DOCTORAL DISSERTATION

A study of top quark pairs production in association with a W boson

Trần Tự Thông
Universiteit Gent & Université Catholique de Louvain

Belgium
November 2023

Supervisors:

Prof. Didar Dobur	Ghent University
Dr. Andrea Giammanco	Catholic University of Louvain

Jury members:

Prof. Philippe Smet (Chairman)	Ghent University
Prof. Agni Bethani (Secretary)	Catholic University of Louvain
Prof. Barbara Alvarez Gonzalez	University of Oviedo
Prof. Nadjieh Jafari	DESY and Isfahan University of Technology
Dr. Kirill Skovpen	Ghent University

Summary

This dissertation presents a measurement of the inclusive cross section of the production top quark-antiquark pairs in association with a W boson in proton-proton collisions at a centre-of-mass energy of 13 TeV. The analysis is conducted on a data set recorded by the CMS detector at CERN's LHC, corresponding to an integrated luminosity of 138 fb^{-1}.

The cross section is measured across the full phase space using events characterised by the presence of either two same-charge leptons (e or μ) or three leptons with a total charge sum of ± 1. The contribution from τ leptons is included through their decays to e and μ. These events are also required to contain at least two jets, with a minimum of two qualifying as b-tagged jets. Dominant background contributions originate from the nonprompt lepton background, $\text{t}\bar{\text{t}}\text{Z}/\gamma^*$, $\text{t}\bar{\text{t}}\text{H}$, WZ, and ZZ. The nonprompt lepton background is estimated using a data-driven "fake-rate" technique, which is validated in a dilepton validation region. In the dilepton channel, the misidentification of lepton charge contributes as an additional background, "charge misID", and is estimated from data, while the other background processes are modelled using simulations. In events with two leptons, the $\text{t}\bar{\text{t}}\text{W}$ signal is differentiated from backgrounds by a multiclass Neural-Network (NN). This network is trained with the simulated $\text{t}\bar{\text{t}}\text{W}$, $\text{t}\bar{\text{t}}\text{Z}$, $\text{t}\bar{\text{t}}\text{H}$, $\text{t}\bar{\text{t}}\gamma$, and nonprompt lepton events from $\text{t}\bar{\text{t}}$ simulation. The NN algorithm leverages kinematic variables associated with lepton and jets. Trilepton events are classified based on jet and b-tagged jet counts, in addition to the sum of lepton charges. Besides two signal regions, two control regions defined with three or four leptons are employed to constrain the WZ, ZZ, and $\text{t}\bar{\text{t}}\text{Z}/\gamma^*$ backgrounds.

The inclusive cross section is measured as $868 \pm 40\,(\text{stat}) \pm 51\,(\text{syst})$ fb, representing the most precise measurement of $\text{t}\bar{\text{t}}\text{W}$ inclusive cross section to date. It is higher than, but consistent with, the Standard Model (SM) predictions, falling within two standard deviations. Additionally, the cross sections of $\text{t}\bar{\text{t}}\text{W}^+$ and $\text{t}\bar{\text{t}}\text{W}^-$ are also measured. These are reported as $553 \pm 30\,(\text{stat}) \pm 30\,(\text{syst})$ fb and $343 \pm 26\,(\text{stat}) \pm 25\,(\text{syst})$ fb, respectively. The corresponding ratio of these two cross sections is determined to be $1.61 \pm 0.15\,(\text{stat})\,^{+0.07}_{-0.05}\,(\text{syst})$.

Samenvatting

Dit proefschrift presenteert een meting van de inclusieve doorsnede voor de productie van top quark-antiquark paren in associatie met een W boson in proton-proton botsingen bij een massamiddelpuntsenergie van 13 TeV. De analyse is uitgevoerd op een dataset die is opgenomen door de CMS-detector bij de LHC van CERN, wat overeenkomt met een geïntegreerde luminositeit van 138 fb^{-1}.

De doorsnede is gemeten over de volledige faseruimte met gebeurtenissen die gekenmerkt worden door het voorkomen van ofwel twee leptonen met dezelfde lading (e of μ) of drie leptonen met een totale ladingssom van ± 1. De bijdrage van τ leptonen wordt geleverd door hun verval naar e en μ. Deze gebeurtenissen moeten ook ten minste twee jets bevatten, met een minimum van twee als b-getagde jets. Dominante achtergrondbijdragen zijn afkomstig van de niet-prompt leptonachtergrond, $\text{t}\bar{\text{t}}\text{Z}/\gamma^*$, $\text{t}\bar{\text{t}}\text{H}$, WZ en ZZ. De niet-prompt leptonachtergrond wordt geschat met behulp van een gegevensgestuurde "fake-rate" techniek, die wordt gevalideerd in een dilepton validatiegebied. In het dileptonkanaal draagt de foutieve identificatie van leptonlading bij als een extra achtergrond, "charge misID", en wordt geschat op basis van gegevens, terwijl de andere achtergrondprocessen worden gemodelleerd met simulaties. In gebeurtenissen met twee leptonen wordt het $\text{t}\bar{\text{t}}\text{W}$-signaal onderscheiden van achtergronden door een multiklasse Neural-Network (NN). Dit netwerk is getraind met de gesimuleerde $\text{t}\bar{\text{t}}\text{W}$, $\text{t}\bar{\text{t}}\text{Z}$, $\text{t}\bar{\text{t}}\text{H}$, $\text{t}\bar{\text{t}}\gamma$, en niet-prompt lepton gebeurtenissen van $\text{t}\bar{\text{t}}$ monster. Het NN algoritme maakt gebruik van kinematische variabelen geassocieerd met lepton en jets. Trileptongebeurtenissen worden geclassificeerd op basis van jet- en b-tagged jet-tellingen, naast de som van leptonladingen. Naast twee signaalgebieden worden twee controlegebieden gedefinieerd met drie of vier leptonen om de WZ-, ZZ- en $\text{t}\bar{\text{t}}\text{Z}/\gamma^*$-achtergronden te beperken.

De inclusieve doorsnede is gemeten als $868 \pm 40\,(\text{stat}) \pm 51\,(\text{syst})$ fb, de meest nauwkeurige meting van de inclusieve doorsnede van $\text{t}\bar{\text{t}}\text{W}$ tot nu toe. De doorsnede is hoger dan, maar consistent met, de voorspellingen van de Standard Model (SM) en valt binnen twee standaarddeviaties. Daarnaast zijn ook de doorsneden van $\text{t}\bar{\text{t}}\text{W}^+$ en $\text{t}\bar{\text{t}}\text{W}^-$ gemeten. Deze worden respectievelijk gerapporteerd als $553 \pm 30\,(\text{stat}) \pm 30\,(\text{syst})$ fb en $343 \pm 26\,(\text{stat}) \pm 25\,(\text{syst})$ fb. De corresponderende verhouding van deze twee doorsneden wordt bepaald als $1.61 \pm 0.15\,(\text{stat})\,^{+0.07}_{-0.05}\,(\text{syst})$.

Résumé

Cette thèse présente la mesure de la section efficace inclusive de la production des paires de quark top et anti-quark top en association avec un boson W dans des collisions proton-proton à une énergie de centre de masse de 13 TeV. L'analyse est effectuée sur un ensemble de données enregistrées par le détecteur CMS au LHC du CERN, correspondant à une luminosité intégrée de 138 fb^{-1}.

La section efficace est mesurée sur l'ensemble de l'espace des phases en utilisant des événements caractérisés par la presence de deux leptons de même charge (e ou μ) ou de trois leptons avec une somme de charge totale de ± 1. La contribution des leptons τ est incluse par le biais de leurs désintégrations en e et μ. Ces événements doivent également avoir au moins deux jets, dont au moins deux sont qualifiés comme des jets du saveur lourde du quark b. Les contributions dominantes du bruit de fond proviennent des processus avec des leptons nonprompt, $\text{t}\bar{\text{t}}\text{Z}/\gamma^*$, $\text{t}\bar{\text{t}}\text{H}$, WZ, et ZZ. Le bruit de fond leptonique nonprompt est estimé à l'aide d'une technique de "fake rate" basée sur les données, qui est validée dans une région cinematique avec des dileptons. Dans le canal des dileptons, la mauvaise identification de la charge des leptons contribue à un bruit de fond supplémentaire, charge misID, qui est estimé à partir des données, tandis que les autres processus de bruit de fond sont modélisés à l'aide de simulations. Dans les événements avec deux leptons, le signal $\text{t}\bar{\text{t}}\text{W}$ est différencié des événements de fond par un neural network multiclasse Neural-Network (NN). Ce network est entraîné avec les événements simulés $\text{t}\bar{\text{t}}\text{W}$, $\text{t}\bar{\text{t}}\text{Z}$, $\text{t}\bar{\text{t}}\text{H}$, $\text{t}\bar{\text{t}}\gamma$, et les événements leptoniques nonprompt du processus de $\text{t}\bar{\text{t}}$. L'algorithme NN exploite les variables cinématiques associées aux leptons et aux jets. Les événements trileptoniques sont classés en fonction du nombre de jets et de jets b, en plus de la somme des charges leptoniques. Outre deux régions de signal, deux régions de contrôle avec trois ou quatre leptons sont utilisées pour contraindre les processus de WZ, ZZ, et $\text{t}\bar{\text{t}}\text{Z}$.

La section efficace inclusive est mesurée comme $868 \pm 40\,(\text{stat}) \pm 51\,(\text{syst})$ fb, ce qui représente la mesure la plus précise de la section efficace inclusive du processus $\text{t}\bar{\text{t}}\text{W}$ à ce jour. La section efficace mesurée est plus élevée que les prédictions du Standard Model (SM), restant en accord avec celles-ci, à deux écarts-types près. En outre, les sections efficaces de $\text{t}\bar{\text{t}}\text{W}^+$ et $\text{t}\bar{\text{t}}\text{W}^-$ ont également été mesurées et sont respectivement indiquées comme $553 \pm 30\,(\text{stat}) \pm 30\,(\text{syst})$ fb et $343 \pm 26\,(\text{stat}) \pm 25\,(\text{syst})$ fb. Le ratio entre $553 \pm 30\,(\text{stat}) \pm 30\,(\text{syst})$ fb et $343 \pm 26\,(\text{stat}) \pm 25\,(\text{syst})$ fb est mesuré comme $1.61 \pm 0.15\,(\text{stat})\,^{+0.07}_{-0.05}\,(\text{syst})$.

Contents

Introduction

The study of particle physics endeavours to unfold the nature of elementary particles and their interactions by employing mathematical models and conducting experiments to evaluate these models' validity. The concept of what constitutes an "elementary particle" has evolved over time. The ancient Greeks first proposed the idea that the basic constituents of matter were indivisible entities, which they termed "atomos". However, this hypothesis was not empirically confirmed at that time.

The transformation of this ancient philosophical concept into a formal scientific theory commenced in the early 19th century with John Dalton. Dalton's work revealed that chemical compounds consist of atoms in distinct, whole number ratios, thus opposing the notion of fractional or partial atoms. Moreover, he proposed that atoms from different elements possessed unique characteristics.

The structure of the atom was further untangled in the 20th century. Significant discoveries, such as the electron by J.J. Thomson and the atomic nucleus by Ernest Rutherford, revolutionised our understanding of the atom. Contrary to previous notions, it became apparent that atoms consist predominantly of empty space, with a minuscule, dense nucleus surrounded by a cloud of electrons.

The quantum mechanics theory, pioneered by Max Planck and Albert Einstein, and later solidified by the contributions of Niels Bohr, Werner Heisenberg, and Erwin Schrödinger, successfully accounted for the structure and stability of atoms. This era was marked by a series of particle discoveries, including the neutron, positron, muon, pion, and various other hadrons. The mid-20th century also heralded the era of particle accelerators, allowing for a more comprehensive study of subatomic structures.

Alongside these developments, Quantum Field Theories (QFTs) were formulated to describe interactions between particles. These theories were successfully employed to explain the electromagnetic interaction between charged particles, the weak interactions in nuclear decays, and the strong interactions existing between neutrons and protons within the nucleus.

The 1960s and 1970s saw continued efforts to develop the SM of particle physics. Murray Gell-Mann and George Zweig introduced the quark model to shed light on the particle zoo. The existence of quarks was then experimentally confirmed in the 1970s

by the Stanford Linear Accelerator Center (SLAC). In 2012, the discovery of the Brout-Englert-Higgs (BEH) boson by the Compact Muon Solenoid (CMS) and A Toroidal LHC Apparatus (ATLAS) collaborations at the the European Organisation for Nuclear Research (CERN)'s Large Hadron Collider (LHC) provided the last piece of the SM puzzle. However, despite its achievements, the SM has yet to provide an answer to numerous observed phenomena in the Universe. As such, the SM must be measured at exceptional precision, and any deviation from the predicted values could guide us to a New Physics (NP) Beyond Standard Model (BSM).

At the LHC, protons are collided at an enormous energy, of the TeV scale, to try to recreate the conditions right after the Big Bang. The energy level is sufficient to produce top quark-antiquark pairs. The top quark, the heaviest particle in the SM, discovered in 1995 at the Tevatron, is a rather unique particle as its Yukawa coupling to the H boson is close to unity, suggesting a special relationship. Additionally, its mass causes it to decay almost instantaneously into a bottom quark and a W boson without being hadronised. Its unique properties make top quark a promising probe for NP models. A detailed discussion of the SM is presented in Chapter 1.

This dissertation presents a measurement of the cross section of the production of a top quark-antiquark in association with W boson ($t\bar{t}W$) in proton-proton collisions at the LHC, at a centre-of-mass energy of $\sqrt{s} = 13$ TeV. The collision data, collected by the CMS detector between 2016 and 2018, amounts to an integrated luminosity of 138 fb^{-1}. The $t\bar{t}W$ process is particularly interesting due to its potential as a complementary approach for measuring asymmetry of top quark production. Furthermore, $t\bar{t}W$ is an important background in leptonic final states of several SM measurements such as $t\bar{t}H$, $t\bar{t}t\bar{t}$, and BSM searches. However, its cross section has been consistently measured to be higher than predicted. Consequently, a precise measurement would enhance background constraints for other measurements where $t\bar{t}W$ serves as a main background. The measurement of the inclusive cross section of the $t\bar{t}W$ production presented here is performed in the final states with two or three leptons (electron and muon).

Chapter 2 provides a description of the LHC and the CMS detector. Moreover, this chapter provides details of the CMS detector components and their functions. The measurement of the $t\bar{t}W$ inclusive cross section is then presented in Chapter 3, detailing the technical aspects of the analysis, including the modelling of backgrounds, signal discrimination, and the statistical approach. This chapter also provides an interpretation of the results. Finally, Chapter 4 discusses directions for the future research.

