UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia “Galileo Galilei”

Master Degree in Physics

Final Dissertation

Differential cross section measurement of top quark pair production with the CMS experiment using a ML-based kinematic reconstruction

Thesis supervisor	Candidate
Prof. Tommaso Dorigo	Giovanni Celotto
Thesis co-supervisor	
Dr. Olaf Behnke	

Academic Year 2022/2023

Abstract

In this thesis a measurement of the differential cross section of top quark pair ($t\bar{t}$) production in proton-proton (pp) collisions at $\sqrt{s} = 13$ TeV is presented. The measurement is performed using data collected in 2018 with the Compact Muon Solenoid (CMS) detector at the Large Hadron Collider (LHC), corresponding to an integrated luminosity of 59.8 fb^{-1}. The analysis is performed using the $t\bar{t}$ dileptonic different-flavor ($e^{\pm}\mu^{\mp}$) decay channel. The cross section is measured differentially as a function of the invariant mass of the $t\bar{t}$ system ($m_{t\bar{t}}$). The presence of final state neutrinos makes a kinematic reconstruction necessary for the measurement of $m_{t\bar{t}}$. In this work, the observable $m_{t\bar{t}}$ is regressed from the visible detector objects by a neural network, leading to an improvement in the efficiency, resolution and a reduction of the statistical uncertainties in the unfolded cross section compared to results based on analytical reconstruction approaches.

Contents

1 Introduction 1

2 Theoretical Overview 3
2.1 The Standard Model of particle physics 3
2.1.1 Particle content 3
2.1.2 Quantum Electrodynamics 5
2.1.3 Weak interaction and Electroweak theory 6
2.1.4 Quantum Chromodynamics 8
2.2 Hadron collider physics 9
2.2.1 QCD factorization and parton distribution functions 9
2.2.2 Monte Carlo simulation 9
2.3 Top quark physics 10
2.3.1 Top quark production mechanisms at hadron colliders 11
2.3.2 Top quark decay modes 12
2.3.3 Differential cross section measurement of top quark pair production 13

3 Experimental Apparatus 15
3.1 The Large Hadron Collider 15
3.1.1 Luminosity at the LHC 16
3.2 The Compact Muon Solenoid 17
3.2.1 CMS coordinate system 18
3.2.2 The solenoid magnet 18
3.2.3 The tracking system 19
3.2.4 The electromagnetic calorimeter 20
3.2.5 The hadronic calorimeter 20
3.2.6 The muon system 21
3.2.7 Trigger 21

4 Object Reconstruction and Event Selection 23
4.1 The particle-flow algorithm 23
4.1.1 Iterative tracking 23
4.1.2 Calorimeter clustering 24
4.1.3 Link algorithm 24
4.2 Physics objects 25
4.3 Data samples and event simulation 27
4.3.1 Monte Carlo simulation 27
4.3.2 Corrections to simulated events 27

4.4 Event selection . . . 29
4.5 Analytical kinematic reconstructions . . . 31
  4.5.1 Full kinematic reconstruction . . . 31
  4.5.2 Loose kinematic reconstruction . . . 32
  4.5.3 Efficiency and unphysical solutions . . . 32

5 Top Quark Pair Invariant Mass Regression and Differential Cross Section Measurement 35
5.1 Machine Learning and Regression Neural Network . . . 35
5.2 NN architecture . . . 37
  5.2.1 Input features . . . 37
  5.2.2 Data and MC comparison for input features . . . 44
  5.2.3 Training . . . 45
5.3 NN Performance . . . 47
  5.3.1 Unfolding . . . 50
5.4 Application to 2018 data . . . 56
  5.4.1 Unfolding of 2018 data . . . 56
5.5 Systematic uncertainties . . . 60
  5.5.1 Covariance matrix definition for systematic uncertainties . . . 62

6 Conclusions 63

A Control plots for input features 65

B Contributions of Systematic Uncertainties 69

C Regression of Top and Antitop Quark Transverse Momentum 81

Bibliography 87

1

Introduction

Elementary particle physics tries to address the questions: "What is matter made of? How do its ultimate constituents interact to form the world we live in?" Even if the first intuition of a world made of fundamental building blocks dates back to the Greek atomists, the modern particle physics can be considered to be born in 1897, with the discovery of the first elementary particle: the electron by J.J. Thomson [1]. Since then, our understanding of the particles and their interactions has drastically evolved, culminating in the development of the Standard Model (SM) of particle physics. This theory was developed in stages, being tested and refined over time through the collaborative synergy of experimental and theoretical efforts and it currently represents our best description of the intricate structure and the complex dynamics of the constituents of nature. Among others, the observation of the existence of the top quark (1995) [18, 17] and the Higgs boson (2012) [36] are two of the most recent confirmations of its impressive predictive power. Nevertheless, the SM fails to account for certain physical phenomena, preventing it from being a complete theory of fundamental interactions.

With a mass of 172.5 GeV [89], the top quark plays a special role within the SM, being the heaviest particle known. Its discovery was announced in 1995 by the CDF and DØ experiments at Fermilab [17, 18]. The huge amount of data collected during Run 1 and Run 2 at the Large Hadron Collider (LHC), respectively in 2009-2012 and 2015-2018, at Conseil Européen pour la Recherche Nucléaire (CERN) allowed the measurements of its properties with unparalleled precision and are the subject of ongoing studies. The top quark's uniqueness stems from its large mass, which significantly affects the phenomenology of this particle. For example, it is the only quark which does not form bound states, since it decays before the strong force can come into play, hence it represents a unique possibility to study the properties of a bare quark. Moreover, the high Yukawa coupling to the Higgs boson suggests its possible involvement in the production and decay modes of new particles at higher energy scales. At the LHC, the top quark is mostly produced in pairs together with the corresponding antiparticle, the antitop quark. The differential cross section measurements of the production of top quark pairs as functions of event kinematic observables are of key importance, since they represent precision tests of the SM, possibly sensitive to deviations attributable to new physics.

Previous measurements of the differential production cross section of top quark pairs were based on analytical and algebraic methods performed in order to reconstruct the kinematic properties of the top quark and antiquark or of the $t\bar{t}$ system [85, 71]. In this thesis, the reconstruction of the invariant mass of top quark pairs ($m_{t\bar{t}}$) produced in proton-proton (pp) collisions at $\sqrt{s} = 13$ TeV and recorded with the Compact Muon Solenoid (CMS) is studied using a machine-learning (ML) approach. Candidate $t\bar{t}$ events are selected in the dileptonic different-flavour final state ($e^{\pm}\mu^{\mp}$). An exemplary Feynman diagram of this process is shown in Fig. 1.1. Electrons and muons arising from τ decays ($\tau^- \to e^-\overline{\nu}_e$ and $\tau^- \to \mu^-\overline{\nu}_\mu$), where the τ leptons originate from the decays of the W bosons ($W^- \to \tau^-\overline{\nu}_\tau$), are included in the signal definition. The regressed observable is used to measure the differential production cross section of top quark pairs as a function of $m_{t\bar{t}}$. In particular, the invariant mass of the $t\bar{t}$ system is reconstructed at the detector level and the cross section is unfolded to parton

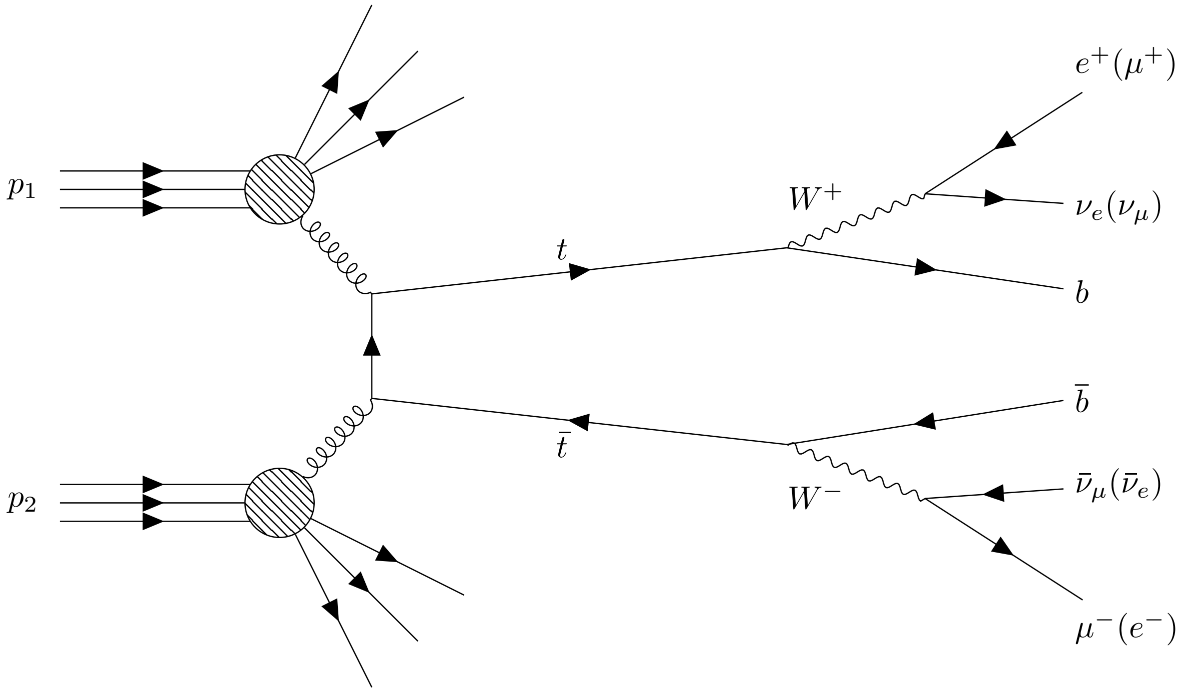


Figure 1.1: An exemplary Feynman diagram for the $t\bar{t}$ production in pp collisions, where both the W bosons decay leptonically resulting in the $e\mu$ final state.

level, correcting for the resolution and efficiency effects via response matrix inversion. The goal of using a ML-based regression is to improve the resolution of the $m_{t\bar{t}}$ reconstruction, which in turn helps to reduce the statistical uncertainties of the differential cross section measurement as well as the statistical correlations between different measurement bins.

The thesis is organized as follows. A brief theoretical introduction to the SM of particle physics, hadron collider physics and top quark physics is presented in Chapter 2. A general description of the experimental setup, including the LHC and the CMS experiment can be found in Chapter 3. Chapter 4 is devoted to the description of the reconstruction of physics objects and the event selection. In addition, the analytical approaches used as standard methods for the kinematic reconstruction of the $t\bar{t}$ system are explained. In Chapter 5 the neural network (NN) used for the regression of $m_{t\bar{t}}$ is introduced and its performance is compared to the analytical methods. The measurement of the cross section is described and compared to theoretical predictions at next-to-leading order (NLO). Finally, some conclusions and future perspectives are summarized in Chapter 6.

2

Theoretical Overview

In this chapter a brief theoretical overview of the fundamental concepts relevant to this thesis is provided. In Sec. 2.1 the Standard Model of Particle Physics is introduced. In Sec. 2.2, some details about *pp* scattering are provided, while specific information about top quark physics, spanning from the production and decay modes at the LHC to the special role played by this particle in the SM and beyond, is given in Sec. 2.3.

2.1 The Standard Model of particle physics

Up to our current knowledge, there are four fundamental interactions at work in nature: electromagnetic, weak, strong and gravitational. While gravity and the electromagnetic interaction have an infinite range, the weak and strong forces are effective only at subatomic scales and they are not experienced in the everyday life. The Standard model of particle physics [87, 19] is a quantum field theory (QFT) that describes the first three of these interactions and classifies all known elementary particles. As a QFT, the SM describes each particle as an excited state of a quantum field that permeates all of space-time and whose dynamics is controlled by a Lagrangian. The terms of the Lagrangian are derived by first postulating the set of symmetries of the system, and later by constructing the most general renormalizable Lagrangian from the field content of the system, that fulfills these symmetries.

As for any relativistic quantum field theories, the SM Lagrangian has to observe the global Poincaré symmetry. In addition, the SM is a gauge theory, meaning that its Lagrangian is invariant under local transformations according to some specific families of operations. The SM Lagrangian is invariant under the following gauge group:

$$\mathrm{G_{SM}} = \mathrm{SU}(3)_\mathrm{C} \otimes \mathrm{SU}(2)_\mathrm{L} \otimes \mathrm{U}(1)_\mathrm{Y} \tag{2.1}$$

Among the gauge groups appearing in Eq. 2.1, the strong interaction is described by the Quantum Chromodynamics (QCD), a non-Abelian gauge theory based on $\mathrm{SU}(3)_\mathrm{C}$, while the electromagnetic and weak interactions are described by the Electroweak theory, which is a non-Abelian gauge theory based on the $\mathrm{SU}(2)_\mathrm{L} \otimes \mathrm{U}(1)_\mathrm{Y}$ gauge group. Through Noether's theorem, they lead to the conservation of the colour charge, weak isospin and weak hypercharge, respectively. The possibility of spontaneous symmetry breaking [4] of the group $\mathrm{G_{SM}}$ into a smaller group $\mathrm{G_{SM}^{SSB}}$ is another key feature of the SM that will be addressed in Sec. 2.1.3. After writing the most general Lagrangian, the dynamics of the fields is found to be dependent on 19 parameters, whose numerical values can only be determined through experimental measurements. Finally, the SM is also renormalizable, i.e. physical quantities extracted from the theory are finite.

2.1.1 Particle content

The particle content of the SM is summarized in Fig. 2.1. It includes 17 particles, divided into two main classes: fermions and bosons. The former group is composed of spin-1/2 particles and they obey the Fermi-Dirac statistics, the latter has particles with integer spins and following Bose-Einstein statistics. At a fundamental level, all matter in the universe is composed of fermions, whose interactions take place through the exchange of bosons.

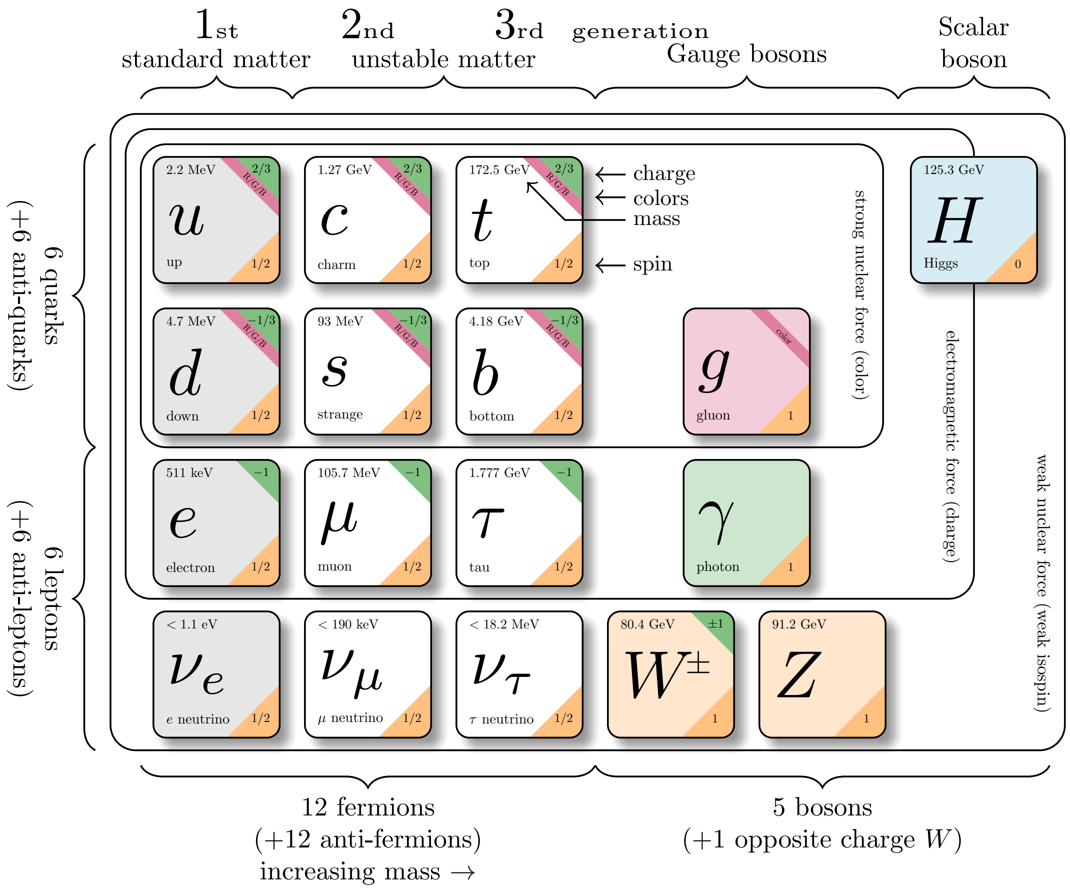


Figure 2.1: Summary of the fundamental particles of the SM, showing fermions, gauge bosons and the Higgs boson. Based on Ref. [84], values from Ref. [89]

Fermions can be further divided into two categories, leptons and quarks, each with three generations:

- **Leptons** are divided into electronic, muonic and tauonic. Each generation includes one charged lepton (e, μ, τ), provided with a negative unitary electric charge and one neutrino (ν_e, ν_μ, ν_τ), electrically neutral. Since leptons obey Fermi-Dirac statistics, for each aforementioned particle there is a corresponding antiparticle, which has opposite quantum numbers (i.e. positive charge for charged leptons) but identical mass. However, the possibility that neutrinos are Majorana particles (i.e. they are their own antiparticles) is not excluded and still under study [90].
- **Quarks** are also divided into three generations. Each generation includes an up-type quark (u, c, t) and a down-type one (d, s, b). The former class has electric charge $+2/3\,e$, the latter has $-1/3\,e$. Being fermions, every quark has a corresponding antiparticle with opposite quantum numbers and identical mass. Quarks are also provided with a color charge, conventionally denoted with red, green, blue, while the color charge of the antiparticles are anti-red, anti-green, anti-blue.

The only property that changes moving from one generation to another is the mass, which increases with the generation. Neutrinos could represent an exception to this rule, since their masses have been ruled out to be zero [20], but have not been precisely measured yet. The fermion fields are split into left- and right-handed, where left-handed fermions and right-handed antifermions transform as doublets under the $\mathrm{SU}(2)_\mathrm{L}$ group (i.e. they carry weak isospin), while right-handed fermions and left-handed anti-fermions are singlets under the same transformation. As a consequence, left-handed fermions (and right-handed antifermions) have a non-zero third component of the weak isospin (T_3) and the weak hypercharge ($\mathrm{Y_W} = 2\,(\mathrm{Q} - \mathrm{T}_3)$) differs between left- and right-handed fermions. Since all the quantum numbers are the same for leptons and quarks of different generations, the behaviour under the gauge group is summarized in Table 2.1 only for the first generation.

Field	$\mathrm{SU(3)_C}$	$\mathrm{SU(2)_L}$	$\mathrm{Y_W}$	$\mathrm{T_3}$	$\mathrm{Q} = \frac{\mathrm{Y_W}}{2} + \mathrm{T_3}$
$L_L = \begin{pmatrix} \nu_L \\ e_L \end{pmatrix}$	**1**	**2**	-1	$\begin{pmatrix} +1/2 \\ -1/2 \end{pmatrix}$	$\begin{pmatrix} 0 \\ -1 \end{pmatrix}$
e_R	**1**	**1**	-2	0	-1
$Q_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix}$	**3**	**2**	$+1/3$	$\begin{pmatrix} +1/2 \\ -1/2 \end{pmatrix}$	$\begin{pmatrix} +2/3 \\ -1/3 \end{pmatrix}$
u_R	**3**	**1**	$+4/3$	0	$+2/3$
d_R	**3**	**1**	$-2/3$	0	$-1/3$

Table 2.1: Transformation behaviour of fermions under the SM gauge group. Number in bold indicate if the field transforms as a singlet, doublet or triplet.

The interactions between particles are described by the exchange of spin 1 gauge bosons. The electromagnetic interaction is mediated by the photon (γ), the weak interaction by the W and Z bosons and the strong interaction by the gluon (g). Besides these particles, the Higgs boson, is the particle associated to the Higgs field which is responsible for giving mass to all the fundamental particles. It is the only spinless fundamental particle in the SM.

2.1.2 Quantum Electrodynamics

All particles with an electric charge different from zero are subjected to the electromagnetic interaction, which is described by the Quantum Electrodynamics (QED). Differently from the other interactions, the gauge group associated to the electromagnetic interaction $\mathrm{U(1)_{EM}}$, arising from the spontaneous symmetry breaking of the $\mathrm{SU(2)_L} \otimes \mathrm{U(1)_Y}$ group (see 2.1.3), is an Abelian group. As a consequence, the photon is electrically neutral and does not interact with itself. The QED Lagrangian density for a spin 1/2 field interacting with the electromagnetic field can be written as:

$$\mathcal{L}_{\mathrm{QED}} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \bar{\psi}(i\gamma^{\mu}\partial_{\mu} - m)\psi - e j^{\mu} A_{\mu}, \tag{2.2}$$

where γ^{μ} are the Dirac matrices, ψ a bispinor field of the spin 1/2 particle, $\bar{\psi}$ its Dirac adjoint and m its mass. The gauge field A_{μ} is the four-potential of the electromagnetic field and it is introduced to ensure the invariance of the Lagrangian under local U(1) transformation. The first term in Eq. 2.2 is the kinetic term of the gauge field and $F_{\mu\nu} = \partial_{\mu} A_{\nu} - \partial_{\nu} A_{\mu}$ is called the electromagnetic tensor. The second term represents the kinetic and massive terms of the spin 1/2 field. The last term is the interaction term between the spin 1/2 field and the gauge field and it depends on the coupling constant e, which is equal to the electric charge of the spin 1/2 field. Finally, j^{μ} is the conserved U(1) current arising from Noether's theorem and it can be written as $j^{\mu} = \bar{\psi}\gamma^{\mu}\psi$. Within the context of QED, the interactions between particles can be described through the exchange of virtual photons and they are usually illustrated by means of Feynman diagrams, which allow calculating the transition probability for a particular process. In these diagrams every vertex corresponds to an interaction term and contributes a factor proportional to the coupling constant. If the coefficients are small enough, the interaction term can be considered a perturbation of the free field term. As a result, a perturbative approach can be employed to compute scattering amplitudes.

2.1.3 Weak interaction and Electroweak theory

The weak interaction has both a charged and a neutral current, which are mediated by the $W^{\pm}$ and Z bosons, respectively. Differently from the photon, these bosons are massive and this determines the short range of this interaction. Since the W boson carries an electric charge it also interacts with the photon. The weak charged current only couples to left-handed fermions, and therefore it is said to maximally violate parity. Conversely, the neutral current couples to both left- and right-handed fermions but with different strengths.

Following the success of QED, several attempts were undertaken to formulate a similar theory for the weak nuclear force. The unification of the electromagnetic and the weak interactions in a consistent theory is one of the greatest successes of the SM [2, 6, 7]. This was achieved by requiring the electroweak Lagrangian to be invariant under local $\mathrm{SU}(2)_{\mathrm{L}} \otimes \mathrm{U}(1)_{\mathrm{Y}}$ transformations. New gauge fields need to be introduced to ensure the invariance of the Electroweak Lagrangian:

$$\mathcal{L}_{\mathrm{EWK}} = -\frac{1}{4}W_a^{\mu\nu}W^a_{\mu\nu} - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} + \overline{Q}_j i\not{D} Q_j + \overline{u}_j i\not{D} u_j + \overline{d}_j i\not{D} d_j + \overline{L}_j i\not{D} L_j + \overline{e}_j i\not{D} e_j, \tag{2.3}$$

where the subscript j sums over the three generations of fermions, Q, u, and d are the left-handed doublet, right-handed singlet up, and right handed singlet down quark fields, L and e are the left-handed doublet and right-handed singlet lepton fields. The covariant derivative D_μ is defined as:

$$D_\mu \equiv \partial_\mu + i\frac{g'}{2}\hat{\mathrm{Y}}_{\mathrm{W}}B_\mu + i\frac{g}{2}\hat{\mathrm{T}}_j W^j_\mu, \tag{2.4}$$

where $\hat{\mathrm{T}}_j$ and $\hat{\mathrm{Y}}_{\mathrm{W}}$ are the three generators of the non-Abelian $\mathrm{SU}(2)_{\mathrm{L}}$ group and the single generator of the Abelian U(1) group respectively, W^j_μ and B_μ are the associated gauge fields, g and g' are the coupling constants of the two interactions. The conserved charge under $\mathrm{SU}(2)_{\mathrm{L}}$ transformations is the third component of the weak isospin (T_3), the one conserved under $\mathrm{U}(1)_{\mathrm{Y}}$ transformations is the weak hypercharge (Y_{W}). The physical states of the gauge fields (i.e. the γ, $W^{\pm}$ and Z bosons) can be expressed by performing a linear combination of W^1_μ and W^2_μ as:

$$W^{\pm}_\mu = \frac{1}{2}\left(W^1_\mu \mp W^2_\mu\right) \tag{2.5}$$

and by performing a rotation of the kind:

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_{\mathrm{W}} & \sin\theta_{\mathrm{W}} \\ -\sin\theta_{\mathrm{W}} & \cos\theta_{\mathrm{W}} \end{pmatrix} \begin{pmatrix} B \\ W_3 \end{pmatrix} \tag{2.6}$$

where θ_{W} is the Weinberg angle, which determines the ratio between the Z and the W boson masses as $\cos\theta_{\mathrm{W}} = m_W/m_Z$ and is related to the electromagnetic charge e as $e = g\sin\theta_W = g'\cos\theta_W$.

The kinetic terms of the gauge bosons in Eq. 2.3 gives rise to self-interaction terms between the W^a_μ fields, which in turn are observed as the γW^+W^- and ZW^+W^- triple gauge interactions and the $W^+W^-W^+W^-$, W^+W^-ZZ and $W^+W^-\gamma\gamma$ quartic gauge interactions after rotating to the physical fields.

The Higgs mechanism

All the fields introduced in the previous section are massless in spite of the finite measured masses of the fermions, W and Z bosons. However, any mass terms would violate the gauge invariance of the Lagrangian. This leads to the introduction of a complex scalar field, the Higgs doublet ϕ

$$\phi = \begin{pmatrix} \varphi^+ \\ \varphi^0 \end{pmatrix} \tag{2.7}$$

which transforms as a doublet under $\mathrm{SU}(2)_{\mathrm{L}}$ and it is provided with a weak hypercharge $\mathrm{Y}_{\mathrm{W}} = 1$. As a consequence, the first component is electrically positively charged and the second one is electrically neutral. The most general gauge invariant potential the Higgs doublet can be subjected to is given by:

$$V(\phi) = -\mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2 \tag{2.8}$$

where the sign of λ, which affects the stability of the Electroweak vacuum, is required to be positive in order to have a potential bounded from below and the sign of μ^2 determines the possibility of having spontaneous symmetry breaking. If the sign of μ^2 is negative, the state of minimum energy will be at $\phi = 0$ and the potential will preserve the symmetries of the Lagrangian. Instead, if the sign of μ^2 is positive, there is an infinite number of degenerate minima satisfying $\phi^\dagger\phi = v^2/2$ (Fig. 2.2) and the field will acquire a vacuum expectation value (VEV)

$$\langle\phi\rangle = \sqrt{\frac{\mu^2}{2\lambda}} \equiv \frac{v}{\sqrt{2}} \tag{2.9}$$

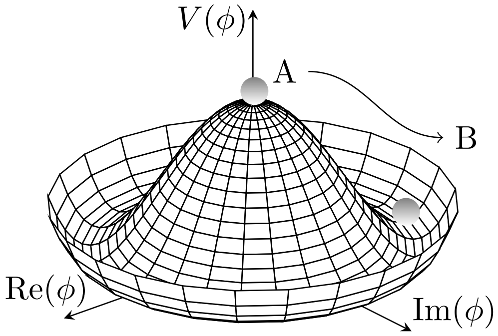


Figure 2.2: Higgs potential, also known as Mexican hat potential, and SSB. Based on Ref. [94]

Since the potential only depends on $\phi^\dagger\phi$, one can arbitrarily choose $\phi = \left(0, v/\sqrt{2}\right)^T$. Expanding the ground-state field around the new minimum of the potential and later exploiting the unitary gauge, the Higgs doublet can be written as:

$$\phi_{ug} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{2.10}$$

By introducing a kinetic term for the Higgs doublet with the covariant derivative defined in Eq. 2.4 and applying the transformations needed to pass from the gauge bosons to the physical states, one can find quadratic (massive) terms for both the W and the Z bosons given by:

$$m_Z = \frac{1}{2}v\sqrt{g^2 + g'^2} \qquad\qquad m_W = \frac{1}{2}vg, \tag{2.11}$$

whereas the photon stays massless. This reflects the fact that after the spontaneous symmetry breaking of the SM symmetry group there is one unbroken generator $\mathrm{Q} = \mathrm{T}_3 + \mathrm{Y_W}/2$:

$$\mathrm{G_{EWK}} = \mathrm{SU(2)_L} \otimes \mathrm{U(1)_Y} \xrightarrow{\mathrm{SSB}} \mathrm{U(1)_{EM}} \tag{2.12}$$

By expanding the terms appearing in Eq. 2.8, one finds that the Higgs field h itself acquires a mass $m_h = \sqrt{2\lambda}\mu$ and trilinear and quartic self-interactions of the Higgs boson proportional to λv and λ, respectively, arise. Moreover, interactions terms of the kind hWW, $hhWW$, hZZ, $hhZZ$ appear from the kinetic term of the Higgs field.

Fermion masses and quark mixing

The mechanism of Electroweak Symmetry Breaking (EWSB) gives an explanation about how bosons acquire masses, but does not generate masses for fermions. This is done by the introduction of the Yukawa Lagrangian:

$$-\mathcal{L}_Y = \bar{Q}_L\tilde{\phi}Y_u u_R + \bar{Q}_L\phi Y_d D_R + \bar{L}_L\phi Y_e E_R + \left(\bar{L}_L\tilde{\phi}Y_\nu N_R\right) + h.c. \tag{2.13}$$

where Y_u, Y_d, Y_e are 3×3 generic matrices, $\tilde{\phi} = i\sigma_2\phi^*$ and a three dimensional notation has been introduced accounting for the three generations:

$$U_{L,R} = \begin{pmatrix} u \\ c \\ t \end{pmatrix}_{L,R} \qquad D_{L,R} = \begin{pmatrix} d \\ s \\ b \end{pmatrix}_{L,R} \qquad E_{L,R} = \begin{pmatrix} e \\ \mu \\ \tau \end{pmatrix}_{L,R} \qquad N_{L(,R)} = \begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix}_{L(,R)} \tag{2.14}$$

and doublets for quarks and leptons are defined as:

$$Q_L = \begin{pmatrix} U \\ D \end{pmatrix}_L \qquad L_L = \begin{pmatrix} N \\ E \end{pmatrix}_L \tag{2.15}$$

After the EWSB the Yukawa Lagrangian in Eq. 2.13 reads in the unitary gauge:

$$\mathcal{L}_Y^{ug} = -\frac{v+h}{\sqrt{2}} \left(\bar{U_L} Y_u U_R + \bar{D_L} Y_d D_R + \bar{E_L} Y_u E_R + (\bar{N_L} Y_u N_R) + h.c. \right) \tag{2.16}$$

At this point the Yukawa matrices are not diagonal. It can be shown that it is always possible to diagonalize generic matrices by means of a biunitary transformation. By rotating the fields with appropriate unitary matrices of the kind

$$\begin{cases} U'_{L,R} = R_{L,R} U_{L,R} \\ D'_{L,R} = S_{L,R} D_{L,R} \\ E'_{L,R} = X_{L,R} E_{L,R} \end{cases} \tag{2.17}$$

one ends up with diagonal Yukawa matrices $\hat{Y}_f = L_f Y_f R_f^\dagger$ and the mass terms can be recognized as $m_f = \hat{y}_f v/\sqrt{2}$. The effect of changing the basis from the flavor basis to the mass basis does not affect the EM and weak neutral interaction terms due to the unitarity of the matrices. However, the weak charged current is affected since this interaction is among up-type and down-type fermions and the unitarity of the rotation matrices cannot be exploited. In the case of quarks, this gives origin to the Cabibbo-Kobayashi-Maskawa (CKM) matrix V_{ij}^{CKM} which describes the probability of a transition from one flavor j to another flavor i [9]. This matrix is predicted to be unitary by construction and its individual elements have been determined experimentally [89]:

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} 0.97435 \pm 0.00016 & 0.22500 \pm 0.00067 & 0.00369 \pm 0.00011 \\ 0.22486 \pm 0.00067 & 0.97349 \pm 0.00016 & 0.04182^{+0.00085}_{-0.00074} \\ 0.00857^{+0.00020}_{-0.00018} & 0.04110^{+0.00083}_{-0.00072} & 0.999118^{+0.000031}_{-0.000036} \end{pmatrix} \tag{2.18}$$

2.1.4 Quantum Chromodynamics

Gluons are massless particles responsible for mediating the strong interaction. Differently from QED, where photons are electrically neutral, in QCD both quarks and gluons carry a so-called color charge. Therefore, gluons can also self interact, as a consequence of the non-Abelian nature of the $\mathrm{SU}(3)_\mathrm{C}$ group. Gluons exist in 8 different states. Two peculiar features characterize the strong interaction: confinement and asymptotic freedom. The first phenomenon refers to the fact that color charged states cannot be observed as free particles, but they are confined in bound states called hadrons, which are singlets under transformations of the $\mathrm{SU}(3)_\mathrm{C}$ group (i.e. they are colorless). The two main types of hadrons are the mesons ($q\bar{q}$), composed of one quark and one antiquark with the corresponding anticolor, and the baryons (qqq), that are bound states of three quarks with different colors. Evidences of tetraquarks and pentaquarks have also been observed, in agreement with QCD predictions [54]. The reason of this phenomenon relies on the terms appearing in the QCD potential, which makes more convenient to create new $q\bar{q}$ pairs, if the distance between a pair of coloured particles is high enough, until all particles are confined into colour-neutral states. The second phenomenon refers to the decreasing value of the coupling constant α_s, when going to higher energies, such that quarks and gluons behave as approximately free particles in this regime.

2.2 Hadron collider physics

2.2.1 QCD factorization and parton distribution functions

When two protons collide at a high center-of-mass energy, their internal structure is resolved and their constituents, commonly referred to as partons (e.g. quarks and gluons), take part into a hard interaction, which leads to the production of other particles. The production cross section for a given process in a pp collision can be factorized into the convolution of two terms, where the first describes the structure of the two protons and the second corresponds to the partonic cross section, which is calculable perturbatively as a power series in α_s [24]:

$$d\sigma_{pp\rightarrow \mathrm{ab+X}} = \sum_{i,j=q,\bar{q},g} f_i(x_1,\mu_F^2) f_j(x_2,\mu_\mathrm{F}^2)\cdot \sigma^{\mathrm{hard}}_{ij\rightarrow \mathrm{ab}}(\alpha_s(\mu_\mathrm{R}^2),\mu_\mathrm{R}^2,\mu_\mathrm{F}^2,\hat{s})dx_1dx_2 \tag{2.19}$$

In Eq. 2.19, the parton distribution functions (PDFs) $f_{i(j)}(x_{1(2)},\mu_F^2)$ give the probability to find the parton i (j) inside the proton 1 (2), carrying the longitudinal momentum fraction $x_i = p_i/p_1$ ($x_j = p_j/p_2$) of the original proton at an energy scale μ_F. The cross section of the hard scattering $\sigma^{\mathrm{hard}}_{ij\rightarrow \mathrm{ab}}$ represents the probability to produce the considered final state (ab) from the partons of the initial state. Besides the strong coupling constant α_s and the factorization scale μ_F, $\sigma^{\mathrm{hard}}_{ij\rightarrow \mathrm{ab}}$ also depends on the partonic center of mass energy, denoted by $\hat{s} = x_1x_2s$ and the renormalization scale μ_R. The choices of both μ_R and μ_F should not affect the calculation. However, since the calculations are available to a certain fixed order of α_s, μ_F and μ_R are typically chosen close to a physical scale of the process. The variation of these parameters by a factor two up and down is usually employed to account for possible effects of missing higher order corrections.

The PDFs depend on both an energy scale factor μ_F and the momentum fraction x. The dependence of the PDFs on μ_F is described by the Dokshitzer–Gribov–Lipatov–Altarelli–Parisi (DGLAP) equations [14, 8, 13], which are a set of integral-differential-equations and can be used to evolve a PDF from one scale to another. Differently, the x-dependence of the PDFs cannot be calculated from first principles yet and has to be determined experimentally, mainly from neutrino-nucleon and electron-nucleon scattering.

2.2.2 Monte Carlo simulation

Proton beams can be accelerated to higher kinetic energy than electron beams in circular colliders. This fact is mainly due to the higher masses of protons, which implies negligible losses from synchrotron radiation [55] and has favored throughout the years hadron colliders over lepton colliders when it comes to setting the record collision energy [52]. However, due to the composite nature of the beam particles, the event structure at hadron colliders is significantly more complex than at lepton colliders. The description of the steps involved in the complex chain occurring in a collision, including the fragmentation, hadronization, and the decay of unstable particles, as well as the detector response requires the use of Monte Carlo simulations.

The simulation of a high energy particle collision can be divided into the following steps, as illustrated in Fig. 2.3:

- **Hard Collision**. The matrix element of the hard scattering process is computed at a fixed order of α_s. The scattering is simulated between the partons of the protons each carrying a momentum accordingly to the PDFs. For this thesis POWHEG (version 2) [25, 28], MG5_AMC@NLO [41] and PYTHIA (version 8.230) [47] are used for the matrix elements (ME) calculations of the signal and background processes, while the set of PDFs used is NNPDF3.1 PDF at next-to-next-to-leading order (NNLO) [60]. In addition, secondary parton-parton interactions and effects of beam remnants are simulated. These interactions are commonly referred to as multi-parton interactions (MPIs) and they lead to soft QCD processes and interference with particles emerging from the hard scattering. All these effects are collectively denoted as underlying events (UE) and they are described by phenomenological models.

- **Parton Shower**. In this step higher order QCD effects are taken into account, like the soft and collinear emissions of partons. Parton shower models are based on the splitting functions $P_{ij}(z)$, which give the probability for a parton i to emit a parton j carrying a momentum fraction z of its longitudinal momentum. The parton shower for the processes of this thesis has been performed using PYTHIA8.
- **Hadronization**. This step describes the production of hadrons starting from the initial color charged partons. There are two classes of hadronization models: the string and the cluster models. In the MC simulation used for this thesis, the Lund string model as implemented in PYTHIA8 is employed. In this model, a color string linking a pair of $q\overline{q}$ is introduced to explain the production of additional quarks and the energy stored in this color string grows linearly with the distance between the pair. Once a $q\overline{q}$ pair is produced in a collision, the increasing distance between the two quarks causes a stretching of the colour string. When the energy stored in the colour string exceeds the amount needed to produce an additional $q\overline{q}$ pair, the string can break into two new quarks, as an effect of the confinement. This procedure is repeated until hadrons are produced. In this model, gluons are described as kinks in the strings affecting the kinematic properties of the hadrons.
- **Particle Decay and Detector response**. After the simulation of the decay of the short-lived particles, the interactions of the long-lived particles with the detectors is simulated. For this thesis, this step is simulated with GEANT4 [22].

Since the emission of partons is accounted in both the fixed-order calculation of the ME and in the PS, a possible double counting can arise. To avoid it, a matching procedure between ME and PS needs to be introduced. As PSs are designed to model soft emission at lower scales in the collinear limit, whereas matrix-element calculations precisely determine hard emissions, the hardest emissions of the PS are corrected. In the POWHEG method [28] a matching scale h_{damp} is introduced to damp the high-p_{T} emissions in the PS, guaranteeing a smooth transition between the ME and PS. The value of h_{damp} is usually tuned using the experimental measurements.

2.3 Top quark physics

Despite the top quark was discovered only in 1995 by the CDF and DØ experiments at Fermilab, its existence had been predicted way before. In 1964 an experiment carried out at the Brookhaven Alternating Gradient Synchrotron (AGS) by Christenson, Cronin, Fitch and Turlay showed that the longer-lived state of the two neutral kaons, the K^0_L, which was thought to decay only to three pions and should therefore be assigned odd parity, rarely decayed into a parity-even two pion state [3]. This process implies the violation of the CP symmetry, which was hard to accept at the beginning but turned out to be one fundamental conditions to explain the baryon asymmetry in the universe [5]. In 1973, nine years after the first CP violation observation and when only three quarks were hypothesized to exist, a theory with three quark generations was introduced by Makoto Kobayashi and Toshihide Maskawa, to explain the source of CP violations as originating from one physical complex phase in the flavor mixing matrix [9]. One year later the discovery of the J/ψ meson came [10, 11] and two years later the τ lepton [12] was discovered breaking the symmetry between the number of known lepton and quark generations. The fifth quark was soon announced in 1977 in an experiment carried out by Lederman at Fermilab [15] and its charge and weak isospin were measured to be compatible with the ones of a down-type quark [16]. The corresponding isopartner was not only missing, but required to preserve the consistency of the SM as a gauge anomaly-free theory. Twenty years of searches, in which the lower bound of the top quark mass was gradually increased, led to the discovery of the sixth quark produced using $p\overline{p}$ collisions at the Tevatron with a center of mass energy of 1.8 TeV [18, 17].

