dirac and majorana masses", *Phys. Lett. B* **94** (1980) 495–498, doi:https://doi.org/10.1016/0370-2693(80)90927-2.

[89] KATRIN Collaboration, "Direct neutrino-mass measurement with sub-electronvolt sensitivity", *Nature Phys.* **18** (2022) 160–166, doi:10.1038/s41567-021-01463-1, arXiv:2105.08533.

[90] M. Gell-Mann, "A schematic model of baryons and mesons", *Phys. Lett.* **8** (1964) 214–215, doi:https://doi.org/10.1016/S0031-9163(64)92001-3.

[91] G. Zweig, "An SU(3) model for strong interaction symmetry and its breaking". 1964.

[92] TASSO Collaboration, "Evidence for planar events in e^+e^- annihilation at high energies", *Phys. Lett. B* **86** (1979) 243–249, doi:https://doi.org/10.1016/0370-2693(79)90830-X.

[93] D. J. Gross and F. Wilczek, "Ultraviolet behavior of non-abelian gauge theories", *Phys. Rev. Lett.* **30** (1973) 1343–1346, doi:10.1103/PhysRevLett.30.1343.

[94] D. J. Gross and F. Wilczek, "Asymptotically free gauge theories", *Phys. Rev. D* **8** (1973) 3633–3652, doi:10.1103/PhysRevD.8.3633.

[95] H. D. Politzer, "Reliable perturbative results for strong interactions?", *Phys. Rev. Lett.* **30** (1973) 1346–1349, doi:10.1103/PhysRevLett.30.1346.

[96] D. Jones, "Two-loop diagrams in yang-mills theory", *Nucl. Phys. B* **75** (1974) 531–538, doi:https://doi.org/10.1016/0550-3213(74)90093-5.

[97] W. E. Caswell, "Asymptotic behavior of non-abelian gauge theories to two-loop order", *Phys. Rev. Lett.* **33** (1974) 244–246, doi:10.1103/PhysRevLett.33.244.

[98] E. Egorian and O. V. Tarasov, "Two Loop Renormalization of the QCD in an Arbitrary Gauge", *Teor. Mat. Fiz.* **41** (1979) 26–32.

[99] O. Tarasov, A. Vladimirov, and A. Zharkov, "The Gell–Mann–Low function of QCD in the three-loop approximation", *Phys. Lett. B* **93** (1980) 429–432, doi:https://doi.org/10.1016/0370-2693(80)90358-5.

[100] S. A. Larin and J. A. M. Vermaseren, "The three loop QCD beta function and anomalous dimensions", *Phys. Lett. B* **303** (1993) 334–336, doi:10.1016/0370-2693(93)91441-O, arXiv:hep-ph/9302208.

[101] K. G. Chetyrkin, J. H. Kühn, and M. Steinhauser, "RunDec: a MATHEMATICA package for running and decoupling of the strong coupling and quark masses", *Comput. Phys. Commun.* **133** (2000) 43–65, doi:10.1016/S0010-4655(00)00155-7, arXiv:hep-ph/0004189.

[102] T. van Ritbergen, J. A. M. Vermaseren, and S. A. Larin, "The four loop beta function in quantum chromodynamics", *Phys. Lett. B* **400** (1997) 379–384, doi:10.1016/S0370-2693(97)00370-5, arXiv:hep-ph/9701390.

[103] M. Czakon, "The four-loop QCD beta-function and anomalous dimensions", *Nucl. Phys. B* **710** (2005) 485–498, doi:10.1016/j.nuclphysb.2005.01.012, arXiv:hep-ph/0411261.

[104] P. A. Baikov, K. G. Chetyrkin, and J. H. Kühn, "Five-loop running of the QCD coupling constant", *Phys. Rev. Lett.* **118** (2017) 082002, doi:10.1103/PhysRevLett.118.082002, arXiv:1606.08659.

[105] CMS Collaboration, "Measurement and QCD analysis of double-differential inclusive jet cross sections in pp collisions at $\sqrt{s} = 8\,\text{TeV}$ and cross section ratios to 2.76 and 7 TeV", *JHEP* **03** (2017) 156, doi:10.1007/JHEP03(2017)156, arXiv:1609.05331.

[106] H1 Collaboration, "Evidence for a narrow anti-charmed baryon state", *Phys. Lett. B* **588** (2004) 17, doi:10.1016/j.physletb.2004.03.012, arXiv:hep-ex/0403017.

[107] LHCb Collaboration, "Observation of $J/\psi\phi$ structures consistent with exotic states from amplitude analysis of $B^+ \to J/\psi\phi K^+$ decays", *Phys. Rev. Lett.* **118** (2017) 022003, doi:10.1103/PhysRevLett.118.022003, arXiv:1606.07895.

[108] LHCb Collaboration, "Observation of $J/\psi p$ resonances consistent with pentaquark states in $\Lambda_b^0 \to J/\psi K^- p$ decays", *Phys. Rev. Lett.* **115** (2015) 072001, doi:10.1103/PhysRevLett.115.072001, arXiv:1507.03414.

[109] A. H. Hoang and I. W. Stewart, "Top mass measurements from jets and the Tevatron top-quark mass", *Nucl. Phys. B Proc. Suppl.* **185** (2008) 220–226, doi:10.1016/j.nuclphysbps.2008.10.028, arXiv:0808.0222.

[110] A. H. Hoang et al., "The MSR mass and the $\mathcal{O}(\Lambda_{\rm QCD})$ renormalon sum rule", *JHEP* **04** (2018) 003, doi:10.1007/JHEP04(2018)003, arXiv:1704.01580.

[111] A. H. Hoang, A. Jain, I. Scimemi, and I. W. Stewart, "Infrared renormalization-group flow for heavy-quark masses", *Phys. Rev. Lett.* **101** (2008) 151602, doi:10.1103/PhysRevLett.101.151602.

[112] J. C. Collins, D. E. Soper, and G. F. Sterman, "Factorization of hard processes in QCD", *Adv. Ser. Direct. High Energy Phys.* **5** (1989) 1–91, doi:10.1142/9789814503266_0001, arXiv:hep-ph/0409313.

[113] Y. L. Dokshitzer, "Calculation of the structure functions for deep inelastic scattering and e^+e^- annihilation by perturbation theory in quantum chromodynamics", *Sov. Phys. JETP* **46** (1977) 641–653.

[114] V. N. Gribov and L. N. Lipatov, "Deep inelastic electron scattering in perturbation theory", *Phys. Lett. B* **37** (1971) 78–80, doi:10.1016/0370-2693(71)90576-4.

[115] G. Altarelli and G. Parisi, "Asymptotic freedom in parton language", *Nucl. Phys. B* **126** (1977) 298–318, doi:10.1016/0550-3213(77)90384-4.

[116] S. I. Alekhin et al., "DGLAP evolution and parton fits", doi:10.5170/CERN-2005-014.119.

[117] A. Vogt, S. Moch, and J. A. M. Vermaseren, "The three-loop splitting functions in QCD: the singlet case", *Nucl. Phys. B* **691** (2004) 129–181, doi:10.1016/j.nuclphysb.2004.04.024, arXiv:hep-ph/0404111.