Author's contribution

The $t\bar{t}W$ measurement detailed in this thesis reflects my significant contributions, spanning from the development phase to the conclusion of the analysis. This analysis was a collaborative effort involving my colleagues from Ghent University (UGent), Catholic University of Louvain (UCLouvain), and the University of Oviedo in Spain. Hereby, I would like to detail my specific contributions chronologically.

Initially, I focused on the measurement of trigger efficiencies and scale factors for the dilepton channels. The same measurements in the trilepton channels were conducted by my colleagues at the University of Oviedo. Subsequently, I proceeded to the study of the charge misidentification (charge misID) background, examining the charge misID rate through both simulation and real data. I handled the measurements and ensured the chosen technique's validity.

When it came to the nonprompt lepton background, I started with the optimisation of the lepton selection — a crucial step for the Fake-rate (FR) measurement. With the optimised lepton section criteria, I measured the FR using both simulation and data-driven techniques. Both approaches are validated with dedicated regions in simulation and data. To constrain some SM background processes, the trilepton and four-lepton control regions were the chosen, my collaborators from Oviedo performed this task.

I then studied the event selection criteria for the dilepton channels to achieve the optimal signal to background ratio. However, it was still challenging to distinguish between signal and background events. As a result, two Multivariate Analysis (MVA) methodologies: Boosted Decision Tree (BDT) and multiclass Neural-Network (NN) were examined. My efforts were concentrated on the BDT approach, whereas the NN was studied by my colleague, Pieter David, from UCLouvain. The comprehensive study on systematic uncertainty saw contributions from all team members across the three institutions.

The results have been presented at various conference since they were first made public in 2022. I personally presented the results at the 15th International workshop on Top-quark physics in Durham, UK. The findings are officially published in the Journal of High Energy Physics on 28 July 2023 as: Measurement of the cross section of top quark-antiquark pair production in association with a W boson in proton-proton collisions at $\sqrt{s} = 13$ TeV, *JHEP* **2307** (2023) 219, doi:10.1007/JHEP07(2023)219, arXiv:2208.06485.

Alongside the $t\bar{t}W$ measurement, I actively participated in several tasks within the CMS collaboration. Between 2019 and 2020, I contributed to the examination of the Jet-MET High-Level Trigger (HLT) system, gearing up for phase-II of the CMS experiment. Subsequently, from 2020 to 2021, I engaged in a study exploring alternative methods for pileup estimation with bunch-by-bunch luminosity. In 2022, I dedicated a month at CERN, serving as the on-call detector expert for the strip tracker system. My responsibilities included monitoring the tracker's performance and ensuring the recording of high-quality data.

In the study of JetMET HLT, I tested the configuration of HLT that was preparing for phase-II upgrade of the CMS. Since the reconstruction done by HLT is most similar to the offline reconstruction by the ParticleFlow (PF) algorithm, jets reconstructed by the HLT-like algorithm (as this is a preparation, it is actually not HLT), are matched to jets reconstructed offline. The transverse momentum (p_{T}) and psuedorapidity (η) of the matched jets are then compared. The efficiencies of HLT-like algorithm are also computed. The reconstructed jets are also compared with generated information

provided in simulation. The test was carried out for different configurations of HLT, tracking system, pileup scenarios as well as different physics processes.

For the pileup study, I investigated an alternative approach to pileup estimation of the Run II data. Currently, pileup is estimated by making a Gaussian approximation to the luminosity distribution. However, the actual distribution of bunch-by-bunch luminosity not exactly Gaussian, as it has a longer tail at its higher end. In this study, I first looked at the relative bunch luminosity distributions with respect to the average luminosity and observer their evolution with time. It is clear that the distribution deviate from Gaussian distribution even more over sometime, especially for long LHC fills. My initial approach was to fit the bunch luminosity distributions using an Exponential Modified Gaussian function. However, this method proved inefficient due to the extensive data set and the challenging nature of the distributions' evolution over time.

Chapter 1

Theoretical background

1.1 The standard model of particle physics

The Standard Model (SM) is the main and most important theoretical framework in particle physics, serving as the foundation of contemporary particle physics. It describes the nature of elementary particles - the basic building blocks of the Universe. Beyond the particle description, the model incorporates three out of four fundamental forces, namely the electromagnetic force, the weak and strong nuclear forces. These forces govern the interactions of the elementary particles, shaping the very structure of the Universe. The framework of Quantum Field Theory (QFT) formulates these fundamental interactions. Over the years, the SM has been thoroughly tested and increasingly fine-tuned. It outlines our current comprehension of the elementary particles and their corresponding interactions.

Within the framework of the SM, elementary particles are characterised by their mass and quantum numbers such as electric charge, colour charge and spin. They are classified into two categories, fermions and bosons, with the bosons serving as the force mediators for the fundamental interactions. Distinguished by their intrinsic angular momentum, or spin, fermions carry half-integer values while bosons are characterised by their integer spin. They are comprehensively summarised as shown in Fig. 1.1.

Fermions

Fermions form the entirety of ordinary matter within the Universe. They are categorised into two groups of six, quarks and leptons. Each of the 12 fermions has a corresponding anti-matter equivalent. Fermions follows to the Fermi-Dirac statistics, they carry a spin of 1/2.

Quarks possess fractional charges of either $-1/3\,\mathrm{e}$ or $+2/3\,\mathrm{e}$ and are made up of six "flavours": up (u), down (d), charm (c), strange (s), top (t) and bottom (b). They exhibit a substantial range in mass, from ≈ 0.002 GeV for a u-quark to $172.13^{+0.76}_{-0.77}\,\mathrm{GeV}$

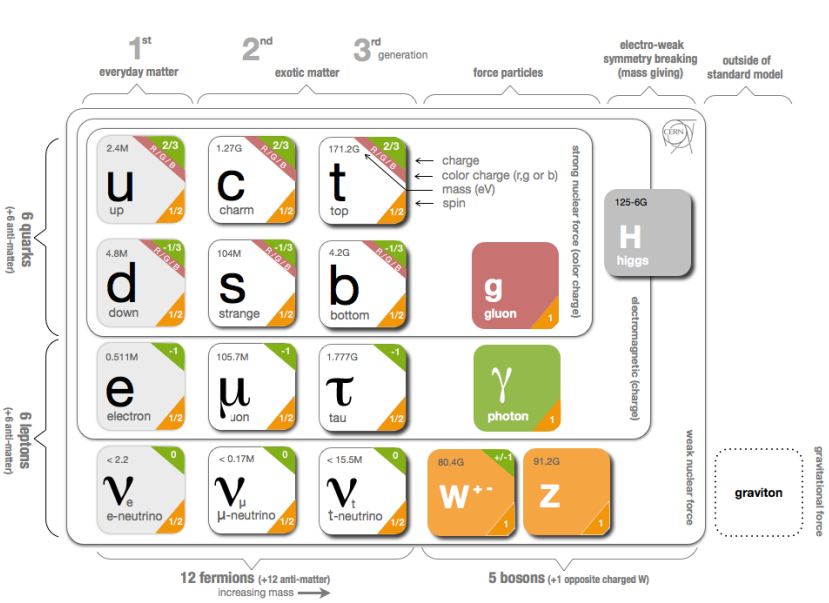

Figure 1.1: Constituent particles of the SM, the first three columns are fermions, ordered into three generations, the fourth column contains four gauge bosons, and finally the H boson. Figure taken from CERN website [1].

for a t-quark [2]. Governed by the Quantum Chromodynamics (QCD) theory, in nature, quarks are not observed in free-state due to quark confinement which binds them together in composite particles known as hadrons. As they also carry colour charges and couple to gluons in strong interactions.

In contrast to quarks, only three out of six types of leptons carry an electric charge of $-1\,\mathrm{e}$, leaving the remaining three with a zero charge. The charged leptons - electron, muon, and tau - each pair with a neutral counterpart, the neutrinos. As such, neutrinos are named after their respective charged pairs: the electron neutrino, muon neutrino, and tau neutrino.

Bosons

Elementary particles interact fundamentally through force exchanges, mediated by gauge bosons, which include the photon (γ), Z boson, W bosons, and gluons. These gauge bosons are of spin-1, they follow the Bose-Einstein statistics. While the photon and gluons are massless, the Z boson and W bosons have masses of 91.1876 ± 0.0021 GeV [3] and 80.377 ± 0.012 GeV [3], respectively. Photons mediate the electromagnetic interaction, while Z and W bosons manage the weak interaction. Gluons, on the other hand, facilitate the strong interaction by coupling to quarks [4].

Within the QFT framework, the strong interaction is characterised by the Quantum Chromodynamics (QCD) theory, whereas the Quantum Electrodynamics (QED) theory describes the electromagnetic interaction of charge particles. At low energy levels, these interactions coexist as distinct entities, but at higher energies, the electromagnetic and weak interactions unify under the Electroweak (EWK) theory to maintain the functionality of the SM. The strong interaction, true to its name, is the most powerful, while the weak interaction is considerably less potent—approximately 10^8 times weaker than the strong interaction [4].

With the discovery of the Brout-Englert-Higgs boson (commonly known as Higgs or H boson) in 2012 [5, 6], the gap in the SM was filled, and once again the validity of the SM was confirmed. Unlike the gauge bosons, the Higgs boson does not mediate the fundamental forces and is the by-product of the Brout-Englert-Higgs mechanism, in which particles are given their mass. The H boson is a massive spin-0 particle, with the mass of $125.11 \pm 0.11\,\text{GeV}$ [7], it is the heaviest boson in SM.

1.1.1 Symmetries in the standard model

The SM theory is a gauge-invariant QFT and built on an unbroken symmetry group,

$$G_{\text{SM}} = \text{SU}(3)_C \otimes \text{SU}(2)_L \otimes \text{U}(1)_Y \tag{1.1}$$

where the SU(3) subgroup describes the strong interaction through QCD theory and the $\text{SU}(2) \otimes \text{U}(1)$ describes the EWK sector. The latter gets spontaneously broken down to U(1) subgroup of the electromagnetic interaction, and the SU(2) represents the weak isospin interactions.

In the QFTs, elementary particles are indivisible and considered as quantum excitations of fields. This contrasts with quantum mechanics, where particles are described by wavefunctions. The dynamics of a QFT can be demonstrated by a Lagrangian density $\mathcal{L}$ - a function of some fields in the space-time coordinate x. A gauge-invariant action S can be constructed from $\mathcal{L}$ as $S = \int \mathcal{L} d^4x$. The complete Lagrangian density for the SM is constructed from the fundamental interactions:

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{QCD}} + \mathcal{L}_{\text{EWK}} + \mathcal{L}_{\text{BEH}} + \mathcal{L}_{\text{Yukawa}} \tag{1.2}$$

The terms $\mathcal{L}_{\text{QCD}}$, $\mathcal{L}_{\text{EWK}}$, $\mathcal{L}_{\text{H}}$ and $\mathcal{L}_{\text{Yukawa}}$ represent the QCD interaction, the EWK interaction, the Brout-Englert-Higgs (BEH) field, and the Yukawa interaction.

Quantum Electrodynamics

The QED theory is the first known QFT where agreement between quantum mechanics and special relativity was achieved. It is an Abelian gauge theory with the symmetry group U(1). The mathematical formulation of QED can be described via the Lagrangian

density of the massless photon field coupled to a free fermion field, ψ, with mass m and is as followed:

$$\mathcal{L}_{QED} = \bar{\psi}(i\gamma^{\mu}D - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{1.3}$$

where $D_{\mu} = \partial_{\mu} + iA_{\mu}$ is covariant derivative, dependent of the electromagnetic field A_{μ}. $F_{\mu\nu}$ is the electromagnetic field tensor and is defined as $F_{\mu\nu} = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu}$.

Since it is an Abelian group, QED doesn't allow self-interactions of photon to occur, thus photons can only couple to charged particles.

Quantum Chromodynamics

In QCD theory, the strong nuclear interactions between quarks and gluons, are described by a local non-Abelian SU(3) group, with eight gluon field generators in the form of Gell-Mann matrices T^a, $a = 1, \ldots, 8$. The QCD Lagrangian density contains the three-colour quark and antiquark fields, ψ and $\bar{\psi}$, and the gluon fields G^{μ}_a. The gluon fields G^{μ}_a are Lorentz vectors and defined as:

$$G^a_{\mu\nu} = \partial_{\mu}G^a_{\nu} - \partial_{\nu}G^a_{\mu} + g_s f^{a,b,c}G^b_{\mu}G^c_{\nu} \tag{1.4}$$

where g_s is the strong coupling constant, $f^{a,b,c}$ are the structure constants for the group, b and c also take integer values from 1 to 8. The structure constants $f^{a,b,c}$ are defined by the commutation relations of the group generators:

$$[T^a, T^b] = if^{a,b,c}T^c \tag{1.5}$$

The Lagrangian is constructed as:

$$\mathcal{L}_{\text{QCD}} = \bar{\psi}_j(i\gamma^{\mu}D_{\mu} - m)\psi_k - \frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a \tag{1.6}$$

The covariant derivative $D_{\mu} = \partial_{\mu} - ig_s G^a_{\mu}T_a$ ensures the local colour gauge invariance of the gluon fields. The indices j and k present the colour charge of quarks, and a summation over the six quark flavours is implied. The fact that QCD is based on a non-Abelian SU(3) group means that self-interactions are possible for gluon, and this is shown in the last term of Equation 1.6.

Electroweak theory

The EWK theory gives a unified description of the electromagnetic and the weak nuclear interactions at high energy. It is invariant under gauge transformation and governed by the $\text{SU}(2)_L \otimes \text{U}(1)_Y$ symmetry group. The subscript L refers to the left-handed structure of SU(2). The group contains three isospin weak generators I^a ($a = 1, 2, 3$) and a weak hypercharge generators Y. In EWK, forces are carried by three gauge bosons fields W^a_{μ}

associated with the weak generator $I^a = \sigma_a/2$ and a B_μ field. The term σ_a represents the Pauli matrices. The Lagrangian of EWK describes interactions of left-handed and right-handed Weyl spinors ψ_L and ψ_R. As a result, the Lagrangian is expressed as follows:

$$\mathcal{L}_{\text{EWK}} = i\bar{\psi}_L\gamma^\mu D_\mu\psi_L + i\bar{\psi}_R\gamma^\mu D_\mu\psi_R - \frac{1}{4}W^a_{\mu\nu}W^{\mu\nu}_a - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} \tag{1.7}$$

The covariant derivative for the EWK is defined as:

$$D_\mu = \partial_\mu + ig'\frac{1}{2}YB_\mu + igW^a_\mu I^a \tag{1.8}$$

where the factors g' and g are coupling constants of the electromagnetic and the weak interaction, respectively. The hypercharge Y is a function of fermion's electric charge Q and the third component of the isospin, $Y = 2(Q - I_3)$. The isospin I_3 is $\pm 1/2$ for left-handed fermions and zero for right-handed fermions. As such, the EWK force only acts on left-handed fermions and right-handed anti-fermions. This is reflected in the EWK Lagrangian, where left-handed and right-handed fermions appear in two separate terms to allow different treatments for left-handed and right-handed particles. Out of all the fundamental interactions, the weak interaction is the only interaction that doesn't follow all conservation laws, in fact, parity violation is maximal.