The top quark is the heaviest particle in the SM and its large mass ($m_{\text{t}} = 172.5 \pm 0.7$ GeV [89]) characterizes its phenomenology. Among several implications, the top quark is the only quark which decays before hadronizing, representing a unique possibility to study the properties of a bare quark, like charge and polarization. Besides this, the large mass of the top quark plays a role in several virtual

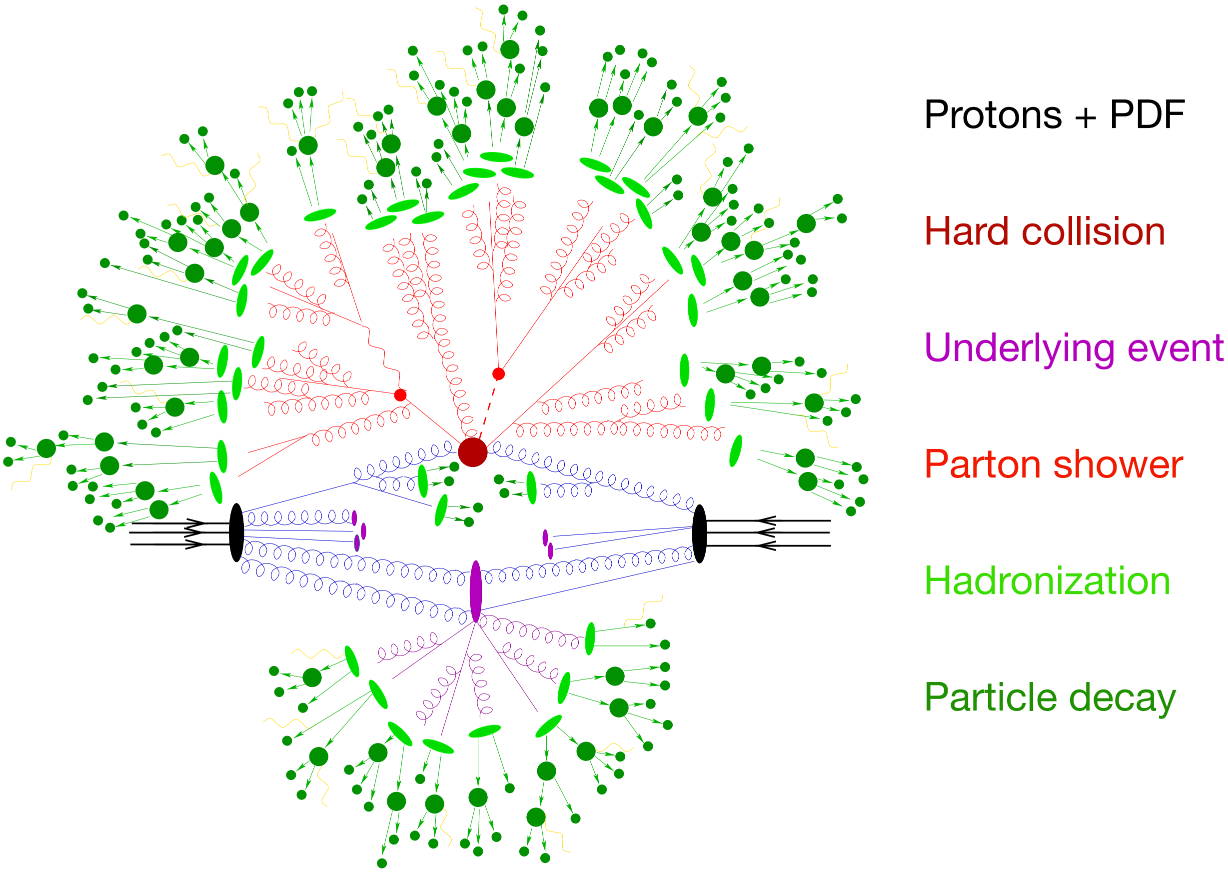


Figure 2.3: Illustration of the MC simulation of a pp collision. Partons (blue) from the initial protons (black) contribute to the hard interaction (dark red) according to their structure as given by the PDFs. An underlying-event interaction is visible in purple. The parton shower branching is shown in red. After the hadronization of partons into hadrons (light green), the decay into final state particles takes place (dark green). Additional photon radiations are shown in yellow. As a final step, the interactions of the generated events with the CMS detector is simulated (not shown). Figure from Ref. [92].

processes. For example, the Higgs boson mass receives quadratically divergent radiative corrections depending on the top quark mass. Similarly, loop diagrams involving top quark pairs tend to drive the quartic Higgs coupling λ to negative values at higher energies, making the electroweak vacuum unstable [35]. Based on the current experimental measurements on m_{t} and m_h, if the SM is extrapolated up to the Planck scale and assuming that there are no couplings to new particles, then the Higgs vacuum sits very close to the border between stability and metastability [37, 79]. Both these phenomena are shown in the plots in Fig. 2.4.

2.3.1 Top quark production mechanisms at hadron colliders

The top quark can either be produced in pairs ($t\bar{t}$) with its antiparticle via the strong interaction or as a single-top via a weak process. The former mechanism is the dominant one at hadron colliders and the main Feynman diagrams contributing to this process are presented in Fig. 2.5 at leading order (LO) in QCD. These processes can be further classified according to their initial state into gluon-gluon fusion ($gg \to t\bar{t}$) and quark-antiquark annihilation ($q\bar{q} \to t\bar{t}$). Among these processes, the main production mechanism in pp collisions at the energies available at the LHC ($\sqrt{s} = 13$ TeV for Run 2) is given by the gluon-gluon fusion, which contributes to about 90% of the total cross section. Conversely, at the Tevatron, where protons and antiprotons collided at $\sqrt{s} = 1.96$ TeV until it was shut down in 2011, the top quark was mainly produced by $q\bar{q}$ annihilation. The reason of this difference relies not only on the presence of the valence antiquarks in the antiprotons which can annihilate with the corresponding quarks in the proton, but also on the fact that at higher energies the proton PDFs are dominated by the gluons. In Figure 2.6 a summary of the $t\bar{t}$ production cross section measurements

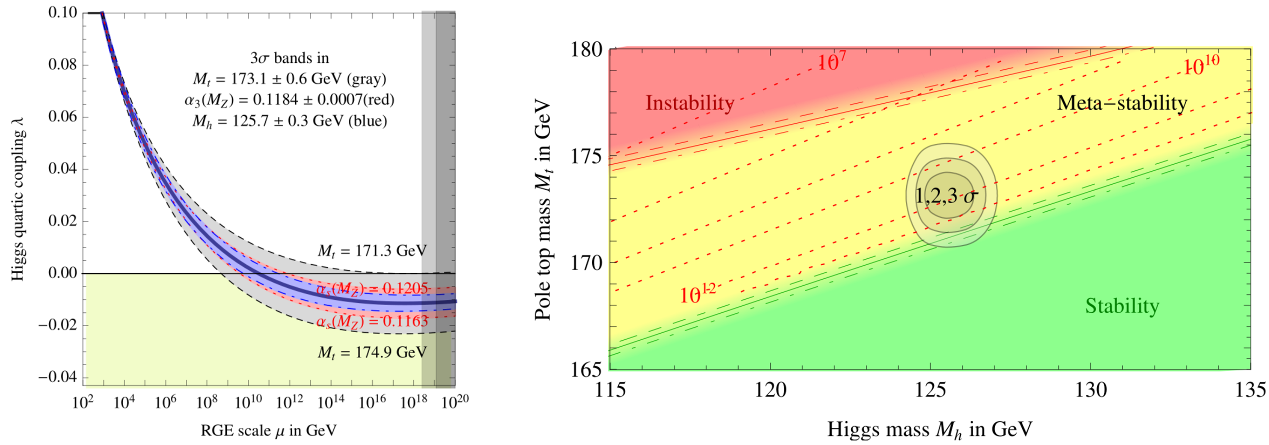


Figure 2.4: Left: evolution of Higgs quartic coupling λ considering $\pm 3\sigma$ variations of the values m_t (black), m_h (blue) and α_s (red) as a function of the renormalization scale μ_{R}. Right: regions of stability, meta-stability, and instability of the SM EWK vacuum as a function of the top quark and Higgs boson masses [37].

at different centre-of-mass energies is shown compared to the SM theoretical predictions at NNLO + next-to-next-to-leading-logarithmic (NNLL) soft gluon resummation, showing an excellent agreement over a range of several TeV. The production cross section at the LHC at $\sqrt{s} = 13$ TeV is predicted to be $\sigma_{t\bar{t}} = 831^{+19.8+35.1}_{-29.2-35.1}$ pb, where the first uncertainty is from the scale dependence and the second one from the PDFs [89].

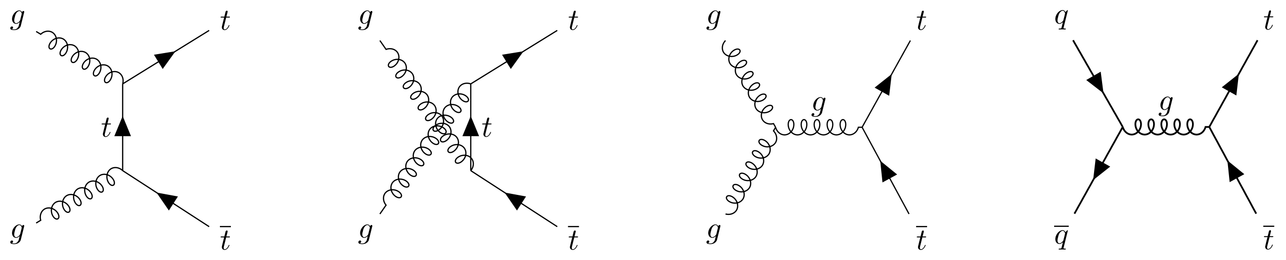


Figure 2.5: Feynman diagrams for $t\bar{t}$ production at LO in QCD. The first three diagrams represent the gluon-gluon fusion production, the last diagram represents the quark-antiquark annihilation.

As far as the single top production is concerned, it can take place in three different channels, summarized in Fig. 2.7. In the s- and t-channels a virtual W boson is involved, while in the tW associated production, the final state presents both a top quark and a W boson. Differently from the Tevatron, at the LHC the s- and t-channel cross sections for top quarks are larger than those for antitop quarks due to the charge-asymmetric initial state. The single top production cross section is proportional to the quark mixing element $|V_{tb}|^2$, which allows for a direct measurement of this CKM matrix element.

2.3.2 Top quark decay modes

Due to its large mass, the top quark is said to decay semi-weakly into a real W boson and a down-type quark, before hadronization can play a role. Since the CKM matrix element $|V_{tb}|$ is close to unity, the top quark mainly decays into a W boson and a b quark ($t \to Wb$), and the decays of the top quark are hence classified according to the decay modes of the W boson. The possible decays of the W boson are the hadronic one, when it decays into one quark and one antiquark, and the leptonic one, when it decays into a charged lepton and the corresponding antineutrino ($W^- \to \ell\overline{\nu}$). Therefore, when dealing with top quark pairs, each decaying into a W boson and b (anti)quark, three possible scenarios can take place, summarized in Fig. 2.8:

- **Fully Hadronic:** Both W bosons decay to quarks, and six final state jets are expected, of which two are b jets. This channel has the largest branching ratio (45.7%) and does not present missing

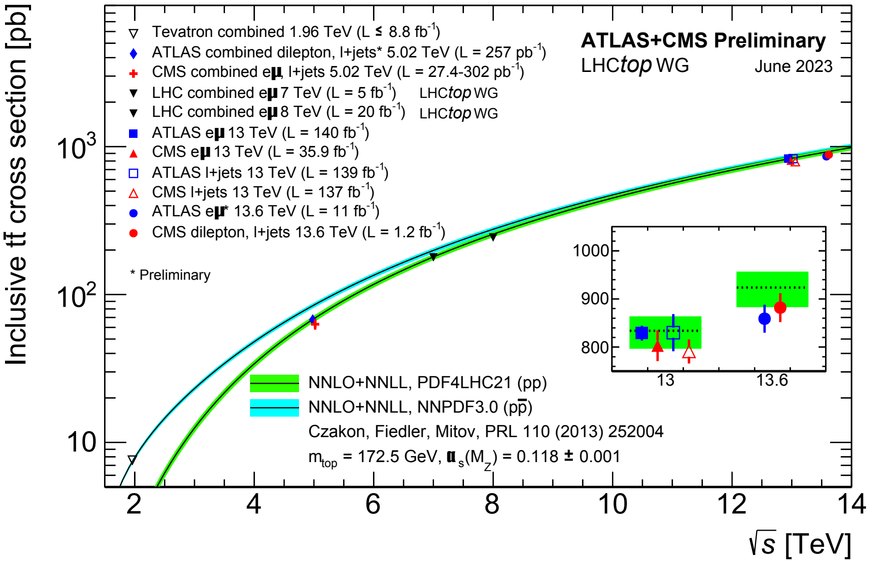


Figure 2.6: Summary of LHC and Tevatron measurements of the top-pair production cross-section as a function of the centre-of-mass energy compared to the NNLO QCD calculation complemented with NNLL resummation calculated using the TOP++ framework [45]. Measurements made at the same centre-of-mass energy are slightly offset for clarity [93].

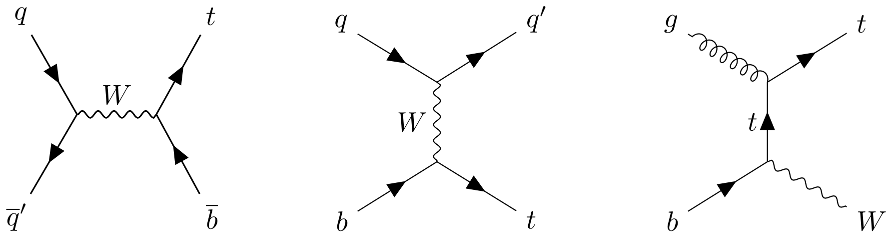


Figure 2.7: Feynman diagrams for the single top quark production via the weak interaction. From left to right: the first diagram corresponds to the s-channel, the middle diagram to the t-channel, and the last diagram to the tW associated production

transverse energy (MET) due to neutrinos, but suffers from the presence of large background from QCD multi-jet processes, which have cross sections that are orders of magnitude larger than the $t\bar{t}$ production.

- **Semileptonic:** In this case one of the W boson decays leptonically and the other hadronically. It has the second largest branching ratio (43.8%) and it is characterized by one final state isolated lepton, three jets, of which one b tagged, and the presence of MET. Typically only events involving decays into prompt electrons or muons are used, while rejecting the τ leptons due to difficulties in the kinematics. This cut lowers the branching ratio to 29.2%.

- **Dileptonic:** Both the W bosons decay leptonically, leading to two final state isolated opposite-signed leptons, large missing transverse energy due to the two neutrinos, and two b jets. It has the smallest branching ratio (10.8% including τ, 4.8% otherwise), but it has the cleanest final state, given by the unique signature of the leptons.

2.3.3 Differential cross section measurement of top quark pair production

The large number of events collected during Run 1 and Run 2 at the LHC allows precise measurements of differential top quark pair production cross sections $d\sigma/dX$, where X can be any relevant observable, such as $p_T(t)$ or $m_{t\bar{t}}$ among others [40]. These distributions can be used to validate MC models as well as to check higher order QCD effects. Deviations from the theoretical predictions can hint at

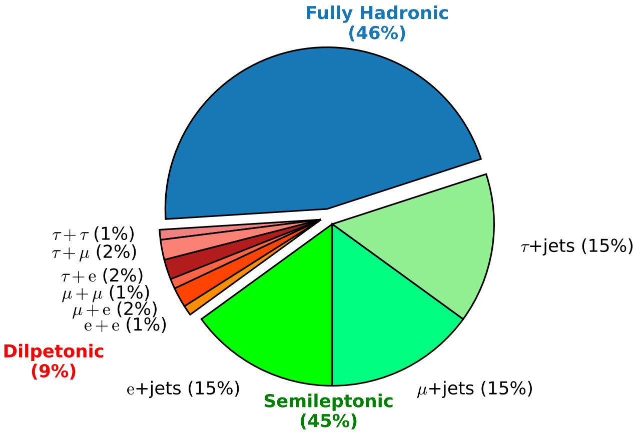


Figure 2.8: Top quark pairs decay modes. Values from [89].

new physics. For instance, in several beyond SM (BSM) models, new particles (e.g. spin-1 Z') are hypothesized to have enhanced couplings to the top quark, resulting in a resonance in the $m_{t\bar{t}}$ distribution [72]. Therefore, it is of fundamental importance to have a precise kinematic reconstruction of the $t\bar{t}$ system. In particular, in Ref. [85] several single- and multi-differential cross sections of $t\bar{t}$ production cross sections have been measured at $\sqrt{s} = 13$ TeV in pp collisions, including $d\sigma/dm_{t\bar{t}}$. Results are shown in Fig. 2.9. As one can see, the statistical uncertainty is still relevant in the region of high values of $m_{t\bar{t}}$. In general, an improvement of the $m_{t\bar{t}}$ reconstruction and its resolution could be exploited to perform the differential cross section measurement with a finer binning than the one chosen for the analysis shown in Fig. 2.9. Moreover, for a given binning, a better resolution leads to smaller statistical uncertainties of the differential cross section and less statistical correlations between the measurements in different bins.

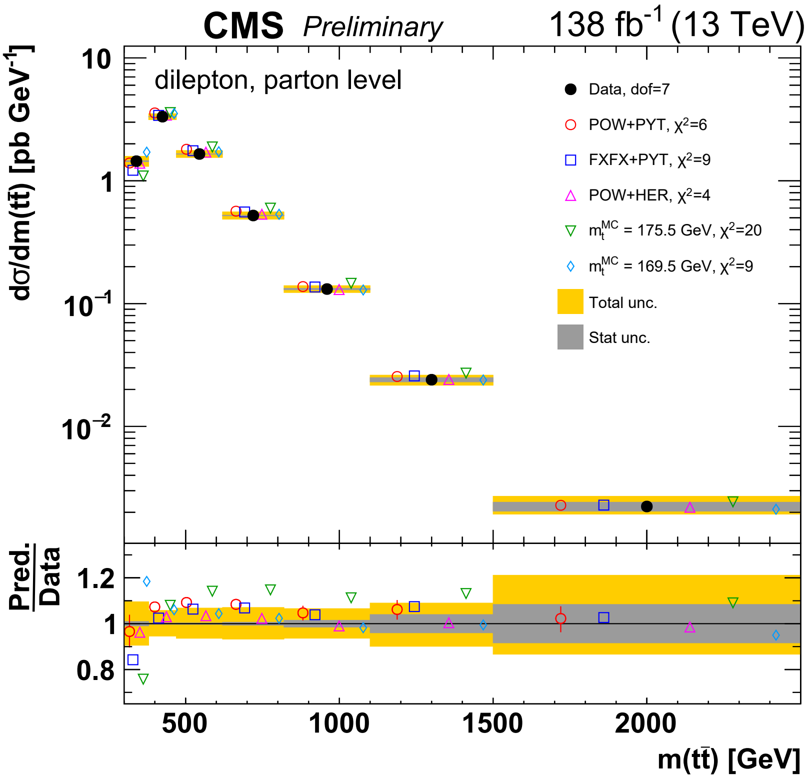


Figure 2.9: Absolute differential $t\bar{t}$ production cross section as a function of $m_{t\bar{t}}$. Data (filled circles) are compared to different MC predictions [85].

3

Experimental Apparatus

In this chapter a description of the experimental facility used to collect data analyzed in this thesis is outlined. In Sec. 3.1 the Large Hadron Collider (LHC) is described. In Sec. 3.2 the Compact Muon Solenoid (CMS) is presented with the most important subdetectors.

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [23] is the world's largest and most powerful particle accelerator. It consists of a 27-kilometer two-ring, superconducting accelerator and collider at a depth ranging from 50 to 175 metres. It was built by the European Organization for Nuclear Research (CERN) between 1998 and 2008 to replace the dismantled Large Electron-Positron Collider (LEP).

The LHC was designed to collide proton beams with a center-of-mass energy up to 14 TeV and a designed peak luminosity of $\mathcal{L} = 10^{34}$ cm^{-2} s^{-1}. It was also designed to accelerate beams of heavy ions: lead-lead (Pb-Pb) collisions and proton-lead (p-Pb) collisions are studied at $\sqrt{s_{NN}} = 2.76$ TeV with a designed peak luminosity of $\mathcal{L} = 10^{27}$ cm^{-2} s^{-1} in the nominal ion scheme.
The beams collide in four crossing points where nine experiments, with different structures and purposes, are positioned around. The four main detectors are ATLAS (A Toroidal LHC ApparatuS), CMS (Compact Muon Solenoid), LHCb (Large Hadron Collider beauty) and ALICE (A Large Ion Collider Experiment). The first two are so-called multi-purpose experiments, optimized for pp collisions and designed for new physics research, LHCb is devoted to the study of CP violation, focusing on hadrons containing a beauty quark and ALICE is an experiment specialized in high multiplicity events in heavy-ion collisions. Additionally, five other smaller experiments are located close to one of the aforementioned detectors and use the same collision points. These are LHCf (Large Hadron Collider forward), TOTEM (TOTal Elastic and diffractive cross section Measurement), MoEDAL (Monopole and Exotics Detector at the LHC), SND (Scattering and Neutrino Detector) and FASER (ForwArd Search ExpeRiment).

The proton beams in the LHC are made up of bunches of protons, spaced 25 ns apart, with each one containing $1.15 \cdot 10^{11}$ particles in the nominal proton beam operation. Before the collisions take place, the two beams inside the accelerator travel at relativistic energies in opposite directions and in separate beam pipes, that are kept at ultrahigh vacuum (10^{-10} mbar). The beams are guided around the accelerators by the presence of a magnetic field of $B = 8.3$ T, produced by superconducting electromagnets. The latter are kept at a temperature of $T = -271.3\,^{\circ}$C, provided by a complex distribution system of liquid Helium.

The collider is supplied with protons from a pre-accelerating system, aiming at gradually increasing the energy of the particles (Fig. 3.1). In particular, this chain involves the LINear ACcelerator 2 (LINAC2), which takes the protons extracted from hydrogen atoms and accelerates them to 50 MeV. Then, they are injected in the Proton Synchrotron Booster (PSB), where they reach the energy of 1.4

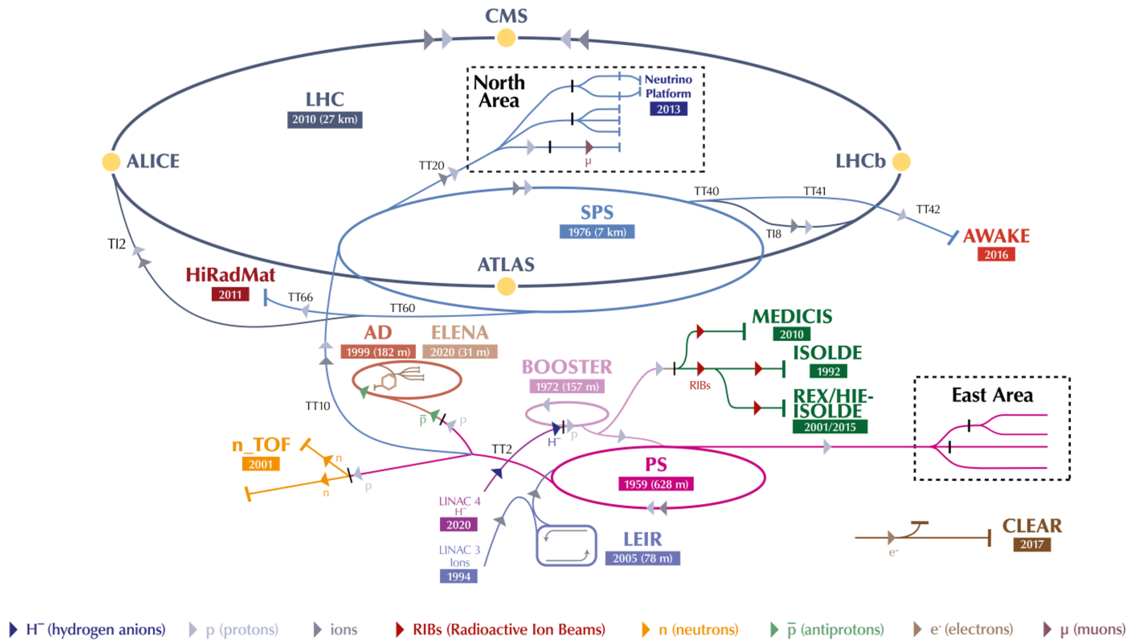


Figure 3.1: LHC accelerator complex. Figure from Ref. [86].

GeV and are delivered to the Proton Synchrotron (PS), which makes them travel at the energy of 25 GeV. The last step before entering in the LHC is in the Super Proton Synchrotron (SPS), a 7 km long machine where they are accelerated up to 450 GeV. Finally, the protons enter the LHC ring, which is designed to accelerate them up to 7 TeV.

3.1.1 Luminosity at the LHC

The instantaneous luminosity ($\mathcal{L}(t)$) and the integrated luminosity ($\mathcal{L}_{\text{int}}$) are two of the most crucial parameters in any particle collider. The instantaneous luminosity is defined as the ratio between the number of events produced per unit of time over the cross section of a given process. Analogously, the integrated luminosity is the integral of the instantaneous luminosity over time and it corresponds to the ratio between the total number of events over the cross section.

$$\mathcal{L}(t) = \frac{1}{\sigma}\frac{dN}{dt} \qquad\qquad \mathcal{L}_{\text{int}} = \int \mathcal{L}(t)\ dt. \tag{3.1}$$

These quantities are used to characterize the performance of a particle accelerator. In particular, the maximization of the integrated luminosity in a particle physics experiment correspond to a larger collection of data to be analyzed.

The instantaneous luminosity depends only on the beam parameters and can be written for a Gaussian beam distribution as [89]:

$$\mathcal{L} = \frac{f_{\text{rev}} N_p^2 n_b}{4\pi\sigma_x^*\sigma_y^*}F = \frac{\gamma f_{\text{rev}} N_p^2 n_b}{4\pi\sqrt{\epsilon_{nx}\beta_x^*\epsilon_{ny}\beta_y^*}}F, \tag{3.2}$$

where γ is the Lorentz factor, f_{rev} is the revolution frequency, n_b is the number of bunches in the ring, N_p is the number of protons per bunch, σ_x^* and σ_y^* characterize the transverse beam size and F is a factor of order 1, accounting for geometric beam overlap effects. The second part of Eq. 3.2 recasts the transverse beam sizes $\sigma_{x,y}^*$ in terms of the normalized transverse emittance $\epsilon_{nx,y}$ and the amplitude beta-functions $\beta_{x,y}^*$ at the interaction point (IP) as ${\sigma_{x,y}^*}^2 = \beta_{x,y}^*\epsilon_{nx,y}/\gamma$. The relevant parameters of the machine and their values are summarized in Tab 3.1 for the LHC design configuration as well as for the configuration achieved in 2018.

Parameter	**2018**	**Nominal**
Energy [TeV]	6.5	7.0
Lorentz factor (γ)	6928	7461
Revolution frequency (f_{rev}) [kHz]	11.245	11.245
Beta-function (β^*) [cm]	$30 \to 25$	55
Half crossing angle [μrad]	$160 \to 130$	142.5
n_b (number of bunches)	2556	2808
N_p (protons/bunch)	$1.1 \cdot 10^{11}$	$1.15 \cdot 10^{11}$
Normalized emittance ϵ_n [μm]	1.9	3.75
Max peak luminosity [cm^2 s^{-1}]	$2.1 \cdot 10^{34}$	$1.0 \cdot 10^{34}$

Table 3.1: Beam and machine parameters in the configuration of the LHC design report [23] and in the achieved 2018 configuration [69]. The arrow indicates a change during the data taking.

3.2 The Compact Muon Solenoid

The Compact Muon Solenoid (CMS) detector [27, 30] is a multi-purpose apparatus operating at the LHC at CERN comprising different subdetectors dedicated to the reconstruction and identification of the particles produced in the *pp* collisions. These subdetectors are organized in a cylindrical onion-like shape around the beam pipe and are geometrically divided into a central "barrel" and two "endcaps" on either side. A sketch of the CMS detector with signatures of various particles passing through it is shown in Fig. 3.2. From the inside out, the inner tracking system including pixel and silicon strip detectors surrounds the beam pipe. The second innermost subdetector is the the electromagnetic calorimeter (ECAL), followed by the hadronic calorimeter (HCAL). All these subdetectors are surrounded by the solenoid magnet. Outside the magnet there are the muon system and the steel flux return yoke.

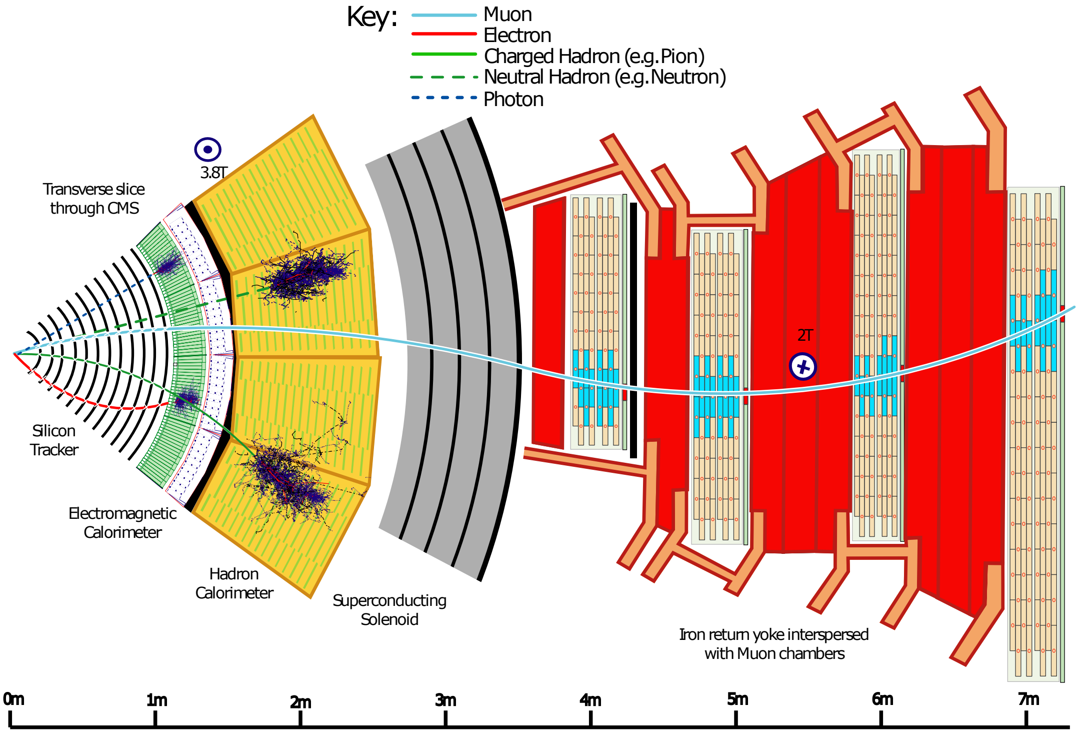


Figure 3.2: A sketch of the different particle tracks in a transverse slice of the CMS detector, from the beam interaction region to the muon detector. Figure from Ref. [61].

3.2.1 CMS coordinate system

The coordinate system adopted by CMS has the origin centred at the nominal collision point inside the experiment, the y-axis pointing vertically upward, and the x-axis pointing radially inward toward the center of the LHC ring. The z-axis points along the direction of flight of the proton beam moving in the counter-clockwise direction. The x and y components of a vector $\vec{p}$ determine its transverse magnitude p_{T}:

$$p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2} \tag{3.3}$$

The polar angle θ is measured from the z-axis, and the azimiuthal one φ is measured from the x-axis in the (x, y) plane. Alternative more common coordinates used in particle physics for the polar angle are the rapidity y and the pseudorapidity η defined by:

$$y = \frac{1}{2}\ln\left(\frac{E + p_z}{E - p_z}\right) \qquad \eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right) = \frac{1}{2}\ln\left(\frac{|\vec{p}| + p_z}{|\vec{p}| - p_z}\right), \tag{3.4}$$

where E is the energy of the particle, $|\vec{p}|$ is the magnitude of the three-momentum $\vec{p} = (p_x, p_y, p_z)$ and p_x, p_y and p_z are the projections along the corresponding axes. Rapidity and pseudorapidity differences ($\Delta y = y_1 - y_2$, $\Delta\eta = \eta_1 - \eta_2$) are invariant under Lorentz transformations along the beam axis. The distance between two points in the (η, φ) plane is usually called angular distance and it is given by:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\varphi)^2}, \tag{3.5}$$

which is also invariant under boost along the beam axis. The transformation between the set of coordinates (p_x, p_y, p_z) of a vector $\vec{p}$ into the equivalent set $(p_{\mathrm{T}}, \eta, \varphi)$ is given by:

$$\begin{aligned} p_x &= p_{\mathrm{T}} \cos(\varphi) \\ p_y &= p_{\mathrm{T}} \sin(\varphi) \\ p_z &= p_{\mathrm{T}} \sinh(\eta) \end{aligned} \tag{3.6}$$

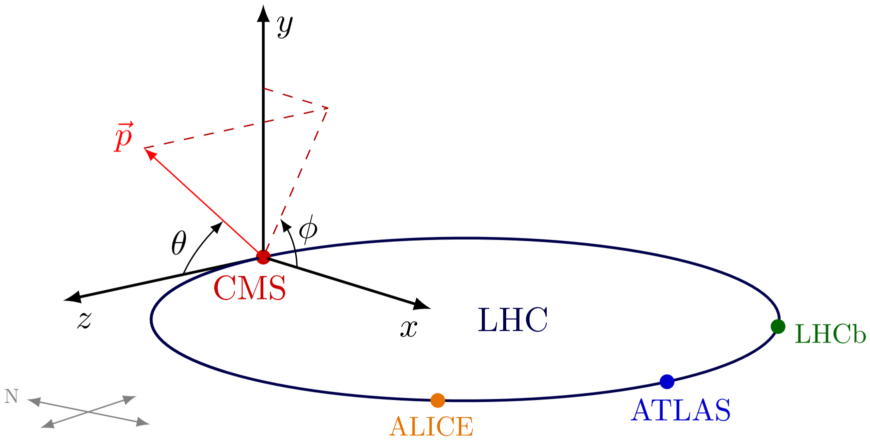


Figure 3.3: CMS coordinate system. Figure from Ref. [94].

3.2.2 The solenoid magnet

A solenoid magnet providing a magnetic field of up to 4 T, limited to 3.8 T during data-taking, is used to bend the trajectories of charged particles as they travel outwards from the collision point. This is done in order to measure the momentum of the charged particles from the position measurements of the tracker, since high-p_{T} particles bend less compared to low-p_{T} particles and to identify the sign of the charge of the particle, since positively and negatively charged particles bend in opposite directions in the same magnetic field.

The magnet is 12.5 m long, it has a diameter of 6.3 m and a weight of 220 t. It is formed by a cylindrical coil of superconducting fibres. The coil is made of Niobium-titanium (NbTi) and is cooled down to 4.8 K to operate in a superconducting regime. The high magnetic field is confined to the volume of the detector by a steel "return yoke". Four muon stations are interspersed with the yoke in both the barrel and endcaps and the field strength in this region outside the solenoid is 2 T with opposite sign.

3.2.3 The tracking system

The CMS tracker is the innermost subdetector of the CMS experiment and it is entirely based on silicon detector technology. It consists of a cylindrical volume of 5.8 m in length and 2.5 m in diameter, covering the pseudorapidity range $|\eta| < 2.5$, subdivided into two main components, a silicon pixel and a silicon strip detector, both designed to provide precise measurements of charged-particles momenta and to identify primary and secondary vertices [30, 43]. Moreover, the information from the tracker is also used for the identification of leptons and hadrons as well as in the High Level Trigger (HLT) to reduce the amount of collected events to be stored. A sketch of the whole CMS tracker is shown in Fig. 3.4.

The pixel detector originally consisted of three barrel layers (BPIX) and of two endcap disks (FPIX) on each end. After an update at the end of 2016 data taking period, the silicon pixel detector has been replaced by a new system, referred to as the CMS Phase-1 pixel detector [80]. The upgraded system consists of four concentric barrel layers (L1-L4) at radii of 29, 68, 109 and 160 mm respectively, and three endcap disks (D1-D3) at either side at distances of 291, 396 and 516 mm from the center of the detector. It provides 3-dimensional space points in the region closest to the interaction point and it has four-hit coverage over the pseudorapidity range $|\eta| < 2.5$. Each of the 124 million pixels has a size of $100 \times 150\ \mu\text{m}^2$, providing good granularity in any direction.

The silicon strip detector [33] includes 9.6 million strips covering an active area of about 198 m^2. It is composed of four subsystems, namely the Tracker Inner Barrel (TIB), Tracker Outer Barrel (TOB), the Tracker Inner Disks (TID) and the Tracker EndCaps (TEC). The TIB and the TOB are composed of 4 and 6 layers, respectively, while the TID and the TEC include 3 and 9 disks. To provide a bi-dimensional measurement, some layers in the barrel and in the endcaps have stereo modules with two silicon strip modules mounted back-to-back and rotated by 100 mrad with respect to each other.

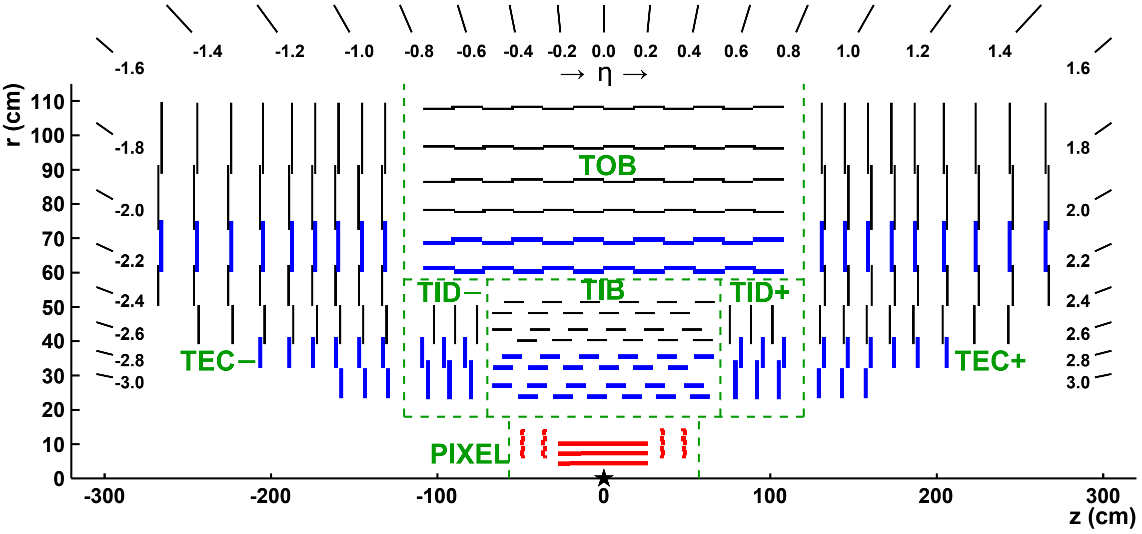


Figure 3.4: Cross section of the CMS tracker in the r-z plane. The centre of the tracker is indicated by a star. The pixel modules are shown by the red lines. Single-sided strip modules are shown by thin black lines, while double-sided strip detectors are shown by thick, blue lines. Figure from Ref. [43].

3.2.4 The electromagnetic calorimeter

The electromagnetic calorimeter (ECAL) [42, 48] is a hermetic, fine grained, homogeneous calorimeter, containing 75848 lead tungstate ($PbWO_4$) scintillating crystals. It surrounds the tracker and it is divided into a barrel ECAL (EB), with an inner radius of 129 cm covering the pseudorapidity range $|\eta| < 1.479$, closed at each end by an endcap ECAL (EE), at a distance of 314 cm from the vertex and covering the pseudorapidity range $1.479 < |\eta| < 3.0$. The main purpose of the ECAL is the measurement of the energy of the electromagnetic showers developed by the passage of electrons and photons. The discrimination between e/γ is possible exploiting the combined information of the tracker and the ECAL. The choice of $PbWO_4$ allows to have extremely dense crystals with small radiation length ($X_0 = 0.89$ cm) and Molière radius ($R_M = 2.2$ cm), good resistance to radiation and fast response time (80% of the light is emitted within 25 ns). However, the light yield is relatively low (30 γ/MeV), requiring the use of photodetectors with intrinsic gain that can operate in a magnetic field. Silicon avalanche photodiodes (APDs) are used as photodetectors in the barrel and vacuum phototriodes (VPTs) in the endcaps. In front of most of the fiducial region of each endcap ($1.635 < |\eta| < 3.0$) there is a preshower device (ES), whose aim is to discriminate between signal from single γ and neutral pions ($\pi^0 \to \gamma\gamma$). A transverse section through ECAL is shown in Fig. 3.5.

The EB is composed of 36 identical supermodules, each covering half the barrel length in η and 20° in φ. The EE consists of identically shaped crystals, grouped in mechanical units of 5 × 5 crystals (supercrystals). Special partial crystals are used on the inner and outer circumferences of the EE. The EB (EE) crystals have a length of 230 mm (220 mm) corresponding to 25.8 X_0 (24.7 X_0) and a front face cross-section of 22×22 mm^2 (28.6×28.6 mm^2). The performance of the the calorimeter has been extensively tested with electron beams. The energy resolution has been parameterized as a function of energy as:

$$\frac{\sigma_E}{E} = \frac{2.8\%}{\sqrt{E\ [\mathrm{GeV}]}} \oplus \frac{12\%}{E\ [\mathrm{GeV}]} \oplus 0.3\% \tag{3.7}$$

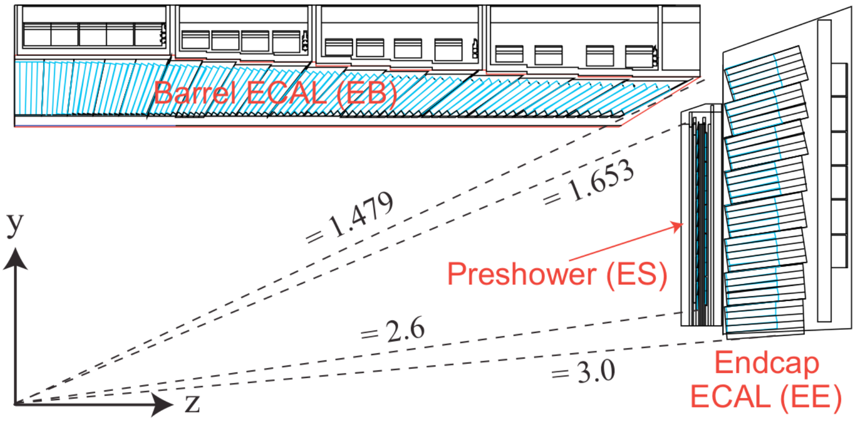


Figure 3.5: Cross section of the CMS ECAL in the y-z plane. Figure from Ref. [42].

3.2.5 The hadronic calorimeter

The hadronic calorimeter (HCAL) is a sampling calorimeter, built of alternating layers of brass absorbers and plastic scintillator tiles read out with embedded wavelength-shifting (WLS) fibres. It is designed to measure the energy of charged and neutral hadrons, providing good containment and hermeticity for the p_T^{miss} measurement. It consists of a hadron barrel (HB) and a hadron endcap (HE), both surrounding the ECAL and being located inside the volume of the magnet, and an additional hadron outer (HO) detector, lining the outside of the coil inside the barrel muon system and a hadron forward (HF) calorimeter. A schematic view of the HCAL is shown in Fig. 3.6.

The HB consists of 2304 towers with a segmentation $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$ covering the pseudorapidity region $|\eta| < 1.4$, while the HE consists of 14 η towers with 5° φ segmentation covering the pseudorapidity region $1.3 < |\eta| < 3.0$. The HO follows the HCAL barrel tower geometry in η and φ covering the pseudorapidity range $|\eta| < 1.26$. It is divided into 5 rings along η, each 2.5 m long in z. The very forward region ($3.0 < \eta < 5.0$) is covered by the steel/quartz fibre HF calorimeter. The front face is located at $z = 11.2$ m and the depth of the absorber is 1.65 m.