[118] S. Moch, J. A. M. Vermaseren, and A. Vogt, "The three loop splitting functions in QCD: the nonsinglet case", *Nucl. Phys. B* **688** (2004) 101–134, doi:10.1016/j.nuclphysb.2004.03.030, arXiv:hep-ph/0403192.

[119] M. Vakili et al., "Nuclear structure functions in the large-x large-Q^2 kinematic region in neutrino deep inelastic scattering", *Phys. Rev. D* **61** (2000) 052003, doi:10.1103/PhysRevD.61.052003.

[120] NuTeV Collaboration, "Precise measurement of neutrino and anti-neutrino differential cross sections", *Phys. Rev. D* **74** (2006) 012008, doi:10.1103/PhysRevD.74.012008, arXiv:hep-ex/0509010.

[121] ZEUS Collaboration, "Measurement of the neutral current cross-section and F(2) structure function for deep inelastic e^+p scattering at HERA", *Eur. Phys. J. C* **21** (2001) 443–471, doi:10.1007/s100520100749, arXiv:hep-ex/0105090.

[122] H1 and ZEUS Collaborations, "Combination of measurements of inclusive deep inelastic $e^{\pm}p$ scattering cross sections and QCD analysis of HERA data", *Eur. Phys. J. C* **75** (2015) 580, doi:10.1140/epjc/s10052-015-3710-4, arXiv:1506.06042.

[123] F. D. Aaron et al., "Combined measurement and QCD analysis of the inclusive $e^{\pm}p$ scattering cross sections at HERA", *JHEP* **01** (2010) 109, doi:10.1007/JHEP01(2010)109, arXiv:0911.0884.

[124] CMS Collaboration, "Measurement of the triple-differential dijet cross section in proton-proton collisions at $\sqrt{s} = 8\,\text{TeV}$ and constraints on parton distribution functions", *Eur. Phys. J. C* **77** (2017) 746, doi:10.1140/epjc/s10052-017-5286-7, arXiv:1705.02628.

[125] S. Höche, "Introduction to parton-shower event generators", in *Theoretical advanced study institute in elementary particle physics: journeys through the precision frontier: amplitudes for colliders*, pp. 235–295. 2015. arXiv:1411.4085. doi:10.1142/9789814678766_0005.

[126] A. Buckley et al., "General-purpose event generators for LHC physics", *Phys. Rept.* **504** (2011) 145–233, doi:10.1016/j.physrep.2011.03.005, arXiv:1101.2599.

[127] J. C. Collins, "Sudakov form-factors", *Adv. Ser. Direct. High Energy Phys.* **5** (1989) 573–614, doi:10.1142/9789814503266_0006, arXiv:hep-ph/0312336.

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

[129] P. Nason, "A New method for combining NLO QCD with shower Monte Carlo algorithms", *JHEP* **11** (2004) 040, doi:10.1088/1126-6708/2004/11/040, arXiv:hep-ph/0409146.

[130] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with parton shower simulations: the POWHEG method", *JHEP* **11** (2007) 070, doi:10.1088/1126-6708/2007/11/070, arXiv:0709.2092.

[131] T. Ježo, J. M. Lindert, N. Moretti, and S. Pozzorini, "New NLOPS predictions for $\mathrm{t\bar{t}}$ + b-jet production at the LHC", *Eur. Phys. J. C* **78** (2018) 502, doi:10.1140/epjc/s10052-018-5956-0, arXiv:1802.00426.

[132] CMS Collaboration, "Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements", *Eur. Phys. J. C* **80** (2020) 4, doi:10.1140/epjc/s10052-019-7499-4, arXiv:1903.12179.

[133] G. Gustafson, U. Pettersson, and P. Zerwas, "Jet final states in WW pair production and colour screening in the QCD vacuum", *Phys. Lett. B* **209** (1988) 90–94, doi:https://doi.org/10.1016/0370-2693(88)91836-9.

[134] T. Sjöstrand and M. van Zijl, "A multiple-interaction model for the event structure in hadron collisions", *Phys. Rev. D* **36** (1987) 2019–2041, doi:10.1103/PhysRevD.36.2019.

[135] S. Argyropoulos and T. Sjöstrand, "Effects of color reconnection on $\mathrm{t\bar{t}}$ final states at the LHC", *JHEP* **11** (2014) 043, doi:10.1007/JHEP11(2014)043, arXiv:1407.6653.

[136] J. R. Christiansen and P. Z. Skands, "String formation beyond leading colour", *JHEP* **08** (2015) 003, doi:10.1007/JHEP08(2015)003, arXiv:1505.01681.

[137] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand, "Parton fragmentation and string dynamics", *Phys. Rep.* **97** (1983) 31–145, doi:https://doi.org/10.1016/0370-1573(83)90080-7.

[138] G. Marchesini et al., "HERWIG 5.1 - a Monte Carlo event generator for simulating hadron emission reactions with interfering gluons", *Comput. Phys. Commun.* **67** (1992) 465–508, doi:https://doi.org/10.1016/0010-4655(92)90055-4.

[139] T. Sjöstrand et al., "An introduction to PYTHIA8.2", *Comput. Phys. Commun.* **191** (2015) 159, doi:10.1016/j.cpc.2015.01.024, arXiv:1410.3012.

[140] C. Peterson, D. Schlatter, I. Schmitt, and P. M. Zerwas, "Scaling violations in inclusive e^+e^- annihilation spectra", *Phys. Rev. D* **27** (1983) 105, doi:10.1103/PhysRevD.27.105.

[141] V. G. Kartvelishvili, A. K. Likhoded, and V. A. Petrov, "On the fragmentation functions of heavy quarks into hadrons", *Phys. Lett. B* **78** (1978) 615–617, doi:10.1016/0370-2693(78)90653-6.

[142] M. G. Bowler, "e^+e^- production of heavy quarks in the string model", *Z. Phys. C* **11** (1981) 169, doi:10.1007/BF01574001.

[143] S. W. Herb et al., "Observation of a dimuon resonance at 9.5 GeV in 400 GeV proton-nucleus collisions", *Phys. Rev. Lett.* **39** (1977) 252–255, doi:10.1103/PhysRevLett.39.252.

[144] A. de Gouvea, D. Hernandez, and T. M. P. Tait, "Criteria for natural hierarchies", *Phys. Rev. D* **89** (2014) 115005, doi:10.1103/PhysRevD.89.115005, arXiv:1402.2658.

[145] A. V. Bednyakov, B. A. Kniehl, A. F. Pikelner, and O. L. Veretin, "Stability of the electroweak vacuum: gauge independence and advanced precision", *Phys. Rev. Lett.* **115** (2015) 201802, doi:10.1103/PhysRevLett.115.201802, arXiv:1507.08833.

[146] J. R. Espinosa, "Implications of the top (and Higgs) mass for vacuum stability", *PoS* **TOP2015** (2016) 043, doi:10.22323/1.257.0043, arXiv:1512.01222.