From the combinations of W^a_μ and B_μ fields, four gauge bosons, which are physical states, can be constructed:

$$\begin{aligned} W^- &= \frac{1}{\sqrt{2}}(W^1_\mu + iW^2_\mu) \\ W^- &= \frac{1}{\sqrt{2}}(W^1_\mu + iW^2_\mu) \\ Z^0 &= W^3_\mu \cos\theta_W - B_\mu \sin\theta_W \\ \gamma &= W^3_\mu \sin\theta_W + B_\mu \cos\theta_W \end{aligned}$$

The mixing angle θ_W depends on the coupling constant, $\theta_W = \arctan(g'/g)$, which is determined experimentally.

Brout-Englert-Higgs (BEH) mechanism

The EWK theory was initially built on the assumption that all four gauge bosons were massless. However, in reality, only photon is massless. If the $W^\pm$ and Z^0 bosons have mass, gauge invariance would not be broken. As a result, the Brout-Englert-Higgs (BEH) mechanism [8–13] was introduced to allow the $SU(2)_L \otimes U(1)_Y$ symmetry to be broken

while retain the gauge invariance. In the BEH mechanism, a complex scalar field ϕ is introduced. The field is presented by the $SU(2)_L$ doublet with zero electric charge, the isospins $I = 1/2$ and $I_3 = -1/2$, and a hypercharge $Y = 1$.

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{1.9}$$

The scalar field is a linear combination of four real scalar fields ϕ_1, ϕ_2, ϕ_3 and ϕ_4. In the Lagrangian of the BEH field, an addition potential is incorporated, defining the Lagrangian as follows:

$$\mathcal{L}_{\text{BEH}} = (D_\mu \phi)^\dagger (D^\mu \phi) - V(\phi^\dagger \phi) \tag{1.10}$$

with covariant derivative D_μ takes the same form as in the EWK Lagrangian in Equation 1.8. The potential

$$V(\phi^\dagger \phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \tag{1.11}$$

contains a mass parameter μ^2 and a BEH coupling constant λ, which must be larger than zero to allow a bound state to exist as $\phi \to \infty$. The potential has a global minimum when $\mu^2 > 0$. On the other hand, when $\mu^2 < 0$, the potential follows a "sombrero" shape with a local maximum at $\phi = 0$. The latter gives a minimum potential at

$$\phi^\dagger \phi = \frac{-\mu^2}{2\lambda} = \frac{v^2}{2} \tag{1.12}$$

where $v = \sqrt{\frac{-\mu^2}{\lambda}}$ is the Vacuum Expectation Value (VEV) of the BEH field. As shown in Fig. 1.2, the potential is clearly unstable in the centre and can spontaneously fall to a more stable state. Due to the symmetry of $V(\phi)$, there is an infinite number of stable states with minimum energy satisfying $\phi^\dagger \phi = v^2/2$.

The complex field ϕ can be written in terms of four real scalar fields, as in Equation 1.9, $|\phi|^2_{min}$ becomes

$$|\phi|_{min} = \frac{1}{2}(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2) = \frac{v}{2}. \tag{1.13}$$

If ϕ is expanded around a specific space where $\phi_1 = \phi_2 = \phi_4 = 0$ and $\phi_3 = v$, the field ϕ is now:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h \end{pmatrix} \tag{1.14}$$

with h being the scalar BEH field. Consequently, the potential term becomes

$$V(\phi) = \lambda v^2 h^2 + \lambda v h^3 - \frac{1}{4}\lambda h^4 \tag{1.15}$$

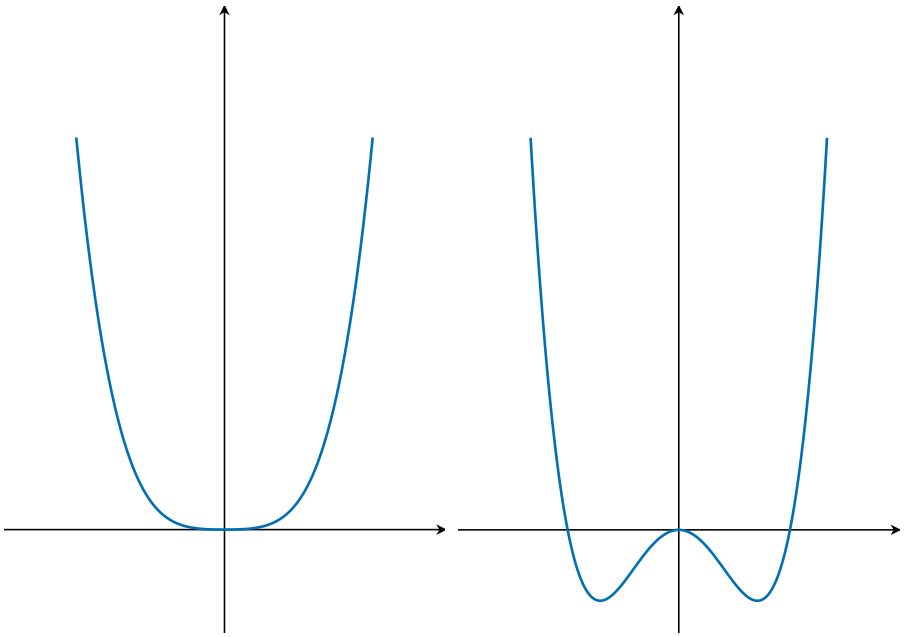

Figure 1.2: The BEH potentials in when $\mu^2 > 0$ (left) and $\mu^2 < 0$ (right).

As a by-product of the BEH field, a new spin-0 BEH boson (H) is introduced. The mass of the H boson is determined from the first term of the potential, $m_h = \sqrt{2\lambda v^2}$. The second and the third terms of the potential reflect the self-interaction property of H, the coupling strength depends on the H mass and the VEV. The BEH mechanism also gives mass to three gauge bosons in the EWK theory, however, photon remains massless. The masses of the W and Z bosons are

$$m_W = \frac{gv}{2} \quad \text{and} \quad m_Z = \frac{v\sqrt{g^2 + g'^2}}{2} = \frac{gv}{2\cos\theta_W}. \tag{1.16}$$

The VEV is related to the Fermi constant G_F, $v = \frac{1}{\sqrt{\sqrt{2}G_F}}$. It is not predicted by the BEH but is determined experimentally with the measured value of ≈ 246 GeV.

Yukawa interaction

The last term in the SM Lagrangian addresses the fermion mass problem, which is not included in the EWK theory. The Yukawa theory describes the interactions between fermions and the BEH field ϕ. As a result, fermions are given their mass. The Lagrangian of the Yukawa interaction is defined as

$$\mathcal{L}_{\text{Yukawa}} = -y_f(\bar{\psi}_L\phi\psi_R + \bar{\psi}_R\phi\psi_L) \tag{1.17}$$

where y_f is the Yukawa coupling constant and $\psi_{L(R)}$ represents the left- (right-)handed fermion. The Yukawa theory predicts the mass of a fermion is proportional to the Yukawa coupling constant y_f and the VEV, $m_f = \frac{y_f v}{\sqrt{2}}$. Hence, the coupling strengths between the BEH field and fermions depend on the fermion masses.

1.1.2 Limitations of the standard model

The SM is a successful theory that can describe a wide range of observed phenomena. It has been extensively tested and validated throughout the years at high precision. Despite being the most comprehensive model of particle physics, there are still many open questions that the SM fails to provide an answer for.

Gravity

Although there are four fundamental forces: gravity, the weak, strong and electromagnetic forces, the SM has so far only described the latter three forces, leaving gravity unexplained. At macroscopic scale, gravity is well established, however, at quantum scale, it is still lacking. Given that the existence of gravitational waves has been confirmed, if gravity would follow a QFT, "Graviton" - the force carrier of gravity should also exist. However, the existence of graviton still hasn't been confirmed.

Matter-antimatter asymmetry

Right after the Big Bang, matter and antimatter were produced in a roughly equal proportion, should their proportion remain unchanged, matter and antimatter would annihilate. However, at some point after the Big Bang, matter outnumbered antimatter by a large fraction. Only about one part per billion of antimatter remains in the Universe. Although the phenomenon of charge conjugation parity (CP) violation, first observed in the decay of neutral Kaons, does contribute to the observed asymmetry between matter and antimatter, the extent of CP violation predicted by the SM is insufficient to fully explain the magnitude of this asymmetry. As such, there must be other sources contributing to this asymmetry which are not described by the SM.

Dark matter and dark energy

Cosmological observations show that particles described by the SM account for only about 5% of the content of the Universe, the remaining contents are dark matter and dark energy. Dark matter contributes to about 25% of the Universe, it does not interact with the electromagnetic force, however, it causes gravitational effects and can be observed through the galactic rotations of galaxies. Evidences supporting the existence of dark matter have been found experimentally through the cosmic microwave background radiation, gravitational lensing and galactic rotation curves. While dark matter causes gravitational effect on galaxies, the "dark energy" cause the accelerated expansion of the Universe. It is the largest composition of the Universe, up to 70%.

Neutrino masses

Neutrinos are treated as massless particles in the SM and only left-handed neutrinos have been observed. The lack of right-handed neutrinos means they do not interact with the BEH field. However, experimental results have shown otherwise, in fact, neutrinos must have a non-zero mass since they oscillate between their flavours. This contradiction implies that the SM is incomplete when it comes to neutrino mass. Nevertheless, the existence of right-handed neutrinos and the origin of their mass remain unknown.

Hierarchy problem

The Spontaneous Symmetry Breaking (SSB) of the EWK force occurs around the expectation value of the Higgs field, in the order of $\mathcal{O}(10^2)$ GeV. However, the scale of this expectation value is far from the next-higher physically relevant energy scale - the Planck scale, which is in the order of $\mathcal{O}(10^{19})$ GeV. If all the quantum corrections to the H mass due to virtual particles are included, the expected mass would be in the Planck scale. For the H mass to be around its measured values 125.11 ± 0.11 GeV [7], it must be significantly fine-tunedqewa. This lead to a question of why there is a severe hierarchy problem in the SM.

Fermion generations and their mass

Both quarks and leptons are subdivided into three generations with the same spin and charge but very different masses. The mass ranges from 0.5 MeV for an electron to 1.8 GeV for a tau, and 2.5 MeV for an up quark to about 172 GeV for a top quark. Furthermore, the masses of particles are not obtained directly from the SM prediction itself, but through experimental measurements. From the theoretical point of view, there is no explanation for why there are three generations of fermions, not two, four, or even more, and why the fermion mass has such a big range.

1.1.3 Beyond the standard model

As highlighted in the previous section, several observed phenomena are not comfortably accommodated by the SM. This prompts the question: do we coexist with extra dimensions, or do we simply need additional BSM physics models to bridge these gaps? In the search for possible explanations to these problems, several theories of physics Beyond Standard Model (BSM), such as the Supersymmetry (SUSY) model, the Grand Unified Theories (GUTs), and the SM Effective Field Theory (SMEFT) have been explored. Among these, the SMEFT is the most related theory to this thesis.

The SMEFT [14–18] is a theoretical framework that enables the integration of the effects of an unknown high-energy scale BSM phenomena into the SM as modifications without the need for a specific model. Rather than detailing individual particles, the SMEFT takes a more general approach by adding higher-dimension operators to the SM Lagrangian. The term “dimension” here relates to the mass dimension in the Lag-

rangian. While the SM is of dimension-4, the supplementary operators in SMEFT are of dimension-6 or higher. Notably, dimension-5 operators are excluded due to their violation of lepton number conservation, they are responsible for generating mass terms for Majorana neutrinos and do not directly involve the top quark sector. The Lagrangian for SMEFT is written as:

$$\mathcal{L} = \mathcal{L}_{\text{SM}} + \sum_i \frac{c_i}{\Lambda^2}\mathcal{O}_i \tag{1.18}$$

In Eq. 1.18, c_i stands for the Wilson coefficients, $\mathcal{O}_i$ for the dimension-6 operators and Λ for the energy scale. The Wilson coefficients quantify the strength of BSM interactions, hence, the existence of some BSM phenomena would result in non-zero Wilson coefficients. Therefore, by making precise measurements on SM processes, constraints on the Wilson coefficients can be made. At the CMS, SMEFT interpretations have been carried out on a number of analyses in the production of top quark-antiquark pair in association with a gauge boson (W, Z or γ) [19–22], where the team at Ghent University played a leading role. The work presented in this thesis focused instead on a precise measurement of $\text{t}\bar{\text{t}}\text{W}$, and left the EFT interpretation to another publication.

1.2 The production of $t\bar{t}W$ at the LHC

1.2.1 Top quark sector

The top quark is the heaviest elementary particle in the SM, at $172.44 \pm 0.49\,\text{GeV}$ [23], it is almost as heavy as a gold atom. Its enormous mass means the top quark couples strongly to the H boson, with a coupling strength close to unity, it provides a test for the SM and any deviation could signal BSM physics.

At the LHC, top quarks are predominantly produced in pairs ($t\bar{t}$) via gluon-gluon fusion and quark-antiquark annihilation, as shown in Fig. 1.3.

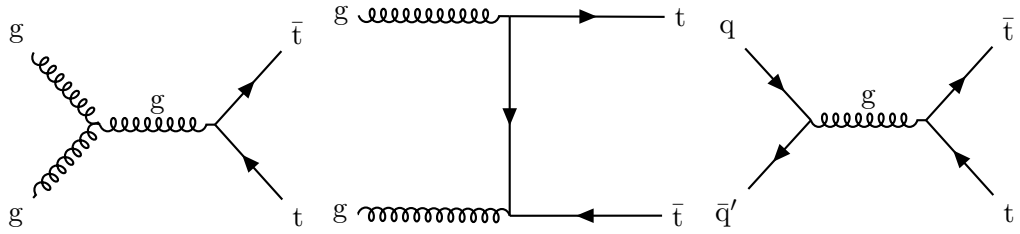

Figure 1.3: The production of top quark-antiquark pairs at the LHC.