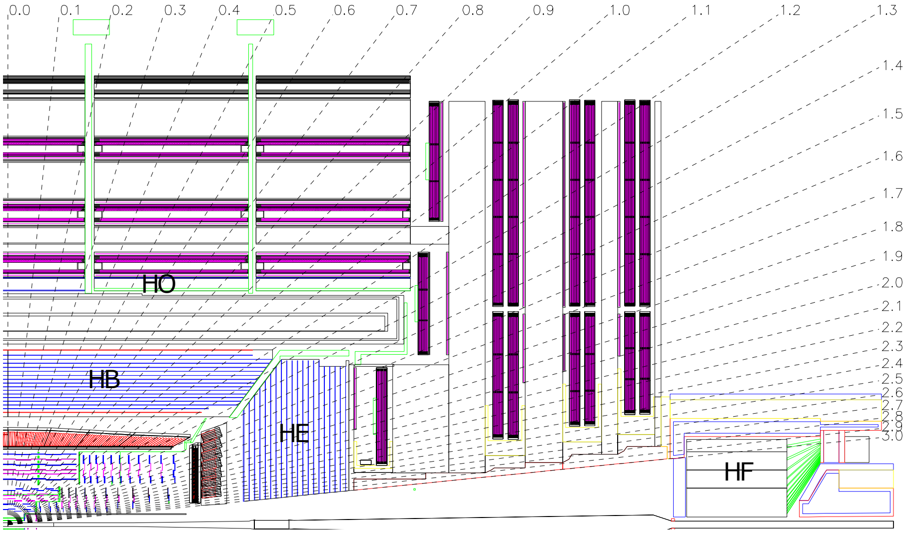


Figure 3.6: Longitudinal view of the CMS detector showing the locations of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters. Figure from Ref. [30].

3.2.6 The muon system

The outermost subdetector of the CMS experiment is the muon system, aiming at the identification and measurement of muons besides playing a relevant role in the Level-1 trigger. It consists of three different types of gaseous detectors, namely drift tube (DT) chambers , cathode strip chambers (CSC) and resistive plate chambers (RPC). A sketch of a quadrant of the CMS detector with a particular focus on the muon system is shown in Fig. 3.7. In the barrel region ($|\eta| < 1.2$) four layers of DTs are used, while in the two endcaps CSCs are deployed and cover the region $0.9 < |\eta| < 2.4$. In addition, RPCs are used in both the barrel and the endcap regions covering a pseudorapidity region of $|\eta| < 1.6$ and providing a fast response with good time resolution but with a coarser position resolution than the DTs or CSCs.

The muon reconstruction and identification, and isolation efficiency is found to be better than 95%, with a misidentification rate less than 1% [67]. For muons with $p_{\mathrm{T}} > 200$ GeV, combining information from the muon system with information from the inner tracker significantly improves the momentum measurement, while for muons with lower p_{T} the momentum measurement is dominated by the performance of the inner tracker.

3.2.7 Trigger

The high proton bunch crossing rate of 25 ns and the peak value of instantaneous luminosity reached in 2018 of $\mathcal{L} = 2.1 \cdot 10^{34}\ \mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ lead to approximately 10^9 interactions per second. However, the total *pp* cross section is dominated by soft QCD processes, which are not of primary interest for the physics program of the CMS experiment. Moreover, the impossibility of recording all the events due to a limited bandwidth makes mandatory the use of a trigger system to select events according to physics-driven choices [68]. The CMS trigger system consists of two consecutive stages: the hardware based Level-1 (L1) trigger system and the software based high level trigger (HLT) system. The L1 trigger systems relies on the information from the calorimeters and the muon system to select interesting events to be transferred to the HLT and to reduce the the event rate from 40 MHz to 100 KHz. Next, the HLT makes use of the information from all the subdetectors of CMS to further reduce the event rate to approximately 1 KHz. Based on the event kinematics and the trigger objects, events are

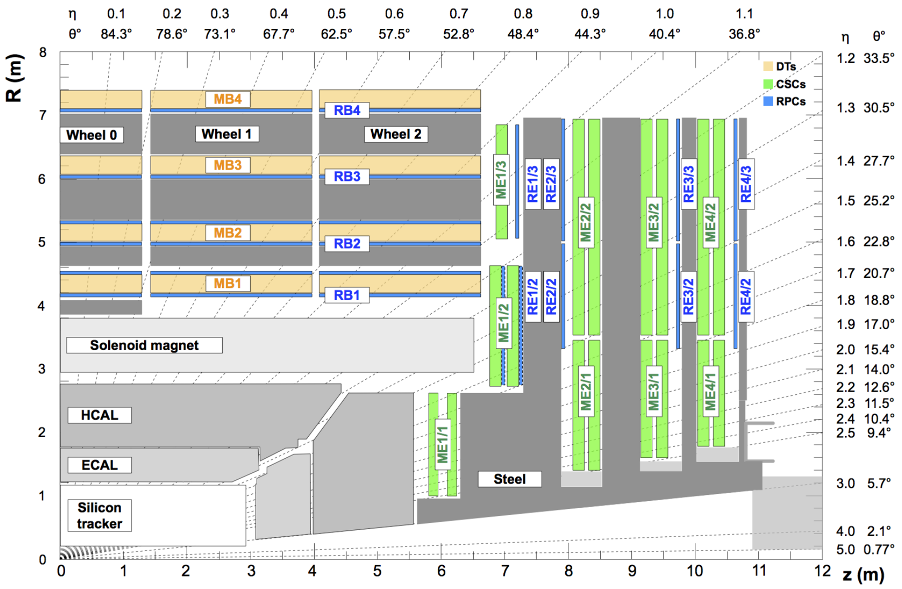


Figure 3.7: An r-z cross section of a quadrant of the CMS detector. The locations of the various muon stations and the steel flux-return disks (dark areas) are shown. The DTs are labeled as MB ("Muon Barrel") and the CSCs are labeled as ME ("Muon Endcap"). RPCs are mounted in both the barrel and endcaps of CMS, where they are labeled as RB and RE, respectively. Figure from Ref. [67].

divided into different subsets, called primary data sets. The set of selection criteria required for an event to be accepted in a primary data set is called trigger path.

4

Object Reconstruction and Event Selection

In this chapter, various aspects fundamental to the purpose of this thesis are covered, including an outline of the physics objects reconstruction used in the CMS experiment, the simulation of the signal and background processes and the event selection. The chapter is structured as follows. In Sec. 4.1 the particle-flow (PF) algorithm, responsible for the global event reconstruction in CMS, is introduced and in Sec. 4.2 the main physics objects are presented. In Sec. 4.3 some details about the data samples used for the analysis and the event simulation are provided. In Sec. 4.4 the event selection is explained and in Sec. 4.5 two analytical approaches for the kinematic reconstruction of the $t\bar{t}$ system are presented.

4.1 The particle-flow algorithm

According to the type of interaction, different particles leave different signs in the detectors, which are converted into electronic signals to be processed. The aim of the PF algorithm [61] is to reconstruct the four-momentum of the stable particles from these electronic signals. The term stable particle refers to all those particles with a lifetime long enough to let them reach the detectors and includes muons, electrons, charged hadrons, neutral hadrons, and photons. Moreover the PF algorithm reconstructs higher level objects like jets and missing transverse energy (MET).

The first step of the reconstruction process consists in the conversion of electronic signals caused by the passage of charged particles through the tracker and muon subdetectors into well-defined locations, generically called "hits" and it is referred to as "local reconstruction" [44, 67]. Different algorithms are used according to the detector technology. In the CSC and DT chambers, which are multi-layer detectors, straight-line track "segments" are built from the reconstructed hits.

4.1.1 Iterative tracking

The track reconstruction is the process of using the hits information from the local reconstruction to obtain the trajectories of the charged particles responsible for the hits [61]. As a first step, tracks are reconstructed independently in the inner tracker and in the muon system [67]. However, while the calorimeters guarantee a high purity environment in the muon chambers by absorbing the majority of the other particles, the multiplicity of charged particles crossing the tracker is much higher. Given an average number of 32 inelastic interactions per bunch crossing in 2018 data taking (Fig. 4.1), the reconstruction of tracks is a computationally challenging task and it is performed iteratively using a combinatorial track finder based on Kalman Filtering (KF) [26], in a way such that the first iterations use tight criteria on the hits, which provide a moderate tracking efficiency but negligible fake rate, in order to search for tracks easy to find (e.g. high-p_{T} tracks). After each iteration, the hits associated with tracks are removed to reduce the combinatorial complexity and to simplify subsequent iterations aimed at searching for more difficult classes of tracks (e.g. low-p_{T} or displaced tracks). Additional quality criteria are required for tracks to be kept for further analysis.

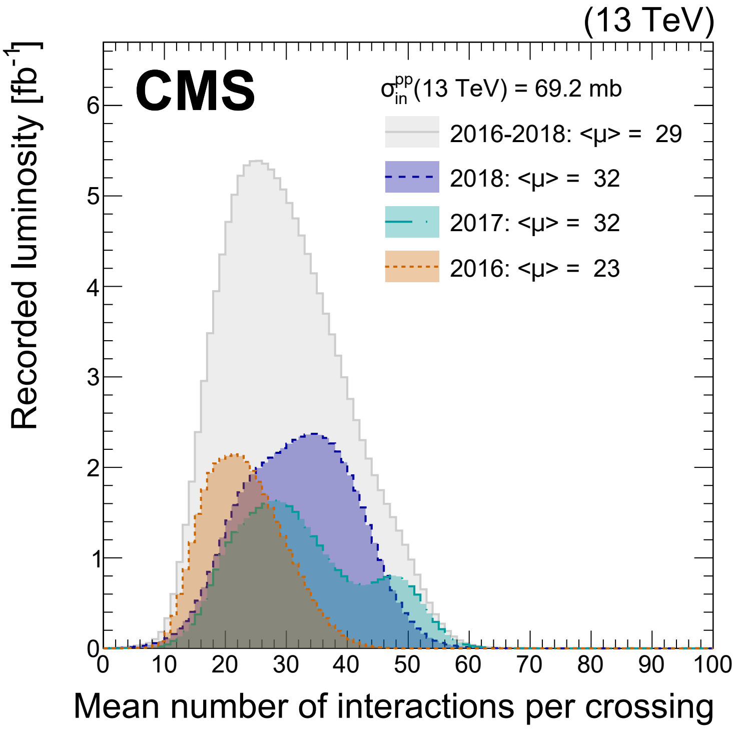


Figure 4.1: Distribution of the mean number of inelastic interactions per crossing (pileup) in data for pp collisions in 2016 (orange), 2017 (teal), 2018 (blue), and integrated over 2016–2018 (grey). A total inelastic pp collision cross section of 69.2 mb is chosen. The mean number of inelastic interactions per bunch crossing is provided in the legend for each year. Figure from Ref. [75].

4.1.2 Calorimeter clustering

The calorimeter clustering [61] is a fundamental step for the determination of the energy and the direction of neutral particles, the separation between neutral and charged particles energy deposits, the identification and reconstruction of electrons and all accompanying bremsstrahlung photons and the improvement of the energy measurements of charged hadrons. The clustering is performed separately in each subdetector: ECAL barrel and endcaps, HCAL barrel and endcaps, and the two preshower layers. Firstly, all the cells with a deposit above threshold and larger than the energy of the neighbouring cells are identified as “cluster seeds”. Secondly, “topological clusters” are built from the seeds by aggregating cells with at least one corner in common with a cell already in the cluster. The aggregated cells are required to have an energy in excess of a cell threshold, whose value depends on the specific detector. Finally, a dedicated algorithm is used to reconstruct the different clusters within a topological cluster, returning their positions and energies.

4.1.3 Link algorithm

In the last step of the reconstruction process, links between signals from the tracker and signals from calorimeters and muon chambers are established. The collected signals from the subdetectors are linked consistently to provide a full reconstruction of the particle and get rid of any possible double counting. For example, links between inner tracks and calorimeter clusters are established by extrapolating the track from its last measured hit in the tracker to the calorimeters at a depth depending on the subdetector (PS, ECAL and HCAL). The inner track is linked to a cluster if its extrapolated position is within the cluster area. If several clusters are linked to the same track or vice versa, only the link with the smallest distance in the (η, φ) plane is kept. Analogously, links between ECAL and HCAL clusters, preshower and ECAL clusters or tracker and muon tracks are established. Elements connected by a link are referred to form a PF block. As a relevant example for the purposes of this thesis, the combined information of tracker and ECAL (muon chambers) allows the reconstruction of electrons (muons).

After all links have been established between signals of different subdetectors, the particle identification is performed in steps. First, muons are identified due to high reconstruction efficiency and low misidentification rate and the corresponding PF elements are removed from the PF block. Then, electrons and prompt photons are identified. Finally, neutral, charged hadrons and non-prompt photons (e.g. originating from meson decays) are identified.

4.2 Physics objects

Primary vertex

Several *pp* collisions take place per bunch crossing. Among all the possible vertices, the primary vertex (PV) of an event is defined as the vertex with the largest sum of associated transverse momenta squared and is required to be reconstructed within $|z| < 24$ cm and $r < 2$ cm in the x–y plane. Events without a PV are rejected. Once the best PV in the event is selected, the rest of the vertices passing some quality criteria but without the largest sum of associated transverse momenta squared are treated as pileup (PU) vertices and the tracks associated to them are removed from the analysis.

Muons

Depending on the approach taken to combine the track information, three types of muon candidates can be defined [67].

- **Standalone muons** are reconstructed using muon chambers information only with a Kalman filter [26]. The track reconstruction makes use of seeds made up of groups of DT or CSC segments.
- **Tracker muons** are built "inside-out" and correspond to tracks propagated from the tracker to the muon system, matching at least one muon segment in the DT or CSC.
- **Global muons** are built "outside-in" by backward propagating standalone muon tracks to the tracker layers. If the standalone muon track matches a tracker track, a combined fit is performed.

Muon are reconstructed in the pseudorapidity range $|\eta| < 2.4$ and are required to pass *tight ID* criteria, aimed at suppressing muons from decay in flight and hadron shower remnants reaching the innermost muon station (punch-through) [67]. Additionally, muons are required to fulfill an isolation criterion: $I_{\mathrm{rel}} < 0.15$, where I_{rel} is the ratio of the sum of transverse energy deposits from charged and neutral hadrons and photons inside a cone of $\Delta R < 0.4$ around the muon in the (η, φ) space to the p_{T} of the muon.

Electrons

Differently from muons, electrons lose energy primarily through bremsstrahlung rather than ionization, due to the significant inner tracker thickness. As a consequence, a dedicated track reconstruction is performed [81]. Energetic ECAL "superclusters" (SCs) are built out of ECAL clusters to account for bremsstrahlung photons emitted due to the bent trajectory under the magnetic field. The energy and position of the SCs are used to infer the position of the hits expected in the tracker layers under the assumptions that the cluster is produced either by an electron or by a positron. For this ECAL-driven tracking approach, a further isolation criterion is required to exclude electrons in jets for which the SCs energy and position are often biased by the overlapping contributions from other particle deposits. Moreover, since photons emitted by low-p_{T} electrons are not covered by one SC, a seeding method similar to the iterative tracking is used, modified to account for bremsstrahlung emissions.

Like muons, electrons are measured if they are within the pseudorapidity range $|\eta| < 2.4$. Moreover, the gap between the barrel and endcap region $(1.44 < |\eta| < 1.57)$ of the ECAL is excluded. Similarly to muons, a relative isolation parameter (I_{rel}) is calculated for each electron candidate. In this case all neutral and charged hadron and photon PF candidates within a cone of $\Delta R = 0.3$ in the (η, φ)

space are considered. Depending on the electron p_{T}, the maximum allowed value for I_{rel} varies from 0.05 to 0.1.

Jets

As explained in Sec. 2.1.4, the quarks and gluons produced during a collision give rise to a large amount of hadrons due to colour confinement. Because of the conservation of momentum, the hadrons are detected as a set of particles, referred to as jets, which are collinear to the original parton.

The clustering of PF candidates into jets in this thesis is done using the infrared- and collinear-safe (i.e. stable with respect to additional radiation of soft partons and insensitive to possible parton splitting into collinear partons) "anti-k$_T$" algorithm [29]. Starting from the left tracks, for each particle i in the event, the distance d_{ij} to the particle j and the distance to the beam d_{iB} are defined as:

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

The parameter R is set to $R = 0.4$ for this analysis. Firstly, the algorithm identifies the smallest distance $\hat{d}_{ij}$ among all the d_{ij}. If $\hat{d}_{ij} < d_{iB}$, the objects i and j are clustered into the same jet, otherwise i is defined as an individual jet and is removed from the list of entities. The algorithm proceeds iteratively until no particles are left, and all jets are identified. Finally, the jet momentum is defined as the vectorial sum of all particle four-momenta in the jet.

Missing transverse momentum

The missing transverse momentum $\vec{p}_T^{\,\mathrm{miss}}$ is the only information available about particles which escape the subdetectors like neutrinos or potential BSM particles. The reconstruction of the $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ is possible under the assumption that the initial state partons have a negligible p_{T}. By exploiting the hermeticity of the CMS detector, the $\vec{p}_T^{\,\mathrm{miss}}$ can be measured as the negative vector sum of the p_{T} of all the PF candidates in an event and its magnitude is denoted as $p_{\mathrm{T}}^{\mathrm{miss}}$ and it is also referred to as missing transverse energy (MET):

$$p_{\mathrm{T}}^{\mathrm{miss}} = \left|\vec{p}_T^{\,\mathrm{miss}}\right| = \left|-\sum_{i=1}^{N_{\mathrm{PF}}} \vec{p}_{T,i}\right| \tag{4.2}$$

Heavy-flavor jet

The identification of jets originating from heavy-flavor quarks (b and c) rather than gluons or light-flavor quarks (u, d, s) can be used to suppress some sources of background and enhance the signal purity of the $t\bar{t}$ events, where one expects the presence of two b jets from the decays of the $t\bar{t}$ system. The tagging of heavy-flavor jets is done by exploiting the peculiar features of this kind of jets. In particular, hadrons containing b quarks are characterized by a large mass and a lifetime ($\tau = 1.5$ ps) long enough such that the distance travelled by the hadron before decaying can be experimentally measured. The decay vertex of a b hadron is identified as a secondary vertex (SV), reconstructed from charged tracks stemming from its decay. By measuring the direction of flight of a particle with a displaced track, an impact parameter (IP) can be defined as the distance of closest approach between the extrapolated track and the PV. The sign of the impact parameter is defined as positive if the angle between the IP vector and the jet axis is smaller than $< \pi/2$, and 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. Information from tracks and SVs are used as input features for a tagging algorithm, called DeepCSV [66], based on NNs to discriminate which flavor category each jet belongs to.

4.3 Data samples and event simulation

In this analysis pp collision data recorded by the CMS experiment in 2018, corresponding to an integrated luminosity of 59.8 fb^{-1} are used. At HLT, events are required to pass either single-lepton or dilepton trigger criteria. The former is included in order to recover interesting dilepton events, which failed to pass the dilepton trigger requirements. In 2018 the requirements for the single lepton and dilepton triggers are the following:

- one electron with $p_{\mathrm{T}} > 32$ GeV
- one muon with $p_{\mathrm{T}} > 27$ GeV
- one electron with $p_{\mathrm{T}} > 23$ GeV and one muon with $p_{\mathrm{T}} > 8$ GeV
- one muon with $p_{\mathrm{T}} > 23$ GeV and one electron with $p_{\mathrm{T}} > 12$ GeV

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

4.3.1 Monte Carlo simulation

All the $t\bar{t}$ processes, including both the ones contributing to the signal and to the background (i.e. dileptonic same-flavor channel, semileptonic, fully hadronic decays) are simulated using POWHEG (version 2) [28] MC event generator at NLO in QCD for the matrix element calculations. In the simulation the top quark mass is set to $m_t^{\mathrm{MC}} = 172.5$ GeV and the proton structure is described using the NNPDF 3.1 PDF at next-to-NLO (NNLO) [34]. The parton showering, hadronization and underlying events are performed with PYTHIA (version 8.230), with the CP5 tune [73]. The h^{damp} parameter in POWHEG, which regulates the matching scale to the parton shower, is set to h^{damp} $= 1.379\ m_t^{\mathrm{MC}}$. The renormalization and factorization scales are both set to the transverse mass of the top quark in the $t\bar{t}$ rest frame: $\mu_{\mathrm{R}} = \mu_{\mathrm{F}} = \sqrt{(m_t^{\mathrm{MC}})^2 + p_{\mathrm{T}}^2}$ [78]. For the simulation of the other background sources, either POWHEG, MG5_AMC@NLO or PYTHIA are used for the ME calculation according to the process, while the parton showering, hadronization, and multiparton interactions are always modelled with PYTHIA. The tW process is simulated using POWHEG at NLO. The Z+jets process is generated using MG5_AMC@NLO at leading order (LO) with up to four additional partons at matrix element (ME) level. W+jets events are simulated using MG5_AMC@NLO at LO with up to four additional partons at ME level. Diboson production is simulated at LO using PYTHIA. The complete list of background processes with the corresponding MC event generator, predicted cross section and order of the prediction is shown in Tab. 4.1. Events generated by POWHEG and MG5_AMC@NLO have non unitary weights which are always taken into account in the analysis and referred to as generator weights in the following.

The comparison between real data and MC simulation is possible after properly normalizing the latter processes to their theoretical cross sections 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) [45] program at NNLO, including resummation of NNLL soft-gluon terms, and assuming $m_t = 172.5$ GeV. The NNPDF3.1 PDF set at NNLO is used, and the resulting cross section value is $830.91^{+20.39}_{-29.96}$ (scale) $^{+3.92}_{-3.83}$ (PDF + α_S) pb, where “scale” refers to the uncertainty assessed by varying the renormalization and factorization scales [91]. All other cross sections are calculated at approximate NNLO for tW production, NLO for Z+jets, NNLO for W+jets, and NLO for diboson production. Exemplary Feynman diagrams for the main background sources are shown in Fig. 4.2

For all simulated samples, the interaction of the final state particles with the CMS detector volume is simulated with the GEANT4 [22] program. To model the effect of pileup, additional minimum bias interactions are added to the simulated events.

4.3.2 Corrections to simulated events

A series of correction factors, usually referred to as Scale Factors (SFs), are used in order to reduce the discrepancy between simulated MC samples and real data. These factors are taken into account

Process	σ **[pb]**	**MC Generator**	**Accuracy**
$t\bar{t}$ (dileptonic)	830.91×0.10706	POWHEG	NLO
$t\bar{t}$ (semileptonic)	830.91×0.44113	POWHEG	NLO
$t\bar{t}$ (hadronic)	830.91×0.45441	POWHEG	NLO
single-t s-channel (t+$\bar{t}$)	10.32	MG5 AMC@NLO	NLO
single-t tW	35.85×0.54559	POWHEG	NLO
single-t t-channel	136.02	POWHEG	NLO
single-$\bar{t}$ t-channel	80.95	POWHEG	NLO
Drell-Yan $10 < m_{\ell\bar{\ell}} < 50$ GeV	22635.1	MG5 AMC@NLO	LO
Drell-Yan $m_{\ell\bar{\ell}} \geq 50$ GeV	6225.4	MG5 AMC@NLO	LO
$W \rightarrow \ell\nu$	61526.7	MG5 AMC@NLO	LO
WW	118.7	PYTHIA	LO
WZ	47.13	PYTHIA	LO
ZZ	16.523	PYTHIA	LO
$t\bar{t}W$ (leptonic)	0.2043	MG5 AMC@NLO	LO
$t\bar{t}W$ (hadronic)	0.4062	MG5 AMC@NLO	LO
$t\bar{t}Z$ (leptonic)	0.2529	MG5 AMC@NLO	LO
$t\bar{t}Z$ (hadronic)	0.5297	MG5 AMC@NLO	LO

Table 4.1: List of simulated samples used for signal and background processes, together with the cross section assigned to the process, the MC event generator and the order of the computation.

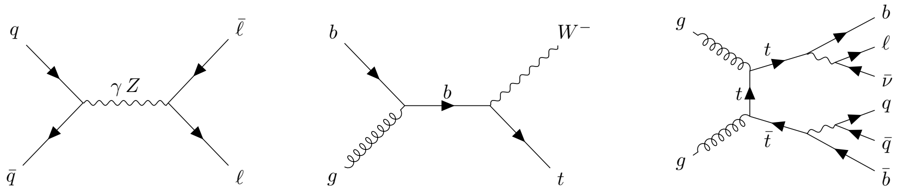


Figure 4.2: Examples of Feynman diagrams for background contributions. From left to right, Drell-Yan process, single top production and $t\bar{t}$ production in a different final state than the considered one.

multiplicatively in the weight of each event and are derived as ratio of efficiency in data to efficiency in simulation (SF $= \epsilon_{\mathrm{data}}/\epsilon_{\mathrm{MC}}$). These corrections include:

- **Lepton SF** accounts for differences in efficiencies for reconstruction, identification and isolation of leptons. This SF is measured using the "tag-and-probe" [67, 83] method starting from Drell-Yan event samples $Z \rightarrow \ell\bar{\ell}$. This method consists in selecting events that contain a dilepton pair with a mass near the Z boson resonance. One lepton is required to satisfy tight selection criteria and the other is used as "probe" lepton. The efficiency is then determined from the number of probe leptons that also pass the tight selection criteria. The SF is computed as a function of the p_{T} and η of the lepton.

- **B-tag SF** corrects the b-tagging efficiency and mistagging rate for c- and light-jets of the simulated sample. It is derived from muon-jet samples as the ratio between the probability to identify all b jets in an event in data and the probability to identify all b jets in a simulated event.

- **Pileup SF** is introduced to reproduce the pileup distribution found in data and it is derived from the bunch-to-bunch instantaneous luminosity and assuming a total inelastic pp cross section of 69.2 mb [75].

- **Prefiring Muons** aims at correcting “trigger prefiring”, an effect which takes place when muon are mistakenly assigned to an earlier or later LHC bunch crossing (BX) [82]. This will lead to the wrong BX being accepted by the L1 trigger system.
- **Trigger SF** accounts for differences in the trigger efficiencies which are measured as a function of the lepton p_{T} using triggers that are only weakly correlated to the dilepton triggers (e.g. $p_{\mathrm{T}}^{\mathrm{miss}}$-based triggers).

4.4 Event selection

The event selection is a set of kinematic requirements applied to the dataset in order to select those events that correspond to the signal that is going to be measured and reject any other physics process. In this thesis, the $t\bar{t}$ production is studied in the $e\mu$ final state, including leptons originating from τ decays. The final state of interest can arise from eight different cases summarized into Tab. 4.2. An example of one Feynman diagram for the $t\bar{t}$ production decaying in the final state of interest is shown in Fig. 1.1. The total branching ratio of the signal is BR($t\bar{t} \to e\mu$) = 3.2% [89].

Top quark	**Antitop quark**
$t \to W^+b \to e^+\nu_e b$	$\bar{t} \to W^-\bar{b} \to \mu^-\bar{\nu}_\mu \bar{b}$
$t \to W^+b \to \mu^+\nu_\mu b$	$\bar{t} \to W^-\bar{b} \to e^-\bar{\nu}_e \bar{b}$
$t \to W^+b \to \tau^+\nu_\tau b \to e^+\nu_e\bar{\nu}_\tau\nu_\tau b$	$\bar{t} \to W^-\bar{b} \to \mu^-\bar{\nu}_\mu \bar{b}$
$t \to W^+b \to \mu^+\nu_\mu b$	$\bar{t} \to W^-\bar{b} \to \tau^-\bar{\nu}_\tau\bar{b} \to e^-\bar{\nu}_e\nu_\tau\bar{\nu}_\tau\bar{b}$
$t \to W^+b \to \tau^+\nu_\tau b \to \mu^+\nu_\mu\bar{\nu}_\tau\nu_\tau b$	$\bar{t} \to W^-\bar{b} \to e^-\bar{\nu}_e \bar{b}$
$t \to W^+b \to e^+\nu_e b$	$\bar{t} \to W^-\bar{b} \to \tau^-\bar{\nu}_\tau\bar{b} \to \mu^-\bar{\nu}_\mu\nu_\tau\bar{\nu}_\tau\bar{b}$
$t \to W^+b \to \tau^+\nu_\tau b \to e^+\nu_e\bar{\nu}_\tau\nu_\tau b$	$\bar{t} \to W^-\bar{b} \to \tau^-\bar{\nu}_\tau\bar{b} \to \mu^-\bar{\nu}_\mu\nu_\tau\bar{\nu}_\tau\bar{b}$
$t \to W^+b \to \tau^+\nu_\tau b \to \mu^+\nu_\mu\bar{\nu}_\tau\nu_\tau b$	$\bar{t} \to W^-\bar{b} \to \tau^-\bar{\nu}_\tau\bar{b} \to e^-\bar{\nu}_e\nu_\tau\bar{\nu}_\tau\bar{b}$

Table 4.2: Possible decays of the $t\bar{t}$ pair contributing to the signal

The signature of this process consists of two isolated opposite-charge high-p_{T} leptons of different flavor, the presence of two b jets and high missing transverse energy. To enhance the purity of the signal and maximize the suppression of background contributions, some kinematic cuts 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 criteria defined in Sec. 4.1. The presence of two opposite-charge different-flavor leptons is required, of which the leading and subleading lepton are required to have $p_{\mathrm{T}} > 25$ and $p_{\mathrm{T}} > 20$ GeV respectively and $|\eta| < 2.4$. Electrons reconstructed in the ECAL barrel-endcap transition region are rejected. If more than two leptons are reconstructed in the event, the two leptons with the highest p_{T} are chosen. The invariant mass of the selected lepton pair is required to be larger than 20 GeV in order to suppress low-mass resonances. The selection criteria discussed up to now reduces the fraction of events to 31% of the total $t\bar{t}$ MC signal dataset. Some requirements are further introduced for the jet multiplicity. Jets and $p_{\mathrm{T}}^{\mathrm{miss}}$ are reconstructed and identified as defined in Sec. 4.1. Jets arising from b quarks are identified using the DeepCSV tagger [66]. At least two jets with $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.4$ are required in the event, of which at least one must be b tagged. The fraction of signal events after this cut is reduced to 21.6% of the total dataset.

Figure 4.3 shows the control plots for four kinematic distributions. The accounted systematic uncertainties are discussed in Sec. 5.5. In general, the MC simulation provides a good description of data. Some discrepancies can be observed in the lepton momentum distributions, where the spectrum predicted by the simulation is harder than the one observed in data. The effect is related to the modeling of the top quark p_{T} in POWHEG + PYTHIA8 $t\bar{t}$ simulation, and similar observations have been made in previous measurements [71, 91]. The jet and b-jet multiplicities from data are well described by the simulation within the uncertainties.

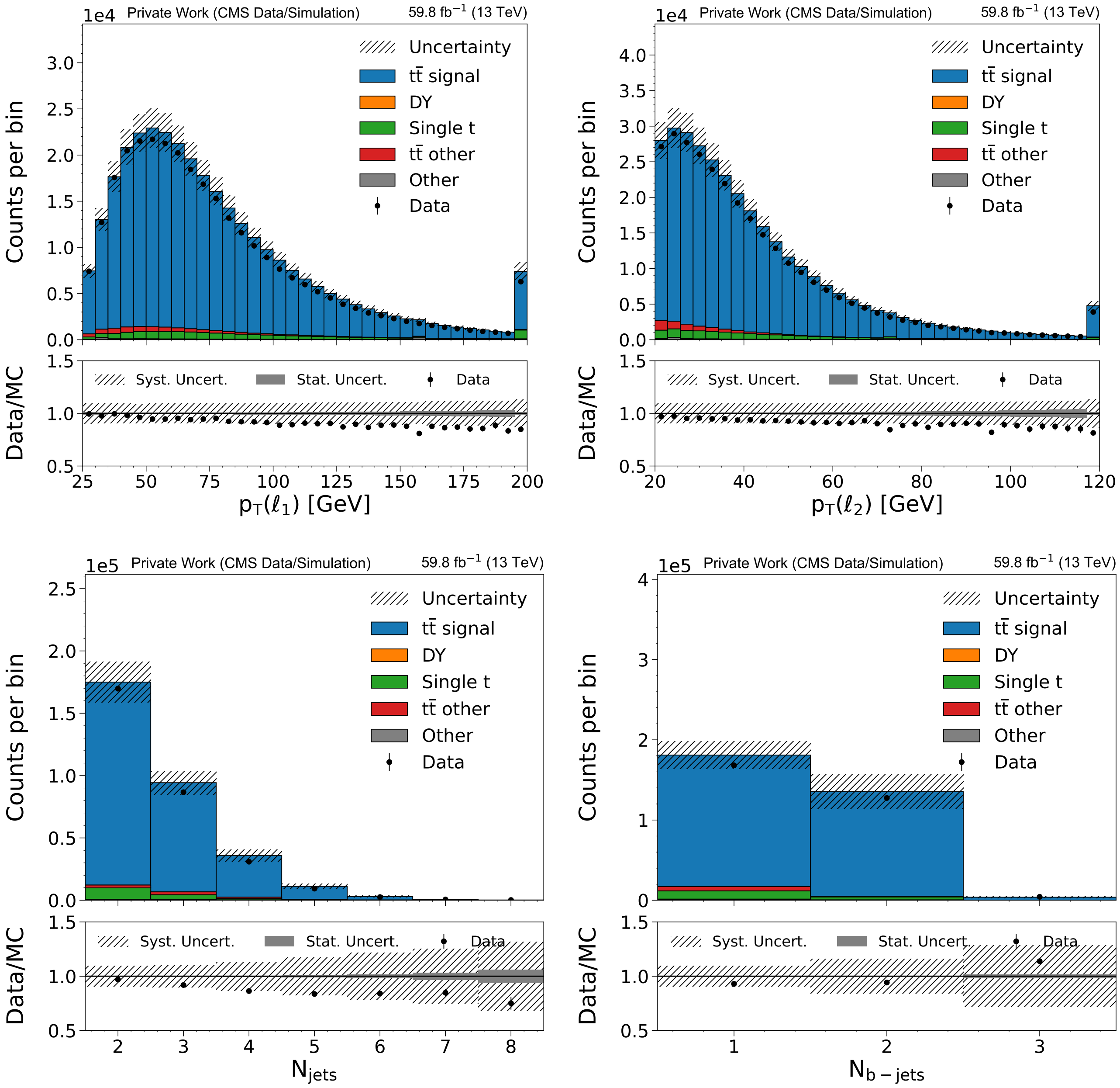


Figure 4.3: Observed (points) and predicted (stacked histograms) kinematic spectra for signal and background as a function of the leading lepton p_T (upper left), subleading lepton p_T (upper right), number of jets (lower left) and number of b jets (lower right). The hatched area indicates the systematic uncertainty on the MC simulation, the grey filled area is the statistical uncertainty on the MC simulation, which is computed assuming Poissonian errors on the MC signal and background counts and by propagating these uncertainties to their sum. Vertical bars represent the statistical uncertainty on the data. The first (last) bin accounts for underflow (overflow) events.

4.5 Analytical kinematic reconstructions

The top quark has a very short lifetime of approximately $\tau_t \approx 10^{-25}$ s, which makes it impossible to directly observe this particle in the detectors. However, the detection of its decay products can be used to perform a kinematic reconstruction and infer the properties either of the top quark and antiquark separately or of the $t\bar{t}$ system. Two alternative methods have been traditionally used, namely the full kinematic reconstruction and the loose kinematic reconstruction. In this thesis, these methods are used as terms of comparison to the NN-based approach, which will be introduced in Chapter 5

4.5.1 Full kinematic reconstruction

The full kinematic reconstruction [50] consists of an analytical and algebraic method aiming at reconstructing the four-momentum of the top and the antitop quarks separately. The momenta of the two final state neutrinos represent six unknowns, assuming them to be massless. Therefore, additional constraints need to be set in order to have a well-constrained system of equations. These constraints are based on the $\mathrm{p_T^{miss}}$, the W boson mass and the top quark mass. The reconstruction relies on the assumption that the whole $\mathrm{p_T^{miss}}$ is due to the two neutrinos only. It follows that:

$$\begin{aligned} E_x^{\text{miss}} &= p_{x,\nu} + p_{x,\bar{\nu}} \\ E_y^{\text{miss}} &= p_{y,\nu} + p_{y,\bar{\nu}} \end{aligned} \tag{4.3}$$

where $p_{x,\nu}$ ($p_{x,\bar{\nu}}$) and $p_{y,\nu}$ ($p_{y,\bar{\nu}}$) are the x and y components of the neutrino (anti-neutrino) momenta. Next, the W boson mass $m_W = 80.4$ GeV is used as a constraint for the invariant mass of its decay products.

$$\begin{aligned} m_{W^+}^2 &= (E_{\bar{\ell}} + E_\nu)^2 - (\vec{p}_{\bar{\ell}} + \vec{p}_\nu)^2 \\ m_{W^-}^2 &= (E_\ell + E_{\bar{\nu}})^2 - (\vec{p}_\ell + \vec{p}_{\bar{\nu}})^2 \end{aligned} \tag{4.4}$$

where $(E_\ell, \vec{p}_\ell)$, $(E_{\bar{\ell}}, \vec{p}_{\bar{\ell}})$, $(E_\nu, \vec{p}_\nu)$ $(E_{\bar{\nu}}, \vec{p}_{\bar{\nu}})$ are the four-momenta of the lepton, neutrino, antilepton and antineutrino respectively. Finally, the top quark mass $m_t = 172.5$ GeV is used to constrain the remaining two unknowns:

$$\begin{aligned} m_t^2 &= (E_{\bar{\ell}} + E_\nu + E_b)^2 - (\vec{p}_{\bar{\ell}} + \vec{p}_\nu + \vec{p}_b)^2 \\ m_{\bar{t}}^2 &= (E_\ell + E_{\bar{\nu}} + E_{\bar{b}})^2 - (\vec{p}_\ell + \vec{p}_{\bar{\nu}} + \vec{p}_{\bar{b}})^2 \end{aligned} \tag{4.5}$$

where $(E_b, \vec{p}_b)$ and $(E_{\bar{b}}, \vec{p}_{\bar{b}})$ are the four-momenta of the jets originating from the b and $\bar{b}$ quarks respectively. Solving the system of equations 4.3 - 4.5 by substitution, a single fourth-order equation in one of the neutrino components (e.g. $p_{x,\nu}$) arises:

$$c_0 + c_1 p_{x,\nu} + c_2 p_{x,\nu}^2 + c_3 p_{x,\nu}^3 + c_4 p_{x,\nu}^4 = 0 \tag{4.6}$$

where the coefficients c_i depends on the E_x^{miss}, E_y^{miss} and on the four-momentum of the visible decay products.

In order to identify the pair of b jets appearing in Eq. 4.6 and to correctly assign the jet originating from the b (anti)quark to the the (anti)top quark decay, the following procedure is used. According to the jet multiplicity n_{jet} of the event, the number of pairs that can be matched to top quark decays is $\frac{n_{\text{jet}}!}{(n_{\text{jet}}-2)!2!}$ and, for each pair, two assignments to the selected leptons are possible, leading to a number of $\frac{n_{\text{jet}}!}{(n_{\text{jet}}-2)!}$ combinations. These combinations are ranked according to the number of b-tagged jets involved (i.e. combinations with a higher b-jet multiplicity are preferred). If no physical solution exists for the maximum b-jet multiplicity, the combinations with the second highest b-jet multiplicity are considered and so on.

For every combination of jets with the same b-tag hierarchy, Eq. 4.6 allows up to four solutions. Moreover, the finite detector resolution can lead to imaginary unphysical solutions, reducing the efficiency of the reconstruction. To cope with this, the reconstruction is performed 100 times. Each time, the measured energies and directions of the reconstructed leptons and jets are randomly smeared in accordance with their respective detector resolutions. For each smearing a weight is computed as

$w = w_{m_{\ell\bar{b}}} \times w_{m_{\bar{\ell}b}}$. The values of $w_{m_{\ell\bar{b}}}$ and $w_{m_{\bar{\ell}b}}$ are taken from the simulated true distributions at the particle level of the observables $m_{\ell\bar{b}}$ and $m_{\bar{\ell}b}$ and if these masses are larger than 180 GeV, the corresponding weights are set to 0. The lepton-jet assignment that gives the largest value of the sum of weights w_s is chosen and the properties of top and antitop quarks are determined as weighted average of the solutions to Eq. 4.6 for every smeared attempt.

$$\vec{p}_t^{\,\mathrm{full}} = \langle \vec{p}_t \rangle = \frac{1}{w_s} \sum_{i=1}^{100} w_i \vec{p}_{t,i}, \quad \text{where} \quad w_s = \sum_{i=1}^{100} w_i \tag{4.7}$$

If a given smearing yields more than one real solution to Eq. 4.6, the one with the smallest invariant mass of the $t\bar{t}$ system is selected as motivated in [50]. If no real solution is present, instead, the weight of the considered smearing is set to zero.

4.5.2 Loose kinematic reconstruction

The loose kinematic reconstruction is an alternative algebraic and analytical approach which reconstructs the top quark pair system as a whole, without giving any information on the four-momentum of the two particles separately. This method does not rely on any assumption on the top quark mass, and therefore can be used for the extraction of this quantity [74]. Similarly to the full kinematic reconstruction, all possible lepton-jet combinations in the event that satisfy the requirement on the lepton-jet invariant mass $m_{\ell b} < 180$ GeV are considered. Combinations are ranked based on the presence of b-tagged jets in the assignments, but among combinations with equal number of b-tagged jets, the ones with the highest-p_{T} jets are chosen. The algorithm reconstructs the x and y components of the $\nu\bar{\nu}$ system from the information on the missing transverse energy as:

$$\begin{aligned} p_{x,\nu\bar{\nu}} &= p_x^{\mathrm{miss}} \\ p_{y,\nu\bar{\nu}} &= p_y^{\mathrm{miss}} \end{aligned} \tag{4.8}$$

Next, the longitudinal momentum and energy of the dineutrino system is set equal to the longitudinal momentum and energy of the dilepton system, namely $p_{z,\nu\bar{\nu}} = p_{z,\ell\bar{\ell}}$ and $E_{\nu\bar{\nu}} = E_{\ell\bar{\ell}}$. Finally, additional constraints are applied to the invariant mass of the neutrino pair, $m_{\nu\bar{\nu}} \geq 0$, and on the invariant mass of the W bosons, $m_{W^-W^+} = m_{\ell\bar{\ell}\nu\bar{\nu}} \geq 2 \cdot 80.4$ GeV. Given the four-momenta of the $\ell\bar{\ell}\nu\bar{\nu}$ system and of the two selected b jets, the four-momentum of the $t\bar{t}$ system can be computed.

4.5.3 Efficiency and unphysical solutions

Both approaches do not have an efficiency equal to 100%, as shown in Fig. 4.4 for the $t\bar{t}$ signal events as simulated by POWHEG + PYTHIA8 (Sec. 4.3.1). The integrated efficiency of the full (loose) kinematic reconstruction for the considered dataset is 93% (95%) after the event selection applied as described in Sec. 4.4. Moreover, for some events the full kinematic resolution has unphysical solutions which predict an invariant mass of the $t\bar{t}$ system greater than the center-of-mass energy of the pp collision. In Fig. 4.5 the events with a $m_{t\bar{t}}^{\mathrm{reco}}$ larger than 13 TeV are shown in an overflow bin above the dashed line.