[147] J. Erler and M. Schott, "Electroweak precision tests of the standard model after the discovery of the Higgs boson", *Prog. Part. Nucl. Phys.* **106** (2019) 68–119, doi:10.1016/j.ppnp.2019.02.007, arXiv:1902.05142.

[148] J. Haller et al., "Update of the global electroweak fit and constraints on two-Higgs-doublet models", *Eur. Phys. J. C* **78** (2018) 675, doi:10.1140/epjc/s10052-018-6131-3, arXiv:1803.01853.

[149] J. de Blas et al., "The global electroweak and Higgs fits in the LHC era", *PoS* **EPS-HEP2017** (2017) 467, doi:10.22323/1.314.0467, arXiv:1710.05402.

[150] M. Czakon, P. Fiedler, and A. Mitov, "Total top-quark pair-production cross section at hadron colliders through $\mathcal{O}(\alpha_S{}^4)$", *Phys. Rev. Lett.* **110** (2013) 252004, doi:10.1103/PhysRevLett.110.252004, arXiv:1303.6254.

[151] NNPDF Collaboration, "Parton distributions for the LHC Run II", *JHEP* **04** (2015) 040, doi:10.1007/JHEP04(2015)040, arXiv:1410.8849.

[152] LHCTopWG, "LHCTopWG Summary Plots". Accessed: 2022-08-18. https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCTopWGSummaryPlots.

[153] M. Czakon and A. Mitov, "Top++: a program for the calculation of the top-pair cross-section at hadron colliders", *Comput. Phys. Commun.* **185** (2014) 2930, doi:10.1016/j.cpc.2014.06.021, arXiv:1112.5675.

[154] CMS Collaboration, "Measurement of CKM matrix elements in single top quark t-channel production in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Phys. Lett. B* **808** (2020) 135609, doi:10.1016/j.physletb.2020.135609, arXiv:2004.12181.

[155] D0 Collaboration, "Useful diagrams of top signals and backgrounds.". Accessed: April, 2022. https://www-d0.fnal.gov/Run2Physics/top/top_public_web_pages/top_feynman_diagrams.html Modified.

[156] A. H. Hoang, S. Plätzer, and D. Samitz, "On the cutoff dependence of the quark mass parameter in angular ordered parton showers", *JHEP* **10** (2018) 200, doi:10.1007/JHEP10(2018)200, arXiv:1807.06617.

[157] ATLAS Collaboration, "A precise interpretation for the top quark mass parameter in ATLAS Monte Carlo simulation", ATLAS Physics Public Note ATL-PHYS-PUB-2021-034, CERN, 2021.

[158] ATLAS Collaboration, "Measurement of the $\text{t}\bar{\text{t}}$ production cross-section and lepton differential distributions in $\text{e}^{\pm}\mu^{\mp}$ dilepton events from pp collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector", *Eur. Phys. J. C* **80** (2020), no. 6, 528, doi:10.1140/epjc/s10052-020-7907-9, arXiv:1910.08819.

[159] M. M. Defranchis, J. Kieseler, K. Lipka, and J. Mazzitelli, "Running of the top quark mass at NNLO in QCD", arXiv:2208.11399.

[160] S. Alioli et al., "Phenomenology of $\text{t}\bar{\text{t}}\text{j} + \text{X}$ production at the LHC", *JHEP* **05** (2022) 146, doi:10.1007/JHEP05(2022)146, arXiv:2202.07975.

[161] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_{T} jet clustering algorithm", *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.

[162] S. Dittmaier, P. Uwer, and S. Weinzierl, "Hadronic top-quark pair production in association with a hard jet at next-to-leading order QCD: phenomenological studies for the Tevatron and the LHC", *Eur. Phys. J. C* **59** (2009) 625–646, doi:10.1140/epjc/s10052-008-0816-y, arXiv:0810.0452.

[163] S. Dittmaier, P. Uwer, and S. Weinzierl, "NLO QCD corrections to $\text{t}\bar{\text{t}}$+jet production at hadron colliders", *Phys. Rev. Lett.* **98** (2007) 262002, doi:10.1103/PhysRevLett.98.262002, arXiv:hep-ph/0703120.

[164] S. Alioli, S. Moch, and P. Uwer, "Hadronic top-quark pair-production with one jet and parton showering", *JHEP* **01** (2012) 137, doi:10.1007/JHEP01(2012)137, arXiv:1110.5251.

[165] G. Bevilacqua, H. B. Hartanto, M. Kraus, and M. Worek, "Off-shell top quarks with one jet at the LHC: a comprehensive analysis at NLO QCD", *JHEP* **11** (2016) 098, doi:10.1007/JHEP11(2016)098, arXiv:1609.01659.

[166] T.-J. Hou et al., "Progress in the CTEQ-TEA NNLO global QCD analysis", (2019). arXiv:1908.11394.

[167] S. Alekhin, J. Blümlein, and S. Moch, "NLO PDFs from the ABMP16 fit", *Eur. Phys. J. C* **78** (2018) 477, doi:10.1140/epjc/s10052-018-5947-1, arXiv:1803.07537.

[168] S. Bailey et al., "Parton distributions from LHC, HERA, Tevatron and fixed target data: MSHT20 PDFs", *Eur. Phys. J. C* **81** (2021) 341, doi:10.1140/epjc/s10052-021-09057-0, arXiv:2012.04684.

[169] NNPDF Collaboration, "Parton distributions from high-precision collider data", *Eur. Phys. J. C* **77** (2017) 663, doi:10.1140/epjc/s10052-017-5199-5, arXiv:1706.00428.

[170] M. Dowling and S.-O. Moch, "Differential distributions for top-quark hadro-production with a running mass", *Eur. Phys. J. C* **74** (2014) 3167, doi:10.1140/epjc/s10052-014-3167-x, arXiv:1305.6422.

[171] O. S. Brüning et al., "LHC Design Report". CERN yellow reports: monographs. CERN, 2004. doi:10.5170/CERN-2004-003-V-1.

[172] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *JINST* **3** (2008) doi:10.1088/1748-0221/3/08/S08002.

[173] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) doi:10.1088/1748-0221/3/08/S08003.

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

[175] LHCb Collaboration, "The LHCb Detector at the LHC", *JINST* **3** (2008) doi:10.1088/1748-0221/3/08/S08005.

[176] E. Mobs, "The CERN accelerator complex - August 2018", 2018. General Photo, Modified. http://cds.cern.ch/record/2636343.

[177] CERN, "Schedule and luminosity forecasts.". Accessed: 2022-07-28. https://lhc-commissioning.web.cern.ch/schedule/HL-LHC-plots.htm.

[178] W. Herr and B. Muratori, "Concept of luminosity", in *Intermediate accelerator physics. Proceedings, CERN Accelerator School, Zeuthen, Germany, September 15-26, 2003.* 2003. doi:10.5170/CERN-2006-002.361.

[179] E. Metral, "LHC: status, prospects and future challenges", *PoS* **LHCP2016** (2016) 002. 20 p, doi:10.22323/1.276.0002.

[180] CMS Collaboration, "CMS physics: Technical Design Report volume 1: Detector performance and software", Technical Design Report CMS-TDR-8-1, CERN, 2006.