The measurement of top quark coupling to bosons is not limited to only H boson, but also extends to vector bosons as $t\bar{t}W$, $t\bar{t}Z$ and $t\bar{t}\gamma$, collectively known as $t\bar{t}V$. Due to its large mass, top quark has a very short lifetime, in the order of $10^{-25}\,\text{s}$. Unlike other quarks, it has never been found in bound states, resulting in top quark decaying through the weak interaction only. Hence, it is sensitive to the EWK coupling. The work in this dissertation centres around the coupling of top quark and W boson, through the measurement of the cross section of the production of top quark-antiquark pair in association with a W boson ($t\bar{t}W$). Together with $t\bar{t}H$ and $t\bar{t}Z$, the measurement of $t\bar{t}W$ provides complementary information on the EWK coupling of the top quark and experimental constraints on the t-H doublet interaction. Moreover, precise measurements of the $t\bar{t}V$ processes not only constrain the Wilson coefficients of the dimension-6 operators within the SMEFT but also serve as a gateway to BSM physics if any deviations from the SM are observed.

1.2.2 The $t\bar{t}W$ phenomena

At the LHC, unlike other $t\bar{t}V$ processes, the W boson in $t\bar{t}W$ production can only be radiated from the initial state quark. At Leading Order (LO) in QCD perturbation theory, the production of $t\bar{t}W$ only occurs via quark-antiquark annihilation. Contributions from gluon initial states are only possible from Next-to-Leading Order (NLO) in QCD via the gluon-(anti)quark channels. The Feynman diagrams of $t\bar{t}W$ production are shown in Fig. 1.4 with the top row being the LO diagrams and the bottom the NLO.

The lack of the gluon fusion initial state in leading terms contributes to a non-negligible

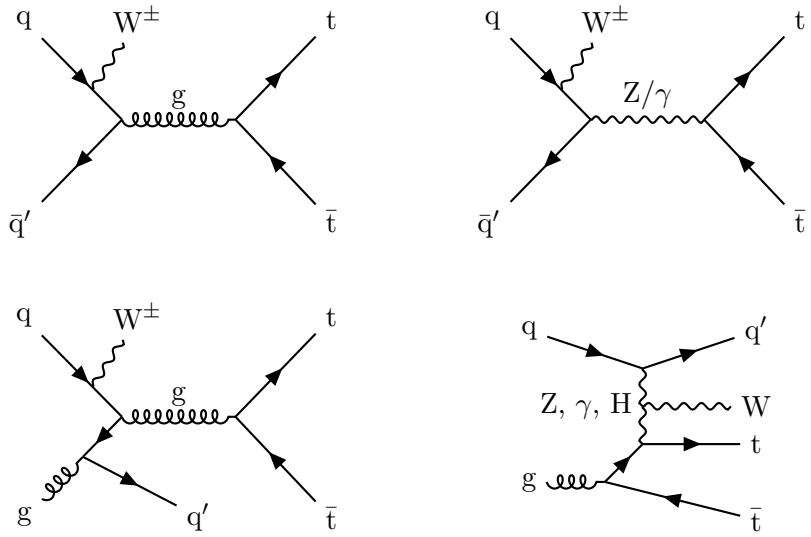

Figure 1.4: Representative Feynman diagrams of $t\bar{t}W$ production at LO (upper row) and NLO (lower row).

difference in the production rates of $t\bar{t}W^+$ and $t\bar{t}W^-$. Since the main production channels are $u\bar{d} \to t\bar{t}W^+$ and $d\bar{u} \to t\bar{t}W^-$, and as a proton contains two u and a d, the abundance of $u\bar{d}$ initial is twice as much as the $d\bar{u}$, hence, the production rate of $t\bar{t}W^+$ is about two times that of $t\bar{t}W^-$, as referenced in Ref. [24].

There have been several measurements of the inclusive cross section of the $t\bar{t}W$ production, carried out by both ATLAS and CMS collaborations at $\sqrt{s} = 7$ [25], 8 [26,27], and 13 TeV [19,28,29]. Prior to the analysis presented in this dissertation, the CMS group at Ghent University played a crucial role in the $t\bar{t}W$ measurement at $\sqrt{s} = 13$ TeV [19]. The $t\bar{t}W$ cross section was observed to be higher than SM predictions. In recent measurements of $t\bar{t}$ production in association with a Higgs boson ($t\bar{t}H$) [30] and the production of four top quarks ($t\bar{t}t\bar{t}$) [31,32], a tension was observed between the observed and predicted yields for the $t\bar{t}W$ process, which represents one of the dominant backgrounds in these results. The main reason for this discrepancy is explained to be the missing higher order corrections in theoretical calculations of the cross section [33–40].

1.2.3 Higher order corrections to $t\bar{t}W$ cross section

The LO is given by the first non-vanishing term of order $\alpha_s^i \alpha^j$, where $i + j$ takes the smallest value and i takes the largest possible value. The parameters α_s and α are strong coupling and fine structure constants, respectively. However, the LO contribution alone generally doesn't give the most accurate prediction, therefore, higher orders need to be taken into account, especially in the case of precision measurements.

At LO, the inclusive cross section of $t\bar{t}W$ is predicted at leading QCD term LO_1 ($\mathcal{O}(\alpha_s^2\alpha)$)

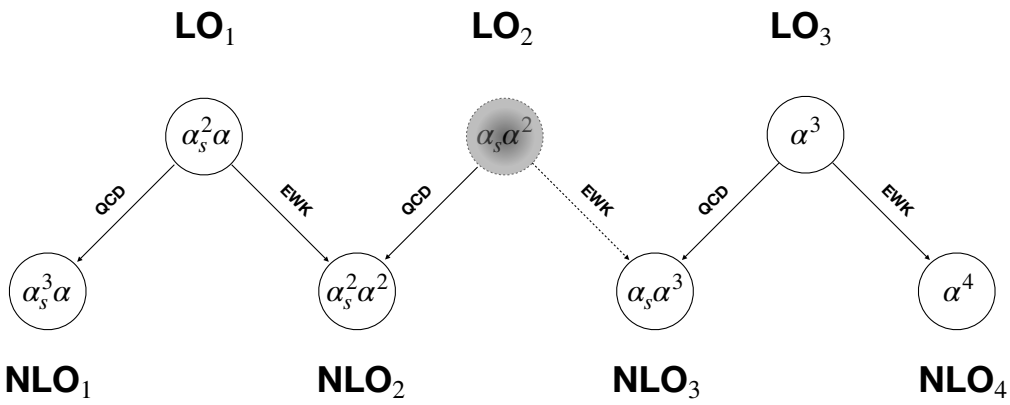

Figure 1.5: Contributions of perturbative correction at LO and NLO to $t\bar{t}W$ production in p-p collisions. The diagram is a modification of Fig. 1 in Ref. [41] with LO and NLO terms labelled based on Ref. [42].

and at sub-leading EWK LO_3 ($\mathcal{O}(\alpha^3)$) at the Born level. The term LO_2, of order $\alpha_s\alpha^2$, arises due to the interference between LO_1 and LO_3. Given that the colour charges are conserved, the term LO_2 vanishes after sum over colours.

Beyond the LO in QCD, the cross section of $t\bar{t}W$ receives several contributions from the NLO in QCD and EWK corrections. Fig. 1.5 shows the perturbative corrections at LO and NLO contribute to the production of $t\bar{t}W$ at tree level. The LO and NLO contributions are labelled with numerical subscript in this section to enhance clarity and ease of explanation. This convention is introduced by the authors of Ref. [42].

At NLO, the term NLO_1 is pure QCD corrections of LO in QCD, it is a leading contribution, approximate 50% of LO QCD. The second term NLO_2 is an EWK correction to LO QCD, it includes QCD corrections to the LO interference LO_2. This term reduces the cross section of $t\bar{t}W$ by about 4%. The term NLO_3 contains pure QCD corrections to LO EWK and doesn't include EWK corrections from LO_2, since LO_2 is zero due to colour structure. NLO_3 is the sub-leading term in NLO, after NLO_1, with a contribution of $\approx 10\%$ to the LO QCD cross section. Finally, the NLO_4 term is an EWK correction to LO in EWK. This term, however, changes the cross section by a very small amount of 0.04%, hence, it can be safely neglected in cross section calculations.

Even though the inclusion of the NLO EWK effects and QCD terms can improve the prediction of $t\bar{t}W$ production cross section, it is still not enough to cover the discrepancy between observations and predictions, a calculation at Next-to-Next-to-Leading Order (NNLO) is needed. However, the computation power needed to perform NNLO calculation is enormous. A NLO QCD with resummation at Next-to-Next-to-Leading Logarithmic (NNLL) is considered instead [34]. Additionally, the NNLL resummation also reduces the theoretical uncertainty with respect to the NLO calculation only.

1.2.4 Charge asymmetries in $t\bar{t}W$ production

The production of $t\bar{t}$ in pp collisions at the LHC is via gluon-gluon fusion at LO, unlike at Tevatron, where $t\bar{t}$ were produced in quark-antiquark annihilation, as a result, $t\bar{t}$ production is highly symmetric. However, QCD theory predicts the charge asymmetry variable A_c^t is non-zero for $t\bar{t}$ production in pp collisions due to $q\bar{q}$ initial state contributions at NLO in QCD. The variable A_c^t is defined as:

$$A_c^t = \frac{N(\Delta_\eta^t > 0) - N(\Delta_\eta^t < 0)}{N(\Delta_\eta^t > 0) + N(\Delta_\eta^t < 0)} \tag{1.19}$$

where $\Delta_\eta^t = |\eta_t| - |\eta_{\bar{t}}|$, and η is the pseudorapidity. On the other hand, at Tevatron, the asymmetry, known as forward-backward asymmetry (A_{FB}), is in fact much higher due to $t\bar{t}$ produced in $p-\bar{p}$ collisions. The results from DZero (D$\emptyset$) [43,44] and Collider Detector at Fermilab (CDF) [45–47] collaborations are consistent with SM predictions.

Unlike $t\bar{t}$, $t\bar{t}W$ production is high asymmetric, which makes it an alternative probe to study $t\bar{t}$ charge asymmetry in pp collisions. Furthermore, the emission of a W in association polarises the initial state quarks, which give the polarisations to $t\bar{t}$. As the result, the decay products of $t\bar{t}$ already exhibit a considerable asymmetry at LO. At $\sqrt{s} = 13$ TeV, the asymmetry A_c^t is predicted to be $0.45^{+0.09}_{-0.06}\%$ for $t\bar{t}$ production and almost five times higher for $t\bar{t}W$ production, at $2.24^{+0.43}_{-0.32}\%$ [24]. Although the $t\bar{t}$ charge asymmetry is higher in $t\bar{t}W$ production, it is still very small, and with the current amount of data collected at the LHC, it is a challenge to perform such measurement. The measurement of charge asymmetry variable is, however, out of scope of this dissertation. As the first step to the charge asymmetry measurement, a precise measurement $t\bar{t}W$ production cross section is presented in this dissertation.

Chapter 2

The Large Hadron Collider and the Compact Muon Solenoid detector

2.1 The Large Hadron Collider

The LHC located at CERN, on the Swiss-Franco border near Geneva, is the largest and most powerful particle accelerator the world has ever built. It consists of a 27-kilometre-long ring of superconducting magnets situated at 100 metre below the ground, it is capable to accelerate beams of protons and nuclei up to the energy scale of tera-electron volt (TeV). For protons, the maximum designed p-p collision energy is $\sqrt{s} =$ 14 TeV, which is equivalent to 7 TeV per proton. At this energy, the protons travel at 99.9999991% the speed of light. During the second run of the LHC, proton beams were collided at a centre-of-mass energy of 13 TeV.

Along the LHC ring, there are four interaction points where the two beam pipes intersect, allowing collisions to occur. Each intersection point features a dedicated detector, namely the Compact Muon Solenoid (CMS) [48], A Toroidal LHC Apparatus (ATLAS) [49], A Large Ion Collider Experiment (ALICE) [50], and LHC Beauty (LHCb) [51] detectors. The CMS and ATLAS detectors are the two high luminosity detectors, they were designed to study a wide range of physics, from the search for Higgs boson to the search for new physics phenomena. Although they are both general-purpose detectors, their technicalities are very different from each other, enabling cross-verification of results, enhancing scientific capabilities, and encouraging diverse approaches to research. On the hand, the ALICE detector was specifically built to study heavy-ion collision, whereas the LHCb detector focuses on the flavour physics sector, they were designed to tolerate a significant lower luminosity than the CMS and ATLAS detectors. In addition to the four main experiments, there are five smaller experiments at the LHC, namely Forward Search Experiment (FASER) [52], LHC-forward (LHCf) [53],

Monopole and Exotics Detector at the LHC (MoEDAL) [54], Total, elastic and diffractive cross-section measurement (TOTEM) [55] and Scattering and Neutrino Detector at the LHC (SND@LHC) [56]. They were designed for the targeted exploration of phenomena that exist within, or extend beyond the framework of the SM.

In order to archive this high collision energy, proton beams are accelerated in a series of accelerators before they are injected into the LHC. Initially, hydrogen gas is ionised by an electric field to produce protons. The protons are then injected into Linear Accelerator 2 (LINAC 2), where they are accelerated to a kinetic energy of 50 MeV. At this energy, the protons enter the Proton Synchrotron Booster (PSB) , then Proton Synchrotron (SP) and Super Proton Synchrotron (SPS). Travelling through these accelerators, protons are progressively accelerated to 1.4 GeV , 25 GeV and 450 GeV, respectively. On leaving the SPS, proton beams are injected into two separate ultra-vacuum beam pipes of the LHC where they are circulated in opposite directions. Fig. 2.1 presents a schematic drawing of the accelerator complex at CERN. In addition to the aforementioned accelerators, the complex also includes several smaller machines that generate beams specifically for conducting smaller-scale experiments.

As proton beams circulate inside the LHC, they pass through 16 Radio-Frequency (RF) cavities where they are accelerated by an electromagnetic field. The beams pass through the RF cavities more than 10 million times and have their energy brought up to 6.5 TeV after about 20 minutes of circulation. The LHC uses 392 quadrupole magnets to precisely steer the beams' trajectory and 1232 dipole magnets to confine the beams inside the beam pipes. The magnets operate at the nominal magnetic field of 8.3 T and a temperature of 1.9 K. Upon entering the interaction points, the beams undergo a process known as "squeezing", which is performed by insertion magnets. Effectively, the diameter of beams is reduced significantly to approximately 16 μm from 0.2 mm. Each insertion magnet system comprises three quadrupole magnets. These insertion magnets are strategically positioned at both ends of each of the four main detectors, resulting in a total of eight insertion magnets distributed throughout the LHC.