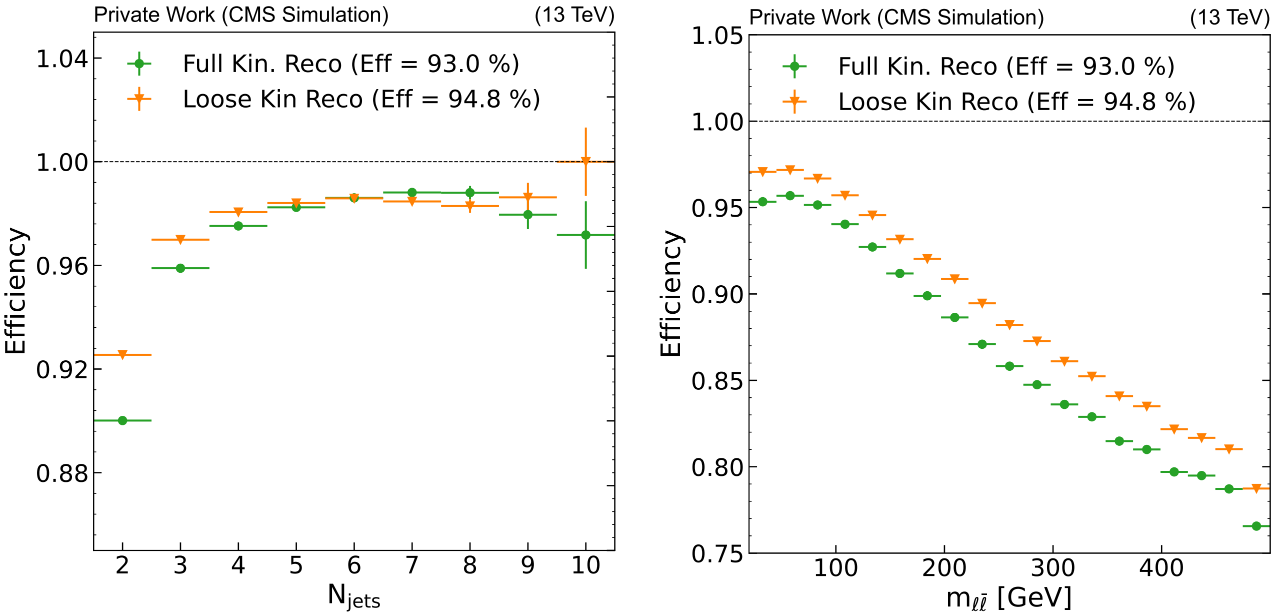


Figure 4.4: Efficiency of loose (orange) and full (green) kinematic reconstructions vs number of jets (left) and dilepton mass (right) after applying the event selection to the $t\bar{t}$ signal events simulated with POWHEG, as described in 4.4.

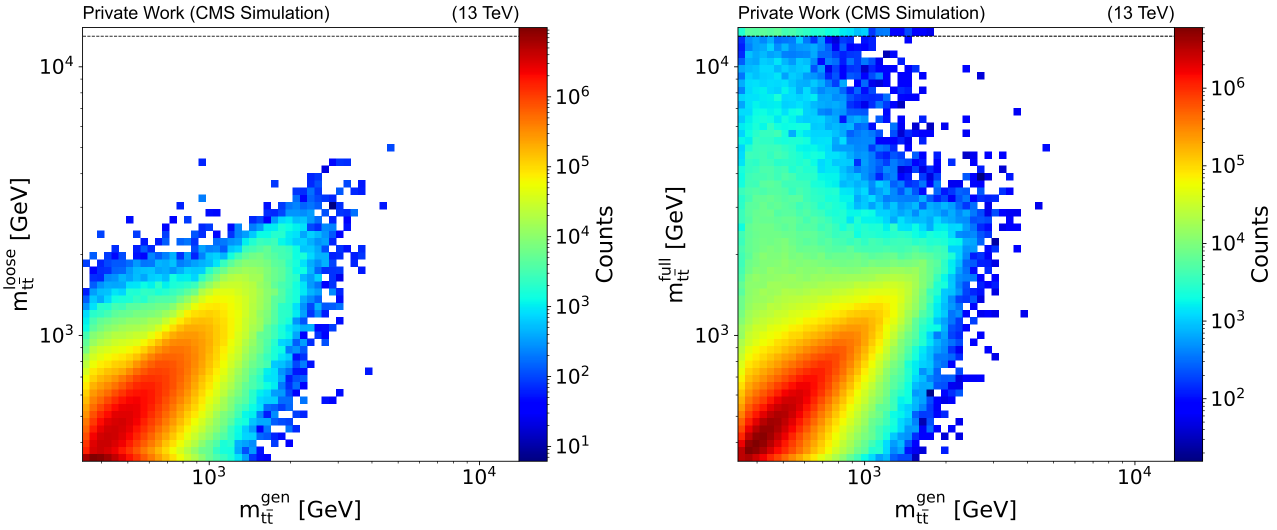


Figure 4.5: Reconstructed invariant mass of the $t\bar{t}$ system by the loose kinematic reconstruction (left) and full kinematic reconstruction (right) versus the generated $m_{t\bar{t}}$ after applying the event selection to the $t\bar{t}$ signal events simulated with POWHEG, as described in 4.4. The dotted line corresponds to $m_{t\bar{t}}^{\text{reco}}$=13 TeV

5

Top Quark Pair Invariant Mass Regression and Differential Cross Section Measurement

In this chapter, the description of the architecture of a neural network (NN) developed for the regression of the invariant mass $m_{t\bar{t}}$ of the $t\bar{t}$ system is described. The regressed observable is used for the measurement of the differential cross section $d\sigma/dm_{t\bar{t}}$ for top quark pair production as a function of $m_{t\bar{t}}$. In Sec. 5.1 a brief introduction to the aspects of machine learning relevant for the scope of this thesis is presented. In Sec. 5.2 the architecture of the NN, the choice of the input features and the hyperparameter optimization are described. In Sec. 5.3 the performance of the NN is presented and compared to the analytical approaches introduced in Sec. 4.5. Finally, in Sec. 5.4 the differential cross section $d\sigma/dm_{t\bar{t}}$ is measured in pp collisions at $\sqrt{s} = 13$ TeV using the data collected with the CMS experiment in 2018 in the $e\mu$ final state, corresponding to a total integrated luminosity of 59.8 fb^{-1}.

5.1 Machine Learning and Regression Neural Network

Machine Learning (ML) [59] is a subfield of artificial intelligence, whose generally aim is to discover patterns in the user data and fit that data into models. Within the context of supervised ML, the two most common tasks addressed with this tool are classification and regression. In the first case, the aim of the algorithm is to guess which class each event belongs to, while in the second case the algorithm is asked to predict the value of a variable. For the purposes of this thesis, the prediction of the invariant mass of the $t\bar{t}$ system is an example of regression problem that can be performed using supervised ML.

Any supervised ML algorithm works with a dataset of N events of the form $\{(\vec{x}_1, y_1), ...(\vec{x}_N, y_N)\}$, where each event is characterized by some variables $\vec{x}$, commonly called input features, and one (or more) target variable y, also referred to as output. Among the different possible ML models, deep feedforward networks, also called feedforward neural networks, or multilayer perceptrons are widely used due to their simplicity and versatility. They consist in a non-linear map $f : X \to Y$ from the input space X to the output space Y, which aims at approximating the underlying unknown relation f^* between $\vec{x}$ and y. These networks have a layerwise structure: one input layer has the starting information, one or more hidden layers process the input information in a sequential way and one final output layer returns the predicted value. The number of layers characterizes the "depth" of the NN and each layer is composed of a fixed number of units, called nodes, which take the output of the nodes of the previous layer and perform a mathematical linear operation. The output of the node n in the layer d can be written as:

$$a_n^d = \sum_i \mathcal{W}_{in}^d a_i^{d-1} + b_n^d, \tag{5.1}$$

where $\mathcal{W}_{in}^d$ is the weight tensor, which gives the multiplicative factor from the node i of the layer $d-1$ to the node n of the layer d, and b_n^d is a bias term. An illustration of the layerwise structure of a neural network is shown in Fig. 5.1.

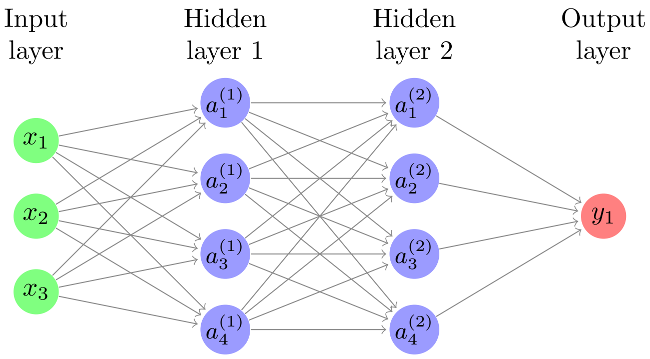


Figure 5.1: Example of a feedforward neural network with three input features (green), two hidden layers with four nodes each (blue) and one output node (red).

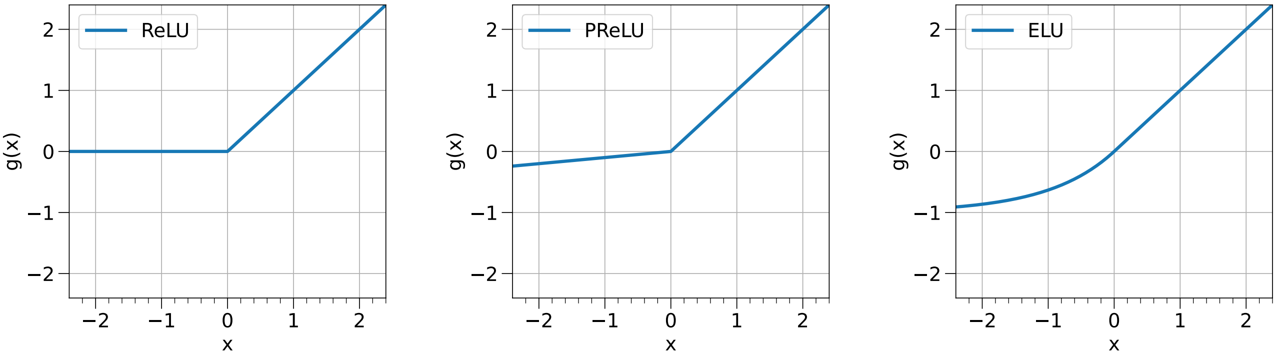


Figure 5.2: Illustration of three activation functions. From left to right: ReLU, Parametric ReLU and ELU.

However, linear combination of linear functions are linear themselves. Thus, the use of non-linear activation functions $g(x)$ is required to let the network learn more complex relations between features. These functions are used at the output of each node, making it read as:

$$g(a_n^d) = g\left(\sum_i \mathcal{W}_{in}^d g(a_i^{d-1}) + b_n^d\right), \tag{5.2}$$

Some common examples of activation functions include Rectified Linear Unit (ReLU) [31], Parametric Rectified Linear Unit (PReLU) [57] and Exponential Linear Unit (ELU) [51]. Their plots are displayed in Fig. 5.2 and their definitions are:

$$\mathrm{ReLU}(x) = \max(0, x) \qquad \mathrm{PReLU}(x) = \begin{cases} \alpha x & \text{if } x \leq 0 \\ x & \text{if } x > 0 \end{cases} \qquad \mathrm{ELU}(x) = \begin{cases} \alpha\,(e^x - 1) & \text{if } x \leq 0 \\ x & \text{if } x > 0 \end{cases} \tag{5.3}$$

The process of tuning the model's weights and biases, from now on collectively denoted with $\mathcal{W}$, in order to obtain a good approximation of the target function is referred to as "training". This is done iteratively, by minimizing the disagreement between the predicted (y) and true (y^*) output, quantified by a loss function. A common loss function used for regression tasks is the Mean Squared Error (MSE):

$$L_{\mathrm{MSE}}(\vec{x}, \mathcal{W}, \vec{y}^*) = \frac{1}{n}\sum_{i=1}^{n} (y_i(\vec{x}, \mathcal{W}) - y_i^*)^2 \tag{5.4}$$

The loss function is not measured over the total dataset, but it usually corresponds to the mean of several evaluations performed on subsets called "batches". An "epoch" corresponds to the entire training dataset being fed to the neural network for learning. The minimization of the loss function is achieved using an iterative technique called stochastic gradient descent [64]. In this method, during each iteration t, weights and biases are shifted in the opposite direction of the gradient:

$$\mathcal{W}^{(t+1)} = \mathcal{W}^{(t)} - \eta \nabla_{\mathcal{W}} L(\mathcal{W}^{(t)}), \tag{5.5}$$

where η is the learning rate and regulates the step size.

The number of layers, nodes, the batch size and the learning rate are a few examples of the so-called hyperparameters of a NN. These parameters are not updated like weights and biases and they need to be set by the user. The optimization of hyperparameters is a step of fundamental importance to achieve the best possible results with NNs [77].

The main aim of any supervised ML algorithm is to perform well on new previously unseen data. The performance of an algorithm can be quantified in terms of two type of errors: the training error and the validation error. The first one is measured by computing the loss function on the training set and the second one on a separate independent dataset. A large training error typically means that the network is not complex enough to approximate the unknown function f^* appropriately. This is also known as underfitting. On the other side, a large validation error and a small training error are indicative of a potential overfitting, meaning that the model is not able to generalize the patterns found in the training samples. This usually happens when the model complexity is too large and it starts fitting the fluctuations in the data. As a result, the model will perform poorly when applied to new and unseen data, unless the complexity is reduced. Alternatively, several regularization techniques have been introduced to prevent overfitting. Among others, the L2 regularization, also known as ridge regression, adds a penalty term in the loss function proportional to the sum of squared weights in the layers.

Finally, another issue related to the training of a NN is the change of the distribution of each layer's inputs as the parameters of the previous layers are updated. This phenomenon is referred to as internal covariate shift [53] and it can slow down the training process or drive the output of nodes far from zero, making the gradient of some activation functions vanish. A common way to tackle this phenomenon is the use of batch normalization layers, which normalize layers inputs. As further advantage, the use of batch normalization can remove the need of regularization techniques, like dropout [46, 53].

5.2 NN architecture

The regression of the observable $m_{t\bar{t}}$ can be performed by using both the full and the loose kinematic reconstruction and differential cross section measurements for the production of top quark pairs as a function of $m_{t\bar{t}}$ have been measured at $\sqrt{s} = 7$, 8 and 13 TeV [39, 50, 85, 74]. In this thesis, a multivariate analysis (MVA) approach is developed based on a regression NN with one output node, which reconstructs the invariant mass of the $t\bar{t}$ system from 73 input features (reduced to 35 after performing feature selection) for each event that has passed the selection criteria described in Sec. 4.4. These input features include the four-momenta of the reconstructed final state particles expressed as $(p_\mathrm{T}, \eta, \varphi, m)^\mathrm{T}$, the four-momenta of some composite systems (e.g. $\ell\bar{b}$ or $\ell\bar{\ell}b\bar{b}$), angular distances between final state particles (e.g. $\Delta R(\ell\bar{\ell})$ and the kinematic properties derived by the analytical solutions described in Sec. 4.5 (e.g. $p_{t\bar{t}}^{\mathrm{loose}}$, $m_{t\bar{t}}^{\mathrm{full}}$ or p_{t}^{full}). However, some events might not have real solutions for the analytical methods. To cope with this, two different NNs are trained, the first one (NN1) accepts all the events in which both the full and the loose kinematic solutions exist, while all the events where at least one of the two approaches fails enter in a second NN (NN2), which makes use of 58 (reduced to 29 after feature selection) out of 73 input features not related to the analytical reconstructions. Moreover, in some cases the full kinematic reconstruction predicts values of $m_{t\bar{t}}$ larger than the center-of-mass energy of the pp collision (see Fig. 4.5). All the events in which $m_{t\bar{t}}^{\mathrm{full}} > 13$ TeV are not considered by the NN1 and are used to train the NN2. The scheme by which the events are split into the two independent NNs is summarized in Fig. 5.3. Both of the NNs are trained using the Tensor-Flow [56] package and the Keras [49] backend.

5.2.1 Input features

The input features used for the training of the NNs can be classified into four main categories:

- **Solutions to analytical approaches**. This class includes the four-momenta of $t\bar{t}$, t and $\bar{t}$ reconstructed by the analytical approaches, when available. The masses of the top and antitop

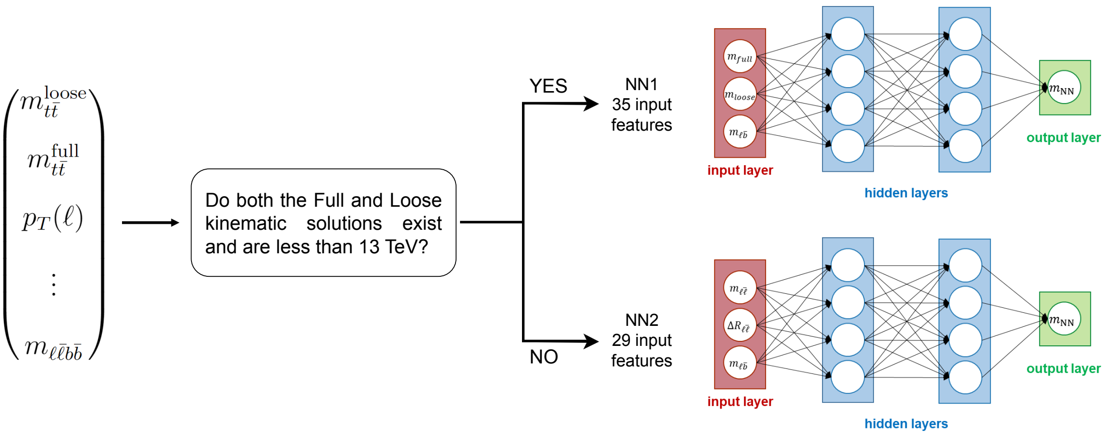


Figure 5.3: Scheme of how events are split to train two NNs

quarks given by the full kinematic reconstruction are not used since they have a unique value equal to $m_t = m_{\bar{t}} = 172.5$ GeV (see Sec. 4.5). In addition, the ΔR distance between the t and the $\bar{t}$ has been used.

- **Individual kinematic properties**. This class includes the $p_{\mathrm{T}}, \eta, \varphi$ and mass of the reconstructed final state particles. A boolean value for the b-tagged status of the jets and the b score given by the DeepCSV algorithm are also used.

- **Composite systems**. This class includes the $p_{\mathrm{T}}, \eta, \varphi$ and mass of the composite systems built on the four-momenta of the previous category. In this case a particular interest is given to the correct assignment of b jets and leptons to construct the four-momenta of systems like $\ell\bar{b}$ and $\bar{\ell}b$. To address this problem, an approach similar to the one described in the Sec. 4.5 for the full kinematic reconstruction has been used. Firstly, the true level distributions of the $m_{\ell\bar{b}}$ variable is considered. Given the two leptons ℓ_1 and ℓ_2, all possible combinations with pairs of jets are considered, starting with the combinations with the highest b-jet multiplicity. Among combinations with the same b-jet multiplicity, a weight is computed according to the values of the invariant masses of the two lepton-jet systems as described in Sec. 4.5. The combination which maximizes the weight is taken. If all the combinations with the highest possible b-jet multiplicity have an invariant mass in the region where the expected frequency from the generated distribution is zero, the second highest b-jet multiplicity is tried and so on. If the weights are zero for all possible combinations, the two jets with the highest p_{T} are taken and the assignment to the two leptons that minimizes the product of the invariant masses is used. As an example, in Fig. 5.4 the true level $m_{\ell\bar{b}}$ distribution is shown and two possible lepton-jet assignments are tried in an event in which one muon, one electron and two b jets are reconstructed.

- **Angular distances**. This class includes some distances in the (η, φ) plane between selected jets and leptons.

In order to limit the phase space and to reduce the redundancy between variables, some rotations are applied to the four-momenta of the final state particles, the composite systems built on them, and the four momenta reconstructed by the analytical approaches. First, the sign of η of all the vectors are multiplied by the sign of η of the leading lepton so that this has always a positive pseudorapidity. This rotation corresponds to an inversion of the z-axis event by event (Fig. 5.5(a)). Second, the φ of the leading lepton is subtracted to all the φ, so that this particle has always $\varphi = 0$. This corresponds to a rotation of the $x - y$ plane (Fig. 5.5(b)). Later, a 2π shift is applied, if necessary, so that all the φ are restored in the domain $[-\pi, \pi]$. Finally, the sign of the φ of all the four-momenta is changed according to the sign of the φ of the subleading lepton, so that this has always a positive sign. This last transformation corresponds to a change in the convention of positive azimuthal angle (Fig. 5.5(c)). Because of these rotations, the φ of the leading lepton was not included as an input feature since it is

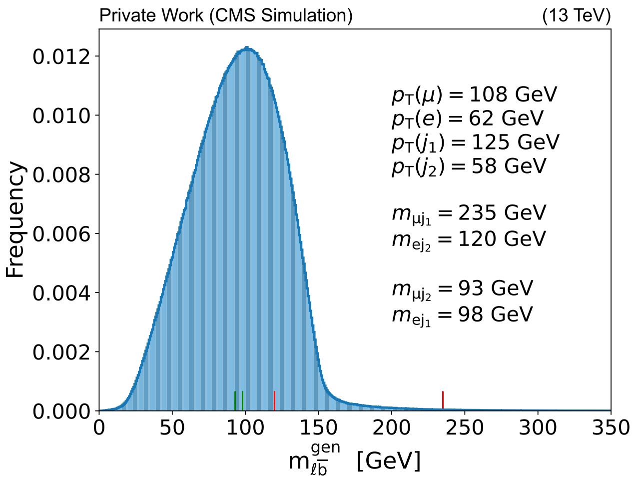
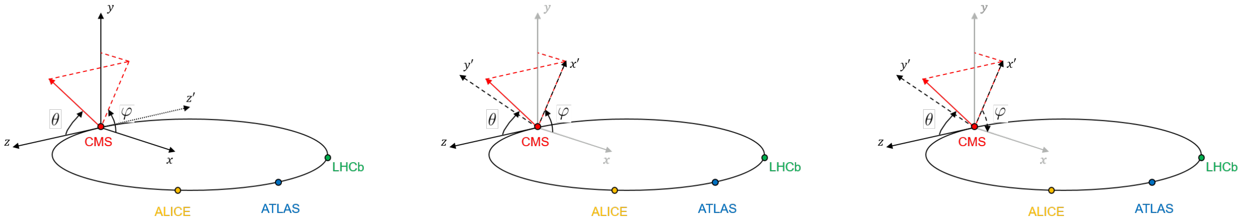


Figure 5.4: True level $m_{\ell\bar{b}}$ distribution with an example of one event in which two b-tagged jets are identified. Two combinations of assignment are considered. The selected assignment is the one that maximizes the product of the two frequencies. In this event the muon is combined with the subleading jet and the electron with the leading one.

always equal to zero and the φ of the subleading lepton becomes the absolute difference between the φ of the two leptons.



(a) (b) (c)

Figure 5.5: Rotations applied to the four momenta of the final state particles. From left to right: inversion of the z-axis according to the pseudorapidity of the leading lepton, rotation of the (x, y) plane such that $\varphi = 0$ for the leading lepton and change of the sign of φ according to the sign of the second lepton.

After applying these rotations a selection of the input features is done in order to select only those variables which can significantly contribute for the regression task. The selection of input variables also reduces the computational cost and might improve the performance of the model. For this analysis the selection has been carried out by considering the following criteria:

- The impact that the variable has on the model prediction
- The correlation between two variables

The impact of each input feature on the output of the NNs is evaluated using the concept of Shapley values, a method from coalitional game theory [63]. Given an event z, a set of variables $\vec{x} = \{x_1, x_2, ..., x_N\}$ and a model output $y(\vec{x})$, the SHAP (SHapley Additive exPlanations) value for the feature x_i in the event z is given by the difference between a baseline prediction, which is the model prediction when all the features are set to their baseline (e.g. average) values $\vec{x}^{(b)}$ and the model prediction when the feature x_i is fixed to the value it has in the event z, while keeping all other features at their baseline values:

$$\text{SHAP}(x_i, z) = y(\vec{x}^{(b)}) - y(x_1^{(b)}, ..., x_i^{(z)}, ..., x_N^{(b)}) \tag{5.6}$$

If this difference is large, it means that by changing the feature x_i, the output is drastically affected, hence the impact of the feature x_i is relevant. Otherwise, if this difference is small, it means that the model does not significantly depend on the feature x_i for its prediction. The average among 1000 events of the absolute SHAP values for the first twelve input features are shown in Fig. 5.6 for both NNs trained after applying the feature selection. As expected, in NN1 the invariant mass of the visible objects $m(\ell\bar{\ell}j_1j_2)$ and the features derived from the analytical solutions have a high score together with the composite systems of the visible decay products of the top and antitop quarks ($m_{\ell\bar{b}}$ and $m_{\bar{\ell}b}$), which have a similar impact. In NN2, features related to the dilepton system play the most relevant roles, despite the invariant mass of the visible objects $m(\ell\bar{\ell}j_1j_2)$ is used. This can be related to the fact that for these events the analytical approaches fail, therefore the naive reconstruction of the invariant mass of complex systems can be compromised (e.g. by the wrong jets selection). Conversely, the angular distance in the (η, φ) plane or the dilepton mass are well measured features, highly correlated with the target, and they are expected to be relevant also in the cases where the analytical reconstructions fail.

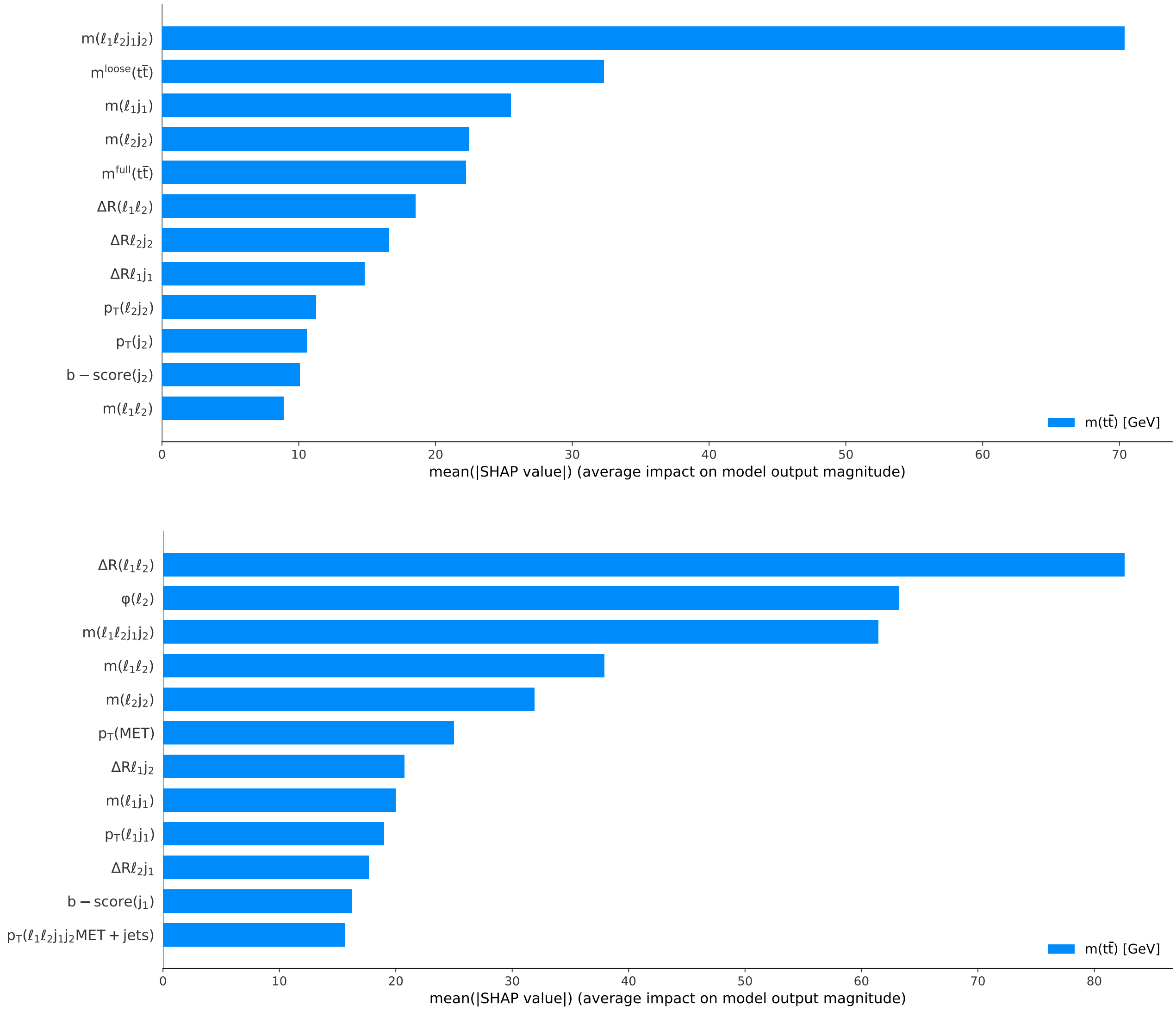


Figure 5.6: SHAP values of the twelve most relevant input features of NN1 (up) and NN2 (down).

As far as the correlation between input features is concerned, ideally it should not be close to 1 or -1. If this happens, it means that the features are perfectly correlated and the value of one can be determined from the other by applying a linear transformation. In other words, the two variables carry similar information. Quantitatively, the correlation between two variables x and y is measured using the weighted Pearson's correlation coefficient:

$$\rho(x, y; w) = \frac{\mathrm{cov}(x, y; w)}{\sqrt{\mathrm{cov}(x, x; w)\,\mathrm{cov}(y, y; w)}}, \tag{5.7}$$

where the weights w_i for each event are given by the generator weights multiplied by the SFs introduced in Sec. 4.3.1 and Sec. 4.3.2 respectively and $\mathrm{cov}(x, y; w)$ is defined as:

$$\mathrm{cov}(x, y; w) = \frac{\sum_i w_i \cdot (x_i - \mathrm{m}(x; w))(y_i - \mathrm{m}(y; w))}{\sum_i w_i}, \quad \text{with} \quad \mathrm{m}(x; w) = \frac{\sum_i w_i x_i}{\sum_i w_i} \tag{5.8}$$

For example, the pseudorapidity of the top quark pair system as reconstructed by the full ($\eta^{\text{full}}(t\bar{t})$), loose ($\eta^{\text{loose}}(t\bar{t})$) and by the system of particles ($\eta(\ell_1\ell_2 j_1 j_2\text{MET}$) are obviously highly correlated, since they measure the same quantity. Hence, only the first one is kept. In the cases in which two features are highly correlated but both carry a high SHAP value, both features are kept, like in the case of $m^{\text{full}}_{t\bar{t}}$ and $m^{\text{loose}}_{t\bar{t}}$. The final number of input features used for the two NNs are 35 and 29 respectively. The complete set of input features used for the training of the NNs is listed in Tab. 5.1-5.4, where a ✓ in the third column indicates if the features has been kept after performing the feature selection.

Finally, the considered features do not have similar scales (e.g. b-score values are in the interval [0-1], while the $m^{\text{full}}_{t\bar{t}} \geq 345$ GeV) and this can degrade the predictive performance of NNs as well as slow down or even prevent the convergence of many gradient-based estimators. For this reason, all the input features are scaled using a standard scaler, which subtracts to each value the mean and normalizes it by the standard deviation of the corresponding original distribution. By doing this, the final distributions will have zero mean and unitary standard deviation. The mean and the standard deviations are always computed from the training set and the same values are used for the testing set and for the rest of the analysis. Different scaling techniques have been tested including non-linear transformations like quantile and power transforms. No significant improvement have been observed compared to standard scaling.

Input feature	**Description**	**Selected**
$p_{\mathrm{T}}^{\text{loose}}(t\bar{t})$	p_{T} of the $t\bar{t}$ system from the loose kinematic reconstruction	✓
$\eta^{\text{loose}}(t\bar{t})$	η of the $t\bar{t}$ system from the loose kinematic reconstruction	✓
$\varphi^{\text{loose}}(t\bar{t})$	φ of the $t\bar{t}$ system from the loose kinematic reconstruction	
$m^{\text{loose}}(t\bar{t})$	mass of the $t\bar{t}$ system from the loose kinematic reconstruction	✓
$p_{\mathrm{T}}^{\text{full}}(t\bar{t})$	p_{T} of the $t\bar{t}$ system from the full kinematic reconstruction	
$\eta^{\text{full}}(t\bar{t})$	η of the $t\bar{t}$ system from the full kinematic reconstruction	
$\varphi^{\text{full}}(t\bar{t})$	φ of the $t\bar{t}$ system from the full kinematic reconstruction	
$m^{\text{full}}(t\bar{t})$	mass of the $t\bar{t}$ system from the full kinematic reconstruction	✓
$p_{\mathrm{T}}^{\text{full}}(t)$	p_{T} of the t system from the full kinematic reconstruction	✓
$\eta^{\text{full}}(t)$	η of the t system from the full kinematic reconstruction	
$\varphi^{\text{full}}(t)$	φ of the t system from the full kinematic reconstruction	
$p_{\mathrm{T}}^{\text{full}}(\bar{t})$	p_{T} of the $\bar{t}$ system from the full kinematic reconstruction	
$\eta^{\text{full}}(\bar{t})$	η of the $\bar{t}$ system from the full kinematic reconstruction	
$\varphi^{\text{full}}(\bar{t})$	φ of the $\bar{t}$ system from the full kinematic reconstruction	
$\Delta R^{\text{full}}(t\bar{t})$	ΔR between the t and the $\bar{t}$ from the full kinematic reconstruction	✓

Table 5.1: First class of input variables (solutions to analytical approaches). The features selected from this class are used for the training of NN1 only.

Input feature	**Description**	**Selected**
$p_{\mathrm{T}}(\ell_1)$	p_{T} of the leading lepton	✓
$\eta(\ell_1)$	η of the leading lepton	
$m(\ell_1)$	φ of the leading lepton	
$p_{\mathrm{T}}(\ell_2)$	p_{T} of the subleading lepton	✓
$\eta(\ell_2)$	η of the subleading lepton	
$\varphi(\ell_2)$	φ of the subleading lepton	✓
$m(\ell_2)$	mass of the subleading lepton	
$p_{\mathrm{T}}(j_1)$	p_{T} of the jet assigned to the leading lepton	✓
$\eta(j_1)$	η of the jet assigned to the leading lepton	
$\varphi(j_1)$	φ of the jet assigned to the leading lepton	
$m(j_1)$	mass of the jet assigned to the leading lepton	✓
b tag(j_1)	boolean b-tagged status of the jet assigned to the leading lepton	
b score(j_1)	b score given by DeepCSV to the jet assigned to the leading lepton	✓
$p_{\mathrm{T}}(j_2)$	p_{T} of the jet assigned to the subleading lepton	✓
$\eta(j_2)$	η of the jet assigned to the subleading lepton	
$\varphi(j_2)$	φ of the jet assigned to the subleading lepton	
$m(j_2)$	mass of the jet assigned to the subleading lepton	✓
j_2 b tag	boolean b-tagged status of the jet assigned to the subleading lepton	
j_2 b score	b score given by DeepCSV to the jet assigned to the subleading lepton	✓
$p_{\mathrm{T}}^{\mathrm{miss}}$	missing transverse momentum	✓
φ^{miss}	φ component of the missing transverse momentum	✓

Table 5.2: Second class of input variables (individual kinematic properties) considered for the training of both NN1 and NN2.

Input feature	**Description**	**Selected**
$p_\mathrm{T}(\ell_1\ell_2)$	p_T of the dilepton system	✓
$\eta(\ell_1\ell_2)$	η of the dilepton system	
$\varphi(\ell_1\ell_2)$	φ of the dilepton system	
$m(\ell_1\ell_2)$	mass of the dilepton system	✓
$p_\mathrm{T}(\ell_1\ell_2 j_1 j_2)$	p_T of the dilepton plus two selected jets	✓
$\eta(\ell_1\ell_2 j_1 j_2)$	η of the dilepton plus two selected jets	✓
$\varphi(\ell_1\ell_2 j_1 j_2)$	φ of the dilepton plus two selected jets	
$m(\ell_1\ell_2 j_1 j_2)$	mass of the dilepton plus two selected jets	✓
$p_\mathrm{T}(\ell_1\ell_2 j_1 j_2\mathrm{MET})$	p_T of the dilepton plus two selected jets and missing energy	
$\eta(\ell_1\ell_2 j_1 j_2\mathrm{MET})$	η of the dilepton plus two selected jets and missing energy	
$\varphi(\ell_1\ell_2 j_1 j_2\mathrm{MET})$	φ of the dilepton plus two selected jets and missing energy	
$\mathrm{m}(\ell_1\ell_2 j_1 j_2\mathrm{MET})$	mass of the dilepton plus two selected jets and missing energy	✓
p_T(extra-jets)	p_T of all the non selected jets	
η(extra-jets)	η of all the non selected jets	
φ(extra-jets)	φ of all the non selected jets	
m(extra-jets)	mass of all the non selected jets	
$p_\mathrm{T}(\ell_1\ell_2 j_1 j_2$ MET extra-jets)	p_T of the whole reconstructed final state	✓
$\eta(\ell_1\ell_2 j_1 j_2$ MET extra-jets)	η of the whole reconstructed final state	
$\varphi(\ell_1\ell_2 j_1 j_2$ MET extra-jets)	φ of the whole reconstructed final state	
$m(\ell_1\ell_2 j_1 j_2$ MET extra-jets)	mass of the whole reconstructed final state	✓
$p_\mathrm{T}(\ell_1 j_1)$	p_T of the leading lepton plus corresponding assigned jet	✓
$\eta(\ell_1 j_1)$	η of the leading lepton plus corresponding assigned jet	
$\varphi(\ell_1 j_1)$	φ of the leading lepton plus corresponding assigned jet	
$m(\ell_1 j_1)$	mass of the leading lepton plus corresponding assigned jet	✓
$p_\mathrm{T}(\ell_2 j_2)$	p_T of the subleading lepton plus corresponding assigned jet	✓
$\eta(\ell_2 j_2)$	η of the subleading lepton plus corresponding assigned jet	
$\varphi(\ell_2 j_2)$	φ of the subleading lepton plus corresponding assigned jet	
$m(\ell_2 j_2)$	mass of the subleading lepton plus corresponding assigned jet	✓
H_T	hadronic transverse energy (sum of the p_T of all the jets)	
N_jets	number of reconstructed jets	
$N_\mathrm{b\ jets}$	number of reconstructed b-tagged jets	

Table 5.3: Third class of input variables (composite systems) considered for the training of both NN1 and NN2.

Input feature	**Description**	**Selected**
$\Delta R(\ell_1\ell_2)$	ΔR between leading and subleading letpon	✓
$\Delta R(\ell_1 j_1)$	ΔR between the leading lepton and the corresponding assigned jet	✓
$\Delta R(\ell_1 j_2)$	ΔR between the leading lepton and the jet assigned to the subleading letpon	✓
$\Delta R(\ell_2 j_1)$	ΔR between the subleading lepton and the jet assigned to the leading letpon	✓
$\Delta R(\ell_2 j_2)$	ΔR between the subleading lepton and corresponding assigned jet	✓
$\Delta R(j_1 j_2)$	ΔR between the two selected jets	✓

Table 5.4: Fourth class of input variables (angular distances) considered for the training of both NN1 and NN2.

5.2.2 Data and MC comparison for input features

The regression of the observable $m_{t\bar{t}}$ by a NN relies on the assumption that the events used for the training simulated by the MC event generators provide a good description of the real data. In particular, a direct comparison of the simulated input features to the measured ones is possible. The systematic uncertainties accounted for this analysis are described in detail in Sec. 5.5. The plots of the comparison between data and MC for four features with a large impact on the NN1 are shown in Fig. 5.7, while all the other features, kept after the feature selection, are shown in the Appendix A. The simulated background and signal processes are normalized to the total luminosity of the 2018 data and to their corresponding cross section predicted by the MC generator model as explained in Sec. 4.3. An excellent agreement between data and simulation is found.

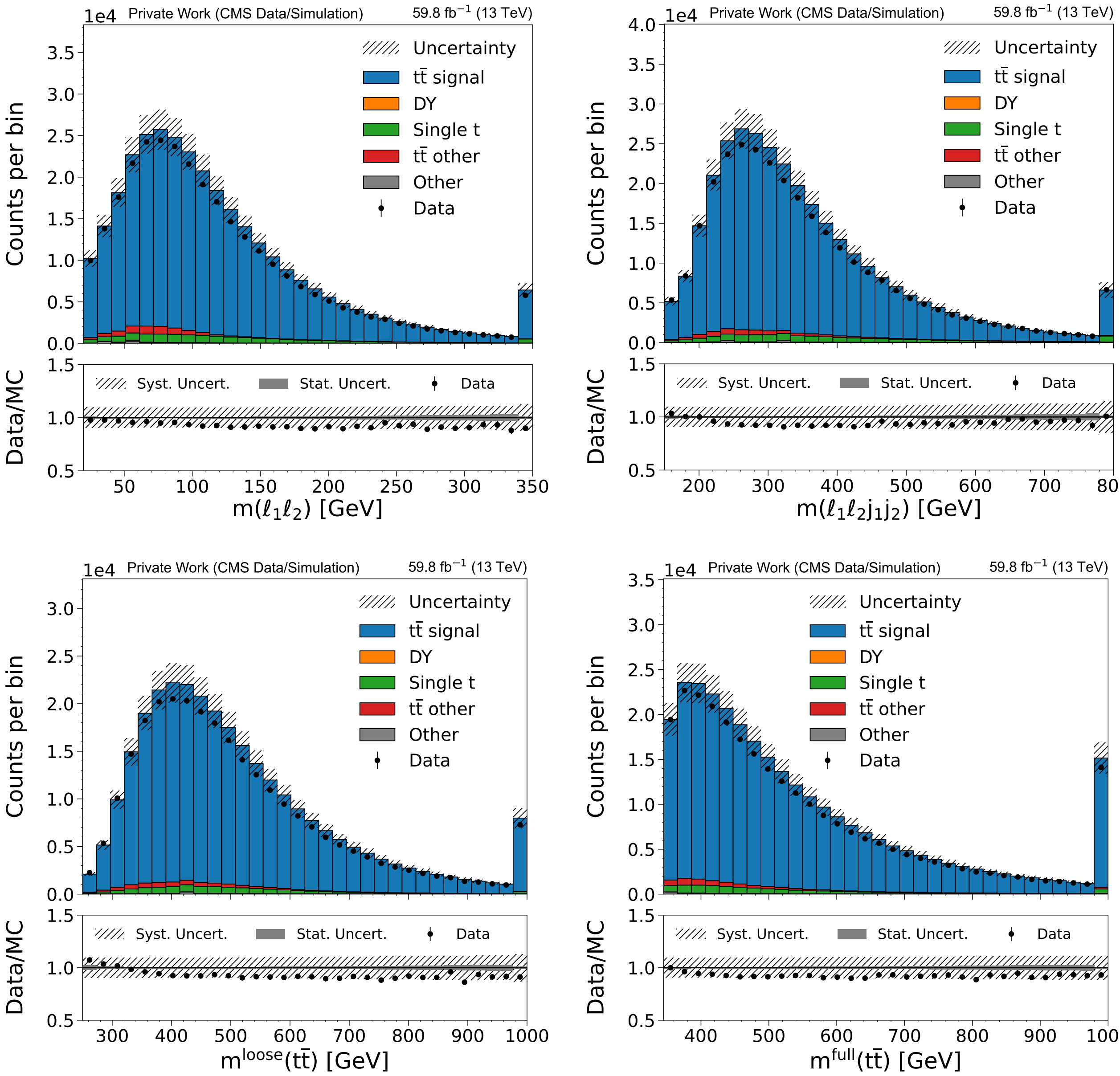


Figure 5.7: Observed (points) and predicted (stacked histograms) kinematic spectra for signal and background as functions of four features: dilepton mass (upper left), invariant mass of leptons and selected jets (upper right), top quark pair mass from the loose (lower left) and full (lower right) kinematic reconstruction. The hatched area indicates the systematic uncertainty on the MC simulation, the grey filled area is the statistical uncertainty on the MC simulation, which is computed assuming Poissonian errors on the MC signal and background counts and by propagating these uncertainties to their sum. Vertical bars represent the statistical uncertainty on the data. The first (last) bin accounts for underflow (overflow) events.