[181] T. Sakuma and T. McCauley, "Detector and event visualization with SketchUp at the CMS experiment", *Journal of Physics: conference series* **513** (2014) 022032, doi:10.1088/1742-6596/513/2/022032.

[182] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector", *JINST* **12** (2017) P10003, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.

[183] I. Neutelings, "CMS coordinate system.". Accessed: 2022-08-01. https://tikz.net/axis3d_cms/.

[184] A. Dominguez et al., "CMS Technical Design Report for the pixel detector upgrade", CMS Technical Design Report CERN-LHCC-2012-016, CMS-TDR-11, CERN, 2012.

[185] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST* **9** (2014) P10009, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.

[186] CMS Collaboration, "CMS Tracking POG Performance Plots for year 2016". Accessed: April, 2022. https://twiki.cern.ch/twiki/bin/view/CMSPublic/TrackingPOGPlots2016.

[187] CMS Collaboration, "Primary vertex resolution in 2016", CMS Detector Performance Note CMS-DP-2016-041, CERN, 2016.

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

[189] CMS Collaboration, "Performance and operation of the CMS electromagnetic calorimeter", *JINST* **5** (2010).

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

[191] CMS Collaboration, "The CMS ECAL performance with examples", CMS Conference Report CMS-CR-2013-430, CERN, 2013.

[192] CMS Collaboration, "Performance of the CMS hadron calorimeter with cosmic ray muons and LHC beam data", *JINST* **5** (2010) T03012, doi:10.1088/1748-0221/5/03/T03012, arXiv:0911.4991.

[193] CMS Collaboration, "HCAL calibration in 2016", CMS Detector Performance Note CMS-DP-2017-017, CERN, 2017.

[194] CMS Collaboration, "HCAL energy reconstruction performance", CMS Detector Performance Note CMS-DP-2016-071, CERN, 2016.

[195] CMS Collaboration, "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7\,\text{TeV}$", *JINST* **7** (2012) P10002, doi:10.1088/1748-0221/7/10/P10002, arXiv:1206.4071.

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

[197] CMS Collaboration, "CMS computing: Technical Design Report", CMS Technical Design Report CMS-TDR-7, CERN, 2005.

[198] CMS Collaboration, "The CMS trigger system", *JINST* **12** (2017) P01020, doi:10.1088/1748-0221/12/01/P01020, arXiv:1609.02366.

[199] CMS Collaboration, "The CMS trigger in Run 2", Technical Report CMS-CR-2017-340, CERN, 2017.

[200] CMS Collaboration, "Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ in 2015 and 2016 at CMS", arXiv:2104.01927.

[201] P. Grafström and W. Kozanecki, "Luminosity determination at proton colliders", *Prog. Part. Nucl. Phys.* **81** (2015) 97–148. 52 p, doi:10.1016/j.ppnp.2014.11.002.

[202] S. van der Meer, "Calibration of the effective beam height in the ISR", Technical Report CERN-ISR-PO-68-31, ISR-PO-68-31, CERN, 1968.

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

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

[205] CMS Collaboration, "Public CMS luminosity information.". Accessed: April, 2022. https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults.

[206] O. Aberle et al., "High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report". CERN Yellow Reports: Monographs. 2020. doi:10.23731/CYRM-2020-0010.

[207] CMS Collaboration, "The Phase-2 upgrade of the CMS tracker", CMS Technical Design Report CERN-LHCC-2017-009, CMS-TDR-014, CERN, 2017. doi:10.17181/CERN.QZ28.FLHW.

[208] CMS Collaboration, "The Phase-2 upgrade of the CMS endcap calorimeter", CMS Technical Design Report CERN-LHCC-2017-023, CMS-TDR-019, CERN, 2017. doi:10.17181/CERN.IV8M.1JY2.

[209] CMS Collaboration, "The Phase-2 upgrade of the CMS muon detectors", CMS Technical Design Report CERN-LHCC-2017-012, CMS-TDR-016, CERN, 2017.

[210] CMS Collaboration, "A MIP timing detector for the CMS Phase-2 upgrade", CMS Technical Design Report CERN-LHCC-2019-003, CMS-TDR-020, CERN, 2019.

[211] N. Bartosik, "PF concept in CMS.". Accessed: 2022-01-10. http://bartosik.pp.ua/.

[212] CMS Collaboration, "Alignment of the CMS tracker with LHC and cosmic ray data", *JINST* **9** (2014) P06009, doi:10.1088/1748-0221/9/06/P06009, arXiv:1403.2286.

[213] R. Mankel, "A concurrent track evolution algorithm for pattern recognition in the hera-b main tracking system", *Nucl. Instrum. Methods Phys. Res. A* **395** (1997) 169–184, doi:https://doi.org/10.1016/S0168-9002(97)00705-5.

[214] P. Billoir and S. Qian, "Simultaneous pattern recognition and track fitting by the kalman filtering method", *Nucl. Instrum. Methods Phys. Res. A* **294** (1990) 219–228, doi:https://doi.org/10.1016/0168-9002(90)91835-Y.

[215] P. Billoir, "Progressive track recognition with a kalman-like fitting procedure", *Comput. Phys. Commun.* **57** (1989) 390–394, doi:https://doi.org/10.1016/0010-4655(89)90249-X.

[216] R. Frühwirth, "Application of kalman filtering to track and vertex fitting", *Nucl. Instrum. Methods Phys. Res. A* **262** (1987) 444–450, doi:https://doi.org/10.1016/0168-9002(87)90887-4.

[217] T. Speer et al., "Vertex fitting in the CMS tracker", CMS Public Note CMS-NOTE-2006-032, CERN, 2006.

[218] K. Rose, "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems", *Proceedings of the IEEE* **86** (1998) 2210–2239, doi:10.1109/5.726788.

[219] R. Frühwirth, W. Waltenberger, and P. Vanlaer, "Adaptive vertex fitting", CMS Public Note CMS-NOTE-2007-008, CERN, 2007.

[220] CMS Collaboration, "Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8\,\text{TeV}$", *JINST* **10** (2015) P06005, doi:10.1088/1748-0221/10/06/P06005, arXiv:1502.02701.

[221] W. Adam, R. Frühwirth, A. Strandlie, and T. Todorov, "Reconstruction of electrons with the gaussian-sum filter in the CMS tracker at the LHC", *J. Phys. G* **31** (2005) N9–N20, doi:10.1088/0954-3899/31/9/n01.

[222] CMS Collaboration, "Electron and photon performance in CMS with the full 2016 data sample.", CMS Detector Performance Note CMS-DP-2017-004, CERN, 2017.

[223] CMS Collaboration, "The Performance of the CMS Muon Detector in Proton-Proton Collisions at $\sqrt{s} = 7\,\text{TeV}$ at the LHC", *JINST* **8** (2013) P11002, doi:10.1088/1748-0221/8/11/P11002, arXiv:1306.6905.

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

[225] CMS Collaboration, "Search for neutral higgs bosons decaying to tau pairs in pp collisions at $\sqrt{s} = 7\,\text{TeV}$", *Phys. Lett. B* **713** (2012) 68–90, doi:https://doi.org/10.1016/j.physletb.2012.05.028.