The proton beams circulating within the LHC are not continuous, but rather consist of tightly packed bunches, each containing approximately 1.15×10^{11} protons. Each beam contains up to 2808 proton bunches, separated by 25 ns, equivalent to a distance of 7.5 m or a bunch crossing frequency of 40 MHz. With this design, the LHC is expected to have a maximum instantaneous luminosity (L) of order $10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ collisions.

The instantaneous luminosity depends on known LHC parameters, namely the number of protons in each bunch (N_b), number of bunches in each beam (n), the bunch crossing frequency (f), the relativistic gamma factor (γ_r), the normalised beam emittance (ϵ_n), the beta function at the interaction point ($\beta*$), and the factor F which is very close to unity, being introduced to account for the bunch crossing angle at the interaction point. The instantaneous luminosity is calculated as following:

$$L = \frac{N_b^2 n_b f \gamma_r}{4\pi\epsilon_n \beta*} \cdot F \tag{2.1}$$

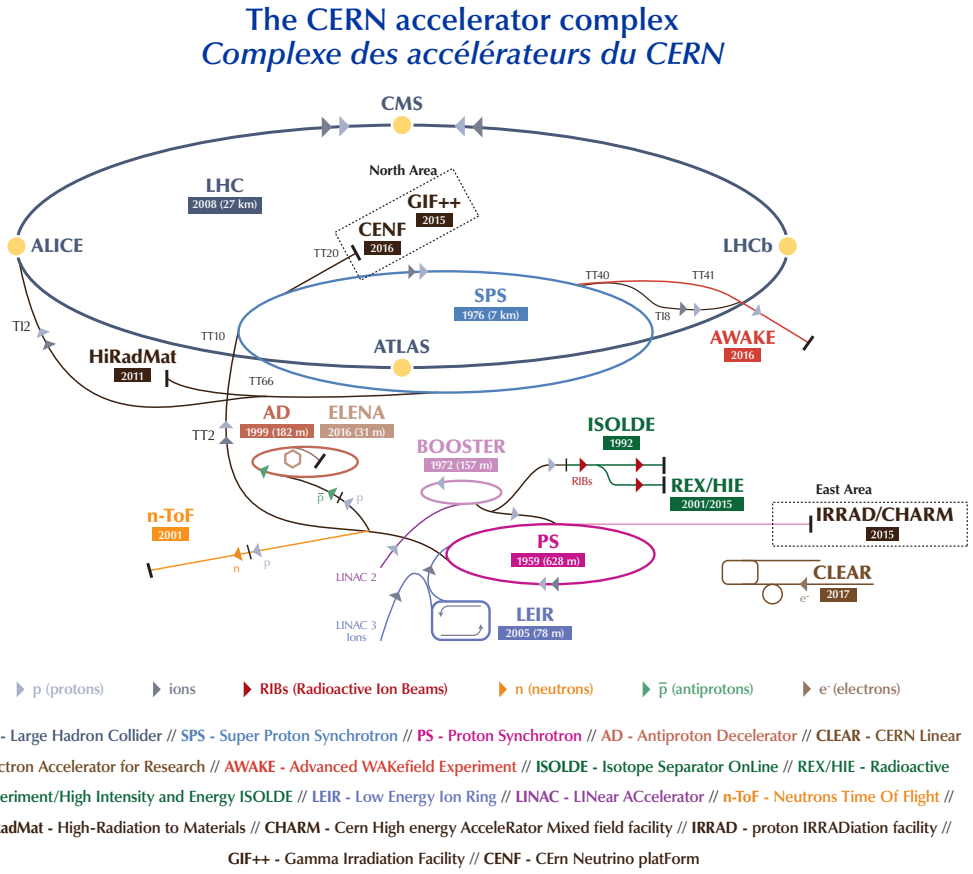

Figure 2.1: The accelerator complex at CERN consists of various smaller accelerators that play a crucial role in preparing particles for injection into the LHC. Additionally, several accelerators are employed in smaller-scale experiments. The directional arrows depicted on the lines symbolise the trajectory of the particles within each accelerator. Image credit: Esma Mobs, CERN [57].

Based on this equation, it is evident that luminosity can be increased with a reduction in emittance by minimising beam spread in both position and angle, and a decrease in $\beta*$ value by tightening focus at interaction points. The instantaneous luminosity is not constant after each injection, however, it drops steadily as the beams collide. Intra-beam scattering and a slow emittance blow-up also contribute to the drop in instantaneous luminosity.

Due to its high luminosity, at each bunch crossing, it is not uncommon to observe more than one p-p collision at the same time. Usually, only one of the collisions is interesting physics' event, it is also known as primary vertex, while other collisions are inelastic collisions, these overlapping events are known as "pileups". During the Run II of LHC, on average, there were about 34 pileup events per bunch crossing. These pileups resulting in particles from multiple collisions mixed together, making it harder to identify particles

arising in the primary vertex. In the CMS experiment, multiple algorithms are used to mitigate the impact of pileup on measurements, and they are discussed in detail in Chapter 2.2.

For most of Run II, the LHC operated with 25 ns bunch spacing, accommodating a maximum of 2556 bunches per beam, with the peak instantaneous luminosity of more than $2 \times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. It delivered $160\,\mathrm{fb}^{-1}$ worth of collision data to the CMS detector. Out of this, the CMS successfully recorded $138\,\mathrm{fb}^{-1}$ of high-quality data, with all sub-detectors fully operational.

2.2 The Compact Muon Solenoid experiment

The Compact Muon Solenoid (CMS) detector, one of the two high luminosity detectors at the LHC, is adeptly designed to scrutinise a vast spectrum of physics phenomena in both proton and ion collisions. Constructed in a cylindrical shape, the detector boasts dimensions of 21 metres in length and 15 metres in height, with a hefty weight of approximately 14000 tonnes. This considerable size positions it as the second-largest detector at the LHC, following the ATLAS detector.

The components of the CMS are arranged in an onion-like configuration, as shown in Fig. 2.2. At the very core of this structure, proximate to the beam pipes, is the tracker system responsible for measuring the trajectories of charged particles. Encircling the trackers is the Electromagnetic Calorimeter (ECAL), a vital component for capturing the energy levels of electrons and photons. Adjacent to the ECAL is the Hadron Calorimeter (HCAL), designed for hadron energy measurement. Surrounding the HCAL is the CMS's solenoid magnet, a cylindrical coil of superconducting fibres, generating a potent magnetic field of 3.8 T.

Finally, the outermost layer accommodates the muon system, encompassing three different types of detectors: Drift Tubes (DTs), Cathode Strip Chambers (CSCs), Resistive Plate Chambers (RPCs), and Gas Electron Multipliers (GEMs). However, the GEM system was installed in the CMS during the Long Shutdown 2 (LS2), spanning 2019 to 2021, and it commenced data acquisition starting from Run III in 2021. The muon system is a crucial feature of the CMS detector, facilitating the exceptional detection and precise measurement of muons. Its strategic placement allows it to capture muons that have traversed through all preceding layers, thus optimising the CMS detector's proficiency in accurately detecting and measuring muons.

The CMS detector is conventionally described by a right-handed coordinate system, wherein the x-axis is oriented towards the centre of the LHC ring, the y-axis ascends vertically upwards, and the z-axis aligns itself with the direction of the beam. Simultaneously, a cylindrical coordinate system is also in place, within which the radial distance r and the azimuthal angle ϕ are defined in relation to the x-y plane. Conversely, the polar angle θ is associated with the r-z plane. The pseudorapidity, symbolised as η, is calculated by the formula $\eta = -ln[tan(\frac{\theta}{2})]$.

2.2.1 The superconducting magnet

The CMS detector's solenoid magnet is a substantial cylindrical structure, with a diameter of 6 metres, a length of 12.5 metres, and a weight tipping the scales at 220 tonnes. With a stored energy of 2.6 GJ, it is capable of generating a uniform magnetic field of up to 3.8 T within its coil. Crafted from superconducting fibres, the coil of the magnet experiences no resistance when electricity courses through it. To preserve this zero-resistance state, the coil is kept at an ultra-low temperature of 4.5 K. This refrigeration enables the magnet to generate a robust magnetic field while employing a relatively

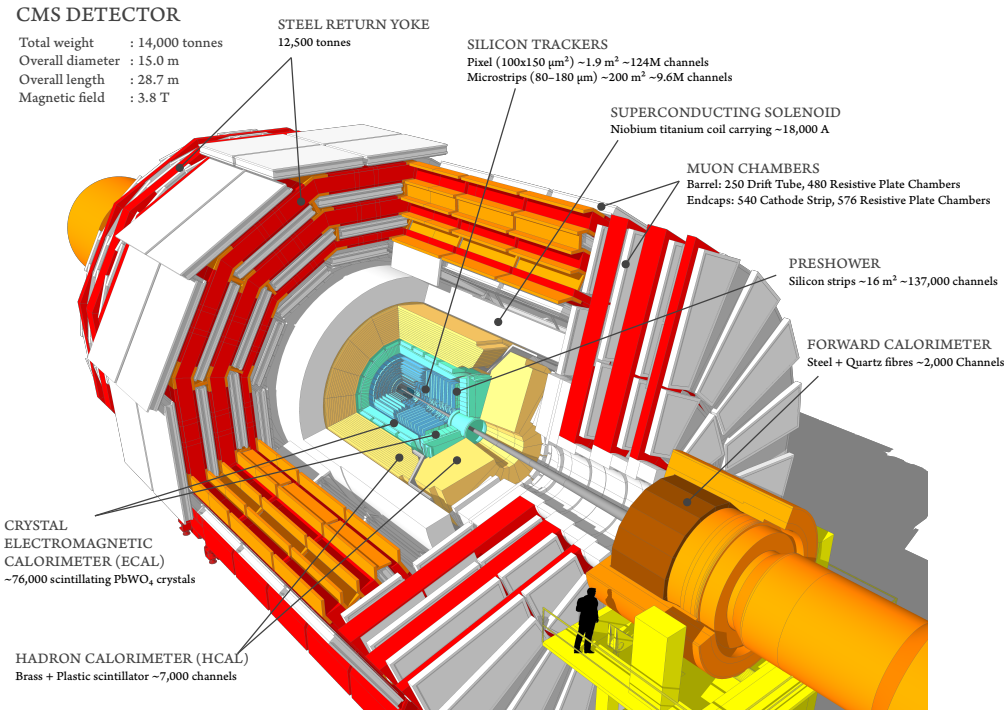

Figure 2.2: A detailed schematic illustration, revealing the comprehensive structure and individual components of the CMS detector [48].

minimal amount of power.

Enveloping the magnet are the iron return "yokes", which contribute significantly to the detector's total weight, with one barrel yoke divided into three layers and two endcap yokes each partitioned into two. These yokes facilitate a path for the magnetic field line to circulate from one end of the magnet to the other, effectively closing the magnetic loop. Fig. 2.3 illustrates the magnetic field within the CMS detector, highlighting that the field is most powerful within the coil, diminishing considerably outside of it.

2.2.2 The tracking system

Situated just a few centimetres from the beams, the tracker stands as the CMS detector's innermost component, serving as the first point of detection for particles resulting from collisions. Enveloped well within the solenoid magnet, the tracker benefits from a uniform magnetic field of 3.8 T, equipping it to reconstruct the paths of charged particles accurately, and precisely measure their charge and momentum.

Due to its proximity to the beam, the tracker endures an extreme environment filled with high radiation and a substantial influx of particles. Silicon detector technology was chosen to rise to this challenge, boasting a high granularity, a rapid response, and robust resistance to intense radiations. The tracker comprises two key elements: a pixel

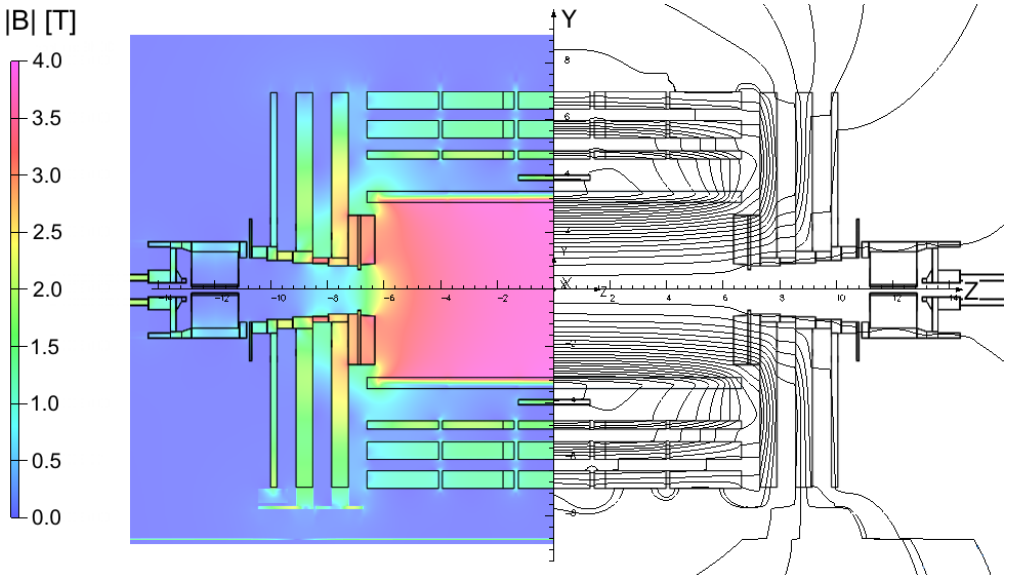

Figure 2.3: Illustration of the magnet field confinement within the structure of the CMS detector [58].

detector and a silicon strip detector.

The pixel tracker, as the innermost part of the CMS tracker, originally consisted of three barrel layers (BPIX) at radii of 4.4, 7.3 and 10.2 cm, along with two disks (FPIX) at each end, distanced 345 and 465 mm from the interaction point. This arrangement allowed the pixel to cover a pseudorapidity range of $-2.5 < \eta < 2.5$. The pixel detector is made up of an array of tiny silicon pixels, each functioning as an individual sensor. In the original designed, approximately 66 million pixel cells of $100 \times 150\,\mu\text{m}^2$ were scattered across the barrel and endcap regions, facilitating extremely precise positional measurements.

Originally designed to operate at a peak luminosity of $1 \cdot 10^{35}\,\text{cm}^2\text{s}^{-1}$, but with the upgrade of the accelerator during the Long Shutdown 1 (LS1) in the years 2013-2014, the pixel detector expected to experience a peak luminosity that exceeded this figure in Run II. To accommodate this increase in luminosity and maintain its efficient, the pixel was replaced with an upgraded model [59] during the technical stop at the end of 2016 and early 2017. In comparison to its predecessor, the improved pixel detector, as depicted in Fig. 2.4, comprises four barrel layers with three endcap disks, collectively housing 124 million pixels. This upgrade enhances the pixel detector's tracking performance and radiation tolerance. Its high spatial resolution is instrumental in identifying closely spaced tracks, a necessity for the study of particles produced in dense environments.