5.2.3 Training

The whole dataset of simulated signal events is split into a training set, a validation set and a testing set with the following proportions: 0.7, 0.15, 0.15. Overall, more than 1.5 million simulated events are used for the training, validation and testing, while an equal number of events is used for the rest of the analysis. Both NNs are trained to minimize the mean squared error (MSE) between the reconstructed value of the invariant mass of the $t\bar{t}$ system ($m_{t\bar{t}}^{\text{reco}}$) and the generated value ($m_{t\bar{t}}^{\text{gen}}$):

$$\text{Loss} = \frac{1}{N_{\text{batch}}} \sum_{i \in \text{batch}} w_i^{\text{train}} \left(m_{t\bar{t}}^{\text{reco}} - m_{t\bar{t}}^{\text{gen}}\right)_i^2 , \tag{5.9}$$

where w^{train} is used to weight the loss function and is defined as:

$$w_i^{\text{train}} = w_i^{\text{MC}} \cdot \text{SFs} \cdot \frac{1}{f_i} \cdot \left(\frac{1}{m_{t\bar{t},i}^{\text{gen}}}\right)^2 , \tag{5.10}$$

where w_i^{MC} is the weight given by the generator weight (Sec. 4.3.1), SFs are the scale factors (Sec. 4.3.2), $\frac{1}{f_i}$ is a factor inversely proportional to the frequency of the binned $m_{t\bar{t}}^{\text{gen}}$ distribution and the last term is used to normalize the MSE into a percentage mean squared error. The reason of the introduction of the third term is the imbalance between different regions of the $m_{t\bar{t}}^{\text{gen}}$ distribution due to its non-uniformity, as can be seen in Fig. 5.8. Forty-three bins are used with different widths from 320 to 2100 GeV for the NN1, while only thirteen for the NN2 due to the limited statistics. Events in underflow (overflow) receive the same correction used for the first (last) bin. Finally, the last term in Eq. 5.10 has been tested to be fundamental for a good regression of the region of low $m_{t\bar{t}}$ values. All the weights are normalized by their mean value so that on average the effective weighted size of one batch is equal to the value set for the batch size.

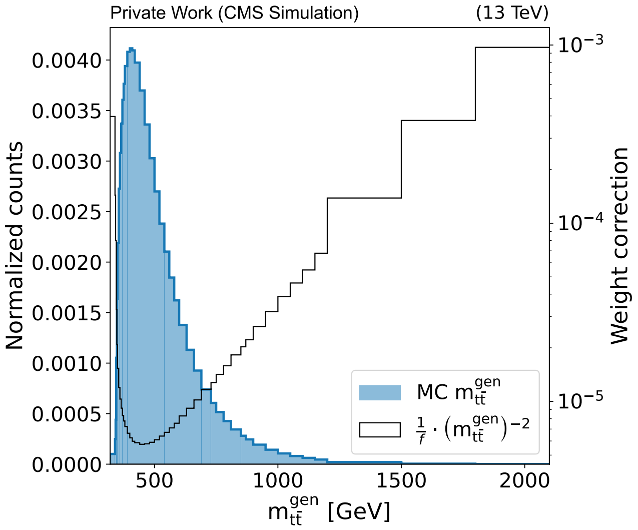


Figure 5.8: Distribution of $m_{t\bar{t}}^{\text{gen}}$, as predicted by the POWHEG MC generator introduced in Sec. 4.3.1, and the corrections applied in the definition of the training weights. The value of $m_{t\bar{t}}^{\text{gen}}$ used in the denominator of the weight correction is taken to be the midpoint of each bin.

The NN1 makes use of one input layer with 35 input features, four hidden layers respectively with 256, 128, 8, 8 nodes and one output node. ELU is used as activation function for the hidden layers with a parameter $\alpha = 1$, whereas a linear activation function is used for the output node. All layers, except for the output layer, are complemented with batch normalization layers [53], which standardize the values of each batch. The batch sizes for the training and the validation are equal to $N_{\text{batch}} = 4096$ events. In order to avoid overfitting, one method of regularization is used, namely the L2 regularization with a parameter $\lambda_{L2} = 2.92 \cdot 10^{-2}$. Thus, the total loss function reads:

$$\text{Loss} = \frac{1}{N_{\text{batch}}} \sum_{i \in \text{batch}} w_i^{\text{train}} \left(m_{t\bar{t}}^{\text{reco}} - m_{t\bar{t}}^{\text{gen}}\right)_i^2 + \lambda_{\text{L2}} \sum_{i,j,d} (W_{ij}^d)^2 \tag{5.11}$$

Moreover, an early stopping procedure is used to avoid overfitting with a patience parameter of 40 epochs. This means that the training is stopped after 40 epochs if no improvement in the validation loss function is observed and the weights are restored to the best epoch. The training and validation loss functions of both NNs are shown in Fig. 5.10, where the vertical arrow indicates the best epoch. The total number of epochs is chosen sufficiently large (3500 epochs) to let the early stopping work. The learning rate is set to $1.5 \cdot 10^{-3}$ and it is updated according to the average first and second moments of the gradients using the ADAM algorithm [62]. The exponential decay rate for the first and second moment estimates are set to default values 0.9 and 0.999, respectively.

A similar but simpler architecture is used for the NN2 since it deals with less data. One hidden layer is employed with 20 nodes and complemented with a batch normalization layer. The same loss function is used with the following regularization parameter $\lambda_{\mathrm{L2}}^{\mathrm{NN2}} = 7.7 \cdot 10^{-3}$. The other hyperparameters are kept unchanged. All the hyperparameters of the two NNs are summarized in Tab. 5.5.

The hyperparameters of the two NNs are the results of an optimization performed following a random search, whose final aim is the minimization of the Jensen-Shannon distance (JSD) between the reconstructed $m_{t\bar{t}}^{\mathrm{reco}}$ and the simulated $m_{t\bar{t}}^{\mathrm{gen}}$ distributions. The JSD is a measure of the similarity between two probability distributions, based on the Kullback–Leibler divergence (KLD), also referred to as relative entropy. Given two discrete probability distributions P and Q defined on the same sample space $\mathcal{X}$, the KLD from P to Q is defined as:

$$D_{\mathrm{KL}}(P \parallel Q) = \sum_{x \in \mathcal{X}} P(x) \log\left(\frac{P(x)}{Q(x)}\right) \tag{5.12}$$

The JSD is a symmetrized version of the KLD with the further advantage of being always finite. The JSD can be computed as:

$$\mathrm{JSD}(P \parallel Q) = \sqrt{\frac{1}{2} D_{\mathrm{KL}}(P \parallel M) + \frac{1}{2} D_{\mathrm{KL}}(Q \parallel M)} \tag{5.13}$$

where $M = \frac{1}{2}(P + Q)$ is an average distribution.

Firstly, the hyperparameters to be optimized are chosen and a range is defined for each of them. The number of layers is varied from 2 to 4. Each hidden layer can have different number of nodes ranging from 8 to 256 units in multiples of 2, the regularization parameters λ_{L2} is varied between 10^{-6} and 10^{-1}. The training and validation batch size are constrained to be equal and are varied in the interval $[2^7 - 2^{13}]$ in multiples of 2. A random search is done in this multi-dimensional space by generating uniformly the number of layers, the number of nodes and the batch size. The regularization parameter λ_{L2} is generated with a log-uniform distribution in the aforementioned interval. Only a limited fraction of events equal to one third of the dataset, split with the same proportions between training, validation and testing, is used for the random search. For each sampled point, an objective function is evaluated. For this analysis, the JSD is the scalar result of the objective function. Each model is trained and tested three times and the average over the three JSDs is considered. The result of the random search and the partial dependence of the JSD with respect to the choice of the hyperparameters is shown in Fig. 5.9, where the selected model, which provides the best JSD, is highlighted in red. Additionally, two scatter plots of the JSD versus the normalized weighted root mean squared error (nRMSE) and the determinant of the covariance matrix of the differential cross section distribution $d\sigma/dm_{t\bar{t}}$, that will be defined in Sec. 5.3.1, are illustrated, highlighting how different objective functions converge to the same choice of the best model. The nRMSE is defined as:

$$\mathrm{nRMSE} = \frac{1}{\sum_i w_i} \sum_i w_i^{\mathrm{train}} \left(\frac{m_{t\bar{t}}^{\mathrm{reco}} - m_{t\bar{t}}^{\mathrm{gen}}}{m_{t\bar{t}}^{\mathrm{gen}}}\right)_i^2 \tag{5.14}$$

An analogous search has been done for the parameters of the NN2. The only differences consist in a smaller number of layers, limited between 1 and 2, and number of nodes, limited between 4 and 64, not necessarily being multiple of 2. All the available statistics has been used for the optimization of the NN2.

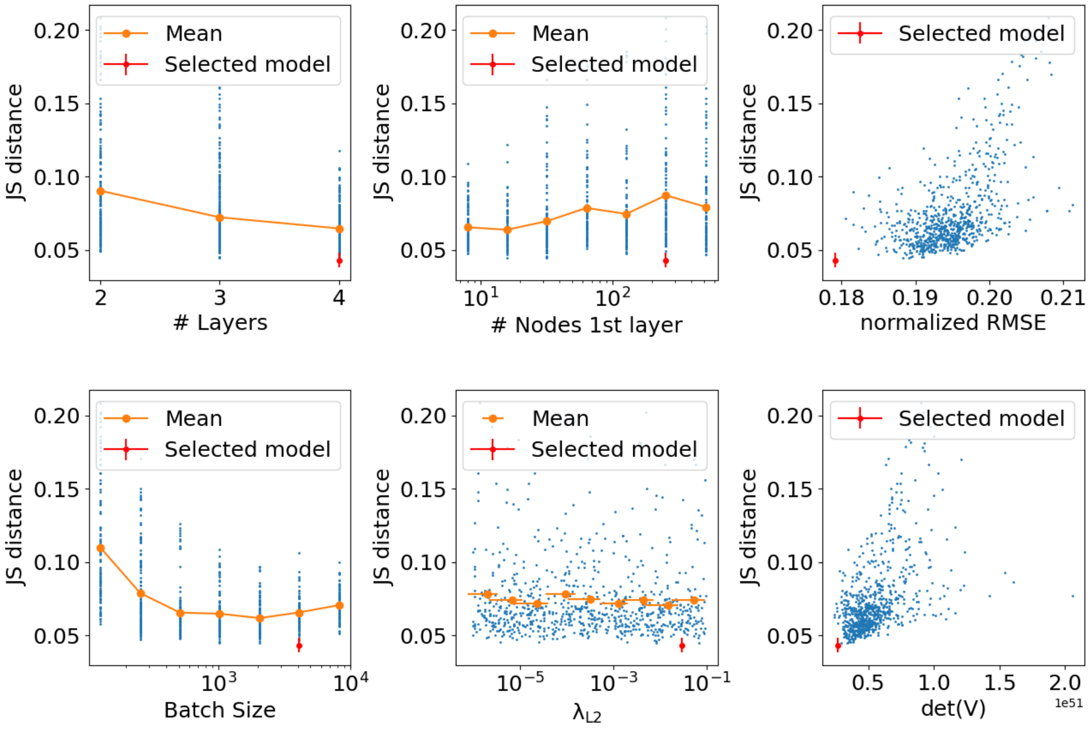


Figure 5.9: Optimization of the parameters of NN1 by randomly generating the number of layers, nodes, the batch sizes and the L2 parameter. The partial dependence (orange) has to be meant as the average value of the JSD as a function of one variable without considering all the other hyperparameters. The upper and lower right subplots show the relations between the JSD and other two objective functions, namely the normalized RMSE and the determinant of the covariance matrix of the differential cross section distribution. The selected model is highlighted with a red dot.

5.3 NN Performance

All the results shown in this section are meant to be obtained using the testing dataset of the $t\bar{t}$ signal MC simulation, if not specified otherwise. The output of the two NNs are combined in order to have one reconstructed invariant mass $m_{t\bar{t}}^{\text{reco}}$ for each event. In this way the efficiency of the NN-based approach is 100% to be compared to 93.0% and 94.8% for the full and loose kinematic reconstructions, respectively (see Fig. 4.4).

In order to quantitatively exclude the possibility of overfitting, the Kolmogorov-Smirnov (KS) test is used between the predictions performed on the training and testing sets. This test can be either used to compare one sample to a reference probability distribution (one-sample KS test), or evaluate if two samples originate from the same distribution (two-sample KS test). In the latter case, given two samples $\vec{m}_{t\bar{t}}^{\text{train}} = \{m_1^{\text{train}}, m_2^{\text{train}}, ..., m_N^{\text{train}}\}$ and $\vec{m}_{t\bar{t}}^{\text{test}} = \{m_1^{\text{test}}, m_2^{\text{test}}, ..., m_R^{\text{test}}\}$, the following empirical cumulative distributions are defined:

$$F_N(x) = \frac{n(x)}{N} \qquad G_R(y) = \frac{r(y)}{R} \tag{5.15}$$

where $n(x)$ is the number of events of $\vec{m}_{t\bar{t}}^{\text{train}}$ with value less than x and the same applies to $r(y)$ and $\vec{m}_j^{\text{test}}$. The KS statistic is defined as:

$$D_{KS} = \sup_i |F(i) - G(i)| \tag{5.16}$$

Once, the KS statistic has been determined, a two-tailed p-value can be computed under the null hypothesis that the two samples originate from the same distribution. The computation of the em-

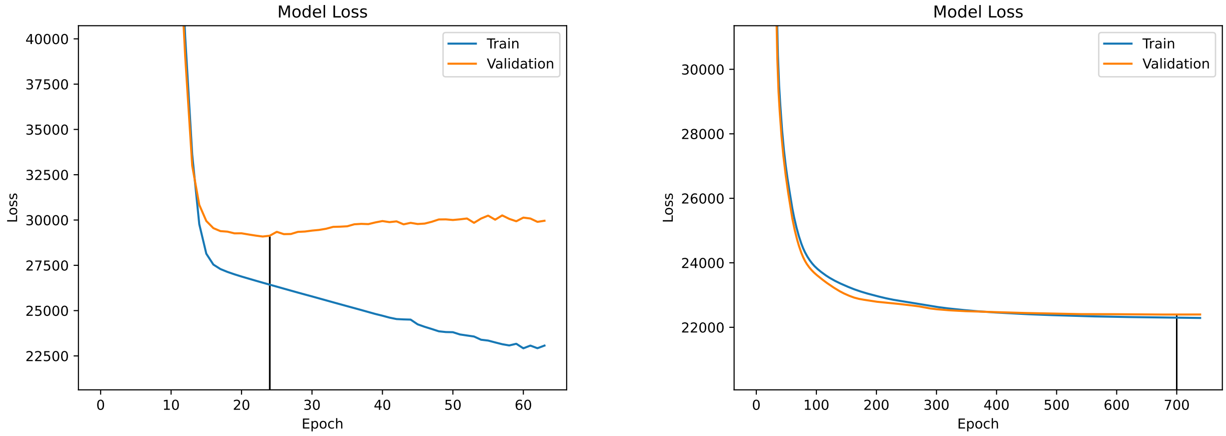


Figure 5.10: Training and validation loss functions for the NN1 (left) and NN2 (right). The arrow indicates when the early stopping took place.

Hyperparameter	**NN1**	**NN2**
learning rate	$1.5 \cdot 10^{-3}$	$1.5 \cdot 10^{-3}$
L2 parameter	$2.92 \cdot 10^{-2}$	$7.72 \cdot 10^{-3}$
Layers	4	1
Nodes	256, 128, 8, 8	20
Activation (hidden layers)	ELU	ELU
patience Early Stopping	40	40
Number of epochs	3500	3500
Training Batch Size	4096	2048
Validation Batch Size	4096	2048

Table 5.5: Summary of the values of the hyperparameters of the two NNs used for the analysis

pirical cumulative distributions and the p-value have been modified to account for the weight of the events starting from the source of the function `scipy.stats.ks_2samp` as implemented in Ref. [88], following recommendations in Ref. [32]. The empirical cumulative distributions of the prediction of training and testing datasets are shown in Fig. 5.11. A p-value of 0.33 is obtained, indicating a good agreement between the two distributions and ruling out the possibility of overfitting.

The reconstructed distributions using the three methods are shown in Fig. 5.12 compared to the true distribution. All the histograms are normalized to unity to compensate the inefficiencies of the analytical methods. The full kinematic reconstruction shows a large a overflow bin as already explained in Sec. 4.5.3 and a sharp rise due to the kinematic threshold ($m_{t\bar{t}} \geq 345$ GeV) imposed by the kinematic constraint $m_t = m_{\bar{t}} = 172.5$ GeV. The loose kinematic reconstruction has a wider distribution close to the threshold production but it shows a tail at high energy closer to the true distribution. The NN-based reconstruction well reproduces the rise at $250 - 300$ GeV and the peak of the distribution.

In Fig. 5.13 the correlation between the difference $m_{t\bar{t}}^{\text{reco}} - m_{t\bar{t}}^{\text{gen}}$ of the three reconstruction methods and $m_{t\bar{t}}^{\text{gen}}$ is shown. The colormap gives information on the number of events. Again, the full kinematic reconstruction is seen to have a sizable number of events at small $m_{t\bar{t}}^{\text{gen}}$ but large $m_{t\bar{t}}^{\text{reco}}$. By projecting these distributions on the y-axis and taking seven slices in the x-axis, the resolution of the three methods can be evaluated in different regions of $m_{t\bar{t}}^{\text{gen}}$. The results are shown in Fig. 5.14 together

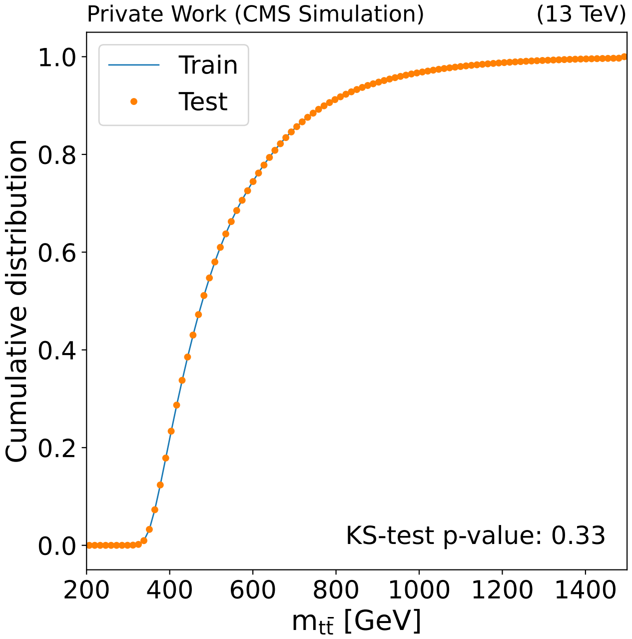


Figure 5.11: Empirical cumulative distribution of the $m_{t\bar{t}}^{\text{reco}}$ as reconstructed by the NN for the training and testing dataset.

with the root mean square (RMS) values, defined as:

$$\text{RMS}(\vec{x}) = \sqrt{\frac{1}{N}\sum_{i}^{N}(x_i)^2}, \tag{5.17}$$

which is sensitive to both the squared distance of the mean value from zero and the spread of the distribution around the mean, since it holds:

$$\text{RMS}(\vec{x}) = \langle x\rangle^2 + \sigma_x^2, \tag{5.18}$$

where $\langle x\rangle$ indicates the mean value of $\vec{x}$ and σ_x its standard deviation. The NN approach has the smallest RMS in all the seven considered regions. In the case of the full kinematic reconstruction, the RMS is drastically affected by the presence of the aforementioned outliers even after removing unphysical events with events $m_{t\bar{t}}^{\text{full}} > 13$ TeV. In any case, the width of the NN-based distribution is comparable if not better than the one based on the full kinematic approach. Moreover, in the plots of the second row of Fig. 5.14, a more symmetric distribution is obtained in the NN case, considering the shape of the histograms and the number of events in underflow and overflow bins. An additional analyzed value is the centroid of the histograms. The means of the seven distributions appearing in Fig. 5.14 are plotted for the same intervals in Fig. 5.15. A clear trend favoring negative values of $\langle m_{t\bar{t}}^{\text{reco}} - m_{t\bar{t}}^{\text{gen}}\rangle$ as $m_{t\bar{t}}^{\text{gen}}$ increases can be observed. The means of the NN-based method are within 50 GeV from zero, except for energies above 1500 GeV. The reason is that the weight-based corrections explained in Eq. 5.10 become less accurate in the region of large $m_{t\bar{t}}$ values, where the binwidths are chosen to be larger due to the small statistics. Still, the mean in this region is closer to zero than the one reconstructed by the loose approach, while showing a net improvement in the RMS, compared to the two analytical methods.

Finally, some statistical coefficients are computed to further compare the three reconstruction approaches. First, the weighted Pearson's correlation coefficient between the generated and the reconstructed masses, as defined in Eq. 5.7 is found to be larger for the NN approach rather than the two analytical reconstruction methods. Second, the Jensen-Shannon distance (JSD) as defined in Eq. 5.13, and last, the normalized root mean squared error as defined in Eq. 5.14 between $m_{t\bar{t}}^{\text{gen}}$ and $m_{t\bar{t}}^{\text{reco}}$ from the total testing set, are computed, showing a net improvement in the NN-based approach. The Tab. 5.6 shows the results for all these coefficients.

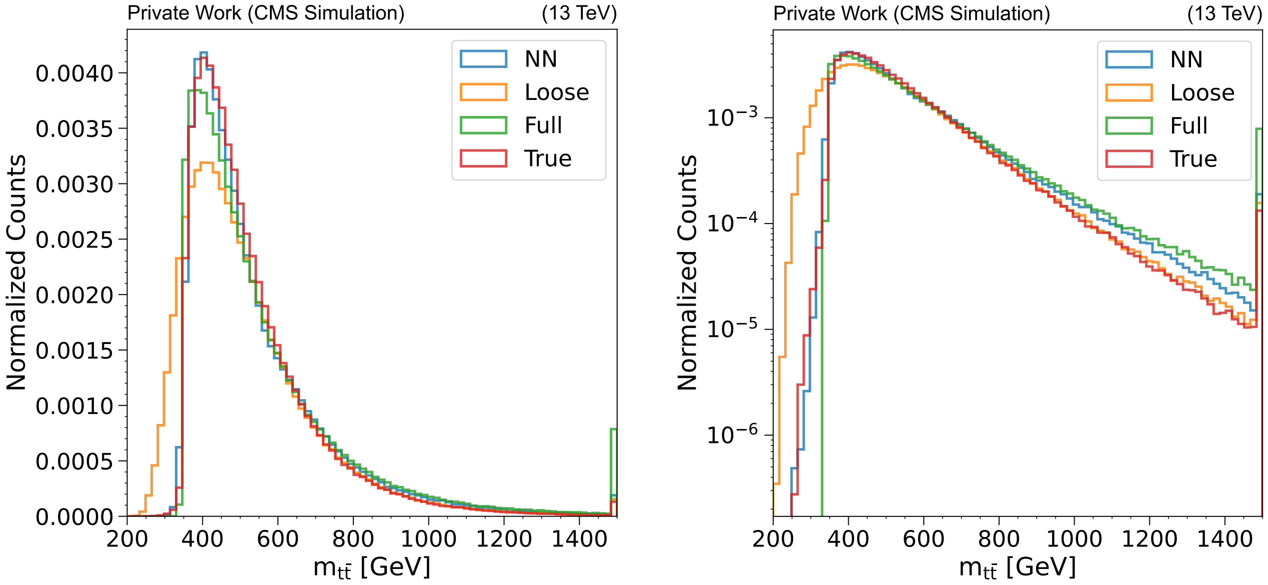


Figure 5.12: Comparison of the invariant mass distributions obtained using the three reconstruction approaches with the true distribution in linear (left) and logarithmic (right) scale.

	NN	**Loose**	**Full**
ρ	0.822	0.746	0.009
JSD	0.037	0.177	0.077
nRMSE	0.19	0.22	0.32

Table 5.6: Statistical performance for the three reconstruction methods.

5.3.1 Unfolding

One of the ultimate goals of this analysis is to perform a measurement of the differential cross section as a function of the invariant mass of the $t\bar{t}$ system. When dealing with differential cross section measurements, the events observed in a given bin differ from the generated ones due to different detector effects, such as acceptances, inefficiencies and observable smearings. As a consequence, to correct for these effects, several procedures have been introduced in order to extract the true distributions, given the observed measurements. These procedures are usually referred to as "unfolding" methods [65]. In this analysis, the matrix inversion method is used. This method starts by defining a folding equation:

$$m_i^{\text{reco}} = \sum_{j=1}^{\text{nGen}} A_{ij} m_j^{\text{gen}} \qquad A_{ij} = \frac{N_{ij}^{\text{reco}\wedge\text{gen}}}{N_j^{\text{gen}}}, \tag{5.19}$$

where m_i^{reco} is the number of reconstructed events in the bin i, m_j^{gen} is the number of generated events in the bin j and A_{ij} is defined as the ratio between the number of events generated in bin j and reconstructed in bin i to the number of events generated in bin j. This matrix is referred to as "response matrix" and in practice it is obtained from the $t\bar{t}$ signal MC. In the ideal case, the response matrix corresponds to the identity matrix. However, inefficiencies and migration effects are inevitably present. In this analysis the unfolding procedure is done from the reconstruction level after the full event selection, explained in Sec. 4.4, to the space in which only the selection on leptons is applied. The extrapolation to the full phase space can be performed by dividing the observable by an acceptance factor determined from the MC events which are either not reconstructed or rejected due to preliminary cuts (e.g. on lepton p_{T}). The response matrices obtained from the $t\bar{t}$ signal MC samples are shown in Fig. 5.16. The NN-based response matrix shows greater diagonal elements, compared to the corresponding entries of the matrices based on analytical methods with the only exception of the first bin, and a reduction of migrations between different bins.

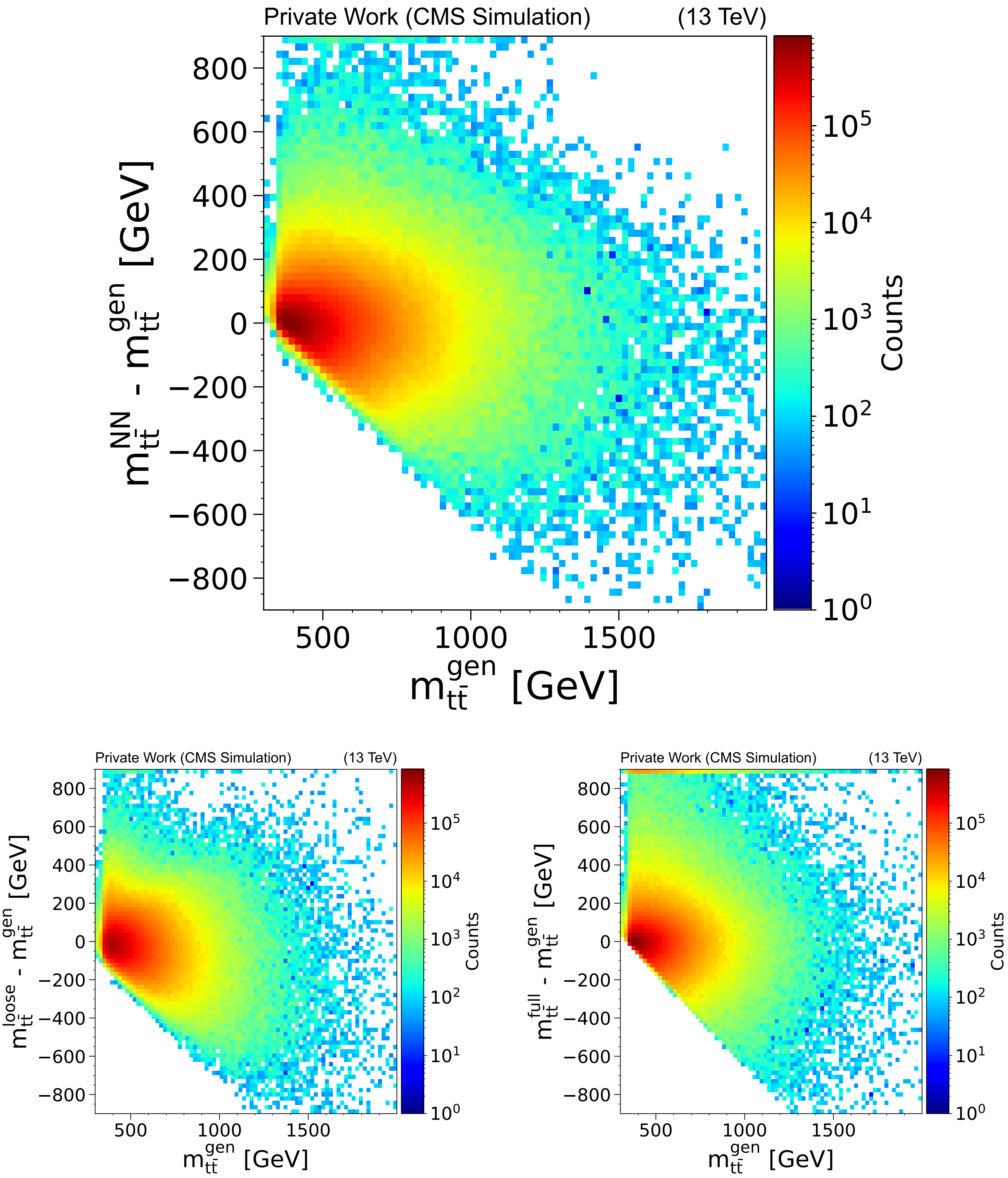


Figure 5.13: Correlations between reconstructed and true value for the three approaches. From top middle via bottom left to right: NN-based approach, loose kinematic reconstruction and full kinematic reconstruction.

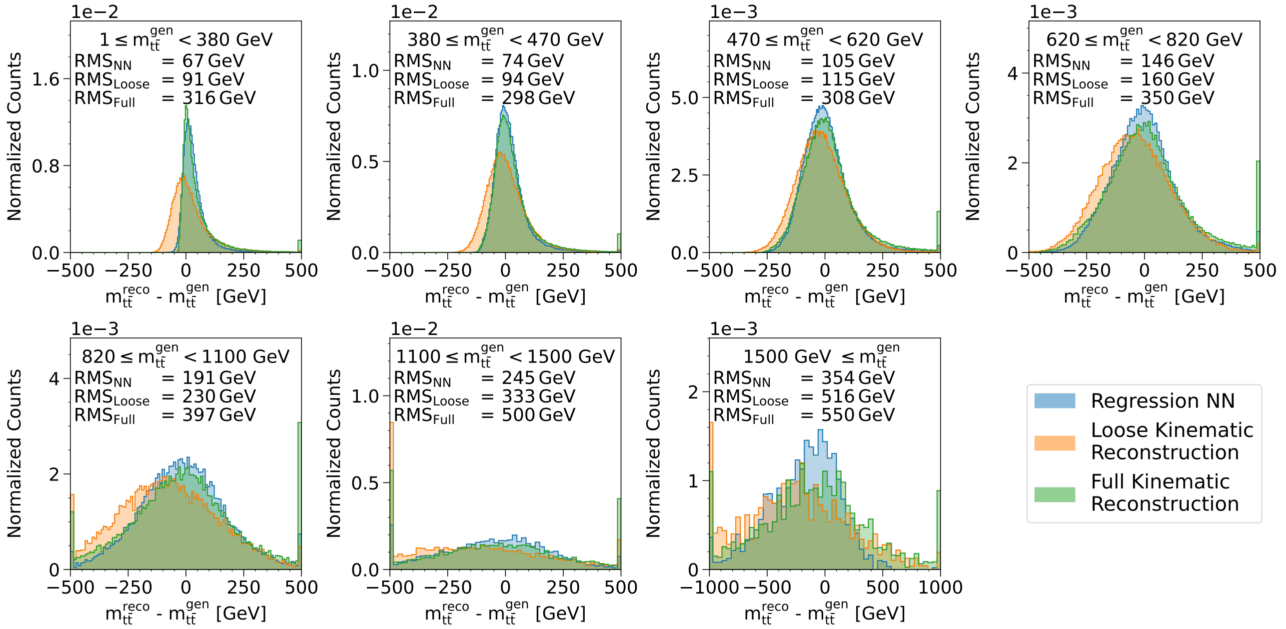


Figure 5.14: Differences between $m_{t\bar{t}}^{\text{reco}}$ and $m_{t\bar{t}}^{\text{gen}}$ for the three reconstruction approaches in seven bins of $m_{t\bar{t}}^{\text{gen}}$. The RMS values of the distributions are also shown. Events with $m_{t\bar{t}}^{\text{full}}$> 13 TeV are removed from the full kinematic distribution.

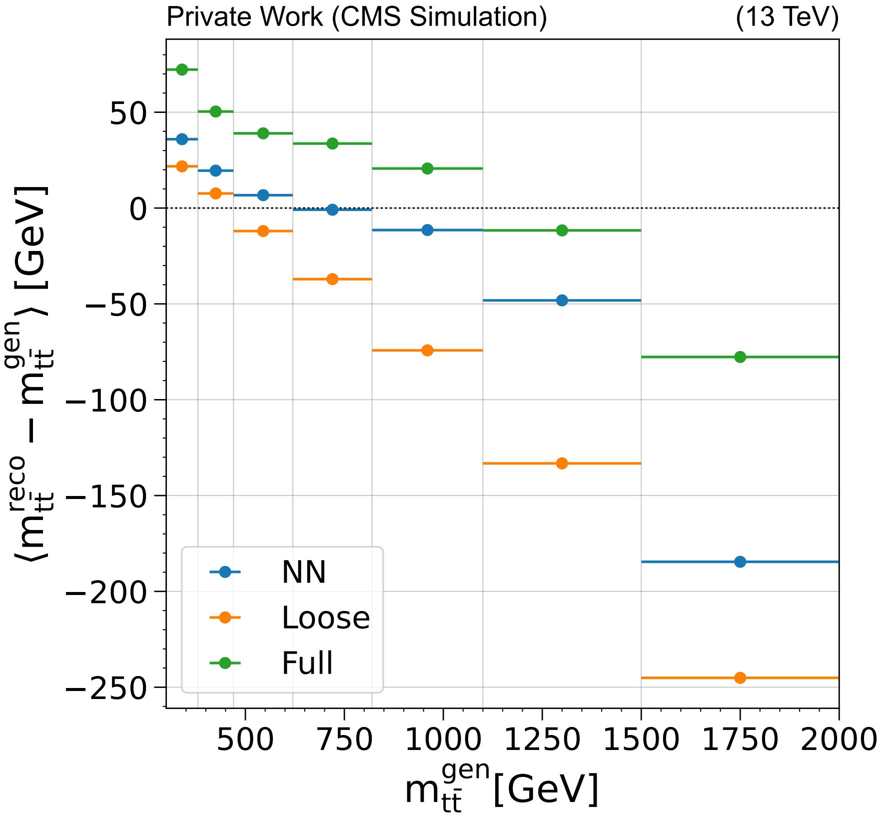
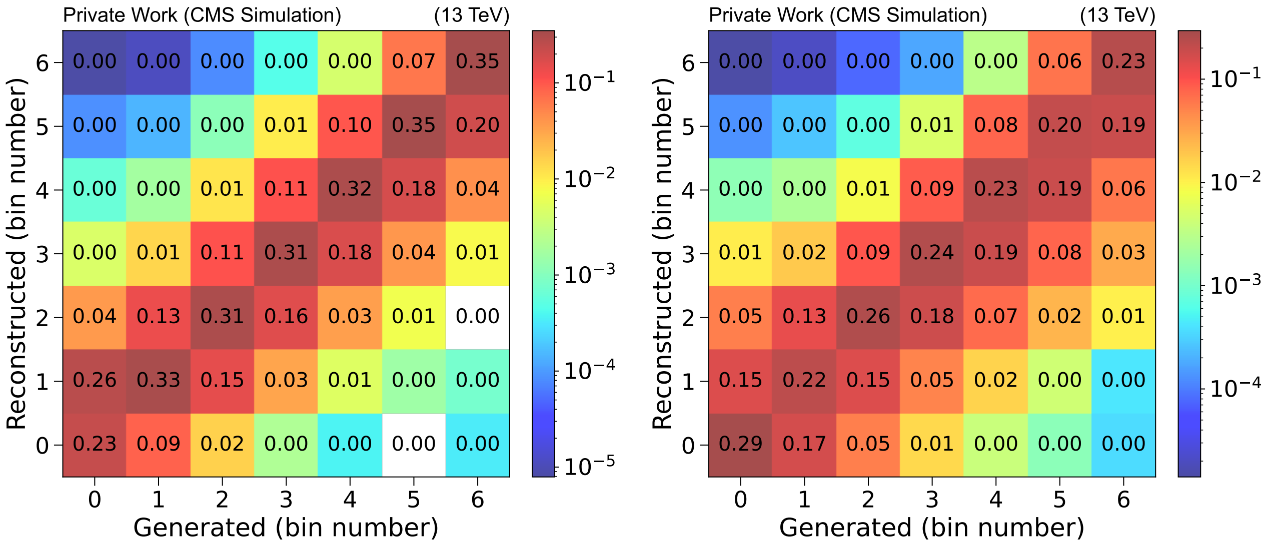


Figure 5.15: Means of the difference between predicted and reconstructed values of $m_{t\bar{t}}$. Events with $m_{t\bar{t}}^{\text{full}}$> 13 TeV are removed from the full kinematic distribution.

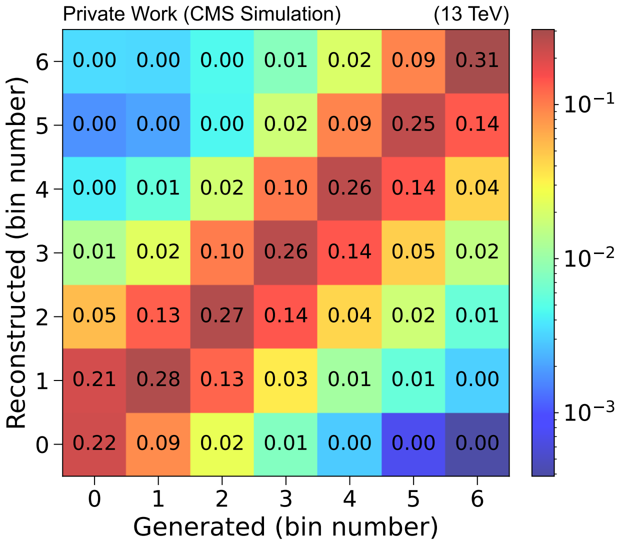
For each response matrix the migration effects can be quantified by defining two quantities, the purity (p_i) and stability (s_i) of each bin, as:

$$\mathrm{p}_i = \frac{N_i^{\text{reco}\wedge\text{gen}}}{N_i^{\text{gen}}} \qquad \mathrm{s}_i = \frac{N_i^{\text{reco}\wedge\text{gen}}}{N_i^{\text{reco}}}, \tag{5.20}$$



(a) NN-based kinematic reconstruction response matrix

(b) Loose kinematic reconstruction response matrix



(c) Full kinematic reconstruction response matrix

Figure 5.16: Response matrices for the three reconstruction approaches

where N_i^{gen}, N_i^{reco} and $N_i^{\text{reco}\wedge\text{gen}}$ are the numbers of events generated, reconstructed and both generated and reconstructed in bin i, respectively. Purity and stability for the seven bins used in this analysis are shown in Fig. 5.17. Again, with few exceptions, the purity and stability of the NN-based regression are found to be superior to those of the two analytical approaches.

After defining a response matrix, the Eq. 5.19 can be inverted to infer the generated distribution from the reconstructed one.

$$m_j^{\text{gen}} = \sum_i A_{ji}^{-1} m_i^{\text{reco}} \tag{5.21}$$

By error propagation, a covariance matrix can be defined as:

$$\text{Cov}(m_j^{\text{gen}}, m_j^{\text{gen}}) = V_{jk} = \sum_i \left(A^{-1}\right)_{ji} (\delta m_i^{\text{reco}})^2 \left(A^{-1}\right)_{ki} \tag{5.22}$$

where $(\delta m_i^{\text{reco}})^2$ is the variance matrix, a diagonal matrix with elements equal to the squared errors on the reconstructed counts in bin i. Considering only statistical uncertainties, these correspond to the Poissonian errors, that are taken to be equal to the square root of the observed number of events.