[226] A. Bodek et al., "Extracting Muon Momentum Scale Corrections for Hadron Collider Experiments", *Eur. Phys. J. C* **72** (2012) 2194, doi:10.1140/epjc/s10052-012-2194-8, arXiv:1208.3710.

[227] CMS Collaboration Collaboration, "Muon Reconstruction and Identification Performance with Run-2 data", CMS Detector Performance Note CMS-DP-2020-040, CERN, 2020.

[228] CMS Collaboration Collaboration, "Muon reconstruction performance during Run II", CMS Detector Performance Note CMS-DP-2019-022, CERN, 2019.

[229] CMS Collaboration, "Performance of electron and photon reconstruction in Run 2 with the CMS experiment", CMS Detector Performance Note CMS-DP-2020-037, CERN, 2020.

[230] S. Catani, Y. Dokshitzer, M. Seymour, and B. Webber, "Longitudinally-invariant k_{T}-clustering algorithms for hadron-hadron collisions", *Nucl. Phys. B* **406** (1993) 187–224, doi:https://doi.org/10.1016/0550-3213(93)90166-M.

[231] CMS Collaboration, "A Cambridge-Aachen (C-A) based Jet Algorithm for boosted top-jet tagging", CMS Physics Analysis Summary CMS-PAS-JME-09-001, CERN, 2009.

[232] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet user manual", *Eur. Phys. J. C* **72** (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.

[233] CMS Collaboration, "Pileup Removal Algorithms", CMS Physics Analysis Summary CMS-PAS-JME-14-001, CERN, 2014.

[234] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", *JINST* **12** (2017) P02014, doi:10.1088/1748-0221/12/02/P02014, arXiv:1607.03663.

[235] CMS Collaboration, "Determination of jet energy calibration and transverse momentum resolution in CMS", *JINST* **6** (2011) P11002, doi:10.1088/1748-0221/6/11/P11002, arXiv:1107.4277.

[236] M. Cacciari, G. P. Salam, and G. Soyez, "The catchment area of jets", *JHEP* **2008** (2008) 005–005, doi:10.1088/1126-6708/2008/04/005.

[237] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas", *Phys. Lett. B* **659** (2008) 119–126, doi:10.1016/j.physletb.2007.09.077, arXiv:0707.1378.

[238] UA2 Collaboration, "Measurement of production and properties of jets at the CERN $\bar{\text{p}}$p collider", *Z. Phys. C* **20** (1983) 117–134, doi:10.1007/BF01573214.

[239] CMS Collaboration, "Jet algorithms performance in 13 TeV data", CMS Physics Analysis Summary CMS-PAS-JME-16-003, CERN, 2017.

[240] CMS Collaboration, "Pileup Jet Identification", CMS Physics Analysis Summary CMS-PAS-JME-13-005, CERN, 2013.

[241] CMS Collaboration, "Performance of the pile up jet identification in CMS for Run 2", CMS Detector Performance Note CMS-DP-2020-020, CERN, 2020.

[242] CMS Collaboration, "A deep neural network for simultaneous estimation of b jet energy and resolution", *Comput. Softw. Big Sci.* **4** (2020) 10, doi:10.1007/s41781-020-00041-z, arXiv:1912.06046.

[243] CMS Collaboration, "CMS MET performance in events containing electroweak bosons from pp collisions at $\sqrt{s} = 7$ TeV", CMS Physics Analysis Summary CMS-PAS-PFT-10-002, CERN, 2010.

[244] D. Bertolini, P. Harris, M. Low, and N. Tran, "Pileup per particle identification", *JHEP* **10** (2014) 059, doi:10.1007/JHEP10(2014)059, arXiv:1407.6013.

[245] CMS Collaboration, "How CMS weeds out particles that pile up.". Accessed: April, 2022. https://cms.cern/news/how-cms-weeds-out-particles-pile.

[246] CMS Collaboration, "Pileup-per-particle identification: optimisation for Run 2 Legacy and beyond", CMS Detector Performance Note CMS-DP-2021-001, CERN, 2021.

[247] CMS Collaboration, "Pileup mitigation at CMS in 13 TeV data", *JINST* **15** (2020) P09018, doi:10.1088/1748-0221/15/09/P09018, arXiv:2003.00503.

[248] Particle Data Group Collaboration, "Review of Particle Physics", *Prog. Theor. Exp. Phys.* **2020** (2020) 083C01, doi:10.1093/ptep/ptaa104.

[249] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV", *JINST* **13** (2018) P05011, doi:10.1088/1748-0221/13/05/P05011, arXiv:1712.07158.

[250] T. Müller, C. Piasecki, G. Quast, and C. Weiser, "Inclusive secondary vertex reconstruction in jets", CMS Public Note CMS-NOTE-2006-027, CERN, 2006.

[251] W. Waltenberger, R. Frühwirth, and P. Vanlaer, "Adaptive vertex fitting", *J. Phys. G* **34** (2007) N343–N356, doi:10.1088/0954-3899/34/12/n01.

[252] CMS Collaboration, "Identification of b-Quark Jets with the CMS Experiment", *JINST* **8** (2013) P04013, doi:10.1088/1748-0221/8/04/P04013, arXiv:1211.4462.

[253] D. Guest et al., "Jet flavor classification in high-energy physics with deep neural networks", *Phys. Rev. D* **94** (2016) 112002, doi:10.1103/PhysRevD.94.112002, arXiv:1607.08633.

[254] E. Bols et al., "Jet flavour classification using DeepJet", *JINST* **15** (2020) P12012, doi:10.1088/1748-0221/15/12/P12012, arXiv:2008.10519.

[255] S. Hochreiter and J. Schmidhuber, "Long short-term memory", *Neural Comp.* **9** (1997) 1735–1780, doi:10.1162/neco.1997.9.8.1735, arXiv:https://direct.mit.edu/neco/article-pdf/9/8/1735/813796/neco.1997.9.8.1735.pdf

[256] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition", *Proc. IEEE* **86** (1998) doi:10.1109/5.726791.

[257] CMS Collaboration, "Performance of the DeepJet b tagging algorithm using 41.9 fb^{-1} of data from proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ with Phase 1 CMS detector", CMS Detector Performance Note CMS-DP-2018-058, CERN, 2018.

[258] CMS Collaboration, "b-jet trigger performance in Run 2", CMS Detector Performance Note CMS-DP-2019-042, CERN, 2019.

[259] M. Pieri and A. Rizzi, "CMS experiment prepares for Run 3". Accessed: 2022-05-17. https://ep-news.web.cern.ch/content/cms-experiment-prepares-run-3.

[260] A. Bocci et al., "Heterogeneous reconstruction of tracks and primary vertices with the CMS pixel tracker", *Frontiers in Big Data* **3** (2020) doi:10.3389/fdata.2020.601728.

[261] CMS Collaboration, "Patatrack heterogeneous computing 2018 demonstrator: Pixel tracks", CMS Detector Performance Note CMS-DP-2018-059, CERN, 2018.