Surrounding the pixel tracker is the silicon strip detector, featuring an array of microstrip detectors that covers the radial region between 20 cm and 116 cm. It emulates the pixel detector's barrel and endcap modules design, allowing it to cover the pseudorapidity range $-2.5 < \eta < 2.5$. The strip detector is segmented into four subsystems: Tracker

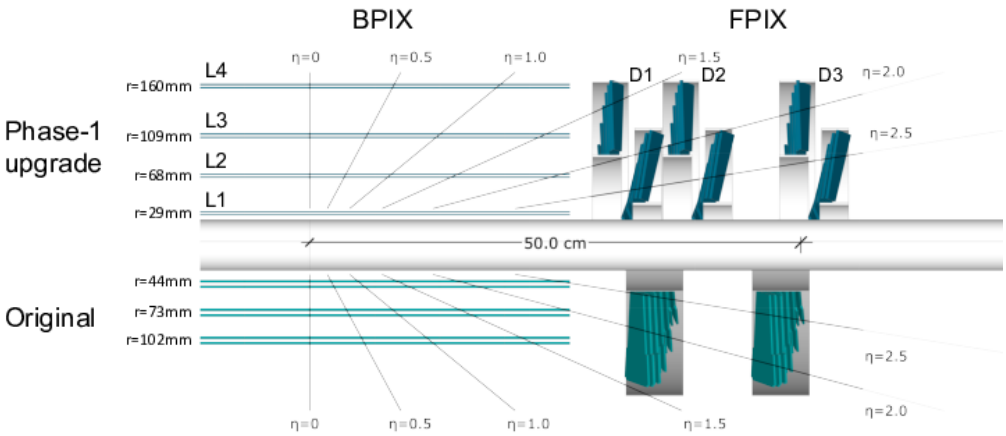

Figure 2.4: Depiction of the pixel detector's layout before (bottom) and after (top) the 2016-2017 technical stop period upgrade [48].

Inner Barrel (TIB), Tracker Inner Disk (TID), Tracker Outer Barrel (TOB) and Tracker End-cap (TEC), as shown in Fig. 2.5. Each subsystem consists of numerous layers of silicon strip detectors, strategically positioned to permit module overlap, thus enabling multiple trajectory measurements for each particle. This intelligent design optimises particle track resolution and provides hit detection redundancy, offering a backup for potential module failure or underperformance.

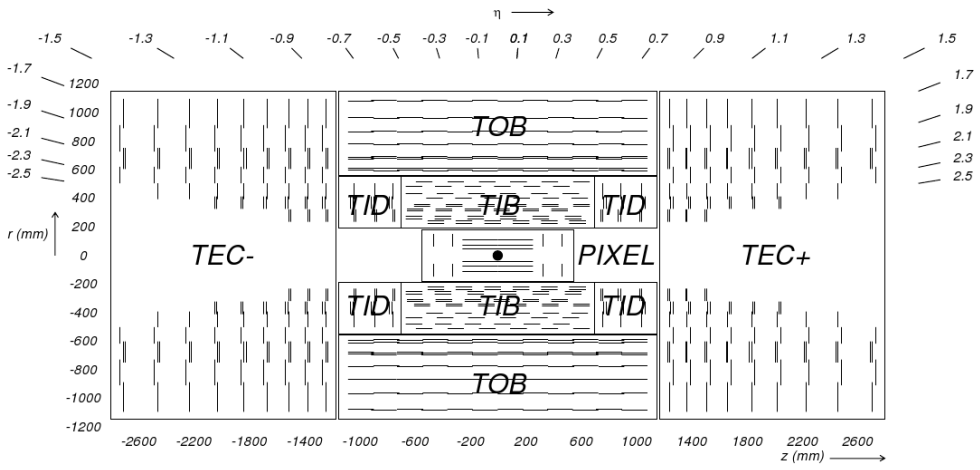

Figure 2.5: Tracker system schematic in the $r - z$ plane, with each line representing a detector module [48].

The size of the silicon micro-strips varies according to their location. In the tracker's inner region, they measure 10 cm in length and 80 μm in width, with a thickness of

320 μm. However, in the outer region of the tracker, these micro-strips are larger, with a length of 25 cm, a width of 180 μm, and a thickness of 500 μm. This design was chosen to provide robust pattern recognition and a momentum measurement with a precision of 1-2%.

The tracker was designed to withstand extreme radiation conditions for up to 10 years, a requirement that necessitated careful consideration of various radiation effects during the design phase. These effects include surface damage to the pixel and strip detectors, modifications to the crystal lattice, increase in detector leakage current, and overheating of the detectors. To mitigate these effects, the whole tracker system has to operate under $-10°$C, this operating temperature is expected to steadily decrease to $-27°$C over the years. As such, all component of the tracker must be able to survive temperature cycles between room temperature and $-30°$C.

2.2.3 The electromagnetic calorimeter

Built in a barrel-and-endcap structure, the Electromagnetic Calorimeter (ECAL) was designed to align with the tracker's coverage, reaching up to $|\eta| < 3.0$. Its primary function is to detect electrons and photons with high energy and positional resolutions. Fig. 2.6 illustrate each component of the ECAL. It was designed to be compact and have a high granularity with more than 75800 lead tungstate ($PbWO_4$) crystals. When a high-energy particle enters the ECAL, it interacts with the lead tungstate crystals, causing it to lose energy in the form of a shower of secondary particles. These particles then cause the crystals to scintillate. The intensity of the scintillation light is proportional to the original energy of the particle. The scintillation light is then detected by Avalanche Photodiodes (APDs) detectors in the barrel region and Vacuum Photo-triodes (VPTs) in the endcap regions, and the signal is then amplified, digitised, and sent to the data acquisition system. The different type of detectors used in barrel and endcap regions are motivated by the difference levels of radiation and its energy. In the endcap, the levels of gamma and neutron radiation are much higher than that in the barrel region, making APDs not a suitable type of detector, which is optimised for lights with a wavelength of about 420 nm [60, 61].

Lead tungstate crystals, transparent to their entire scintillation emission spectrum, have a high density of 8.28 g/cm^3, an exceptionally short radiation length of 0.89 cm, and a small Molière radius of 2.2 cm. As a result, the ECAL is a homogenous calorimeter with a high granularity, allowing it to archive an excellent energy and position resolution. The crystal scintillation decay time coincides with the order of the LHC bunch crossing time, with 80% of the light emitted in 25 ns at a peak wavelength of about 420 nm. The crystals are arranged in a quasi-projective geometry to avoid cracks aligned with trajectories of particles, resulting in their axes forming a small angle with respect to the vector from the interaction vertex. In the barrel region, the crystals span 230 mm in length and posses a cross section of 22×22 mm^2, while they manifest a dimension of 220 mm in length and a cross section of 28.6×28.6 mm^2. Overall, the energy resolution of ECAL ranges between 2% in the barrel region to 5% in the endcap. This resolution

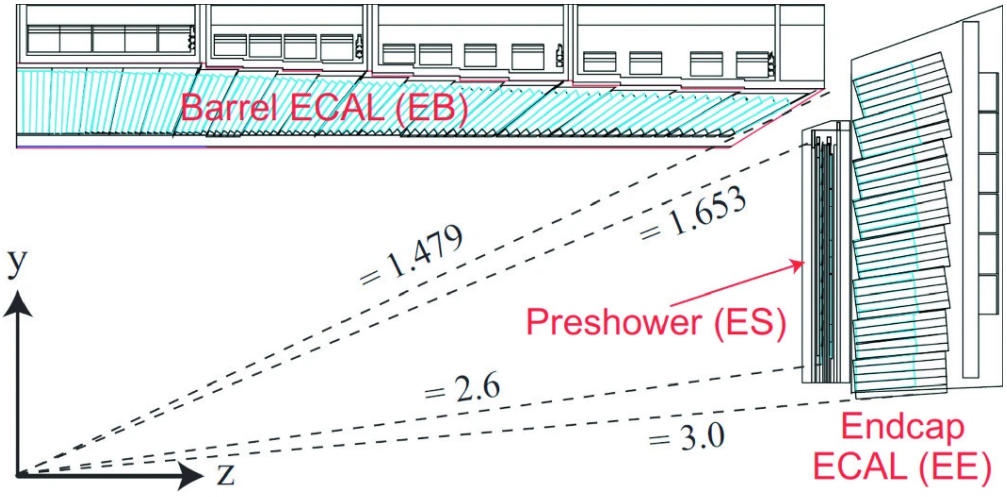

Figure 2.6: A visualisation of the ECAL on the $z - \eta$ plane, with each component depicted. The cyan blue lines denote the positions of the detector crystals [62].

is measured in $Z \to e^+e^-$ events, and it remained stable through the whole Run II data taking period of the CMS [63].

Located in front of each endcap, a Preshower detector serves the primary function of identifying neutral pions within a specified absolute pseudorapidity range of $1.653 < |\eta| < 2.6$. It also assists in differentiating signals from neutral pions and improve the precision of position determination for electrons and photons with high granularity. The Preshower is a 20 cm thick sampling calorimeter, comprising two planes of silicon strip sensors, interleaved with a lead radiator. The lead radiators initiate electromagnetic showers from incoming photons and electrons, whilst the silicon strip sensors, located after each radiator, measure the deposited energy and the transverse shower profiles. Each silicon strip sized $60 \times 1.9\,\text{mm}^2$ and 300 μm thick—compared to the ECAL crystal's $28.6 \times 28.6\,\text{mm}^2$ dimension—establishes the Preshower as a fine-grain photon-sensitive detector. It can, therefore, detect photons and electrons from π^0 decays, which typically exhibit small separation, about 8 mm, in the endcap region.

2.2.4 The hadronic calorimeter

Following the ECAL, the Hadron Calorimeter (HCAL), another layer of the CMS detector, comes into play. Its unique role lies in detecting and measuring the energy and position of high momentum hadrons. The HCAL, which envelops the ECAL, is engineered to maximise the amount of energy deposition by strongly interact with particles, including neutral hadrons that pass through the ECAL without any interaction. This function is crucial in analyses where jets appear in the final states.

Comprising multiple components, the HCAL includes a Barrel Hadron Calorimeter (HB) and a End-cap Hadron Calorimeter (HE) situated within the solenoid magnet, a Outer

Hadron Calorimeter (HO) enveloping the solenoid magnet, and a Forward Hadron Calorimeter (HF) positioned in the farthest region of the overall CMS detector. The longitudinal view of these individual HCAL elements is depicted in Fig. 2.7.

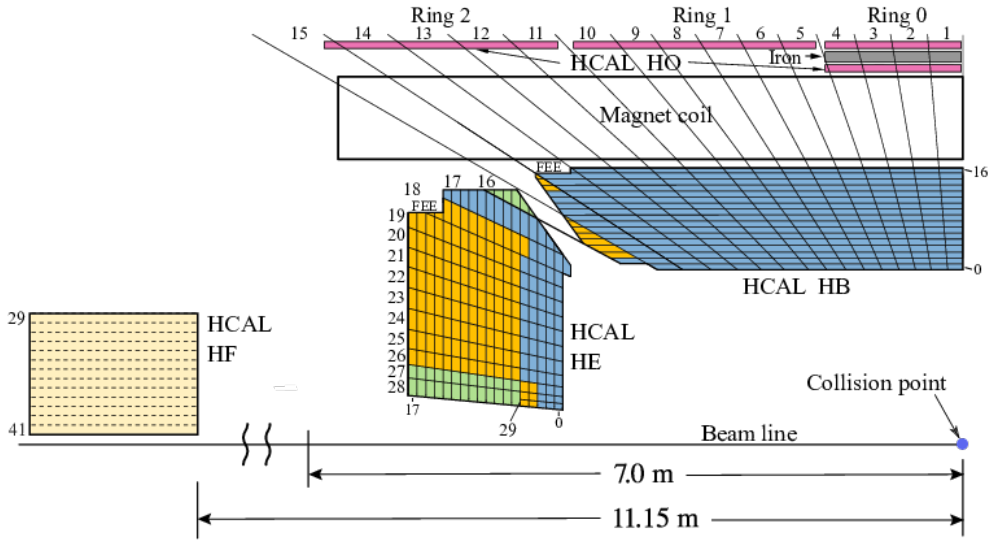

Figure 2.7: Longitudinal schematic of the HCAL, indicating the locations of HB, HE, HO, and HF relative to the magnet and the beam line [62].

The HB covers a pseudorapidity range of $|\eta| < 1.3$, and is composed of 36 identical half-barrel wedges, arranged parallel to the beam line. These wedges consist of alternating layers of absorber plates and plastic scintillators. The scintillators are divided into 16 η sectors, providing the HB a granularity of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. Two of the absorber plates, located on the outermost and innermost layers, are composed of steel, with the remaining plates made of brass. When a hadron strikes the absorber plates, it interacts strongly with nuclei of the absorber material, resulting in a cascade of secondary particles. These secondary particles further interact with other layers of the absorber, triggering additional hadronic showers. Charge particles in theses hadronic showers then interact with scintillators, producing light signals that are captured by photodetectors. Moreover, an additional scintillator layer, known as layer 0, is placed in front of the first absorber plate to sample hadronic showers developing between the ECAL and HCAL.

The HEs cover pseudorapidity range of $1.3 < |\eta| < 3.0$, where they detect about 34% of particles produced in final states. Each HE consists of 36 megatiles that together form a disc on the $x - y$ plane, with the Preshower and the endcap of the ECAL attached to it. The design of the HEs is similar to the HB, with alternating layers of absorber plates and plastic scintillators. However, due to its position at the end of the magnet, non-magnetic material is used for the absorber plates, leading to all the plates being made of brass. The granularity of the HEs is the same as the HB for $|\eta| < 1.6$, and

changes to $\Delta\eta \times \Delta\phi = 0.17 \times 0.17$ for $|\eta| \geq 1.6$.

The HO, located in the barrel region outside the magnet coil, is designed to measure the energy of particles that have passed through the HB without being detected, due to the spatial constraint of the HB. The HO covers the same pseudorapidity acceptance as the HB, $1.3 < |\eta|$. It contains plastic scintillator tiles with wavelength shifting fibres. The central ring features two layers of scintillators sandwiching the first layer of the return yoke, while the rest of the barrel has one layer of scintillators behind the first layer of the return yoke. The first layer of the return yoke, with a thickness of 30 cm, also serves as an absorber for the HO. Signals from the scintillators are collected by the wavelength shifting fibres and processed by a silicon photomultiplier.