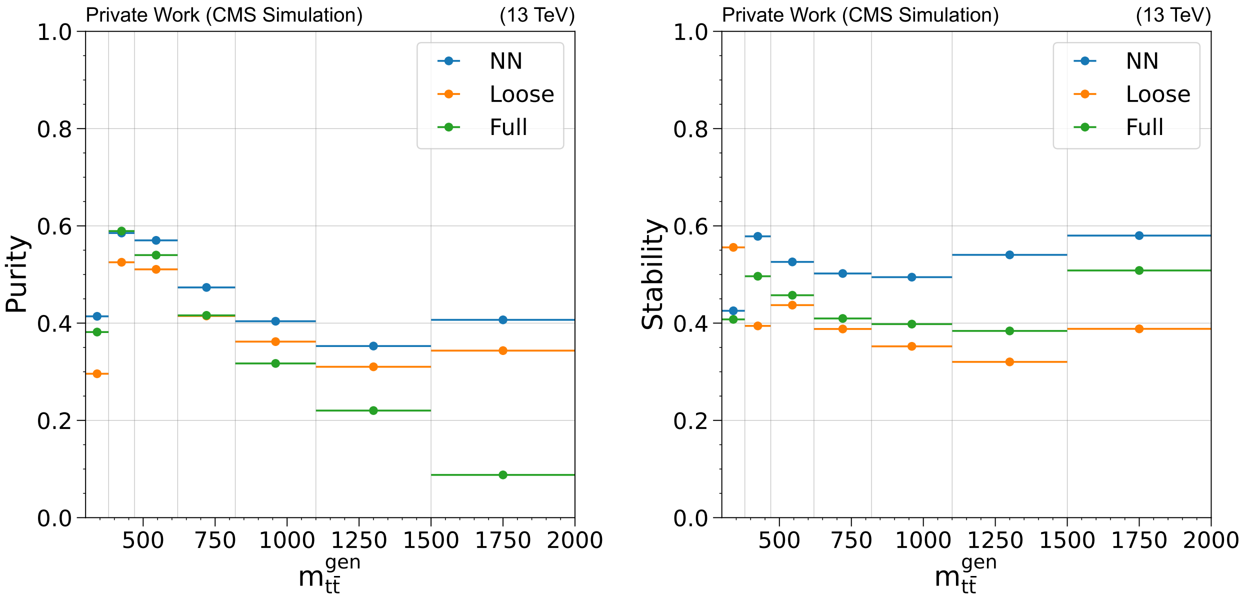


Figure 5.17: Purity and stability for the reconstruction methods.

The errors on the unfolded distributions are given by the square root of the diagonal elements of the covariance matrix:

$$\delta x_j = \sigma_j = \sqrt{V_{jj}} \tag{5.23}$$

As a cross-check, the same testing dataset used for the construction of the response matrices can be unfolded using the three different methods. The results are shown are in Fig. 5.18. As expected, the unfolded distributions are identical to the true distributions since the response matrices were defined using the same dataset. Like in the other cases, the first (last) bin accounts for underflow (overflow) events. A significant improvement in the statistical uncertainties of the unfolded cross section has been achieved using the NN approach with respect to the analytical methods. This is due to the improved efficiency and resolution as encoded in the response matrix. The improved resolution also leads to smaller bin-to-bin statistical correlations or, in other words, to a stronger reduction of the off-diagonal elements V_{jk} in Eq. 5.22. A general description of unfolding including a comprehensive discussion of binning and resolution effects can be found in Ref. [38].

The determinant of the covariance matrix of the differential cross section defined in Eq. 5.22 was introduced during the hyperparameter optimization as an alternative objective function to be minimized. The reason of this choice relies on the fact that, given a covariance matrix V, its determinant, denoted with $|V|$ or $\det(V)$, represents the multivariate analog of the one-dimensional variance σ^2, accounting also for correlation and anti-correlation effects [21]. Moreover, the determinant of the covariance matrix appears in several likelihood ratio criteria for testing hypotheses. The determinants of the covariance matrices in the three reconstruction methods are presented in Tab. 5.7, showing a relevant improvement in the NN-based approach.

	NN	**Loose**	**Full**
det(V) ($\times 10^{53}$)	3.96	780	248

Table 5.7: Determinants of the covariance matrices of the unfolded $m_{t\bar{t}}$ distribution when using the NN, loose and full kinematic reconstructions.

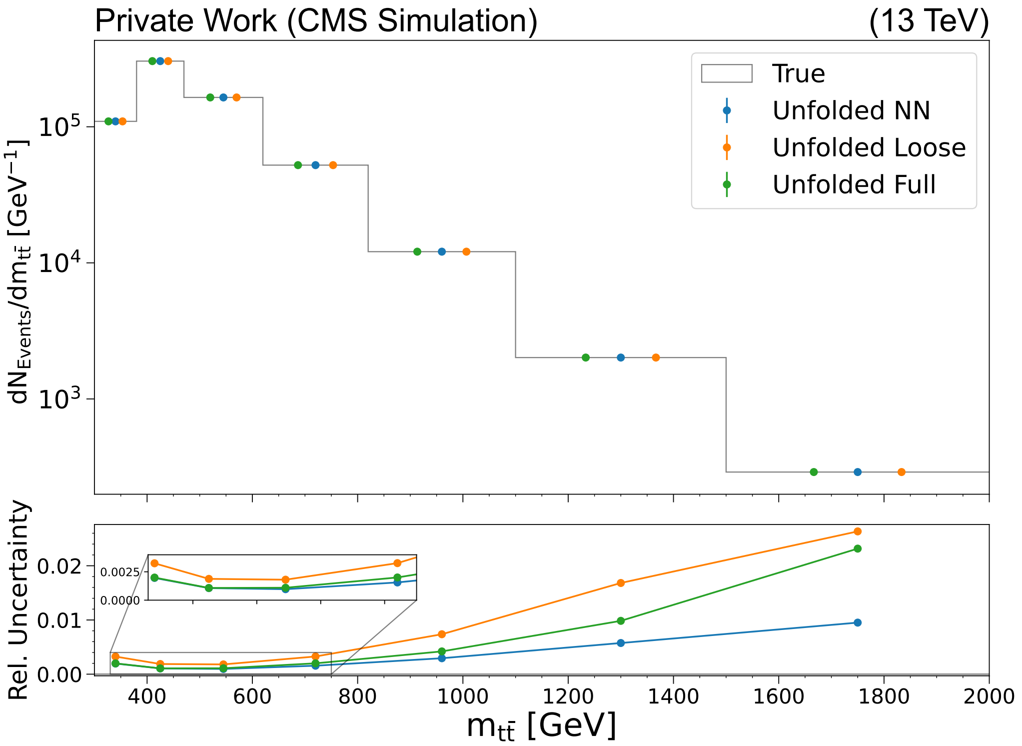


Figure 5.18: Cross-check for the definition of the response matrices. The differential cross sections are computed by unfolding the $m_{t\bar{t}}^{\text{reco}}$ distributions as reconstructed by the NN (blue), loose (orange) and full (green) kinematic reconstruction after the event selection from the same simulated $t\bar{t}$ signal MC events as used for the definition of the response matrices.

5.4 Application to 2018 data

In this section data collected by the CMS experiment in 2018 are analyzed for a total integrated luminosity of $\mathcal{L} = 59.8$ fb^{-1}. After training and testing the NN using MC signal events, the model is used to predict the values of $m_{t\bar{t}}^{\text{reco}}$ for real data events. Real data and simulated signal and background events undergo the same event selection used for the training dataset and the same scaling operation mentioned in Sec. 5.2.1 is applied to the input features of the three datasets. For each event the NN output consists of one value of $m_{t\bar{t}}^{\text{reco}}$. These values are histogrammed and simulated events are normalized to the total luminosity of data and to their corresponding cross sections as predicted by the MC generator. Results of this procedure are shown in Fig. 5.19, where the vertical bars represent the statistical uncertainties on data and the hatched area represents the systematic uncertainties on the MC simulation, that will be discussed in Sec. 5.5. Excellent compatibility between data and MC predictions is found.

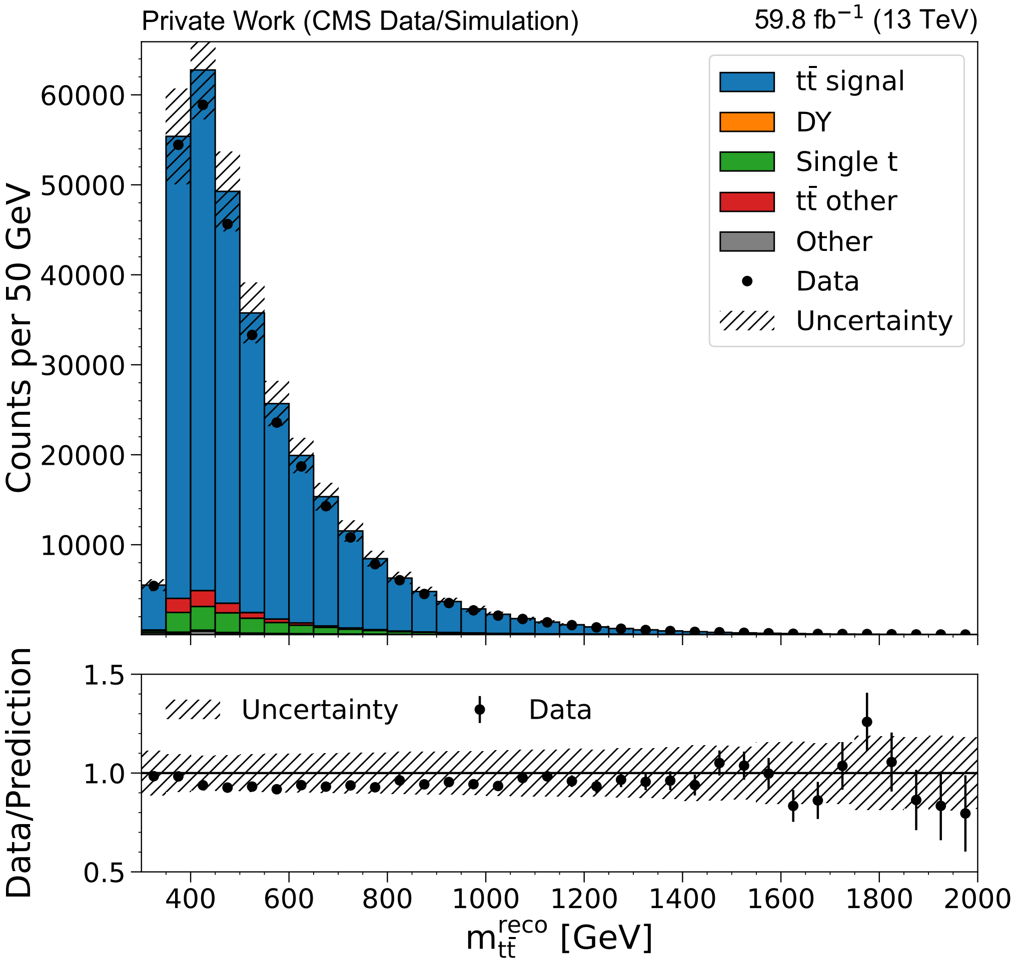


Figure 5.19: Observed $m_{t\bar{t}}$ distributions at detector level for real data (points) and estimated signal and background contributions as taken from the MC simulation. The reconstruction of $m_{t\bar{t}}$ is performed using the NN approach.

5.4.1 Unfolding of 2018 data

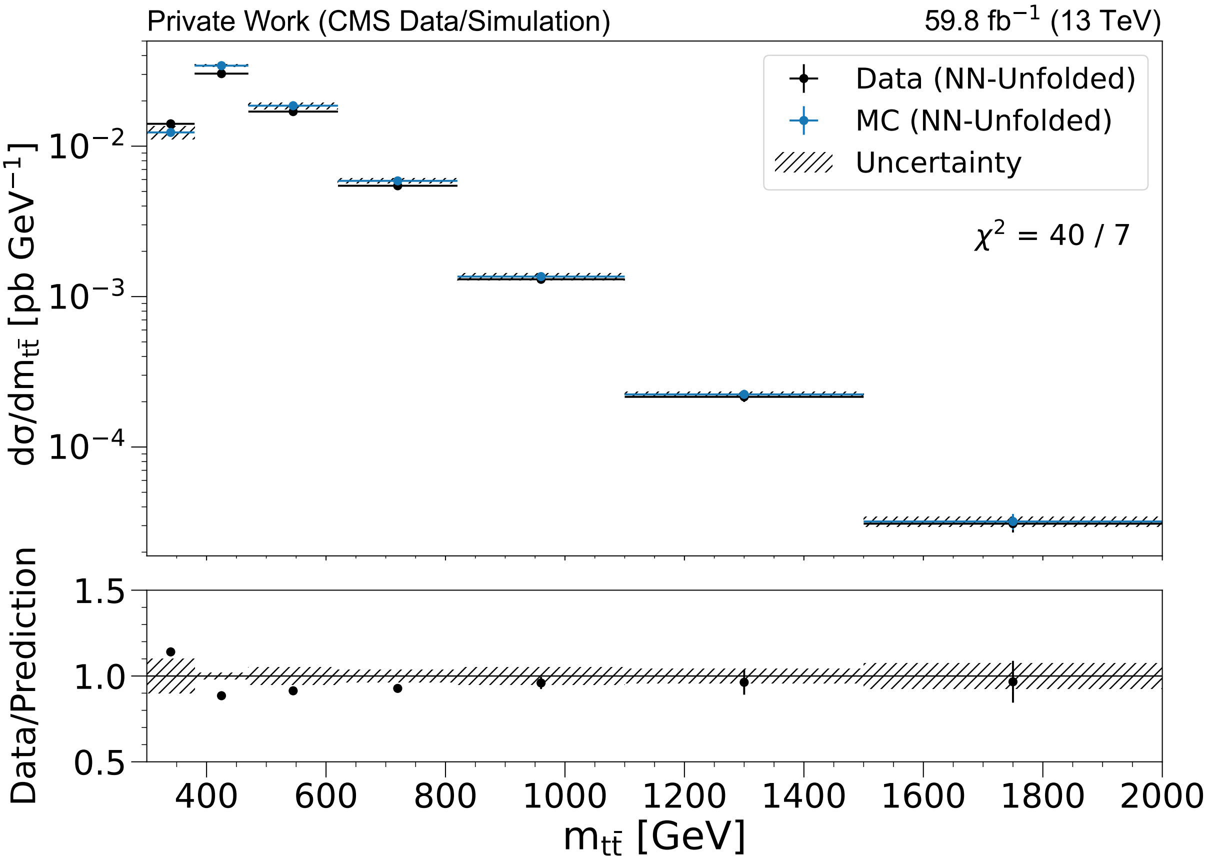
Using the response matrices defined with the MC samples, the differential cross section for $t\bar{t}$ production can be measured at the parton level (i.e. after parton showering with on-shell top quarks). As a first step, the $m_{t\bar{t}}$ distribution is reconstructed at the detector level using real data events with the same binning employed for the definition of the response matrices. Next, since the response matrices are defined for unfolding a distribution of signal events only from the detector level to the parton level, the estimated background contributions are computed from MC samples and subtracted. Finally, the distribution can be unfolded by inverting the response matrices. To obtain a differential cross section, the number of counts in each bin is normalized to its binwidth and to the luminosity:

$$\frac{d\vec{\sigma}}{dm_{t\bar{t}}} = \frac{A^{-1}\left(\vec{m}_{t\bar{t}}^{\text{reco}} - \vec{b}\right)}{\Delta_m\ \mathcal{L}}, \tag{5.24}$$

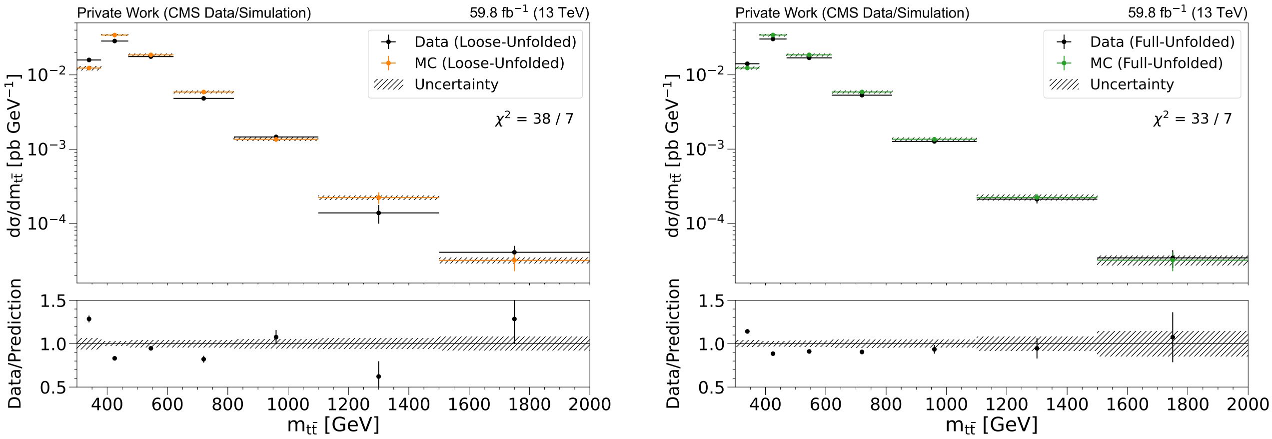
where Δ_m is the vector of bin widths. In Fig. 5.20 the unfolded differential cross sections using the three kinematic reconstruction approaches are compared to the differential cross sections obtained using the MC samples. In all three cases the background contribution to be subtracted from real data is computed with the corresponding reconstruction approach and the unfolding of the MC differential cross section is done with the same response matrix used for real data. The dataset used for the differential cross section predicted by the MC simulation is different from the dataset used for the definition of the response matrix. The errors δy appearing in the Eq. 5.22 are computed by error propagation considering the subtraction of background counts to the total number of collected events, accounting for the statistical component only. The systematic uncertainties in the unfolded cross section will be discussed in Sec. 5.5. A very good agreement is found with the theoretical predictions in all three cases. In particular, a significant reduction of the statistical uncertainty is found in the last bins of the NN-based approach compared to the other methods. The relatively large values of the χ^2, whose computation is explained in Sec. 5.5, are justified by the fact that the systematic uncertainties taken into account are not complete. More details are provided in Sec. 5.5. Finally, the extraction of the differential cross section to the full phase space can be done by dividing the obtained unfolded cross section by the branching ratio of the considered process and by an acceptance factor accounting for the kinematic cuts on the lepton reconstruction, out-of-acceptance and other rejected events (e.g. primary vertex not satisfying quality criteria). Overall, the unfolded cross section to the full phase space can be determined as:

$$\frac{d\vec{\sigma}}{dm_{t\bar{t}}} = \frac{A^{-1}\left(\vec{m}_{t\bar{t}}^{\text{reco}} - \vec{b}\right)}{\Delta_m\ \mathcal{L}\ \text{BR}\ \epsilon} \tag{5.25}$$

where BR is the branching ratio and ϵ the acceptance factor. The fraction has to be evaluated component by component of the vector at the numerator and denominator. The obtained results for the three reconstruction approaches and the comparison of the relative statistical uncertainties are shown in Fig. 5.21, where a clear improvement can be observed in the NN method, especially in the region of high $m_{t\bar{t}}$ values. As a final cross-check, by integrating the differential cross section using the MC samples, the original total inclusive cross section is found.



(a) Unfolded differential cross section using the NN-based kinematic reconstruction



(b) Unfolded differential cross section using the loose kinematic reconstrucion

(c) Unfolded differential cross section using the full kinematic reconstrucion

Figure 5.20: Comparison of the unfolded cross sections using the three considered kinematic reconstructions. The black points represent the data with vertical bars showing the statistical uncertainties, obtained from the diagonal elements of the covariance matrices of the unfolding procedure, as defined in Eq. 5.22. The hatched area represents the systematic uncertainties, which are obtained from the diagonal elements of the systematic covariance matrices, as explained in Sec. 5.5.1. The coloured points show the predictions from the POWHEG+PYTHIA MC generator as obtained from applying a similar unfolding procedure as used for the data to the MC datasets.

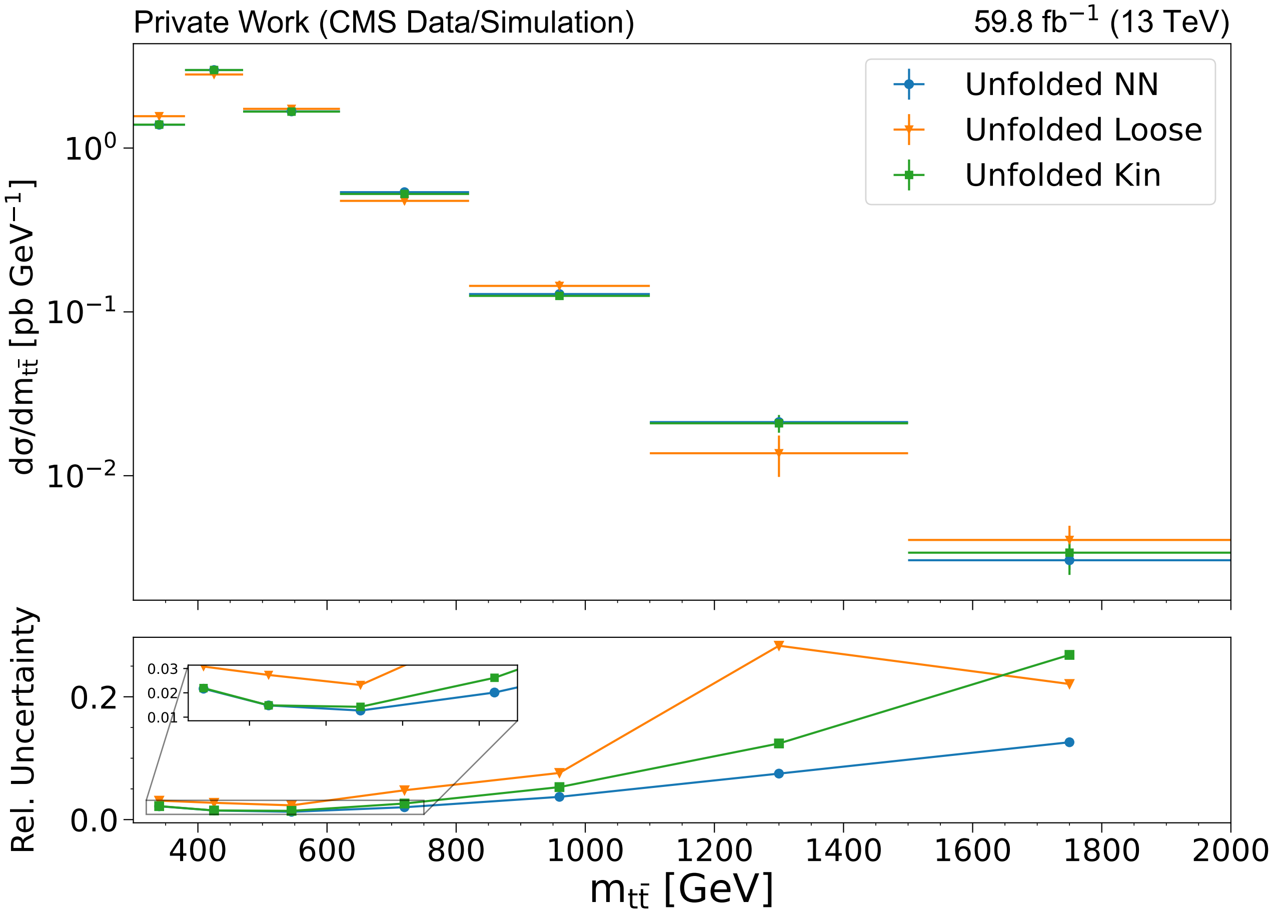


Figure 5.21: Comparison of the unfolded cross sections using the three considered kinematic reconstructions and the relative uncertainties.

5.5 Systematic uncertainties

The systematic uncertainties accounted for this analysis consist in a reweighting of the MC samples either by varying the SFs defined in Sec. 4.3.2 or by applying additional multiplicative factors in the weight definition. The list of the systematic uncertainties is here presented:

- **Pileup**. This uncertainty is computed by varying the simulated pileup distribution by performing a ±4.6% variation of the minimum bias cross section. This variation would also affect the total cross section for $t\bar{t}$ production, leading to an overestimation of the uncertainty. As a consequence, for this and other uncertainties in the following, the total $t\bar{t}$ production cross section is kept fixed by applying a correction factor to the varied distributions.

- **Trigger efficiency**. The corresponding SF defined to mitigate the differences between trigger efficiencies in data and simulation is varied within its uncertainty.

- **Lepton identification, isolation, reconstruction**. To measure the systematic uncertainties in the determination of the SFs, each SF is varied within its uncertainties individually for both the lepton flavor (e/μ) and type (identification/isolation/reconstruction).

- **L1 muon prefiring**. The corresponding SF derived for this correction is shifted up and down by its uncertainty.

- **b tagging**. The SFs accounting for the differences of b-tagging efficiencies and mistagging rates between data and simulation are varied within their estimated uncertainties. Additionally, separated uncertainties are considered for heavy-flavor (b and c) jets and light-flavor jets. Heavy-flavor uncertainties are split into four subsources which are: *PILEUP*, *JES*, *TYPE3*, and *STATISTIC*, where the *TYPE3* uncertainty accounts for SF derivation method and the modeling.

- **PDF**. The uncertainty on the parton distribution functions (PDF) is propagated by shifting up and down the eigenvectors of the NNPDF3.1 NNLO PDF set by the corresponding uncertainties. Moreover, the value of α_{S} is also independently shifted up and down by its uncertainty.

- **Renormalization and factorization scales**. Renormalization μ_{R} and factorization μ_{F} scales are varied independently by a factor two up and down with respect to their nominal values. Simultaneous variations are also considered.

- **b quark fragmentation**. The fragmentation of a b quark into b hadrons is described by the Bowler-Lund function, which depends on the Bowler-Lund parameter set to 1.056 [58]. This parameter is varied within its uncertainty. Alternatively, the use of the Peterson fragmentation function is used and the difference between the nominal case is considered as an additional systematic uncertainty.

- **B semileptonic BR**. The uncertainties originating from the semileptonic branching fractions of b hadrons are estimated by varying those within their uncertainties. Values and uncertainties are taken from [76].

- **Top quark p_{T} reweighting**. Differential $t\bar{t}$ measurements at $\sqrt{s} = 13$ TeV have shown that the predicted kinematic spectrum by the MC generator POWHEG is harder than what observed [85]. This additional uncertainty, which is estimated by reweighting the simulation to data, accounts for this difference.

- **ME-PS matching** The uncertainty on the matching of the ME to PS calculations is evaluated by means of dedicated POWHEG + PYTHIA 8 simulations by shifting up and down the h_{damp} value by its uncertainty: $h_{\mathrm{damp}} = 1.379^{+0.926}_{-0.5052}\ m_t$.

- **Luminosity uncertainty**. The uncertainty on the integrated luminosity corresponding to 2018 data samples is 2.5% [70].

- **Initial- and Final-State Radiation**. The variation of the 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 is considered to account for uncertainties in the modeling of the PS.
- **Correction due to varying b-tagging efficiency**. The variation of some scale factors might affect the b-tag efficiency, making it different from the one computed in the nominal $t\bar{t}$ MC simulation. For a given variation of a systematic uncertainty j, an effective b-tag SF ($\mathrm{SF}_{\mathrm{eff}}$) is introduced as a function of the jet flavor, jet p_T and jet η as:

$$\mathrm{SF}^{j}_{\mathrm{eff}} = \mathrm{SF}_{\mathrm{b\text{-}tag}} \frac{\epsilon^{\mathrm{nominal}}_{\mathrm{MC}}}{\epsilon^{\mathrm{syst},\,j}_{\mathrm{MC}}} \tag{5.26}$$

  where $\epsilon^{\mathrm{nominal}}_{\mathrm{MC}}$ and $\epsilon^{\mathrm{syst},\,j}_{\mathrm{MC}}$ are the b-tag efficiencies in the nominal case and considering a given variation.

In all the control plots presented in this thesis, the contribution of a single variation of a systematic source is computed as a bin-by-bin difference of the histogram in the nominal case and the histogram considering the variation of the weights. The listed systematic uncertainties are taken into account for the signal only, while the background is kept in the nominal case. All the systematic contributions are assumed to be uncorrelated between the sources discussed in the above list, while for each single source the systematic effect on the cross sections is treated as fully correlated between different bins. The total systematic uncertainty shown in the control plots is computed as the square root of the sum of the squares of the individual contributions.

The effects of the systematic sources on the unfolded cross section depend on the kinematic reconstruction approach and on the bin of measurement. However, the most dominant contributions can be identified to be the b quark fragmentation, which is responsible for the larger systematic uncertainty in the first bin of the NN-based differential cross section, b-tag efficiency, initial and final state radiations, especially for the analytical approaches in the region of high $m_{t\bar{t}}$, μ_{R} and μ_{F} variations, and top p_{T}.

Some relevant additional systematic uncertainties not considered in this analysis are:

- **JES** (Jet Energy Scale) refers to the correction applied to the measured energy of a jet in order to account for various effects that can lead to discrepancies between the measured energy and the true energy of the jet. These discrepancies can stem from different origins, including uncertainties in calibration, and deviations from linearity. These corrections are usually parameterized as a function of the jet p_{T}, η and flavor. The variations of these corrections within their uncertainties allow to estimate the systematic contribution arising from JES, which is expected to be one of the most dominant contributions according to Ref. [85].
- **JER** (Jet Energy Resolution). Similarly to JES, JER determines the precision with which the energy of a jet can be measured and the variation of a corresponding SF accounts for its uncertainty.
- **Color reconnection** (CR) is the process by which color charges are exchanged and rearranged between quarks and gluons during the process of hadronization. Variations in the model assumptions used in PYTHIA 8 are considered as source of systematic uncertainties.
- **Unclustered $p_{\mathrm{t}}^{\mathrm{miss}}$** refers to the recalculation of $p_{\mathrm{T}}^{\mathrm{miss}}$ varying the deposited energy from the charged and neutral hadrons and photons according to the corresponding energy resolutions.
- **Underlying event (UE) tune**. The parameters characterizing the CP5 tune used to model the UE are varied up and down within their uncertainties.
- **Background Normalization**. The normalization of the background depends on the inclusive cross section of the considered process. The uncertainty on the normalization of background processes from $t\bar{t}$ events resulting in a different final state than the studied one can be evaluated by varying the inclusive $t\bar{t}$ production cross section within its uncertainty. Following previous

analyses [71], the uncertainty on the normalization of all other background processes are obtained by varying the corresponding normalization up and down by ±30%.

5.5.1 Covariance matrix definition for systematic uncertainties

In the determination of the unfolded cross section the statistical uncertainties are taken from the square root of the diagonal elements of the covariance matrix as defined in 5.22. In order to include the systematic uncertainties in the analysis a systematic covariance matrix $\mathrm{Cov}_{ij}^{\mathrm{syst}}$ is defined as:

$$\mathrm{Cov}_{ij}^{\mathrm{syst}} = \sum_{k,l} = \frac{1}{N_k} C_{j,k,l} C_{i,k,l}, \qquad 1 \leq i \leq N, \qquad 1 \leq j \leq N, \tag{5.27}$$

where the term $C_{j,k,l}$ represents the difference of the counts in the bin j of the unfolded cross section between the nominal case and the case in which the response matrix was built after reweighting the generated and reconstructed distributions according to a variation l (e.g. up or down) of the systematic uncertainty source k. When a systematic uncertainty source has more than one variation, the term N_k denotes the total number of variations of the source k. For example, all the systematic uncertainties which are computed by varying up and down a factor within its uncertainty have $N_k = 2$. The hatched area shown in Fig. 5.20 corresponds the square root of the diagonal elements of the systematic covariance matrix. The plots of the relative difference in the unfolded cross sections for each systematic variation are shown in Appendix B.

After computing the systematic covariance matrix, the theory to data comparison has been evaluated with χ^2 values, defined as:

$$\chi^2 = D^T \mathrm{Cov}^{-1} D, \tag{5.28}$$

where D is a vector of differences between the measured cross section and the corresponding prediction given by the MC simulation in all the considered bins and the matrix Cov is defined as:

$$\mathrm{Cov} = \mathrm{Cov}^{\mathrm{unf}} + \mathrm{Cov}^{\mathrm{syst}}, \tag{5.29}$$

where $\mathrm{Cov}^{\mathrm{unf}}$ is the covariance matrix resulting from the unfolding procedure as defined in Eq. 5.22, while $\mathrm{Cov}^{\mathrm{syst}}$ is the systematic covariance matrix defined in Eq. 5.27.

6

Conclusions

In this thesis, the differential cross section of the $t\bar{t}$ production has been measured using data collected in pp collisions at $\sqrt{s} = 13$ TeV with the CMS detector in 2018, corresponding a total integrated luminosity of 59.8 fb^{-1}. The cross section is measured as a function of $m_{t\bar{t}}$ using the reconstructed signal distribution at detector level and unfolding it to the parton level via matrix inversion. Events where the top quark pairs decay into the dileptonic different-flavor final state ($e^{\pm}\mu^{\mp}$) are considered.

A novel kinematic reconstruction is developed for the measurement of $m_{t\bar{t}}$ from the visible detector objects based on the use of a fully connected neural network (NN). The model has been trained using MC signal events, generated with POWHEG at NLO, after verifying the compatibility between simulations and data. The performance of the NN-based kinematic reconstruction are found to improve the efficiency, the resolution and the statistical uncertainties in the unfolded cross section compared to analytical kinematic reconstructions. In particular, a net improvement in the unfolded cross section has been observed in the region of large $m_{t\bar{t}}$ values, where the contribution from statistical uncertainties is still relevant, as verified from previous analyses.

A non exhaustive list of sources of systematic uncertainties has been taken into account in the computation of the differential cross section and additional sources of systematic contributions have been identified to be included. As a future perspective, the use of the NN-based kinematic reconstruction should be tested with a different ME generator to verify the robustness of the regression versus the underlying physics model of the generator.

A

Control plots for input features

Here, all the control plots for the comparison of data and MC are presented for the 35 input features used in this analysis. The angular variables η and φ have to be meant after the rotation explained in Sec. 5.2.1. The MC simulated distributions have been normalized to the luminosity of data and to the cross section of the corresponding process.

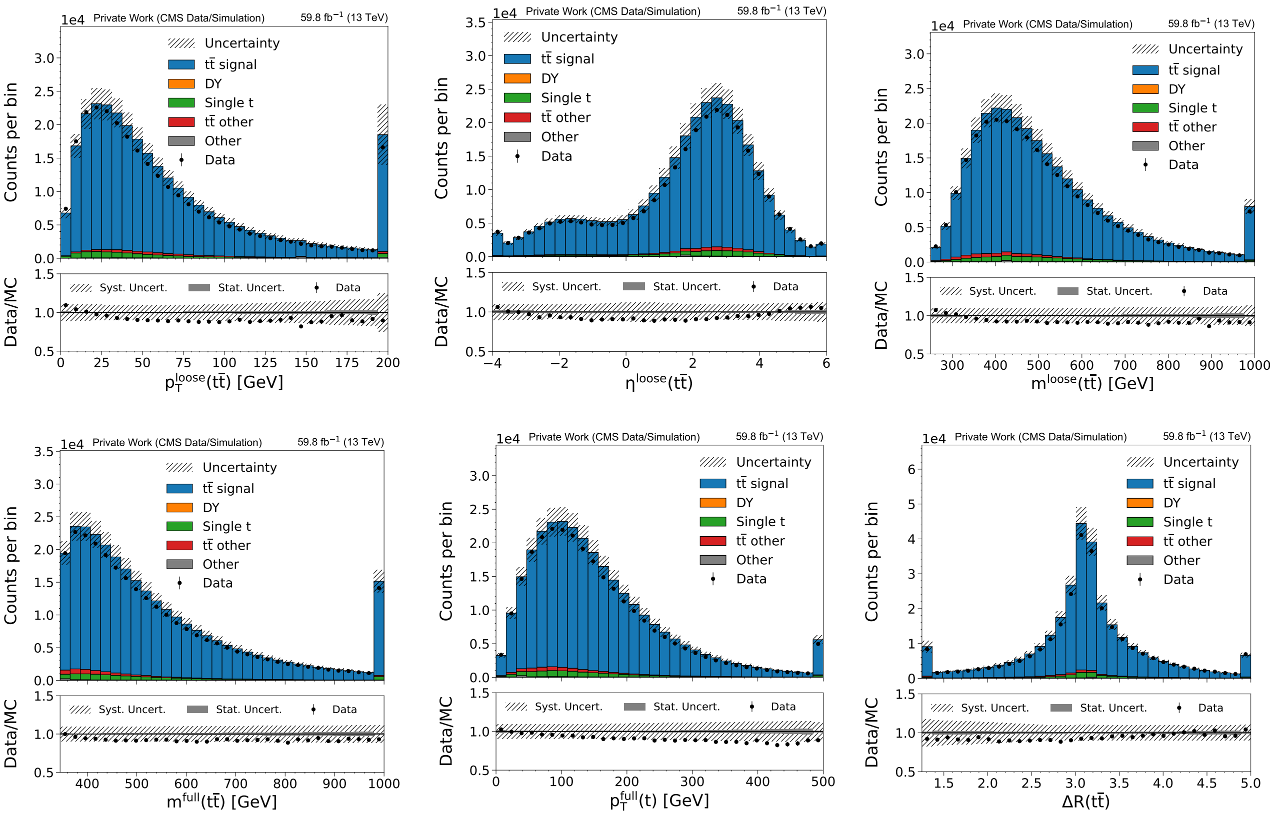


Figure A.1: Control plots of input features (first part). The hatched area indicates the systematic uncertainty on the MC simulation, the grey filled area is the statistical uncertainty on the MC simulation, which is computed assuming Poissonian errors on the MC signal and background counts and by propagating these uncertainties to their sum. Black dots with vertical error bars indicate the data counts with the statistical uncertainties.

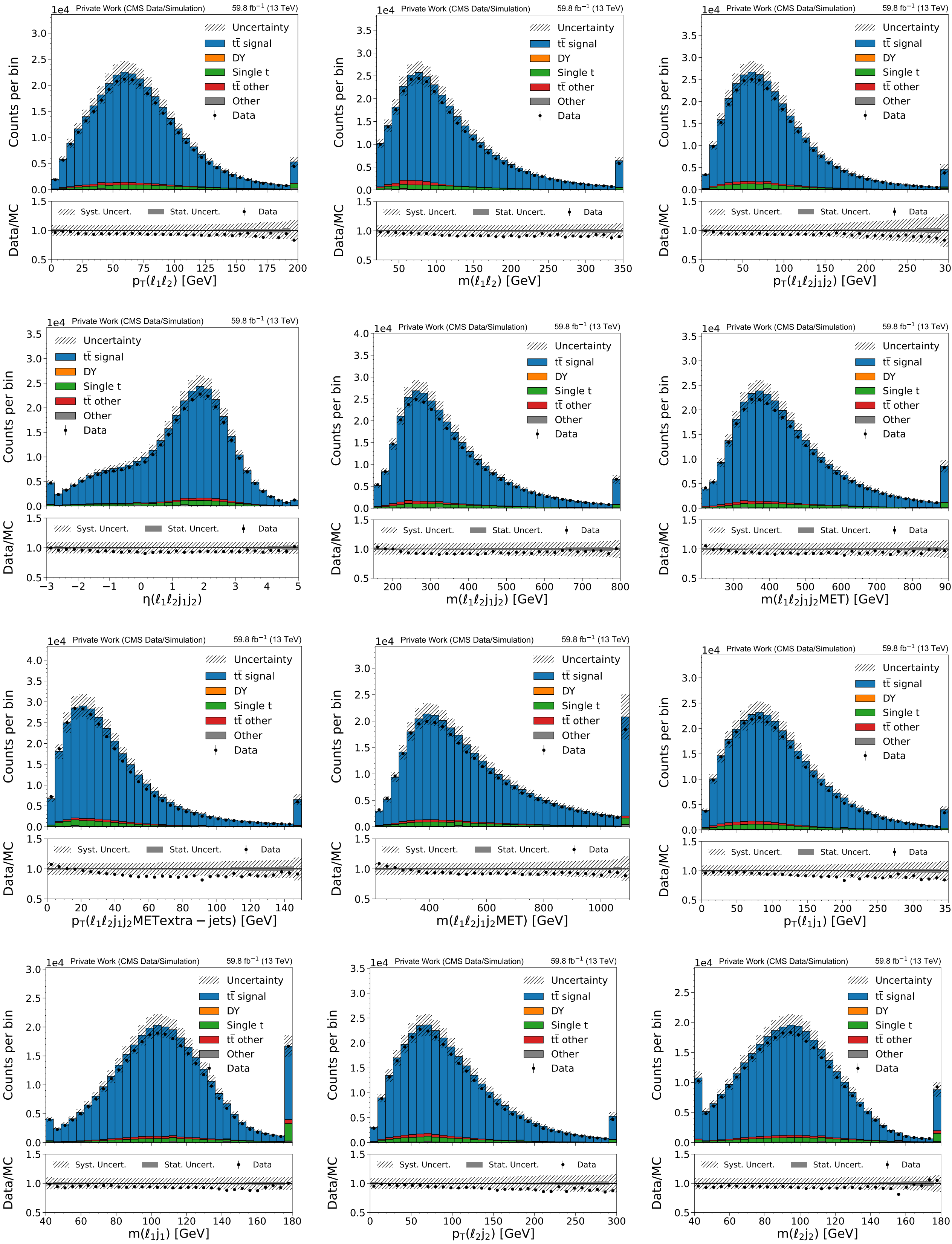


Figure A.2: Control plots of input features (second part). The hatched area indicates the systematic uncertainty on the MC simulation, the grey filled area is the statistical uncertainty on the MC simulation, which is computed assuming Poissonian errors on the MC signal and background counts and by propagating these uncertainties to their sum. Black dots with vertical error bars indicate the data counts with the statistical uncertainties.