[262] CMS Collaboration, "PF Jet Performances at High Level Trigger using Patatrack pixel tracks", CMS Detector Performance Note CMS-DP-2021-005, CERN, 2021.

[263] J. Duarte, "Fast reconstruction and data scouting", in *4th International Workshop Connecting The Dots 2018*. 2018. arXiv:1808.00902.

[264] CMS Collaboration, "Data scouting and data parking with the CMS High Level Trigger", *PoS* **EPS-HEP2019** (2020) 139. 6 p, doi:10.22323/1.364.0139.

[265] F. James, "Statistical Methods in Experimental Physics". World Scientific, 2nd edition, 2006. doi:10.1142/6096.

[266] K. Pearson, "X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling", *Lond. Edinb. Dublin philos. mag.* **50** (1900) 157–175, doi:10.1080/14786440009463897, arXiv:https://doi.org/10.1080/14786440009463897.

[267] W. G. Cochran, "The χ^2 test of goodness of fit", *Ann. Math. Stat.* **23** (1952) 315 – 345, doi:10.1214/aoms/1177729380.

[268] R. B. D'Agostino and M. A. Stephens, "Goodness-of-Fit Techniques". USA, 1986. ISBN 0824774876.

[269] Y. Avni, "Energy spectra of X-ray clusters of galaxies", *ApJ* **210** (1976) 642–646, doi:10.1086/154870.

[270] S. Baker and R. D. Cousins, "Clarification of the use of chi-square and likelihood functions in fits to histograms", *Nucl. Instrum. Methods Phys. Res.* **221** (1984) 437–442, doi:https://doi.org/10.1016/0167-5087(84)90016-4.

[271] M. Baak, S. Gadatsch, R. Harrington, and W. Verkerke, "Interpolation between multi-dimensional histograms using a new non-linear moment morphing method", *Nucl. Instrum. Meth. A* **771** (2015) 39–48, doi:10.1016/j.nima.2014.10.033, arXiv:1410.7388.

[272] R. Barlow and C. Beeston, "Fitting using finite Monte Carlo samples", *Comput. Phys. Commun.* **77** (1993) 219, doi:10.1016/0010-4655(93)90005-w.

[273] J. S. Conway, "Incorporating nuisance parameters in likelihoods for multisource spectra", in *Proc. 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding (PHYSTAT 2011): Geneva, Switzerland, January 17–20, 2011*. 2011. arXiv:1103.0354. doi:10.5170/CERN-2011-006.115.

[274] M. M. Defranchis, "First Measurement of the Running of the Top Quark Mass". PhD thesis, Hamburg University, 2022. doi:10.1007/978-3-030-90376-3.

[275] J. Neyman and E. S. Pearson, "On the Problem of the Most Efficient Tests of Statistical Hypotheses", *Philosophical Transactions of the Royal Society of London Series A* **231** (1933) 289–337, doi:10.1098/rsta.1933.0009.

[276] R. D. Cousins, "Generalization of chisquare goodness-of-fit test for binned data using saturated models, with application to histograms", 2013. http://www.physics.ucla.edu/~cousins/stats/cousins_saturated.pdf.

[277] J. K. Lindsey, "Parametric statistical inference", 1996.

[278] A. Kolmogorov, "Sulla determinazione empirica di una legge di distribuzione", *Inst. Ital. Attuari, Giorn.* **4** (1933) 83–91.

[279] N. V. Smirnov, "On the estimation of the discrepancy between empirical curves of distribution for two independent samples", *Bull. Math. Univ. Moscou* **2** (1939) 3–14.

[280] F. Rademakers et al., "root-project/root: Release v6-11/02", 2017. doi:10.5281/zenodo.1003159, https://doi.org/10.5281/zenodo.1003159.

[281] I. Goodfellow, Y. Bengio, and A. Courville, "Deep Learning". MIT Press, 2016. http://www.deeplearningbook.org.

[282] S. J. Russell and P. Norvig, "Artificial intelligence: a modern approach (4th edition)". Pearson, 2020. ISBN 9780134610993.

[283] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, "Self-normalizing neural networks", doi:10.48550/ARXIV.1706.02515.

[284] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning Representations by Back-propagating Errors", *Nature* **323** (1986) 533–536, doi:10.1038/323533a0.

[285] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization", 2014. doi:10.48550/ARXIV.1412.6980, https://arxiv.org/abs/1412.6980.

[286] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms", (2012). arXiv:1206.2944.

[287] E. Brochu, V. M. Cora, and N. de Freitas, "A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning", (2010). arXiv:1012.2599.

[288] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift", 2015. doi:10.48550/ARXIV.1502.03167, https://arxiv.org/abs/1502.03167.

[289] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks", in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, Y. W. Teh and M. Titterington, eds., volume 9 of *Proceedings of Machine Learning Research*, pp. 249–256. PMLR, Chia Laguna Resort, Sardinia, Italy, 2010.

[290] N. Srivastava et al., "Dropout: a simple way to prevent neural networks from overfitting", *J. Mach. Learn. Res.* **15** (2014).

[291] A. Farahani, S. Voghoei, K. Rasheed, and H. R. Arabnia, "A brief review of domain adaptation", 2020. doi:10.48550/ARXIV.2010.03978, https://arxiv.org/abs/2010.03978.

[292] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation", (2015). arXiv:1409.7495.

[293] CMS Collaboration, "Measurement of the top quark pole mass using $\mathrm{t\bar{t}}$+jet events in the dilepton final state at $\sqrt{s} = 13\,\mathrm{TeV}$", CMS Physics Analysis Summary CMS-PAS-TOP-21-008, CERN, 2022.

[294] S. Frixione, G. Ridolfi, and P. Nason, "A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction", *JHEP* **09** (2007) 126, doi:10.1088/1126-6708/2007/09/126, arXiv:0707.3088.

[295] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP* **06** (2010) 043, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.

[296] NNPDF Collaboration, "Unbiased global determination of parton distributions and their uncertainties at NNLO and at LO", *Nucl. Phys. B* **855** (2012) 153, doi:10.1016/j.nuclphysb.2011.09.024, arXiv:1107.2652.

[297] 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", *JHEP* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.

[298] R. Frederix and S. Frixione, "Merging meets matching in MC@NLO", *JHEP* **12** (2012) 061, doi:10.1007/JHEP12(2012)061, arXiv:1209.6215.

[299] P. Artoisenet, R. Frederix, O. Mattelaer, and R. Rietkerk, "Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations", *JHEP* **03** (2013) 015, doi:10.1007/JHEP03(2013)015, arXiv:1212.3460.

[300] E. Re, "Single-top Wt-channel production matched with parton showers using the POWHEG method", *Eur. Phys. J. C* **71** (2011) 1547, doi:10.1140/epjc/s10052-011-1547-z, arXiv:1009.2450.

[301] S. Alioli, P. Nason, C. Oleari, and E. Re, "NLO single-top production matched with shower in POWHEG: s- and t-channel contributions", *JHEP* **09** (2009) 111, doi:10.1088/1126-6708/2009/09/111, arXiv:0907.4076. [Erratum: doi:10.1007/JHEP02(2010)011].