Finally, the HFs, located at both end of the CMS detector, are specifically engineered to measure particles in the very forward region of the detector, with a pseudorapidity range of $3.0 < |\eta| < 5.0$. They differ in design from the rest of the HCAL due to the significantly high radiation environment in the forward region. For instance, at $|\eta| = 5.0$, the HFs are expected to encounter ≈ 10 MGy after ≈ 10 years of LHC operation. Constructed from a steel absorber structure, they are composed of 5 mm grooved plates filled with quartz fibres, employ the quartz fibre calorimetry technology. When a high energy charged particle, created by the interaction of a hadron with the absorber, traverses through the quartz fibre at a velocity surpassing the speed of light in quartz, it emits photons as a result of the Cherenkov effect. The photons are then detected by photomultipliers, which are shielded behind 40 cm of steel and borated polyethylene slabs.

2.2.5 The muon system

Positioned as the external layers of the CMS detector, the muon system is a crucial component of the CMS detector that is designed to identify muons from collision with remarkable precision of a few ns [64]. Due to their high mass, muons lose significantly less energy through bremsstrahlung effect than elections do, enabling them to traverse and emerge from the inner components of the CMS detector. This characteristic necessitates the positioning of the muon system in the detector's most external layer.

The structural design of the muon system features a series of concentric cylinders that envelop the collision point, complemented by endcaps at both ends to cater for muons produced at a wide range of angles. It encompasses three distinct types of gas detectors: Drift Tubes (DTs), Cathode Strip Chambers (CSCs), and Resistive Plate Chambers (RPCs). These detectors are interlaced between the layers of the return yoke. Their alignment with respect to other CMS sub-detectors is presented in Fig. 2.8.

The DTs are located in the barrel region of the detector, covering the pseudorapidity range of $|\eta| < 1.2$, where both residual magnetic field and neutron-induced background are small. Each DT chamber consists of multiple layers of aluminium tubes filled with a gas mixture, with a charged wire placed inside each tube. As a muon traverses a drift cell, it ionises the gas. The ionised electrons drift towards the wire, gaining energy as they approach. This subsequently triggers additional ionisations, causing an avalanche

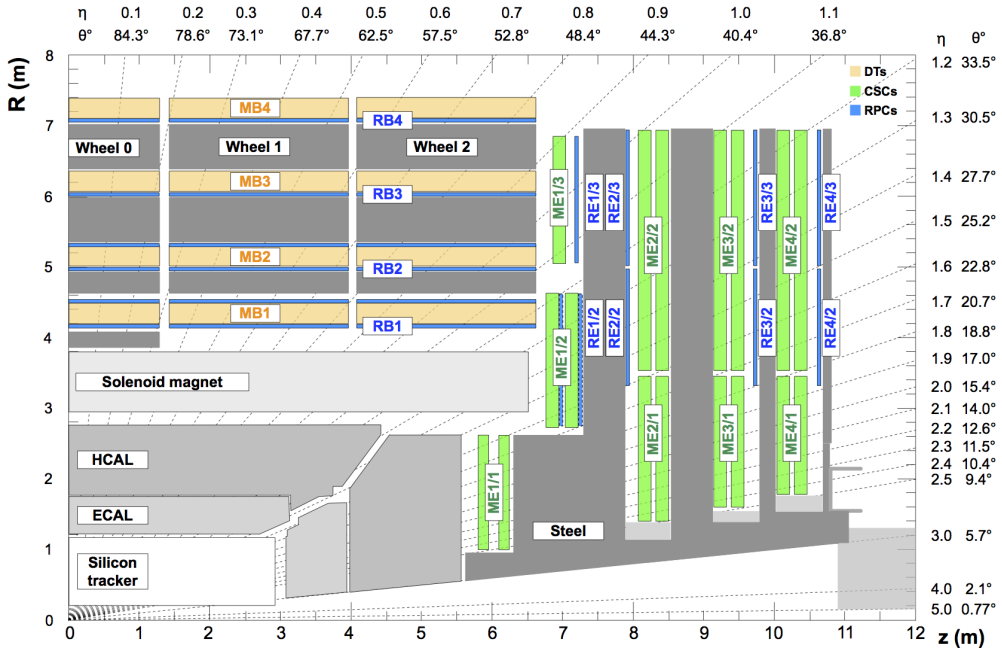

Figure 2.8: Projection of the muon system onto the $z - \eta$ plane, showcasing the spatial arrangement of each component within the CMS detector [64].

of ionisations near the wire. This avalanche induces a noticeable electric current change, forming a signal that is detected by a Time-to-Digital-Converter (TDC) as a drift time. The drift time measurement accounts for the muon's time-of-flight from the interaction point, the signal propagation along the wire, and the trigger. The ionisation position caused by the muon can be determined from the drift time and velocity. The drift velocity, dependent on the wire's potential, the gas composition, and the gas temperature and pressure, can be calculated using the known parameters of the drift tube. The DT detectors boast a small spatial resolution of 200 μm and radial resolution of 100 μm, with an efficiency of 99.9%.

The CSCs are implemented in the endcap regions, covering the pseudorapidity range of $0.9 < |\eta| < 2.4$. In this specific region, the magnetic field is stronger and uneven, and particle rates are higher than in the barrel, rendering the DT an unsuitable type of detector. Each CSC is made up of seven cathode panels interleaved among six anode wire planes. The panels house cathode strips, with the gaps between the panels filled with gas. The CSCs function similarly to the DTs, wherein ionised electrons create a signal on the wires and the cathode strips. The ionisation locations caused by the muons are identified by the grid of cathodes and anodes, thereby eliminating the need for the gas conditions inside the CSCs to be precisely controlled.

Contrary to the DTs and the CSCs, the RPCs are present in both the barrel and endcap regions, covering the full pseudorapidity range of the muon system, $|\eta| < 2.4$. A RPC

unit comprises two parallel plates with high resistance, the gap between the plates is filled with gas. Two units are placed adjacent to each other to form a double-module with a common set of read-out strips in between. A high voltage is applied to the top and the bottom plates to create a strong electric field in the gas to attract ionised electrons, which produced in the interaction between muons and the gas, toward the read-out strips in the middle. While DTs and CSCs exhibit longer response times, the RPCs have much shorter times in which signals can be recorded within a few nanoseconds. This results in an excellent time resolution of a few ns, making them valuable for triggering decisions. For muons with $p_{\mathrm{T}} \leq 100\,\mathrm{GeV}$, the energy resolution is about 1% in the barrel and 3% in the endcap [64].

During the LS2, between 2018 and 2022, the Gas Electron Multipliers (GEMs) were installed in the CMS muon system, and it started to take data from the third run of the LHC, from 2022. GEMs are complementary to the existed muon system, extending its coverage to the very forward region. They play an important role in maintaining the performance of muon reconstruction and trigger system under high background environment during High-Luminosity LHC (HL-LHC).

2.2.6 Data acquisition system and trigger

With a bunch crossing time of 25 ns at the LHC, the CMS detector anticipates approximately 4×10^7 collisions every second. This number is extraordinarily high, it surpasses the current computing capability. However, not all collisions yield interesting physics phenomena, in fact, only a small fraction of collisions proves useful for physics analysis. To select interesting events, the CMS employed a two-tier trigger system: the Level-1 trigger (L1) and the High-Level Triggers (HLTs).

The L1 is a hardware-based trigger system that can make a decision every 25 ns, in sync with the bunch crossing time. It requires the sub-detector to have superb time resolutions, ensuring signals from various sub-detectors are synchronised and assigned to the same event. The L1 uses coarsely segmented data from the muon system (DT, CSC and RPC) as well as from the calorimeters (ECAL and HCAL), and is designed to have an output rate of 100 kHz, equivalent to a decrease of 400 times. It has local (also known as Trigger Primitive Generator (TPG)), regional and global components, as shown in Fig. 2.9.

In the calorimeters, the TPGs search for high-energy deposits in their respective subsystem and pass the information to the regional calorimeter trigger. Here, objects are assigned to the energy deposit, and any information pertinent to muons is forward to the global muon trigger. The global calorimeter trigger does not only identify jets and their kinematic quantities, but it also computes the missing transverse momentum and ranks the objects. The ranks correspond to confidence levels in the identification of objects.

In the muon system, each sub-detector delivers information in the form of track segments to the local trigger, which then forwards this information to the regional muon triggers

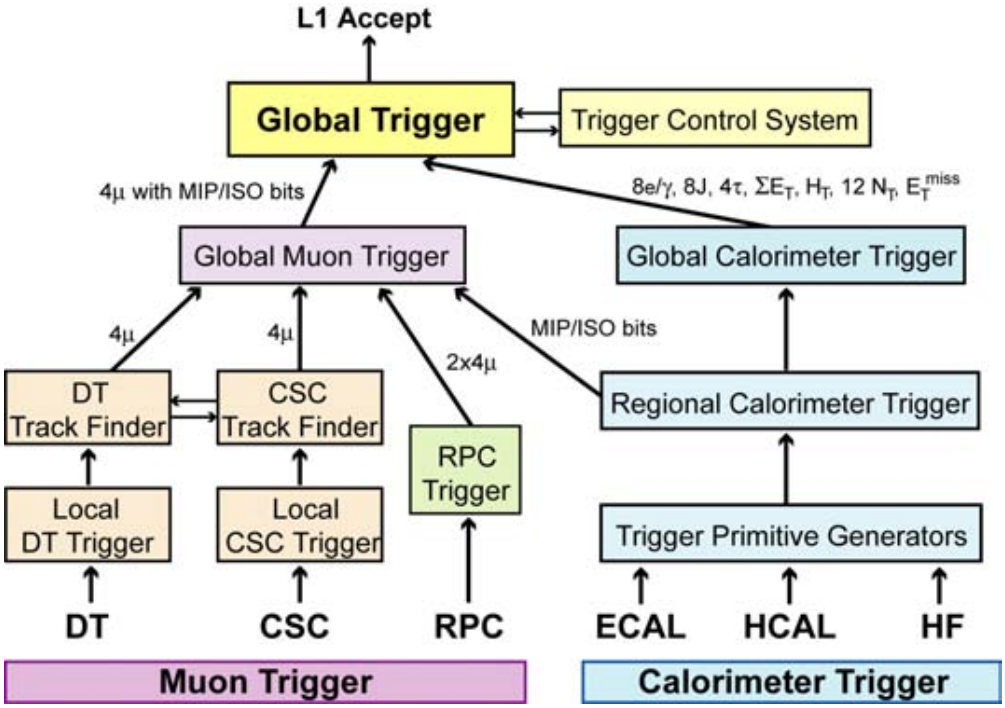

Figure 2.9: A flow chart showing the L1-trigger architecture at the CMS, outlining each component [48].

(DT and CSC track finders). Here, complete tracks are reconstructed with kinematic quantities assigned, and muon candidates are sorted by rank. The global muon trigger aims to enhance the trigger efficiency of muon using information from regional muon triggers and regional calorimeter triggers. Ultimately, using information from global muon and global calorimeter trigger, the global trigger makes a decision to either accept or reject an event.

Upon passing the L1, the selected events are directed to the HLT, a software-based trigger system that can utilise all information from all sub-detectors, including the tracker system, and conduct a comprehensive event reconstruction. The HLT has more time and processing power to make a decision compared to the L1. Therefore, it employs full detector information, including the tracker data, and performs an in-depth reconstruction of the events. The HLT further reduces the event rate to around 1 kHz, which is the rate at which data can be stored for subsequent analysis.

The trigger requirements applied to each event are known as trigger paths, most of which are based on the presence of a physics object such as an electron, a muon, or a jet, and the kinematic quantities associated with these objects. When an event is rejected by the triggers, it is permanently discarded and cannot be retrieved. Therefore, trigger paths are applied using "OR" logic in order to accept a wider variety of events, enhance efficiency, and prevent the rejection of interesting physics events. On the other hand, the

accepted events are saved as “raw” data, containing digital signals from each subsystem. These events must first be reconstructed from raw data before they can be utilised for further analyses.

2.3 Object and event reconstruction

The design of the CMS subsystems is tailored to register signals from distinct types of physics objects such as photons, electrons, muons, and jets. The tracker system gathers information about the trajectories of these physics objects upon their production. The ECAL is dedicated to detecting photons and electrons, while hadrons are detected by the HCAL. Of all these particles, only muons possess the capability to traverse through most layers of the CMS before finally being detected by the outermost layer, the muon system.

To reconstruct each particle from the multitude of signals detected by these subsystems, CMS employs the ParticleFlow (PF) algorithm. This algorithm amalgamates information from all subsystems, thereby creating a comprehensive picture of each proton-proton collision event. It identifies candidate particles with high precision, associates trajectories with each particle, and computes their kinematic observables. For charged particles, their respective charges are discerned from the curvature of their tracks, a result of their trajectories being deflected by the strong magnetic field.

2.3.1 Track and vertex reconstruction

As charged particles traverse the silicon tracker, they deposit some energy. This energy deposit is subsequently amplified and recorded. The CMS employs the Combinatorial Track Finder (CTF) algorithm, which is based on the Kalman Filter (KF) technique [65–67], to discern all the tracks based on the hits recorded by the tracker. The CTF algorithm is a three-tier process: seeding, track finding and fitting. Neutral hadrons, which do not interact with the tracker through ionization nor are affected by the magnetic field, are detected indirectly. This detection occurs through the observation of secondary particles produced when neutral hadrons undergo nuclear interactions with the material of the detector.

The first stage, seeding, involves selecting any two or three hits from two or three consecutive layers that could potentially form tracks of charge particles originating from the interaction region. Despite hits being considered from both the pixel and silicon strip detectors during this process, the pixel detector usually offers superior track seeding due to its specific configuration and placement.

These seeds serve as the starting points for the second stage - track finding, which utilises the KF approach. In this stage, tracks are extrapolated to the next layer of the tracker, whether it's the $r-\phi$ layer in the barrel or the z layer in the endcaps. If a compatible hit is found, it is added to the track before moving on to the next layer. This extrapolation takes the magnetic field into account to compute the tracks' curvature and locate the hit in the next layer. The KF algorithm can bypass a layer if no compatible hit is found, treating it as a missing hit. This track extrapolation yields valuable information such as momentum, curvature, location of the next hits, their uncertainties, and it also assigns a weight to each track indicating its quality.