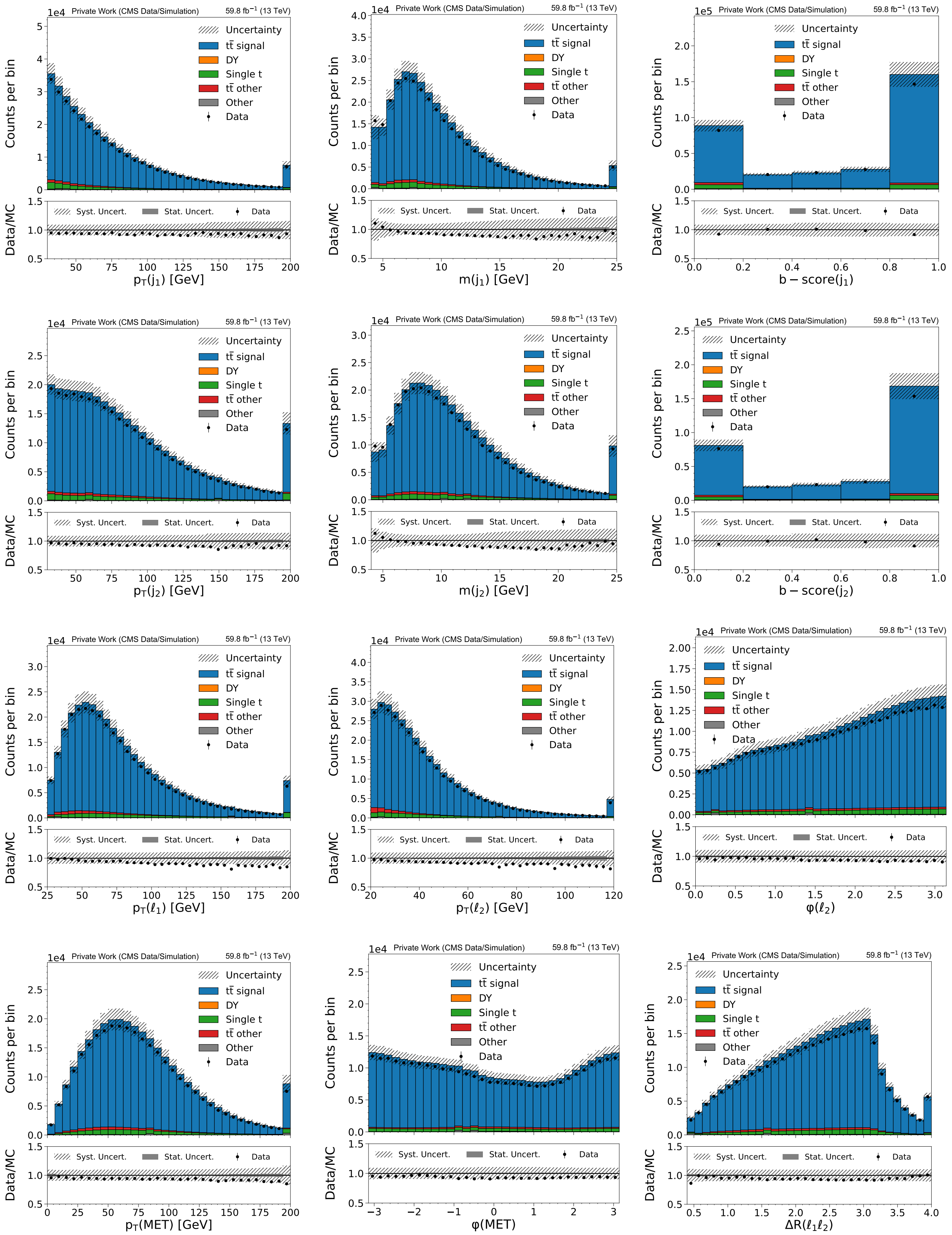


Figure A.3: Control plots of input features (third part). The hatched area indicates the systematic uncertainty on the MC simulation, the grey filled area is the statistical uncertainty on the MC simulation, which is computed assuming Poissonian errors on the MC signal and background counts and by propagating these uncertainties to their sum. Black dots with vertical error bars indicate the data counts with the statistical uncertainties.

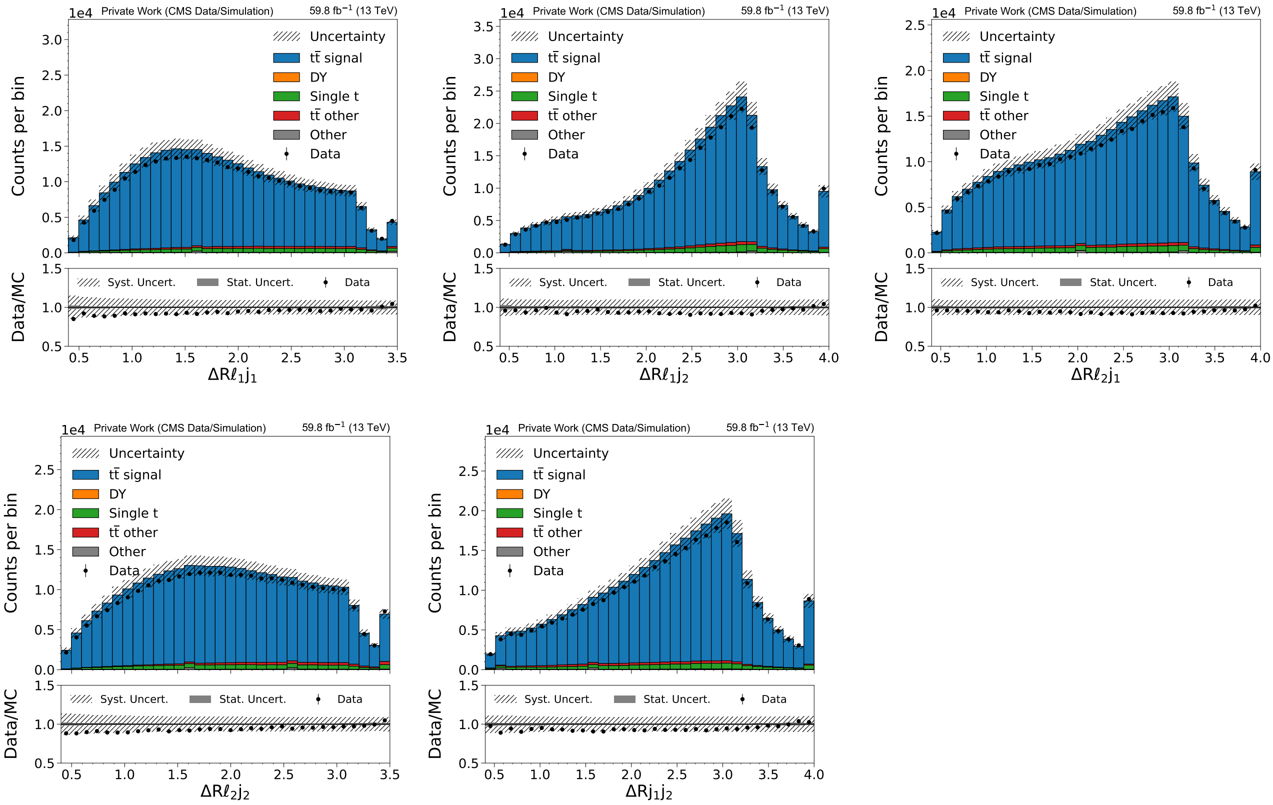


Figure A.4: Control plots of input features (fourth part). The hatched area indicates the systematic uncertainty on the MC simulation, the grey filled area is the statistical uncertainty on the MC simulation, which is computed assuming Poissonian errors on the MC signal and background counts and by propagating these uncertainties to their sum. Black dots with vertical error bars indicate the data counts with the statistical uncertainties.

B

Contributions of Systematic Uncertainties

In this section, the relative differences in the unfolded cross section between the nominal case and the case in which the response matrix was obtained by reweighting the MC simulations according to a given variation are shown.

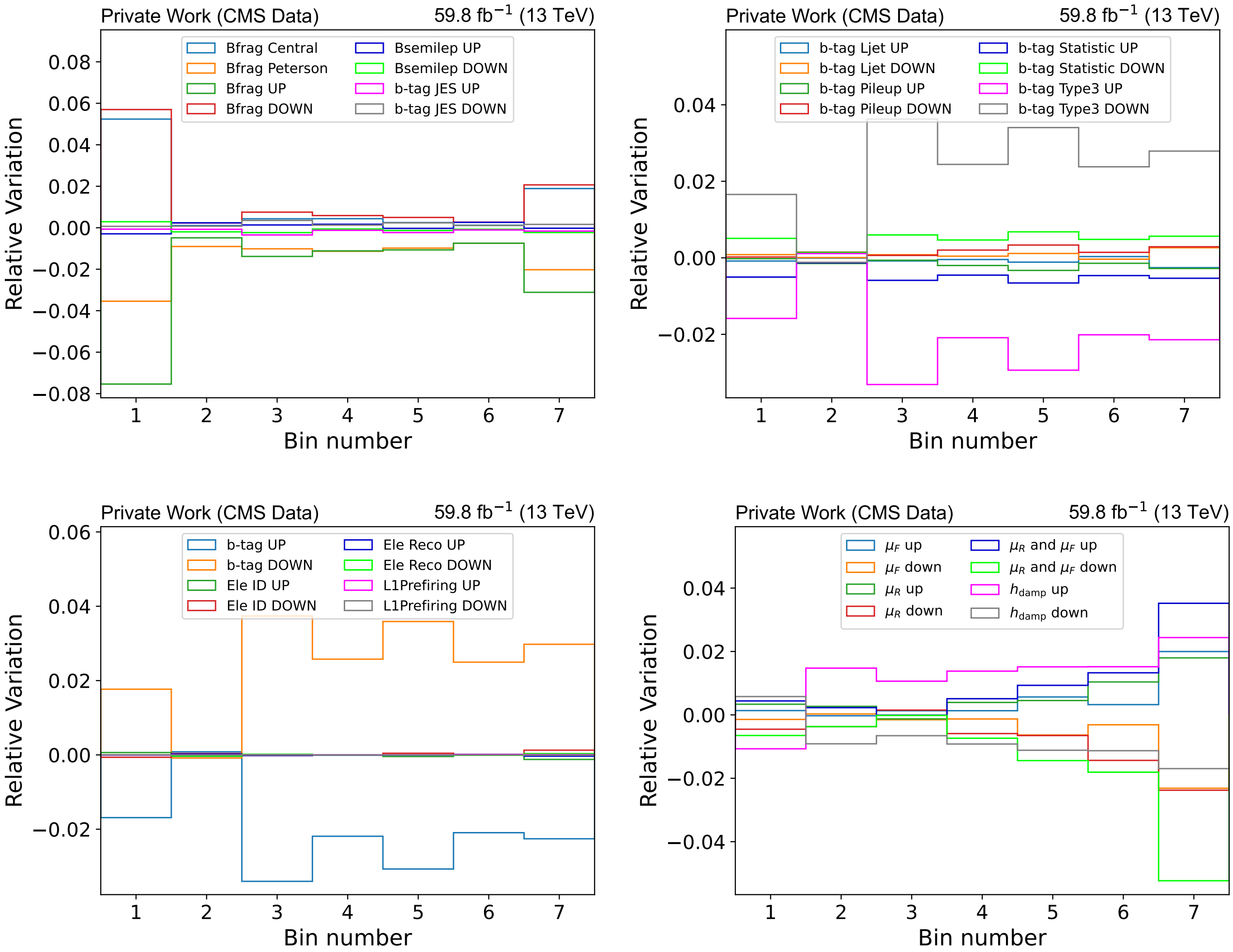


Figure B.1: Relative variations of each bin in the unfolded differential cross section with respect to the nominal case considering the systematic uncertainties using the NN-based method (first part).

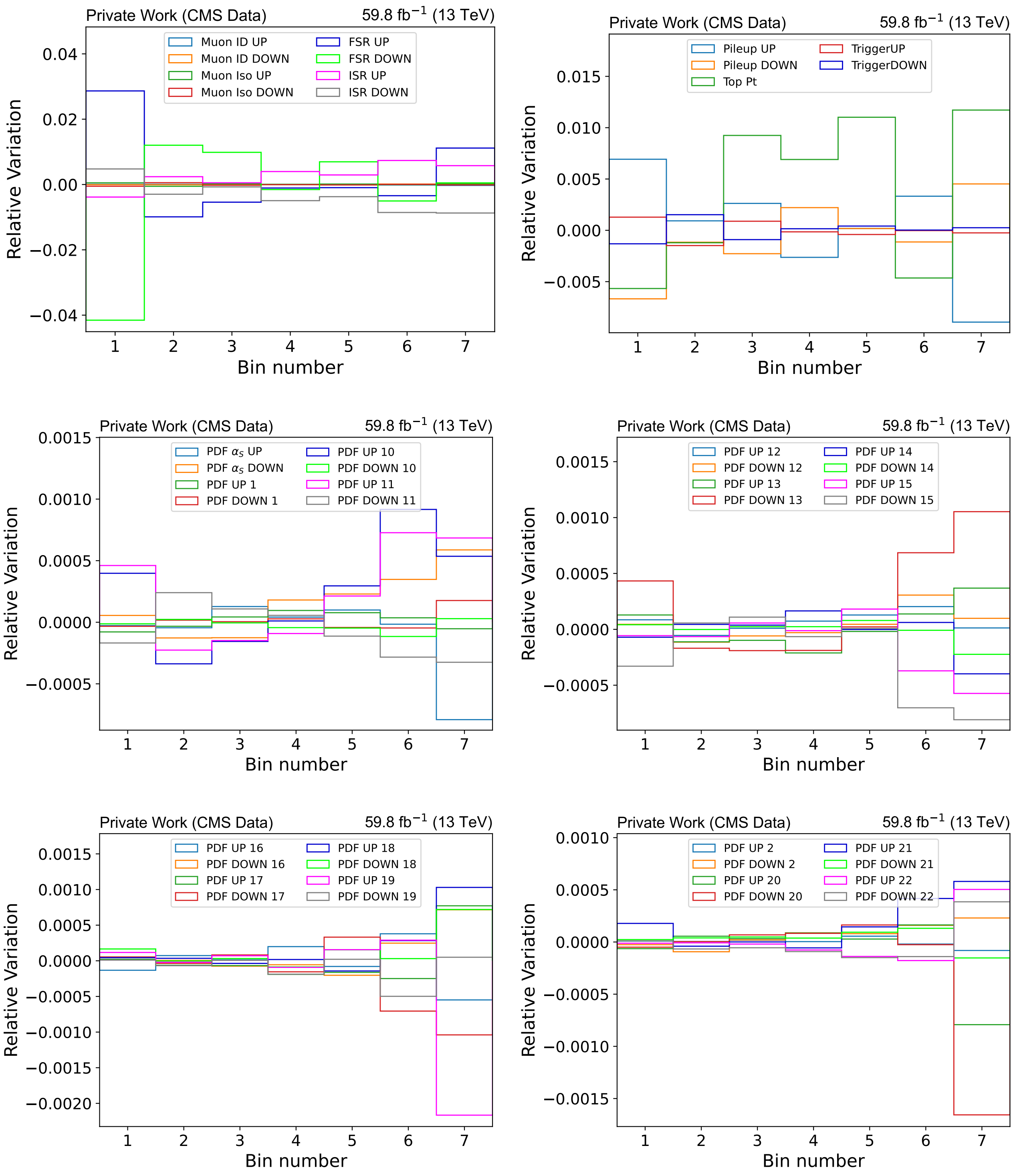


Figure B.2: Relative variations of each bin in the unfolded differential cross section with respect to the nominal case considering the systematic uncertainties using the NN-based method (second part).

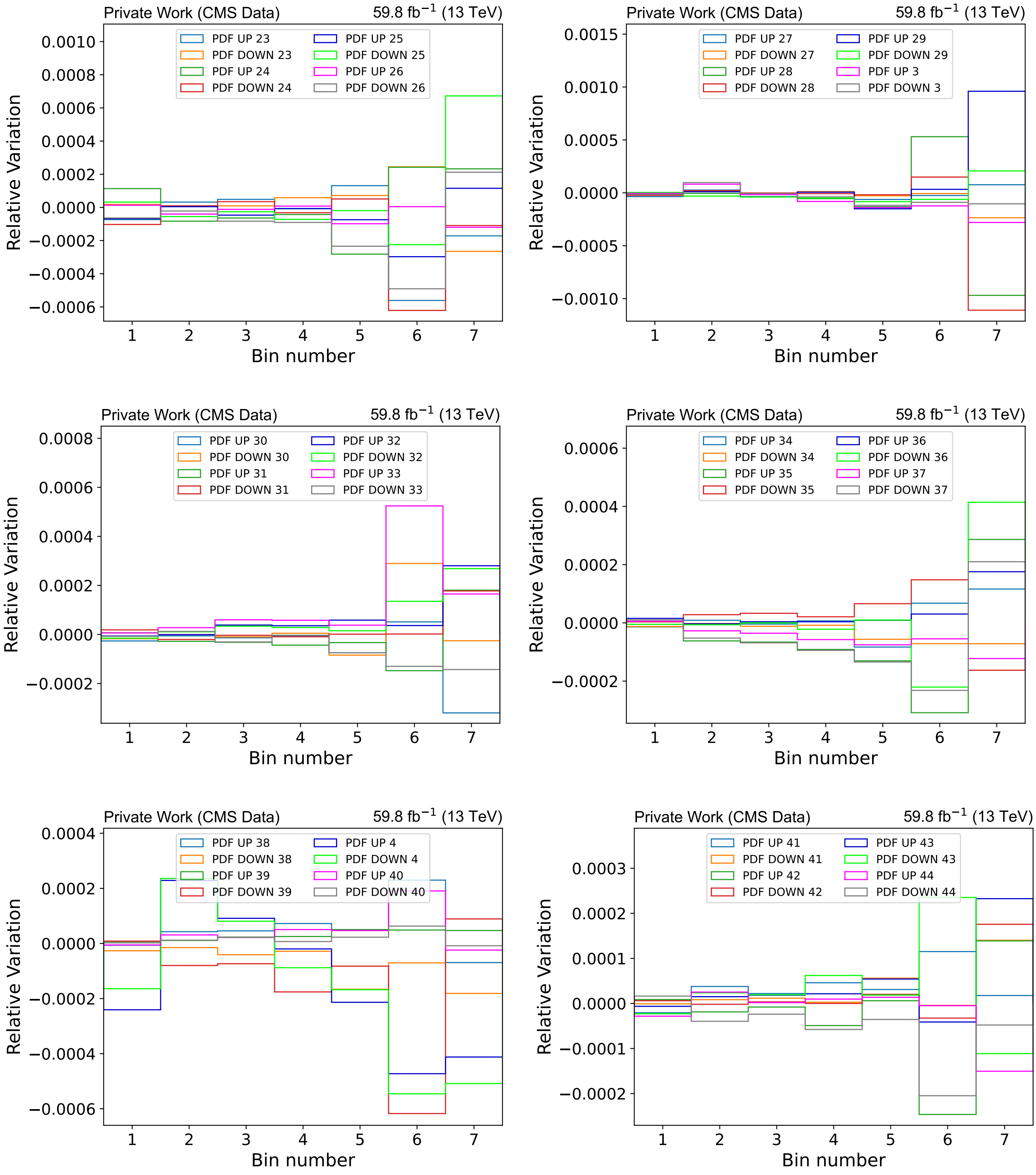


Figure B.3: Relative variations of each bin in the unfolded differential cross section with respect to the nominal case considering the systematic uncertainties using the NN-based method (third part).

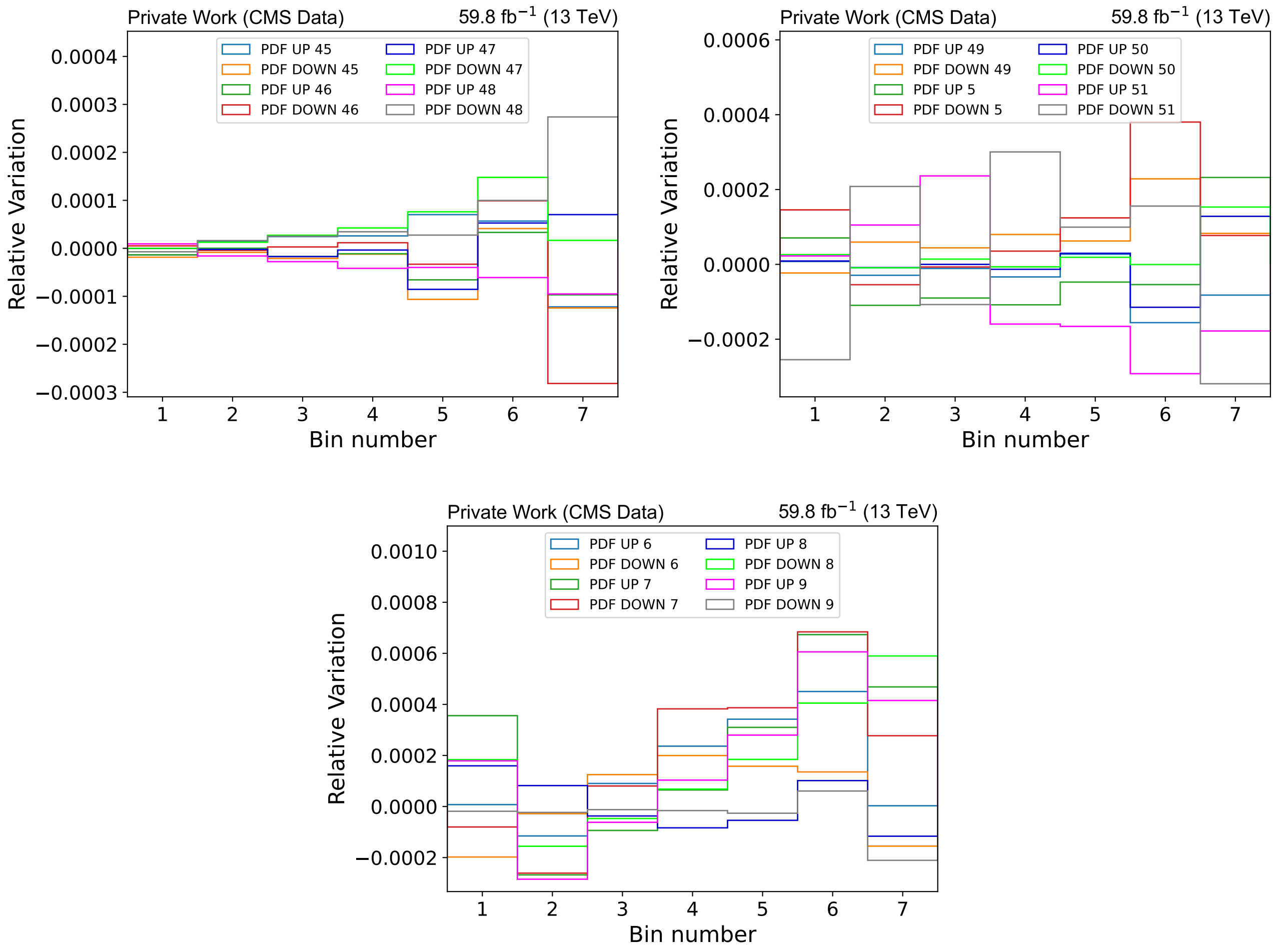


Figure B.4: Relative variations of each bin in the unfolded differential cross section with respect to the nominal case considering the systematic uncertainties using the NN-based method (fourth part).

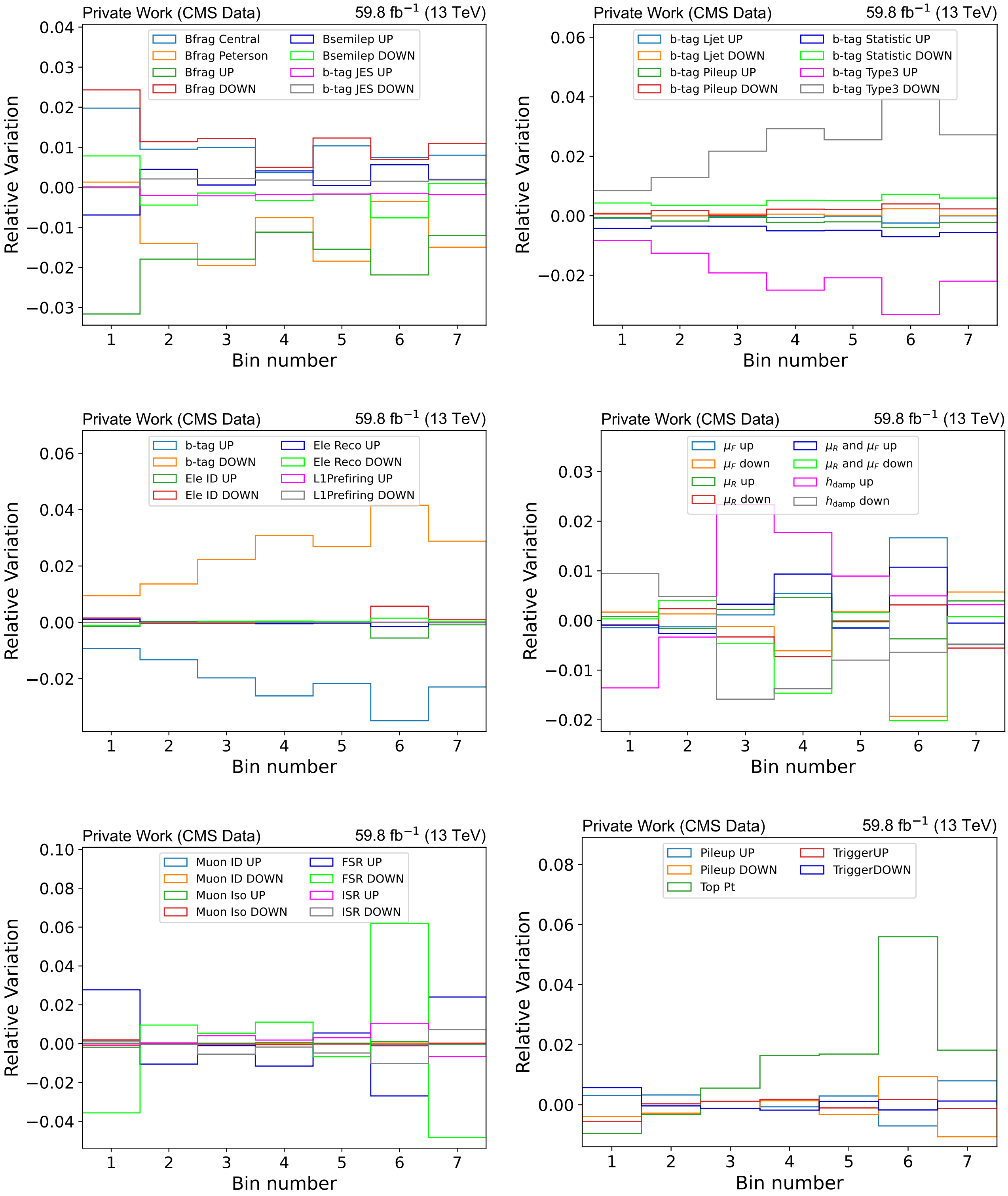


Figure B.5: Relative variations of each bin in the unfolded differential cross section with respect to the nominal case considering the systematic uncertainties using the loose kinematic reconstruction (first part).

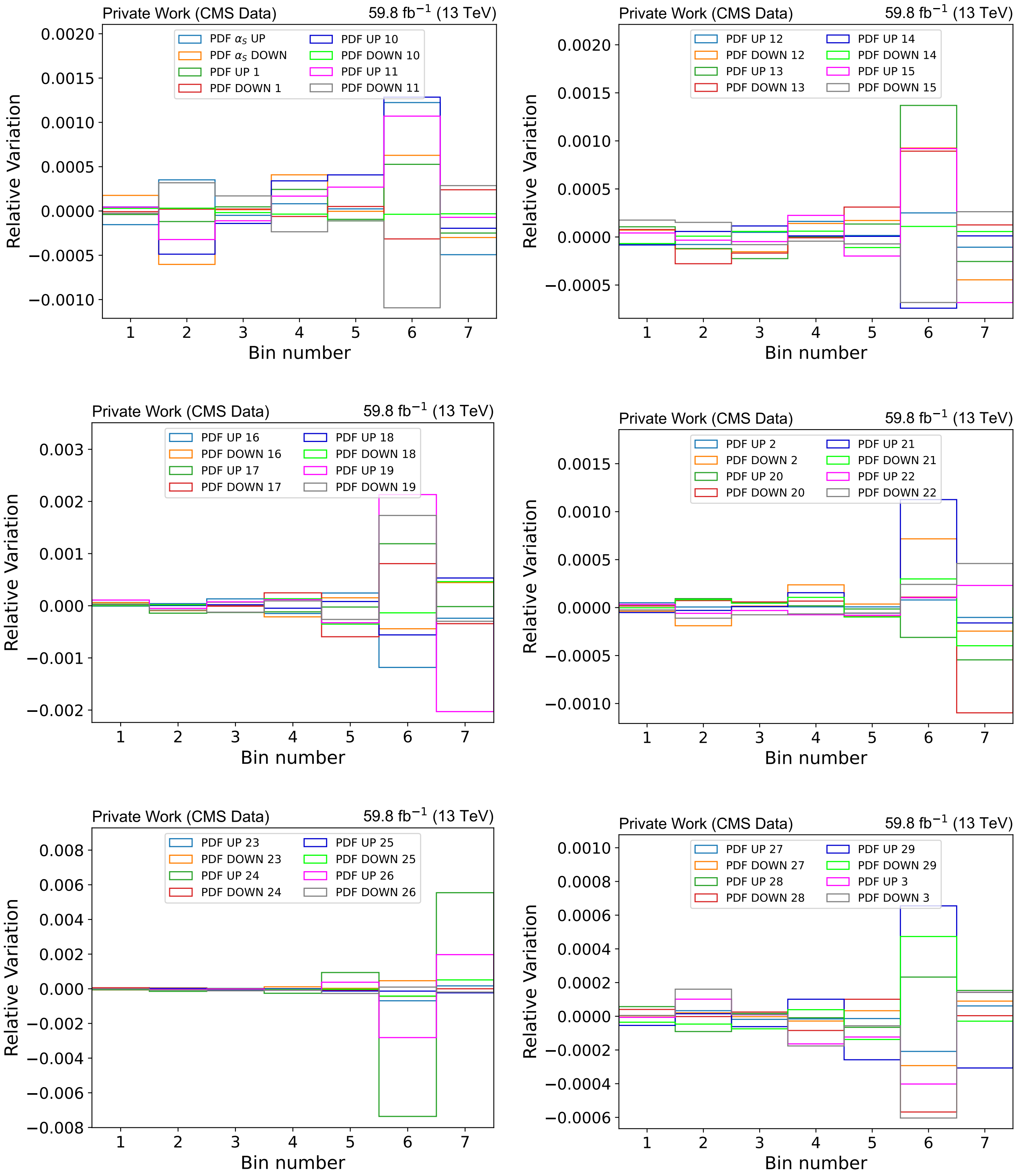


Figure B.6: Relative variations of each bin in the unfolded differential cross section with respect to the nominal case considering the systematic uncertainties using the loose kinematic reconstruction (second part).

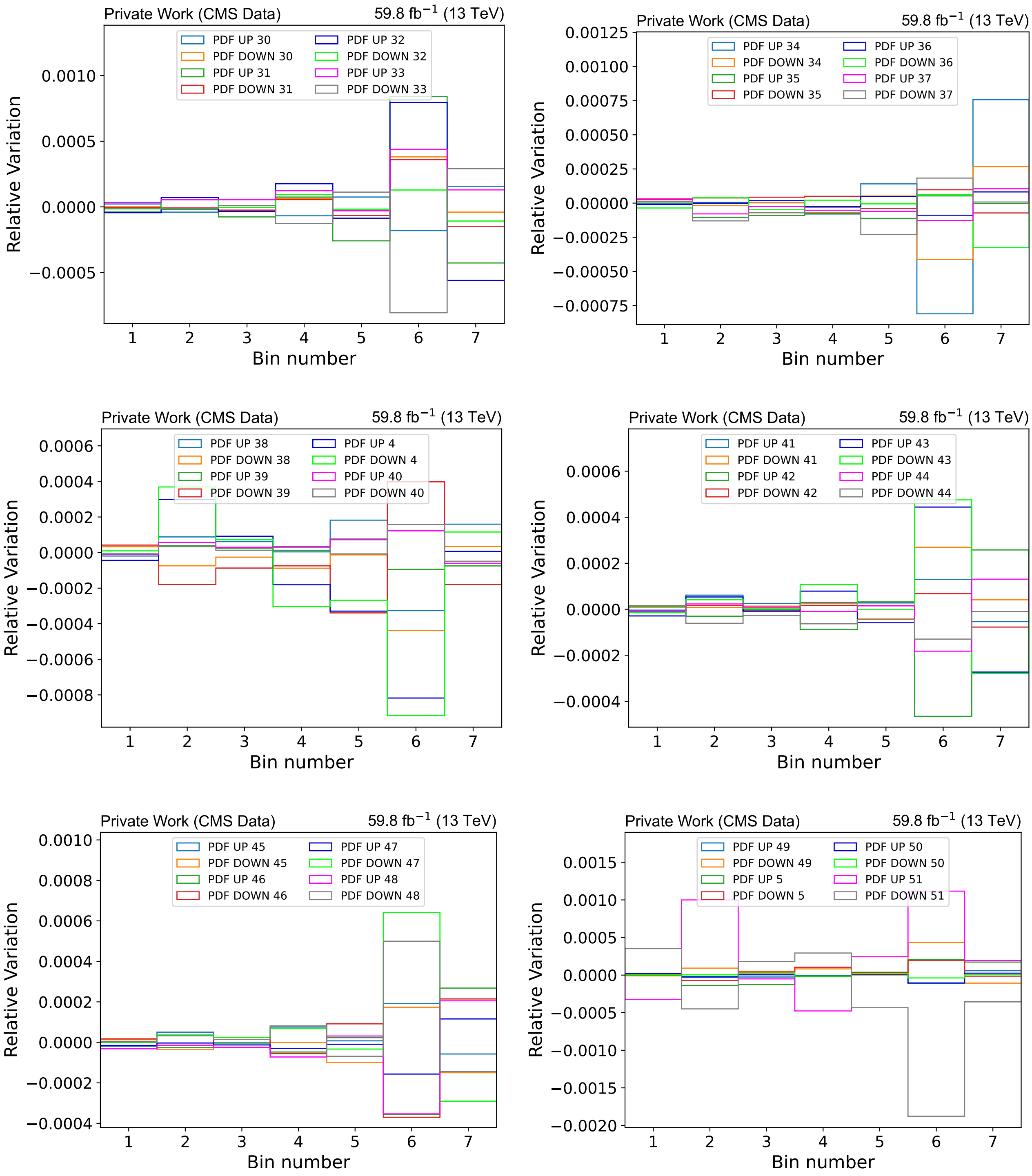


Figure B.7: Relative variations of each bin in the unfolded differential cross section with respect to the nominal case considering the systematic uncertainties using the loose kinematic reconstruction (third part).

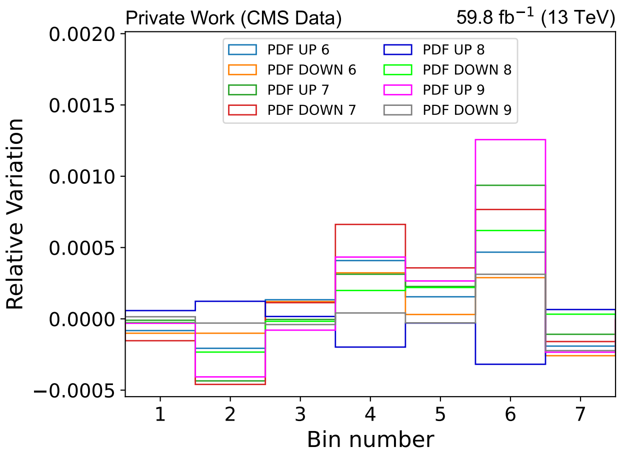


Figure B.8: Relative variations of each bin in the unfolded differential cross section with respect to the nominal case considering the systematic uncertainties using the loose kinematic reconstruction (fourth part).

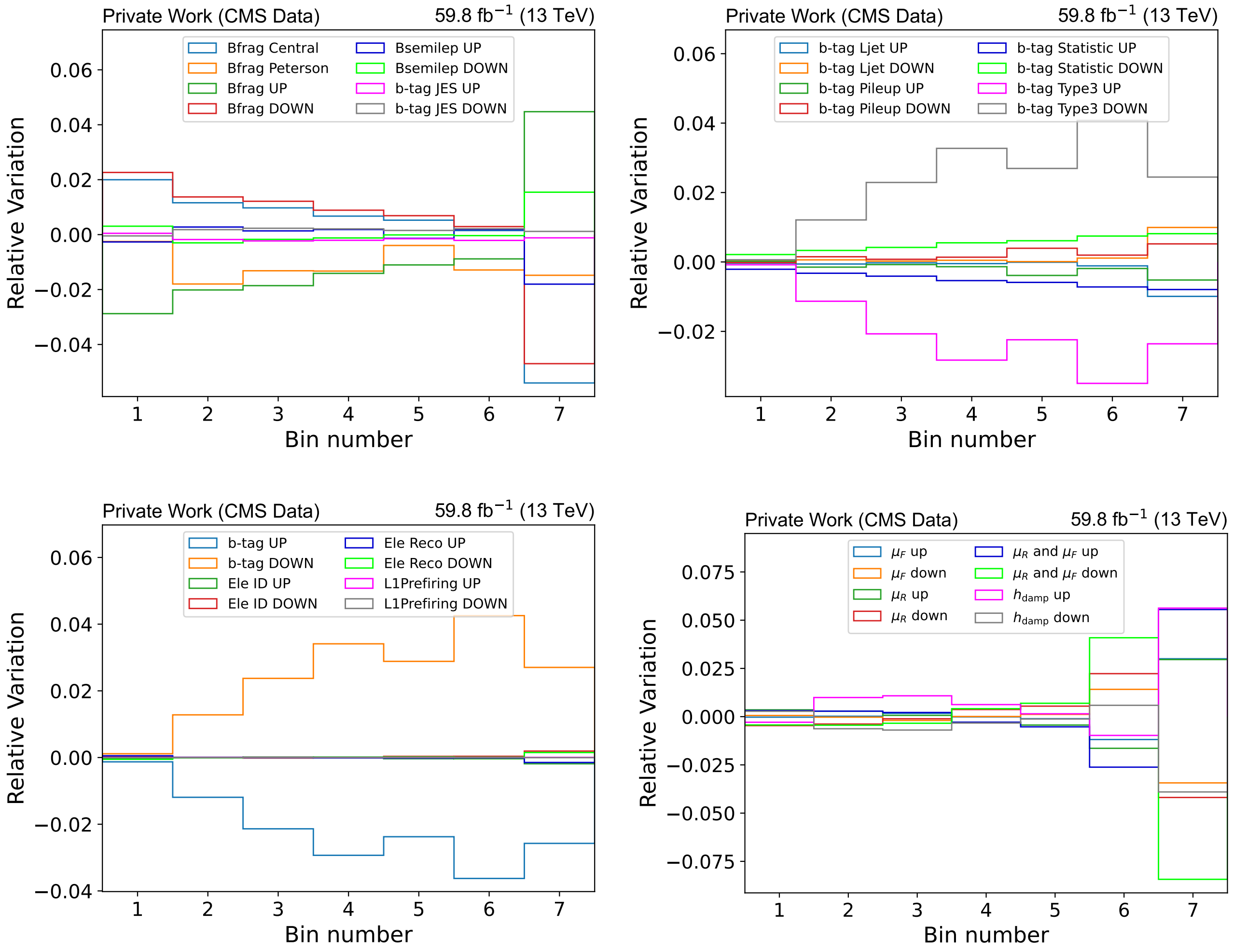


Figure B.9: Relative variations of each bin in the unfolded differential cross section with respect to the nominal case considering the systematic uncertainties using the full kinematic reconstruction (first part).

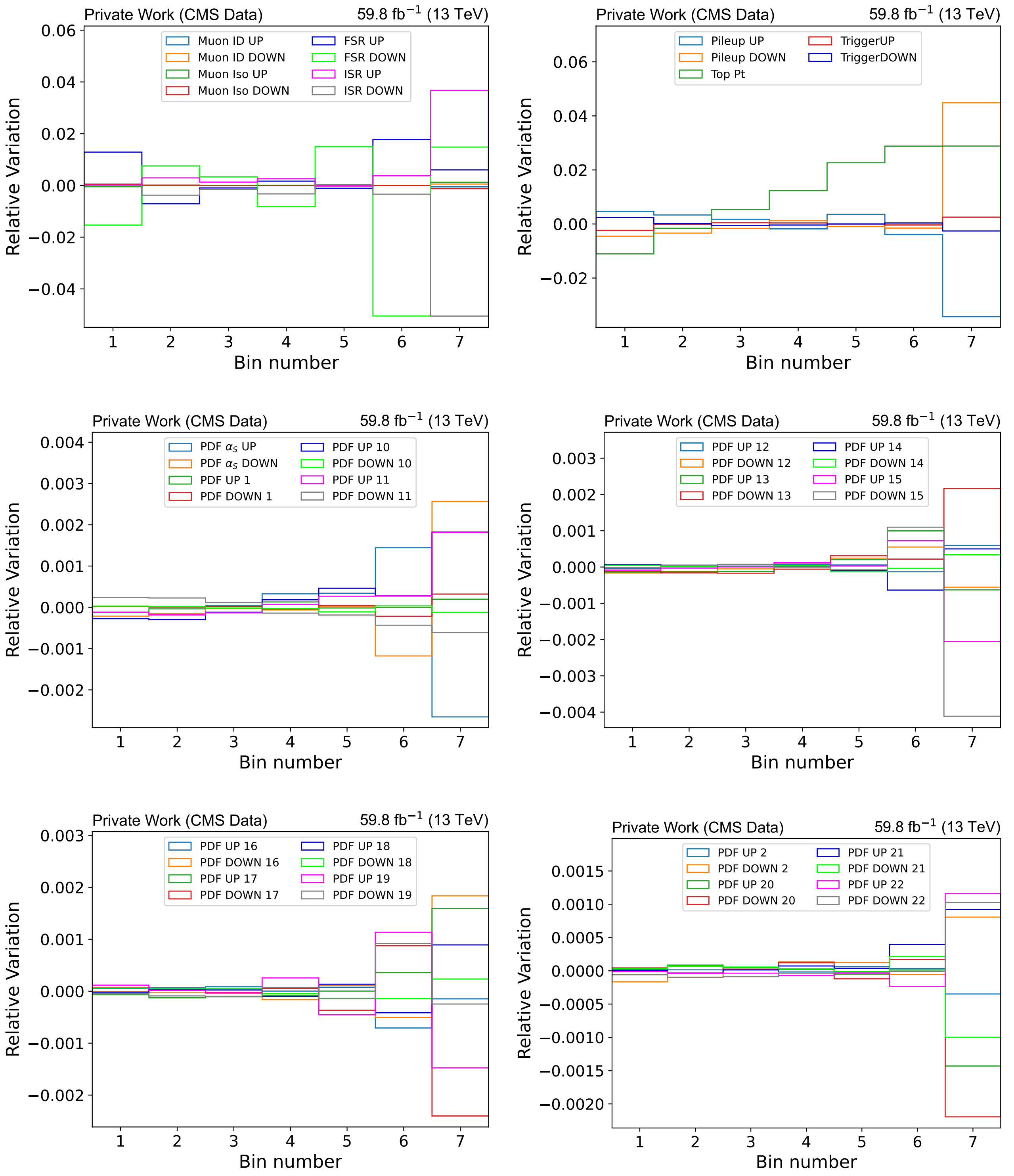


Figure B.10: Relative variations of each bin in the unfolded differential cross section with respect to the nominal case considering the systematic uncertainties using the full kinematic reconstruction (second part).