[302] CMS Collaboration, "Investigations of the impact of the parton shower tuning in PYTHIA8 in the modelling of $\mathrm{t\bar{t}}$ at $\sqrt{s} = 8$ and 13 TeV", CMS Physics Analysis Summary CMS-PAS-TOP-16-021, CERN, 2016.

[303] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements", *Eur. Phys. J. C* **76** (2016) 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.

[304] P. Skands, S. Carrazza, and J. Rojo, "Tuning PYTHIA8.1: the Monash 2013 tune", *Eur. Phys. J. C* **74** (2014) 3024, doi:10.1140/epjc/s10052-014-3024-y, arXiv:1404.5630.

[305] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, "Matching matrix elements and shower evolution for top-pair production in hadronic collisions", *JHEP* **01** (2007) 013, doi:10.1088/1126-6708/2007/01/013, arXiv:hep-ph/0611129.

[306] S. Mrenna and P. Richardson, "Matching matrix elements and parton showers with HERWIG and PYTHIA", *JHEP* **05** (2004) 040, doi:10.1088/1126-6708/2004/05/040, arXiv:hep-ph/0312274.

[307] M. Cacciari et al., "Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation", *Phys. Lett. B* **710** (2012) 612, doi:10.1016/j.physletb.2012.03.013, arXiv:1111.5869.

[308] P. Bärnreuther, M. Czakon, and A. Mitov, "Percent level precision physics at the Tevatron: next-to-next-to-leading order QCD corrections to $\mathrm{q\bar{q}} \to \mathrm{t\bar{t}} + \mathrm{X}$", *Phys. Rev. Lett.* **109** (2012) 132001, doi:10.1103/PhysRevLett.109.132001, arXiv:1204.5201.

[309] M. Czakon and A. Mitov, "NNLO corrections to top-pair production at hadron colliders: the all-fermionic scattering channels", *JHEP* **12** (2012) 054, doi:10.1007/JHEP12(2012)054, arXiv:1207.0236.

[310] M. Czakon and A. Mitov, "NNLO corrections to top pair production at hadron colliders: the quark-gluon reaction", *JHEP* **01** (2013) 080, doi:10.1007/JHEP01(2013)080, arXiv:1210.6832.

[311] M. Beneke, P. Falgari, S. Klein, and C. Schwinn, "Hadronic top-quark pair production with NNLL threshold resummation", *Nucl. Phys. B* **855** (2012) 695, doi:10.1016/j.nuclphysb.2011.10.021, arXiv:1109.1536.

[312] N. Kidonakis, "Two-loop soft anomalous dimensions for single top quark associated production with a W^- or H^-", *Phys. Rev. D* **82** (2010) 054018, doi:10.1103/PhysRevD.82.054018, arXiv:1005.4451.

[313] Y. Li and F. Petriello, "Combining QCD and electroweak corrections to dilepton production in the framework of the FEWZ simulation code", *Phys. Rev. D* **86** (2012) 094034, doi:10.1103/PhysRevD.86.094034, arXiv:1208.5967.

[314] J. M. Campbell, R. K. Ellis, and C. Williams, "Vector boson pair production at the LHC", *JHEP* **07** (2011) 018, doi:10.1007/JHEP07(2011)018, arXiv:1105.0020.

[315] GEANT4 Collaboration, "GEANT4—a simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250, doi:10.1016/S0168-9002(03)01368-8.

[316] ATLAS Collaboration, "Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV with the ATLAS detector at the LHC", *Phys. Rev. Lett.* **117** (2016) 182002, doi:10.1103/PhysRevLett.117.182002, arXiv:1606.02625.

[317] CMS Collaboration, "Performance of muon reconstruction including Alignment Position Errors for 2016 Collision Data", CMS Detector Performance Note CMS-DP-2016-067, CERN, 2016.

[318] CMS Collaboration, "Muon Identification and Isolation efficiency on full 2016 dataset", CMS Detector Performance Note CMS-DP-2017-007, CERN, 2017.

[319] CMS Collaboration, "Performance of the CMS muon trigger system in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *JINST* **16** (2021) P07001, doi:10.1088/1748-0221/16/07/P07001, arXiv:2102.04790.

[320] CMS Collaboration, "Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *JINST* **15** (2020) P10017, doi:10.1088/1748-0221/15/10/P10017, arXiv:2006.10165.

[321] C. J. Clopper and E. S. Pearson, "The use of confidence or fiducial limits illustrated in the case of the binomial", *Biometrika* **26** (1934) 404–413, doi:10.1093/biomet/26.4.404.

[322] R. D. Cousins, K. E. Hymes, and J. Tucker, "Frequentist evaluation of intervals estimated for a binomial parameter and for the ratio of poisson means", *Nucl. Instrum. Methods Phys. Res. A* **612** (2010) 388–398, doi:https://doi.org/10.1016/j.nima.2009.10.156.

[323] CMS Collaboration, "Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8\,\text{TeV}$", *Eur. Phys. J. C* **75** (2015) 542, doi:10.1140/epjc/s10052-015-3709-x, arXiv:1505.04480.

[324] CMS Collaboration, "Measurement of double-differential cross sections for top quark pair production in pp collisions at $\sqrt{s} = 8\,\text{TeV}$ and impact on parton distribution functions", *Eur. Phys. J. C* **77** (2017) 459, doi:10.1140/epjc/s10052-017-4984-5, arXiv:1703.01630.

[325] CMS Collaboration, "Measurement of normalized differential $\mathrm{t\bar{t}}$ cross sections in the dilepton channel from pp collisions at $\sqrt{s} = 13\,\text{TeV}$", *JHEP* **04** (2018) 060, doi:10.1007/JHEP04(2018)060, arXiv:1708.07638.

[326] CMS Collaboration, "Measurements of $\mathrm{t\bar{t}}$ differential cross sections in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ using events containing two leptons", *JHEP* **02** (2019) 149, doi:10.1007/JHEP02(2019)149, arXiv:1811.06625.

[327] L. Sonnenschein, "Analytical solution of $\mathrm{t\bar{t}}$ dilepton equations", *Phys. Rev. D* **73** (2006) 054015, doi:10.1103/PhysRevD.73.054015, arXiv:hep-ph/0603011. [Erratum: doi:10.1103/PhysRevD.78.079902].

[328] I. Korol, "Measurement of Double differential $\mathrm{t\bar{t}}$ production cross sections with the CMS detector". PhD thesis, Universität Hamburg, Hamburg, 2016. doi:10.3204/DESY-THESIS-2016-011.

[329] M. Abadi et al., "TensorFlow: large-scale machine learning on heterogeneous distributed systems", (2016). arXiv:1603.04467. Software available from https://www.tensorflow.org/.

[330] F. Chollet et al., "Keras", 2015. https://keras.io.

[331] F. Nogueira, "Bayesian optimization: open source constrained global optimization tool for Python", 2014. https://github.com/fmfn/BayesianOptimization.

[332] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions", in *Adv. Neural Inf. Process Syst. 30*, I. Guyon et al., eds., pp. 4765–4774. Curran Associates, Inc., 2017.