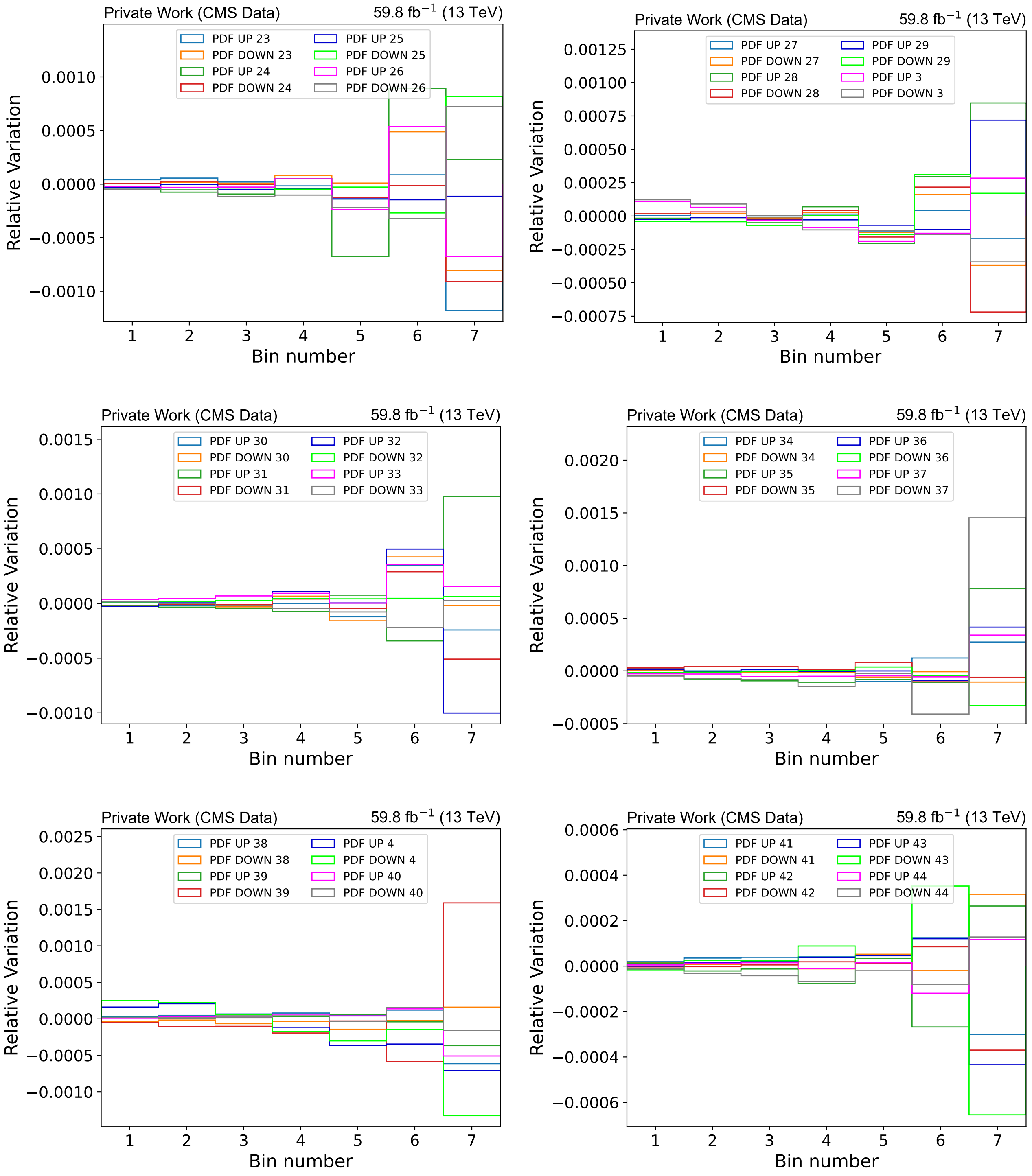


Figure B.11: Relative variations of each bin in the unfolded differential cross section with respect to the nominal case considering the systematic uncertainties using the full kinematic reconstruction (third part).

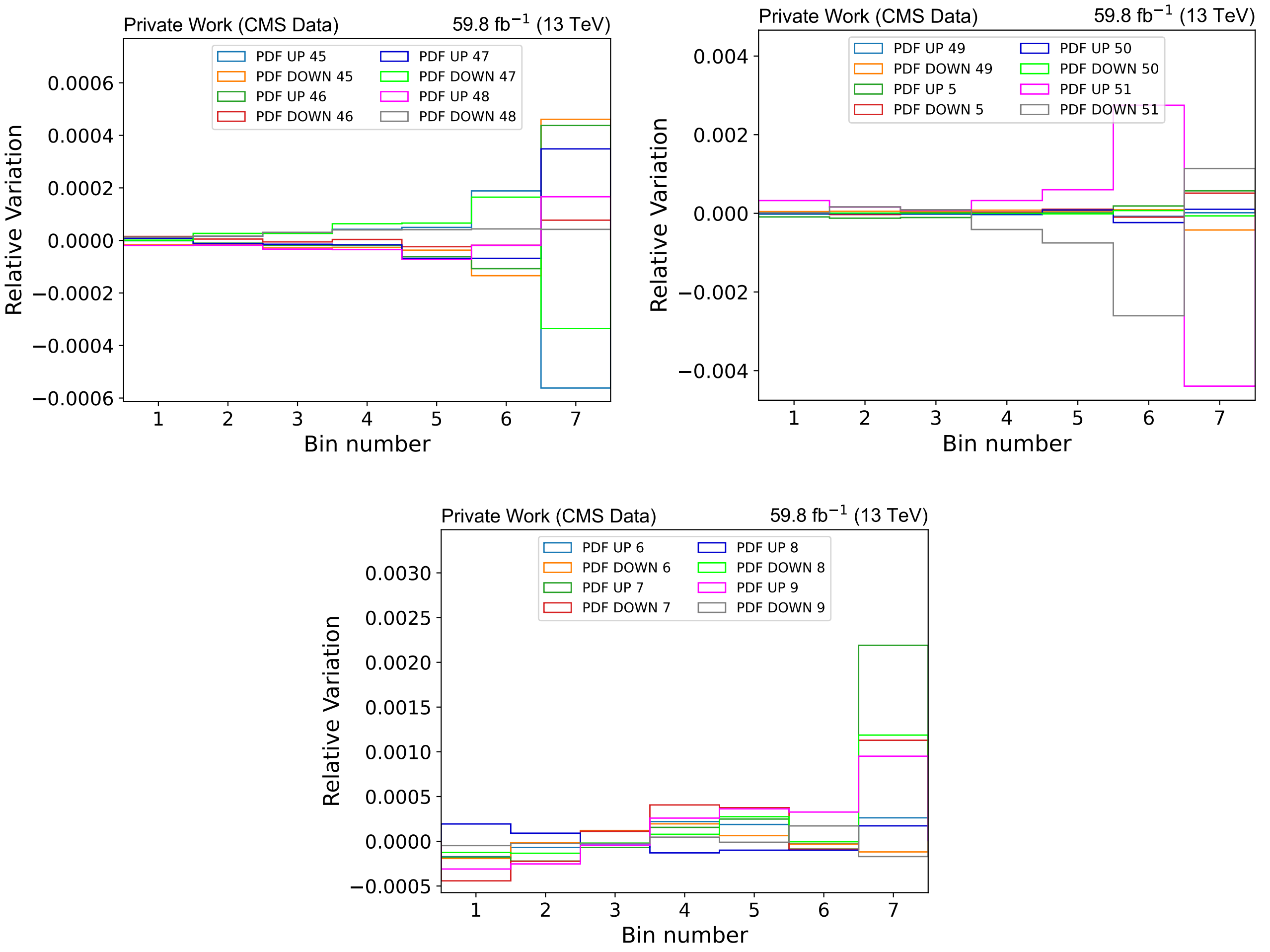


Figure B.12: Relative variations of each bin in the unfolded differential cross section with respect to the nominal case considering the systematic uncertainties using the full kinematic reconstruction (fourth part).

C

Regression of Top and Antitop Quark Transverse Momentum

Some preliminary results concerning a regression NN aimed at reconstructing the p_{T} of both the top and antitop quarks are here presented.

In this case, only events where both solutions do not fail are considered for the training of the NN. All the features listed in Tables 5.1-5.4 are used without performing a feature selection. Additionally, the four-momenta of the b jets from the top quark pair decay as reconstructed by the full kinematic reconstruction are also used as input features. Identically to the case of $m_{t\bar{t}}$ the angular variables are rotated, the input variables are normally scaled and the weights for the training are corrected by a factor proportional to the inverse of the frequency of the binned distribution of the average p_{T} between top and antitop quarks:

$$\frac{p_{\mathrm{T}}(t) + p_{\mathrm{T}}(\bar{t})}{2} \tag{C.1}$$

The architecture of the NN is similar to the one described in Chapter 5.2, with the only difference of having two output nodes: one for $p_{\mathrm{T}}(t)$ and one for $p_{\mathrm{T}}(\bar{t})$. Three layers are used respectively with 128, 64 and 32 nodes. The learning rate is set to 10^{-3}, and the training and validation batch sizes are set to 1024 events. An L2 regularization is used with a parameter $\lambda_2 = 0.003$. The same number of epochs and the same early stopping procedure as explained in Sec. 5.2 are used. As loss function, the mean squared error between the generated and reconstructed values is used. Each of the two output nodes has its own loss function and the total loss function to be minimized is the sum of the two.

In the following, some plots analogous to the ones described for $m_{t\bar{t}}$ are shown, with the only difference that the comparison can be performed only with the full kinematic reconstruction and not with the loose one, since the latter does not provide information on the four-momenta of the top and antitop quarks separately. In Fig. C.1 the distributions of the generated p_{T} and the reconstructed by the NN and full kinematic reconstruction are shown for both the top and the antitop quarks. The peak and the fall of the generated distribution are better reproduced by the NN approach, while at very low-p_{T} the full kinematic seems to better describe the true distribution. In Fig. C.2 the comparison of the resolutions is presented and in Fig. C.3 the response matrices for the two approaches are shown for top p_{T} only. Finally, the unfolded cross section using the same MC samples used for the definition of the response matrices is presented in Fig. 5.18. While keeping the uncertainty almost equal to full kinematic case in the low p_{T} region, the NN-based approach significantly improves the relative uncertainty in the hardest part of the spectrum.

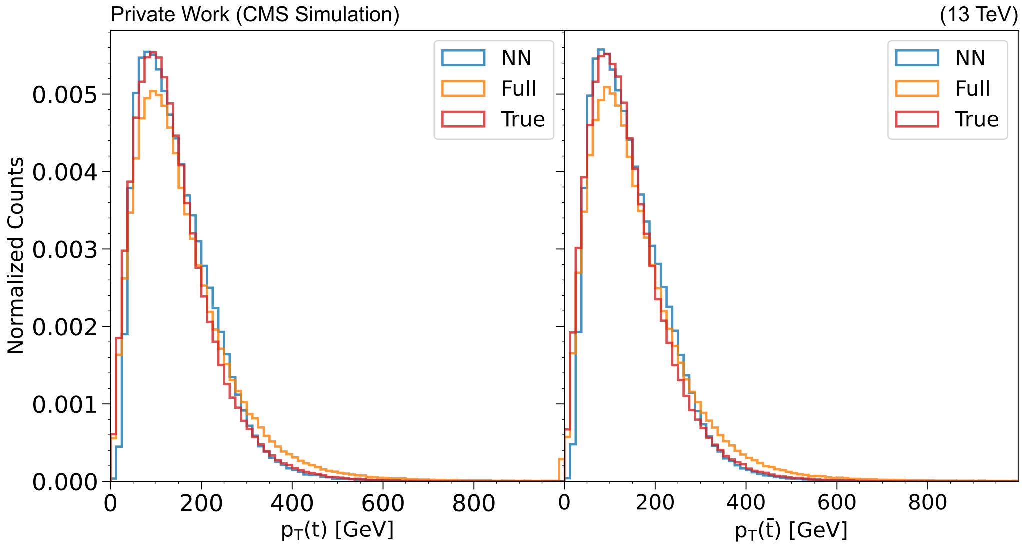


Figure C.1: Comparison of the p_T distributions obtained using the NN and the full kinematic reconstruction with the true distribution for the top (left) and antitop (right) quarks.

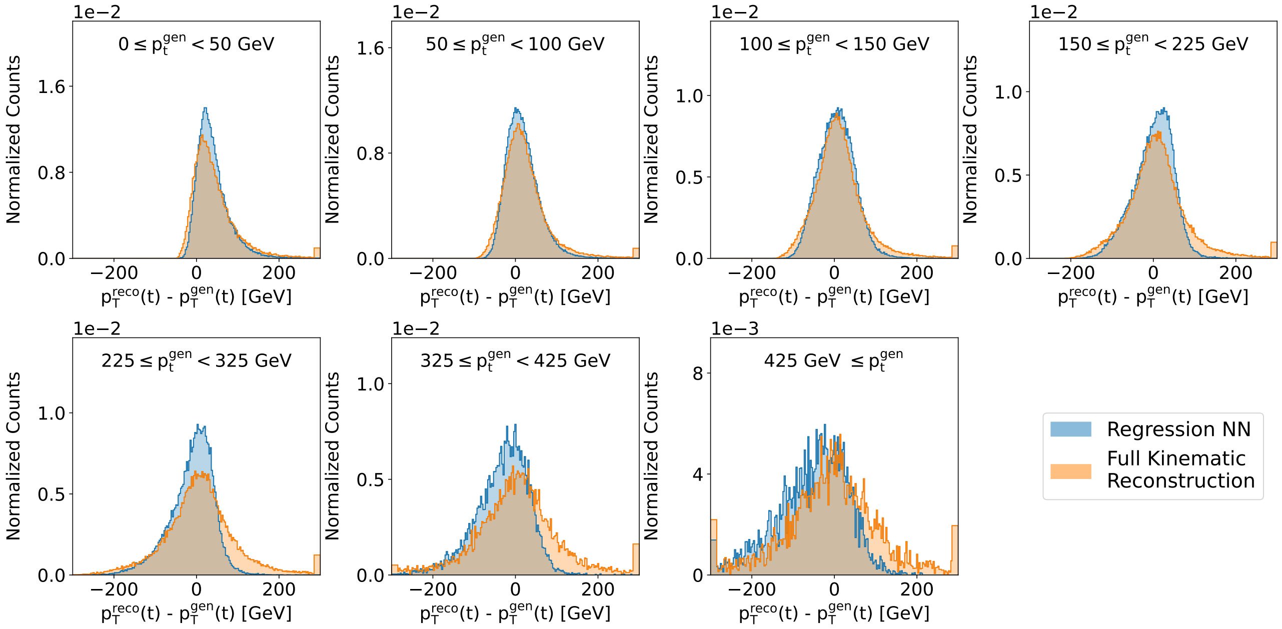
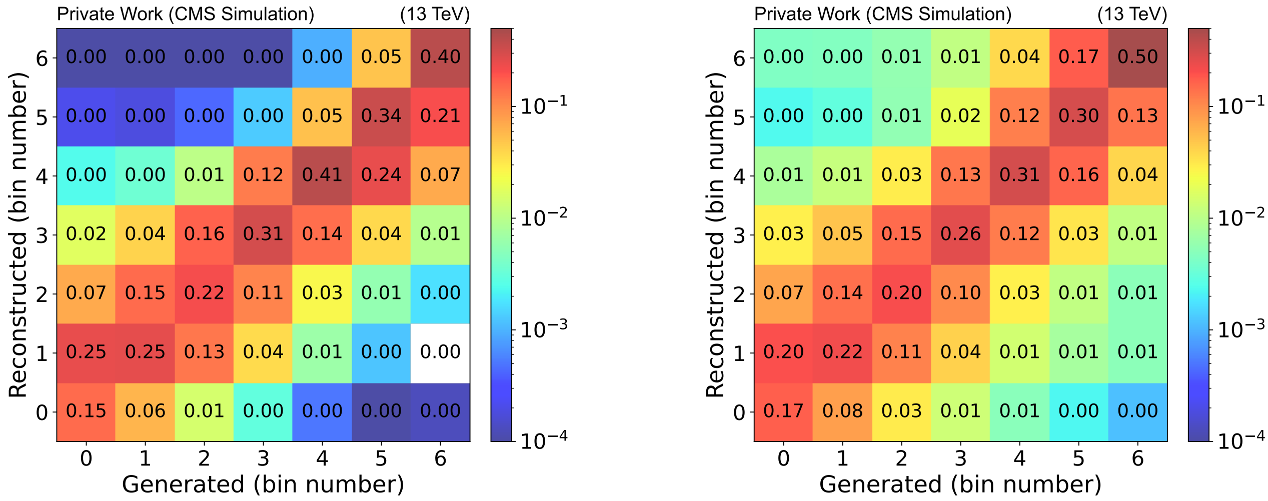


Figure C.2: Differences between $p_\mathrm{T}^\mathrm{reco}$ and $p_\mathrm{T}^\mathrm{gen}$ for the NN and the full kinematic reconstruction in seven bins of $p_\mathrm{T}^\mathrm{gen}$.



Reconstructed \ Generated	0	1	2	3	4	5	6
6	0.00	0.00	0.00	0.00	0.00	0.05	0.40
5	0.00	0.00	0.00	0.00	0.05	0.34	0.21
4	0.00	0.00	0.01	0.12	0.41	0.24	0.07
3	0.02	0.04	0.16	0.31	0.14	0.04	0.01
2	0.07	0.15	0.22	0.11	0.03	0.01	0.00
1	0.25	0.25	0.13	0.04	0.01	0.00	0.00
0	0.15	0.06	0.01	0.00	0.00	0.00	0.00

(a) NN-based kinematic reconstruction response matrix

Reconstructed \ Generated	0	1	2	3	4	5	6
6	0.00	0.00	0.01	0.01	0.04	0.17	0.50
5	0.00	0.00	0.01	0.02	0.12	0.30	0.13
4	0.01	0.01	0.03	0.13	0.31	0.16	0.04
3	0.03	0.05	0.15	0.26	0.12	0.03	0.01
2	0.07	0.14	0.20	0.10	0.03	0.01	0.01
1	0.20	0.22	0.11	0.04	0.01	0.01	0.01
0	0.17	0.08	0.03	0.01	0.01	0.00	0.00

(b) Full kinematic reconstruction response matrix

Figure C.3: Response matrices for the two reconstruction approaches

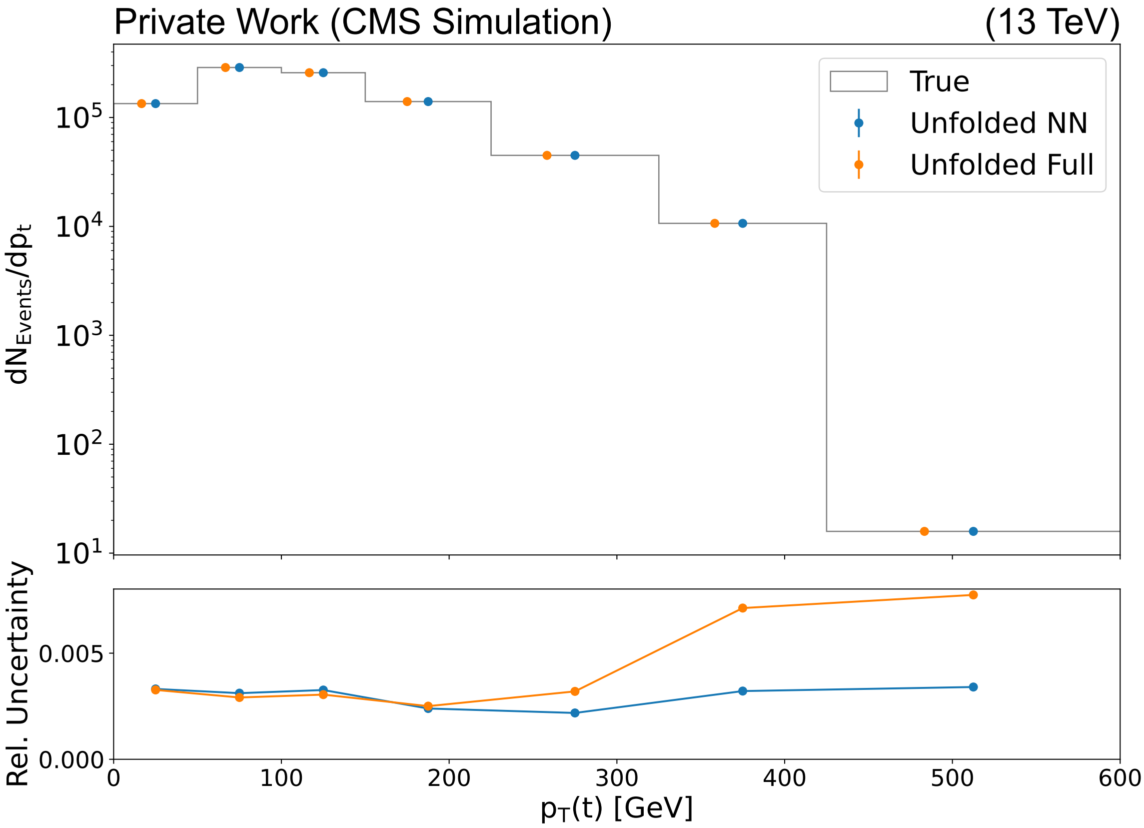


Figure C.4: Cross-check for the definition of the response matrices

Acknowledgements

At the end of this thesis, I would like to spend a few words to express my gratitude towards some people, who have been of fundamental importance for this goal.

First of all, I would like to thank both my supervisors: Dr. Olaf Behnke, for his patience and wisdom with which he has guided my work during my internship at DESY, and Prof. Tommaso Dorigo, for the precious feedback of invaluable usefulness he has provided me with. Special thanks also go to Dr. Sebastian Wuchterl, for the countless times he was available for clarifications and discussions and to the whole DESY CMS group, for being so welcoming and helpful during my stay in Hamburg.

I would also like to thank the friends with whom I shared my last five years in Padua, particularly those I have lived with in my College. You have been like a second family to me in this city and there are no words to express how much I owe you.

Finally, a special acknowledgement goes to my family for the support and love: my parents, Angelo and Tiziana, and my brother, Mattia.

Bibliography

[1] J. J. Thomson. “XL. Cathode Rays”. In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 44.269 (1897), pp. 293–316. DOI: 10.1080/14786449708621070.

[2] S. L. Glashow. “Partial-symmetries of weak interactions”. In: *Nuclear Physics* 22.4 (1961), pp. 579–588. ISSN: 0029-5582. DOI: https://doi.org/10.1016/0029-5582(61)90469-2. URL: https://www.sciencedirect.com/science/article/pii/0029558261904692.

[3] J. H. Christenson et al. “Evidence for the 2π Decay of the K_2^0 Meson”. In: *Phys. Rev. Lett.* 13 (4 July 1964), pp. 138–140. DOI: 10.1103/PhysRevLett.13.138. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.138.

[4] P. W. Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. In: *Phys. Rev. Lett.* 13 (16 Oct. 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.508.

[5] A. D. Sakharov. “Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe”. In: *Pisma Zh. Eksp. Teor. Fiz.* 5 (1967), pp. 32–35. DOI: 10.1070/PU1991D034n05ABEH002497.

[6] S. Weinberg. “A Model of Leptons”. In: *Phys. Rev. Lett.* 19 (21 Nov. 1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264. URL: https://link.aps.org/doi/10.1103/PhysRevLett.19.1264.

[7] A. Salam. “Weak and Electromagnetic Interactions”. In: *Conf. Proc. C* 680519 (1968), pp. 367–377. DOI: 10.1142/9789812795915_0034.

[8] V.N. Gribov and L.N. Lipatov. “Deep inelastic electron scattering in perturbation theory”. In: *Physics Letters B* 37.1 (1971), pp. 78–80. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(71)90576-4. URL: https://www.sciencedirect.com/science/article/pii/0370269371905764.

[9] M. Kobayashi and T. Maskawa. “CP Violation in the Renormalizable Theory of Weak Interaction”. In: *Prog. Theor. Phys.* 49 (1973), pp. 652–657. DOI: 10.1143/PTP.49.652.

[10] J. J. Aubert et al. “Experimental Observation of a Heavy Particle J”. In: *Phys. Rev. Lett.* 33 (23 Dec. 1974), pp. 1404–1406. DOI: 10.1103/PhysRevLett.33.1404. URL: https://link.aps.org/doi/10.1103/PhysRevLett.33.1404.

[11] J. -E. Augustin et al. “Discovery of a Narrow Resonance in e^+e^- Annihilation”. In: *Phys. Rev. Lett.* 33 (23 Dec. 1974), pp. 1406–1408. DOI: 10.1103/PhysRevLett.33.1406. URL: https://link.aps.org/doi/10.1103/PhysRevLett.33.1406.

[12] M. L. Perl et al. “Evidence for Anomalous Lepton Production in $e^+ - e^-$ Annihilation”. In: *Phys. Rev. Lett.* 35 (22 Dec. 1975), pp. 1489–1492. DOI: 10.1103/PhysRevLett.35.1489. URL: https://link.aps.org/doi/10.1103/PhysRevLett.35.1489.

[13] G. Altarelli and G. Parisi. “Asymptotic freedom in parton language”. In: *Nuclear Physics B* 126.2 (1977), pp. 298–318. ISSN: 0550-3213. DOI: https://doi.org/10.1016/0550-3213(77)90384-4. URL: https://www.sciencedirect.com/science/article/pii/0550321377903844.

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

[15] S. W. Herb et al. “Observation of a Dimuon Resonance at 9.5 GeV in 400-GeV Proton-Nucleus Collisions”. In: *Phys. Rev. Lett.* 39 (5 Aug. 1977), pp. 252–255. DOI: 10.1103/PhysRevLett.39.252. URL: https://link.aps.org/doi/10.1103/PhysRevLett.39.252.

[16] D. Schaile and P. M. Zerwas. “Measuring the weak isospin of B quarks”. In: *Phys. Rev. D* 45 (1992), pp. 3262–3265. DOI: 10.1103/PhysRevD.45.3262.

[17] CDF Collaboration. “Observation of Top Quark Production in $\overline{p}p$ collisions”. In: *Physical Review Letters* 74.14 (Apr. 1995), pp. 2626–2631. DOI: 10.1103/physrevlett.74.2626.

[18] D0 Collaboration. “Observation of the Top Quark”. In: *Physical Review Letters* 74.14 (Apr. 1995), pp. 2632–2637. DOI: 10.1103/physrevlett.74.2632.

[19] M. E. Peskin and D. V. Schroeder. *An Introduction to Quantum Field Theory*. Reading, USA: Addison-Wesley (1995) 842 p. Westview Press, 1995.

[20] Y. Fukuda et al. “Evidence for Oscillation of Atmospheric Neutrinos”. In: *Physical Review Letters* 81.8 (Aug. 1998), pp. 1562–1567. DOI: 10.1103/physrevlett.81.1562. URL: https://doi.org/10.1103%2Fphysrevlett.81.1562.

[21] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. 3rd. Wiley, 2003.

[22] GEANT4 Collaboration. “Geant4—a simulation toolkit”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(03)01368-8. URL: https://www.sciencedirect.com/science/article/pii/S0168900203013688.

[23] O. S. Brüning et al. *LHC Design Report*. CERN Yellow Reports: Monographs. Geneva: CERN, 2004. DOI: 10.5170/CERN-2004-003-V-1. URL: https://cds.cern.ch/record/782076.

[24] J. C. Collins, D. E. Soper, and G. Sterman. *Factorization of Hard Processes in QCD*. 2004. arXiv: hep-ph/0409313 [hep-ph].

[25] P. Nason. “A New Method for Combining NLO QCD with Shower Monte Carlo Algorithms”. In: *Journal of High Energy Physics* 2004.11 (Nov. 2004), pp. 040–040. DOI: 10.1088/1126-6708/2004/11/040. URL: https://doi.org/10.1088%2F1126-6708%2F2004%2F11%2F040.

[26] W. Adam et al. *Track Reconstruction in the CMS tracker.* Tech. rep. Geneva: CERN, 2006. URL: http://cds.cern.ch/record/934067.

[27] CMS Collaboration. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software.* Technical design report. CMS. Geneva: CERN, 2006. URL: https://cds.cern.ch/record/922757.

[28] S. Frixione, P. Nason, and C. Oleari. “Matching NLO QCD computations with parton shower simulations: the POWHEG method”. In: *Journal of High Energy Physics* 2007.11 (Nov. 2007), pp. 070–070. DOI: 10.1088/1126-6708/2007/11/070. URL: https://doi.org/10.1088%2F1126-6708%2F2007%2F11%2F070.

[29] M. Cacciari, G. P. Salam, and G. Soyez. “The anti-kt jet clustering algorithm”. In: *Journal of High Energy Physics* 2008.04 (Apr. 2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. URL: https://dx.doi.org/10.1088/1126-6708/2008/04/063.

[30] CMS Collaboration. “The CMS experiment at the CERN LHC”. In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004. URL: https://dx.doi.org/10.1088/1748-0221/3/08/S08004.

[31] V. Nair and G. Hinton. “Rectified Linear Units Improve Restricted Boltzmann Machines Vinod Nair”. In: vol. 27. June 2010, pp. 807–814.

[32] J. F. Monahan. *Numerical Methods of Statistics.* 2nd. Cambridge University Press, 2011. ISBN: 978-0-521-13951-9.

[33] J.-L. Agram. “CMS Silicon Strip Tracker Performance”. In: *Physics Procedia* 37 (2012). Proceedings of the 2nd International Conference on Technology and Instrumentation in Particle Physics (TIPP 2011), pp. 844–850. ISSN: 1875-3892. DOI: https://doi.org/10.1016/j.phpro.2012.02.423. URL: https://www.sciencedirect.com/science/article/pii/S1875389212017683.

[34] R. D. Ball et al. “Unbiased global determination of parton distributions and their uncertainties at NNLO and at LO”. In: *Nuclear Physics B* 855.2 (2012), pp. 153–221. ISSN: 0550-3213. DOI: https://doi.org/10.1016/j.nuclphysb.2011.09.024. URL: https://www.sciencedirect.com/science/article/pii/S0550321311005463.

[35] S. Alekhin, A. Djouadi, and S. Moch. “The top quark and Higgs boson masses and the stability of the electroweak vacuum”. In: *Physics Letters B* 716.1 (2012), pp. 214–219. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2012.08.024. URL: https://www.sciencedirect.com/science/article/pii/S0370269312008611.

[36] CMS Collaboration. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. In: *Physics Letters B* 716.1 (Sept. 17, 2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021.

[37] G. Degrassi et al. “Higgs mass and vacuum stability in the Standard Model at NNLO”. In: *Journal of High Energy Physics* 2012.8 (Aug. 2012). DOI: 10.1007/jhep08(2012)098. URL: https://doi.org/10.1007%2Fjhep08%282012%29098.

[38] V. Blobel. “Unfolding”. In: *Data Analysis in High Energy Physics.* John Wiley & Sons, Ltd, 2013. Chap. 6, pp. 187–225. ISBN: 9783527653416. DOI: https://doi.org/10.1002/9783527653416.ch6. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9783527653416.ch6. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9783527653416.ch6.

[39] S. Chatrchyan et al. “Measurement of Differential Top-Quark Pair Production Cross Sections in *pp* collisions at $\sqrt{s} = 7$ TeV”. In: *Eur. Phys. J. C* 73.3 (2013), p. 2339. DOI: `10.1140/epjc/s10052-013-2339-4`. arXiv: `1211.2220 [hep-ex]`.

[40] M. Gallinaro. “Top quark physics: A tool for discoveries”. In: *Journal of Physics: Conference Series* 447.1 (July 2013), p. 012012. DOI: `10.1088/1742-6596/447/1/012012`. URL: `https://dx.doi.org/10.1088/1742-6596/447/1/012012`.

[41] J. Alwall et al. “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”. In: *Journal of High Energy Physics* 2014.7 (July 2014). DOI: `10.1007/jhep07(2014)079`. URL: `https://doi.org/10.1007%2Fjhep07%282014%29079`.

[42] A. Benaglia. “The CMS ECAL performance with examples”. In: *Journal of Instrumentation* 9.02 (Feb. 2014), p. C02008. DOI: `10.1088/1748-0221/9/02/C02008`. URL: `https://dx.doi.org/10.1088/1748-0221/9/02/C02008`.

[43] CMS Collaboration. “Description and performance of track and primary-vertex reconstruction with the CMS tracker”. In: *Journal of Instrumentation* 9.10 (Oct. 2014), P10009. DOI: `10.1088/1748-0221/9/10/P10009`. URL: `https://dx.doi.org/10.1088/1748-0221/9/10/P10009`.

[44] CMS Collaboration. “Description and performance of track and primary-vertex reconstruction with the CMS tracker”. In: *Journal of Instrumentation* 9.10 (Oct. 2014), P10009–P10009. DOI: `10.1088/1748-0221/9/10/p10009`. URL: `https://doi.org/10.1088%2F1748-0221%2F9%2F10%2Fp10009`.

[45] M. Czakon and A. Mitov. “TOP++: A program for the calculation of the top-pair cross-section at hadron colliders”. In: *Computer Physics Communications* 185.11 (Nov. 2014), pp. 2930–2938. DOI: `10.1016/j.cpc.2014.06.021`. URL: `https://doi.org/10.1016%2Fj.cpc.2014.06.021`.

[46] N. Srivastava et al. “Dropout: A Simple Way to Prevent Neural Networks from Overfitting”. In: *Journal of Machine Learning Research* 15.56 (2014), pp. 1929–1958. URL: `http://jmlr.org/papers/v15/srivastava14a.html`.

[47] T. Sjöstrand et al. “An introduction to PYTHIA 8.2”. In: *Computer Physics Communications* 191 (June 2015), pp. 159–177. DOI: `10.1016/j.cpc.2015.01.024`. URL: `https://doi.org/10.1016%2Fj.cpc.2015.01.024`.

[48] C. Biino and R. Romaniuk. “The CMS Electromagnetic Calorimeter: Overview, lessons learned during Run 1 and future projections”. In: *Journal of Physics Conference Series* 587 (Feb. 2015), pp. 1–45. DOI: `10.1088/1742-6596/587/1/012001`.

[49] F. Chollet et al. *Keras*. 2015. URL: `https://keras.io`.

[50] CMS Collaboration. “Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV”. In: *The European Physical Journal C* 75.11 (Nov. 2015). DOI: `10.1140/epjc/s10052-015-3709-x`. URL: `https://doi.org/10.1140%2Fepjc%2Fs10052-015-3709-x`.

[51] K. He et al. *Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification*. 2015. arXiv: `1502.01852 [cs.CV]`.

[52] Stefan Höche. “Introduction to parton-shower event generators”. In: *Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision Frontier: Amplitudes for Colliders*. 2015, pp. 235–295. DOI: 10.1142/9789814678766_0005. arXiv: 1411.4085 [hep-ph].

[53] S. Ioffe and C. Szegedy. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*. 2015. arXiv: 1502.03167 [cs.LG].

[54] S. Stone. *“Pentaquarks and Tetraquarks at LHCb”*. 2015. DOI: 10.48550/ARXIV.1509.04051. URL: https://arxiv.org/abs/1509.04051.

[55] Richard Talman. “Scaling behavior of circular colliders dominated by synchrotron radiation”. In: *International Journal of Modern Physics A* 30.23 (Aug. 2015), p. 1544003. DOI: 10.1142/s0217751x15440030. URL: https://doi.org/10.1142%2Fs0217751x15440030.

[56] M. Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems*. 2016. arXiv: 1603.04467 [cs.DC].

[57] D.-A. Clevert, T. Unterthiner, and S. Hochreiter. *Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)*. 2016. arXiv: 1511.07289 [cs.LG].

[58] ”CMS Collaboration”. *Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of* $\mathrm{t\bar{t}}$ *at* $\sqrt{s} = 8$ *and 13 TeV*. Tech. rep. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2235192.

[59] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016.

[60] R. D. Ball et al. “Parton distributions from high-precision collider data”. In: *The European Physical Journal C* 77.10 (Oct. 2017). DOI: 10.1140/epjc/s10052-017-5199-5. URL: https://doi.org/10.1140%2Fepjc%2Fs10052-017-5199-5.

[61] CMS Collaboration. “Particle-flow reconstruction and global event description with the CMS detector”. In: *Journal of Instrumentation* 12.10 (Oct. 2017), P10003. DOI: 10.1088/1748-0221/12/10/P10003. URL: https://dx.doi.org/10.1088/1748-0221/12/10/P10003.

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

[63] S. Lundberg and Su-In Lee. *A Unified Approach to Interpreting Model Predictions*. 2017. arXiv: 1705.07874 [cs.AI].

[64] S. Ruder. *An overview of gradient descent optimization algorithms*. 2017. arXiv: 1609.04747 [cs.LG].

[65] S. Schmitt. “Data Unfolding Methods in High Energy Physics”. In: *EPJ Web of Conferences* 137 (2017). Ed. by Y. Foka, N. Brambilla, and V. Kovalenko, p. 11008. DOI: 10.1051/epjconf/201713711008. URL: https://doi.org/10.1051%2Fepjconf%2F201713711008.

[66] CMS Collaboration. “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV”. In: *Journal of Instrumentation* 13.05 (May 2018), P05011–P05011. DOI: 10.1088/1748-0221/13/05/p05011. URL: https://doi.org/10.1088%2F1748-0221%2F13%2F05%2Fp05011.

[67] CMS Collaboration. "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Journal of Instrumentation* 13.06 (June 2018), P06015–P06015. DOI: 10.1088/1748-0221/13/06/p06015. URL: https://doi.org/10.1088%2F1748-0221%2F13%2F06%2Fp06015.

[68] Mia Tosi. *The CMS trigger in Run 2*. Tech. rep. Geneva: CERN, 2018. DOI: 10.22323/1.314.0523. URL: https://cds.cern.ch/record/2290106.

[69] R. Bruce et al. "Review of LHC Run 2 Machine Configurations". In: (2019), pp. 187–197. URL: https://cds.cern.ch/record/2750415.

[70] CMS Collaboration. "CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} =$ 13 TeV". In: (2019).

[71] CMS Collaboration. "Measurements of $\mathrm{t\bar{t}}$ differential cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV using events containing two leptons". In: *JHEP* 02 (2019), p. 149. DOI: 10.1007/JHEP02(2019)149. arXiv: 1811.06625 [hep-ex].

[72] CMS Collaboration. "Search for resonant $\bar{t}t$ production in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Journal of High Energy Physics* 2019.4 (Apr. 2019). DOI: 10.1007/jhep04(2019)031. URL: https://doi.org/10.1007%2Fjhep04%282019%29031.

[73] CMS Collaboration. "Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements". In: *Eur. Phys. J. C* 80.1 (2020), p. 4. DOI: 10.1140/epjc/s10052-019-7499-4. arXiv: 1903.12179 [hep-ex].

[74] CMS Collaboration. "Measurement of ttbar normalised multi-differential cross sections in pp collisions at $\sqrt{}$=13 TeV, and simultaneous determination of the strong coupling strength, top quark pole mass, and parton distribution functions". In: *The European Physical Journal C* 80.7 (July 2020). DOI: 10.1140/epjc/s10052-020-7917-7. URL: https://doi.org/10.1140%2Fepjc%2Fs10052-020-7917-7.

[75] CMS Collaboration. "Pileup mitigation at CMS in 13 TeV data". In: *Journal of Instrumentation* 15.09 (Sept. 2020), P09018–P09018. DOI: 10.1088/1748-0221/15/09/p09018. URL: https://doi.org/10.1088%2F1748-0221%2F15%2F09%2Fp09018.

[76] Particle Data Group. "Review of Particle Physics". In: *Progress of Theoretical and Experimental Physics* 2020.8 (Aug. 2020), p. 083C01. ISSN: 2050-3911. DOI: 10.1093/ptep/ptaa104. eprint: https://academic.oup.com/ptep/article-pdf/2020/8/083C01/34673722/ptaa104.pdf. URL: https://doi.org/10.1093/ptep/ptaa104.

[77] T. Yu and H. Zhu. *Hyper-Parameter Optimization: A Review of Algorithms and Applications*. 2020. arXiv: 2003.05689 [cs.LG].

[78] The ATLAS and CMS Collaborations. *Towards Common $t\bar{t}$ Monte-Carlo Settings for ATLAS and CMS*. Tech. rep. Geneva: CERN, 2021. URL: https://cds.cern.ch/record/2772793.

[79] S. D. Bass, A. De Roeck, and M. Kado. "The Higgs boson implications and prospects for future discoveries". In: *Nature Reviews Physics* 3.9 (July 2021), pp. 608–624. DOI: 10.1038/s42254-021-00341-2. URL: https://doi.org/10.1038%2Fs42254-021-00341-2.

[80] CMS Collaboration. “The CMS Phase-1 pixel detector upgrade”. In: *Journal of Instrumentation* 16.02 (Feb. 2021), P02027. DOI: 10.1088/1748-0221/16/02/P02027. URL: https://dx.doi.org/10.1088/1748-0221/16/02/P02027.

[81] CMS Collaboration. “Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC”. In: *JINST* 16.05 (2021), P05014. DOI: 10.1088/1748-0221/16/05/P05014. arXiv: 2012.06888 [hep-ex].

[82] CMS Collaboration. “Performance of the CMS muon trigger system in proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: *Journal of Instrumentation* 16.07 (July 2021), P07001. DOI: 10.1088/1748-0221/16/07/p07001. URL: https://doi.org/10.1088%2F1748-0221%2F16%2F07%2Fp07001.

[83] CMS collaboration. “Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC”. In: *Journal of Instrumentation* 16.05 (May 2021), P05014. DOI: 10.1088/1748-0221/16/05/P05014. URL: https://dx.doi.org/10.1088/1748-0221/16/05/P05014.

[84] C. Burgard. *Standard Model of Physics.* Accessed: 14 June 2023. 2022. URL: https://texample.net/tikz/examples/model-physics/.

[85] 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.* Tech. rep. Geneva: CERN, 2022. URL: http://cds.cern.ch/record/2803771.

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

[87] P. Paradisi. *Standard Model.* Lecture notes. University of Padua. 2022.

[88] SciPy Contributors. *scipy.stats._stats_py.py.* Version: 1.11.1. 2022. URL: https://github.com/scipy/scipy/blob/v1.11.1/scipy/stats/_stats_py.py#L8644-L8875.

[89] R. L. Workman et al. “Review of Particle Physics”. In: *PTEP* 2022 (2022), p. 083C01. DOI: 10.1093/ptep/ptac097.

[90] R. Brugnera. “Neutrinoless double beta decay search with LEGEND”. In: *PoS* NOW2022 (2023), p. 075. DOI: 10.22323/1.421.0075.

[91] 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”. In: *Journal of High Energy Physics* 2023.7 (July 2023). DOI: 10.1007/jhep07(2023)077. URL: https://doi.org/10.1007%2Fjhep07%282023%29077.

[92] S. Wuchterl. “Top quark pole mass from $t\bar{t}$+jet using a machine learning based reconstruction for $t\bar{t}$ kinematics”. Presented 10 Feb 2023. 2023. URL: https://cds.cern.ch/record/2850510.

[93] LHCTopWG. *LHCTopWG Summary Plots.* Accessed: 2023-06-25. URL: %7Bhttps://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCTopWGSummaryPlots%7D.

[94] I. et al. Neutelings. *Higgs Potential Visualization.* Accessed: 14 June 2023. URL: https://tikz.net/category/physics/.