[333] E. Štrumbelj and I. Kononenko, "Explaining prediction models and individual predictions with feature contributions", *Knowl. Inf. Syst.* **41** (2013) 647–665, doi:10.1007/s10115-013-0679-x.

[334] D. Smilkov et al., "Smoothgrad: removing noise by adding noise", 2017. doi:10.48550/ARXIV.1706.03825, https://arxiv.org/abs/1706.03825.

[335] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks", 2017. doi:10.48550/ARXIV.1703.01365, https://arxiv.org/abs/1703.01365.

[336] F. Nielsen, "On the Jensen–Shannon symmetrization of distances relying on abstract means", *Entropy* **21** (2019), no. 5, doi:10.3390/e21050485.

[337] A. Hocker and V. Kartvelishvili, "SVD approach to data unfolding", *Nucl. Instrum. Meth. A* **372** (1996) 469–481, doi:10.1016/0168-9002(95)01478-0, arXiv:hep-ph/9509307.

[338] Schmitt, Stefan, "Data unfolding methods in high energy physics", *EPJ Web Conf.* **137** (2017) 11008, doi:10.1051/epjconf/201713711008.

[339] G. Choudalakis, "Fully bayesian unfolding", 2012. doi:10.48550/ARXIV.1201.4612, https://arxiv.org/abs/1201.4612.

[340] J. Kieseler, K. Lipka, and S. Moch, "Calibration of the top-quark Monte Carlo mass", *Phys. Rev. Lett.* **116** (2016) 162001, doi:10.1103/PhysRevLett.116.162001, arXiv:1511.00841.

[341] F. James and M. Roos, "MINUIT—a system for function minimization and analysis of the parameter errors and correlations", *Comput. Phys. Commun.* **10** (1975) 343, doi:10.1016/0010-4655(75)90039-9.

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

[343] CMS Collaboration, "Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV", *Eur. Phys. J. C* **75** (2015) 212, doi:10.1140/epjc/s10052-015-3351-7, arXiv:1412.8662.

[344] ATLAS and CMS Collaborations, "Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV", *JHEP* **08** (2016) 045, doi:10.1007/JHEP08(2016)045, arXiv:1606.02266.

[345] F. Spano, "Unfolding in particle physics: a window on solving inverse problems", *EPJ Web Conf.* **55** (2013) 03002, doi:10.1051/epjconf/20135503002.

[346] O. Behnke, K. Kröninger, G. Schott, and T. Schörner-Sadenius, "Data analysis in high energy physics: a practical guide to statistical methods". Wiley-VCH, Weinheim, 2013. doi:10.1002/9783527653416.

[347] CMS Collaboration, "Measurement of differential cross sections for top quark pair production using the lepton+jets final state in proton-proton collisions at 13 TeV", *Phys. Rev. D* **95** (2017) 092001, doi:10.1103/PhysRevD.95.092001, arXiv:1610.04191.

[348] CMS Collaboration, "Measurement of the $\mathrm{t\bar{t}}$ production cross section in the all-jets final state in pp collisions at $\sqrt{s} = 8\,\mathrm{TeV}$", *Eur. Phys. J. C* **76** (2016) 128, doi:10.1140/epjc/s10052-016-3956-5, arXiv:1509.06076.

[349] CMS Collaboration, "Measurement of the differential cross section for $\mathrm{t\bar{t}}$ production in the dilepton final state at $\sqrt{s} = 13\,\mathrm{TeV}$", CMS Physics Analysis Summary CMS-PAS-TOP-16-011, CERN, 2016.

[350] CMS Collaboration, "CMS tracker performance results for full Run 2 legacy reprocessing", CMS Detector Performance Note CMS-DP-2020-012, CERN, 2020.

[351] CMS Collaboration, "b-tagging performance of the CMS legacy dataset 2018", CMS Detector Performance Note CMS-DP-2021-004, CERN, 2021.

[352] P. F. Monni, E. Re, and M. Wiesemann, "MiNNLO$_{\mathrm{PS}}$: optimizing $2 \to 1$ hadronic processes", *Eur. Phys. J. C* **80** (2020) 1075, doi:10.1140/epjc/s10052-020-08658-5, arXiv:2006.04133.

[353] P. F. Monni et al., "MiNNLO$_{\mathrm{PS}}$: a new method to match NNLO QCD to parton showers", *JHEP* **05** (2020) 143, doi:10.1007/JHEP05(2020)143, arXiv:1908.06987.

[354] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J. C* **71** (2011) 1554, doi:10.1140/epjc/s10052-011-1554-0, arXiv:1007.1727. [Erratum: Eur.Phys.J.C 73, 2501 (2013)].

List of publications

Main publications

CMS Collaboration, Measurement of the top quark pole mass using $\mathrm{t\bar{t}}$+jet events in the dilepton final state in proton-proton collisions at $\sqrt{s}$ =13 TeV, Submitted to the Journal of High Energy Physics (JHEP), [arXiv:2207.02270].

Other publications

CMS Collaboration, Measurement of differential cross sections for the production of top quark pairs and of additional jets in pp collisions at $\sqrt{s}$ =13 TeV, CMS Physics Analysis Summary CMS-PAS-20-006, [https://cds.cern.ch/record/2804974].

CMS Collaboration, The Phase-2 Upgrade of the CMS Data Acquisition and High Level Trigger, CMS Technical Design Report CMS-TDR-022, [https://cds.cern.ch/record/2759072].

CMS Collaboration, Expected Performance of Run-3 HLT b-quark jet identification, CMS Detector Performance Note CMS-DP-2022/030, [https://cds.cern.ch/record/2825704].

Conference proceedings

S.Wuchterl [for the CMS Collaboration], Measurement of the top quark pole mass using $\mathrm{t\bar{t}}$+jet events in the dilepton final state at $\sqrt{s}$ =13 TeV, *Proceedings of the 41st International Conference on High Energy physics - ICHEP 2022*, [arXiv:2207.13413].

S.Wuchterl [for the ATLAS and CMS Collaborations], Top quark mass and properties in ATLAS and CMS, *Proceedings of the The Tenth Annual Conference on Large Hadron Collider Physics - LHCP 2022*, [arXiv:2208.07153].

S.Wuchterl [for the CMS Collaboration], Measurement of the top quark pole mass using $\mathrm{t\bar{t}}$+jet events in the dilepton final state at $\sqrt{s}$ = 13 TeV with the CMS experiment, *Proceedings of the The Tenth Annual Conference on Large Hadron Collider Physics - LHCP 2022*, [arXiv:2207.13422].

S.Wuchterl [for the ATLAS and CMS Collaborations], Standard model parameters from top quark measurements at LHC with ATLAS and CMS, *Proceedings for the 14th International Workshop on Top Quark Physics (TOP) 2021*, [arXiv:2201.0694].

S.Wuchterl [for the CMS Collaboration],Recent results on top quark mass and properties and rare/anomalous top quarks interactions in CMS, *Proceedings for the XXVIII International Workshop on Deep-Inelastic Scattering and Related Subjects (DIS)*, SciPost Phys. Proc. 8, 120 (2022), [arXiv:2106.03478].