Finally, once the search for compatible hits reaches the tracker's outermost layer, each track is then fitted with all the associated hits, beginning from the innermost layer and moving outward. This fit also eliminates any hit that has been unambiguously assigned during the track finding step. The tracks are then refitted, but this time in the opposite direction - from the outermost layer inwards - yielding the best estimates of tracks. Tracks possessing at least eight hits and of the highest quality - in cases where multiple tracks share the same hit - are retained for further analysis. Tracks are discarded if they originate outside a cylinder of a few millimetres radius centred around the beamline or if they have a transverse momentum (p_{T}) of less than 0.9 GeV. This process also provides information on multiple vertices that are consistent with the tracks. However, only the vertex with only the vertex that has the highest sum of p_{T} of its associated tracks, designated as the Primary Vertex (PV), is considered for analysis. While the other vertices, which are mostly due to pileup, are not relevant to the analysis. The CTF algorithm achieves a reconstruction efficiency of 99% for isolated muons, and between 70-80% for electrons and charged pions.

2.3.2 Muon reconstruction

Muon tracking relies not only on the tracker but also on the muon sub-detectors (DT, CSC, and RPC), ensuring efficient identification of muons. Depending on how their tracks are recognised, muons can be classified into three categories:

- **Standalone muons** are those with tracks reconstructed exclusively within the muon system. Track segments are formulated by assembling hits within each sub-detector and then fitted to generate the final standalone-muon track.
- **Global muons** are standalone muons whose tracks are compatible with those identified in the tracker. This matching is performed for all standalone muons. Subsequently, hits from tracker tracks and standalone muon tracks are combined and fitted to yield the global-muon tracks. At $p_{\mathrm{T}} \gtrsim 200\,\mathrm{GeV}$, the global-muon fit enhances momentum resolution.
- **Tracker muons** are identified by an inner track with $p_{\mathrm{T}} > 0.5\,\mathrm{GeV}$ and $p > 2.5\,\mathrm{GeV}$, which is extrapolated and matched with at least one track segment in the muon system. The extrapolation and the segment are deemed matched if the distance between the x-components of their coordinates lies within 3 cm, or if the ratio of the distance to its uncertainty is less than 4.

If a global muon and a tracker muon share an identical track in the tracker, they are consolidated into a single muon candidate.

Within the acceptance of muon detectors, approximately 99% of muons are reconstructed as either a global muon or a tracker muon, and often as both. As a result, muon identification is based on the properties of global and tracker muons. Variables pertinent to muon reconstruction such as track fit χ^2, number of hits per track, and muon segment compatibility are computed by the PF algorithm. The muon segment compatibility

score reflects the quality of the match between tracker track and muon segments, it is influenced by the number of matched segments when a tracker track is extrapolated to the muon system, as well as the closeness of the match. This compatibility score ranges from 0 to 1, with 1 representing the highest compatibility.

A set of selection criteria is defined to identified muons, this identification is standard for all CMS analyses, and each analysis may decide to apply more specific selection criteria, in the case of this analysis. The PF muon selection has three widely-used working points [68]:

- Loose muon identification includes all muons that are either tracker muons or global muons. With such a broad criterion, loose muon identification achieves an efficiency of over 99.8% [68].
- Medium muon identification includes loose muons that hit at least 80% of inner tracker layers. If they are tracker muons, their segment compatibility must exceed 0.451; if global muons, it must be above 0.303. However, for global muons, the goodness-of-fit per degree of freedom (χ^2/dof) must be less than 3. Furthermore, the position match between the tracker muon and standalone-muon must have $\chi^2 < 12$, and the maximum χ^2 computed by the kink-finding algorithm [69, Section 6.4.3] must be less than 20. This working point is optimised for prompt muons and muons from heavy flavour decays, with a selection efficiency exceeding 98%.
- Tight muon identification requires a loose muon to qualify as both a tracker muon and a global muon. Its inner track must be consistent with hits in at least 6 layers, one of which must be a pixel layer. The muon must also be compatible with the PV with the transverse impact parameter ($|d_{xy}|$) being less than 0.2 cm and the longitudinal impact parameter ($|d_z|$) less than 0.5 cm. In addition, the tracker muon track must match with at least 2 segments in the muon system, while the global muon fit must have $\chi^2/\text{dof} < 10$. This working point is aimed at eliminating muons from decays in flight and hadronic punch-through. It boasts an efficiency of 96% in the region with $|\eta| < 0.9$ and 97.5%, otherwise.

In this analysis, PF-medium muons serve as an initial baseline upon which more specific selection criteria are applied. The details of these additional selection parameters will be elaborated upon in Sec. 3.2.

2.3.3 Electron and photon reconstruction

The reconstruction of electrons and photons relies on information derived from the tracker and the ECAL, specifically the reconstructed tracks in the inner tracker and energy hits in the ECAL. As electrons travel through the detector, the intense magnetic field bends their trajectory in the $x - y$ plane. Electrons interact with the detector material and emit bremsstrahlung photons, resulting in a considerable energy loss. At $|\eta| \approx 1.5$, where the tracker is thickest, there is an 85% [70] chance of bremsstrahlung photons being emitted. This, coupled with conversions of photons

and further bremsstrahlung effects, makes the accurate reconstruction of electrons and photons a challenging task. Because electrons and photons leave similar signals in the ECAL, their differentiation relies on the absence of interactions between photons and the tracker material.

To overcome this, energetic ECAL clusters with a transverse energy (E_{T}) of at least 4 GeV are leveraged to account for energy hits from both the electron and bremsstrahlung photons. A supercluster contains ECAL clusters within a small η window and a broader ϕ window, which accounts for bremsstrahlung photons spreading predominantly along the ϕ direction. The seed cluster, being the one with the highest energy in the region, initiates the supercluster formation, and its transverse energy must be at least 1 GeV.

The supercluster is subsequently extrapolated to the pixels, where compatible trajectory seeds and the supercluster seed initiate the Gaussian-Sum filter (GSF) tracking step [70, 71]. Simultaneously, all reconstructed tracks in the event are also extrapolated to the ECAL to evaluate their compatibility with an electron trajectory. If compatibility is established, these tracks are included in the initiation of the GSF tracking step.

In an effort to identify electron-positron pairs originating from photon conversions, generic reconstructed tracks from the tracker are used in an iterative tracking procedure, which aims at finding track from displaced vertices. This procedure is described in Ref. [72].

The PF algorithm then uses ECAL clusters, superclusters, GSF tracks, generic tracks to associate with electrons, and conversion tracks as a foundation to segregate electrons and photons. A set of selection criteria is then applied to the potential electron/photon candidates to further differentiate them. An object that passes both the electron and photon criteria is classified as electron if it has a GSF track that contains a hit in the first layer of the pixels. To optimise the selection of electrons and photons, the CMS employs a Multivariate Analysis (MVA) technique [70], which classifies the objects based on their kinematic observables and energy deposit in the ECAL and the HCAL. Three working points: loose, medium and tight are defined based on the MVA output with efficiencies of 90%, 80%, and 70%, respectively.

Unlike muons, the measurement of electron charge is effected by bremsstrahlung effect and photon conversions. If bremsstrahlung happens early inside the tracker, and the bremsstrahlung photon further converts to an electron-positron pair, the electron and the positron can leave hits in the tracker, leading to the wrong reconstruction of the trajectory of the original electron. Charge of electrons can be measured using three different approaches at the CMS. The first approach is based on the GSF track, however, this approach can be altered by the presence of conversions, with a misidentification rate up to 10% in the region $|\eta| > 2$. The second approach is based on the associated track that is matched to a GSF track, where both tracks share at least one hit in the tracker. Finally, the ϕ angle difference between the vector joining the nominal interaction point to the supercluster position and the vector connecting the same interaction point to the

innermost hit of the GSF track. The charge is assigned when at least two out of the three measurements agree.

2.3.4 Jet reconstruction

Following the reconstruction and identification of muons, electrons and photons, the remaining particles are charged and neutral hadrons. Within the CMS, hadrons are primarily detected in the HCAL in form of hadronic showers. It's important to note that hadrons also impart some energy to the ECAL. Similar to electrons and muons, charged hadrons leave signatures as they traverse the tracker. However, neutral hadrons do not leave tracks until they reach the HCAL, where the hadronic showers commence. Collectively, these hadrons are clustered into jets.

Jet reconstruction, somewhat analogous to electron reconstruction, begins with energy clusters from both the HCAL and the ECAL. The anti-k_T algorithm [73, 74], with a distance parameter of R = 0.4, is utilised for jet clustering. This algorithm functions by identifying pairs of objects and calculating a distance parameter for each pair. The pairs with the smallest distance parameters are clustered together, and this process is iterated until no more particles can be clustered. The final clusters are recognised as jets. Moreover, due to pileup, reconstructed jets can be contaminated by particles from pileups. At the CMS, the Charged-Hadron Subtraction (CHS) technique [75] is used to remove charged-hadrons associated with pileup vertices from jets.

Even though a jet comprises various hadrons, they all originate from a quark or a gluon. The flavour of the originating hadron can be deduced based on the kinematic signatures of the jet. At the CMS, this is accomplished via an algorithm known as DeepJet [76–78]. DeepJet is a multiclass Deep Neural-Network (DNN) that uses approximately 650 input variables, including information about tracks, vertices, energy, jet direction, and so on. The algorithm produces probabilities that a jet originated from a b-quark, a c-quark, a light-flavour quark, or a gluon. DeepJet also establishes three working points: loose, medium, and tight, with the rates of falsely identifying a light jet as a b jet being 10%, 1%, and 0.1%, respectively.

Missing transverse momentum

In numerous SM and BSM processes, the final states often include neutrinos or unidentified particles that do not interact with the CMS detector material, hence, these particles do not deposit energy within the detector. Consequently, there exists an observable missing of energy, often referred to as "missing energy" in the final states. Related to this missing energy is the missing transverse momentum, denoted as $\overrightarrow{p_T}^{\text{miss}}$. It is computed as the negative sum of the transverse momentum vectors of all PF candidates. To improve the precision of the p_T^{miss} estimation, corrections made to the jet p_T are also incorporated, as demonstrated in Eq.2.2 [79].

$$\overrightarrow{p_{\mathrm{T}}}^{\mathrm{miss}} = -\sum_{i}^{N_{PF}} \overrightarrow{p_{\mathrm{T}}}_{,i} - \sum_{j}^{N_{PF,jets}} (\overrightarrow{p_{\mathrm{T}}}_{,j}^{corr} - \overrightarrow{p_{\mathrm{T}}}_{,j}) \tag{2.2}$$

where N_{PF} is the number of PF candidates, $N_{PF,jets}$ is the number of PF jets, and $\overrightarrow{p_{\mathrm{T}}}_{,j}^{corr}$ is the corrected jet p_{T}.

Chapter 3

Measurements of the $t\bar{t}W$ process at $\sqrt{s} = 13\,\text{TeV}$

In this chapter, a detail presentation is given on the analysis of the inclusive cross section measurement of $t\bar{t}W$ production at the LHC. The measurement targets the final states with two same-sign or three leptons (either e or μ). The final states with tau are included through their decays to e and μ. In the same-sign dilepton channel, the associated W boson decays into a lepton and a neutrino. The t quark, which has the same charge as the associated W boson, decays into a b-quark, a lepton and a neutrino. While the other t quark undergoes hadronic decay. In the trilepton channel, both t quarks and the associated W decays leptonically. The result presented in this chapter has been published in the Journal of High Energy Physics on 28 July 2023 [80].

This chapter begins with an introduction of the data simulation samples used. The simulation event generators as well calibrations and corrections to simulations are elaborated upon. In Sec. 3.2, criteria for the selection of muons, electrons, and jets are outlined. The section following dedicates to the modelling of background processes, including a study of the data-driven techniques used to estimate contributions from the misidentification of lepton charge and nonprompt leptons.

Section 3.4 introduces the techniques to distinguish signal from background events in the signal regions. These techniques include the multivariate analysis algorithms for signal discrimination for same-sign dilepton channel and a single variable discriminant for events in trilepton channel. Section 3.5 details the measurement of the inclusive cross section of $t\bar{t}W$ production, describing of the statistical method employed in the fit and systematic uncertainties that could modify the yield and shape of the signal and background distributions. Finally, the results of the analysis results are presented, along with comparisons to theoretical predictions and other recent measurements.

3.1 Data and simulation samples

3.1.1 Data sets

During the Run II of LHC operations, spanning from 2016 to 2018, an impressive amount of p-p collision data was delivered, reaching approximately 160 fb^{-1} at a centre-of-mass energy of 13 TeV. Within this period, the CMS detector successfully recorded nearly 138 fb^{-1} of high quality data. To provide a breakdown of the data-taking years, the integrated luminosities recorded were approximately 36.33 fb^{-1} in 2016, 41.53 fb^{-1} in 2017, and 59.74 fb^{-1} in 2018.

Recorded data are sorted in data streams based on the HLTs that each event fires. They form the starting points for subsequent data analysis. Each data stream contains events with similar properties, such as a presence of specific physical objects or pass some kinematic requirements. In this analysis, data streams with the presence of at least a muon or an electron are used throughout the entire study. Additionally, the JetMET and JetHT data streams are specifically selected for the measurements of trigger efficiency and scale factor.

Although the collision energy remained constant throughout Run II, it is important to note that the data-taking conditions experienced slight variations due to upgrades in the pixel detector. As a result, different sets of simulations were generated to correspond to each specific data-taking year.

3.1.2 Simulation samples

The analysis relies on essential simulation samples generated using advanced event generators to model the expected signal and background contributions to the $t\bar{t}W$ process. At the CMS, simulation samples are produced using the Monte-Carlo (MC) method and mostly produced centrally using the common software and computing infrastructure.

The event generation is the first step of the simulation production. It includes everything from the point where two protons collide up until the final decay products are formed, following SM predictions or other theories. The production also takes into account the Matrix Element (ME) corrections, Parton Distribution Functions (pdfs), Parton Shower (PS), hadronisations, and topography of the processes. After the generation of simulated events, the next step involves passing these events to the GEANT4 [81] software for further processing. During this stage, the intricate characteristics of the CMS detector and the interactions of particles with each of its sub-detectors are meticulously simulated, resulting in the creation of complete events.

Most signal and background samples are simulated at NLO accuracy in QCD using the MADGRAPH5_aMC@NLO [82] event generator, incorporating the FxFx merging [83] scheme to prevent double counting between ME and PS calculations. However, for certain samples, the POWHEG v2 [84–86] generator is employed instead. The PYTHIA8 [87]

event generator is used to simulate the showering, hadronisation and underlying events. For the majority of the 2016 samples, as well as all the 2017 and 2018 samples, PYTHIA8 [87] incorporates the NNPDF3.1 [88] pdf sets, and the CP5 [89, 90] tuning parameters for simulating the underlying events. Conversely, other 2016 samples employ the NNPDF3.0 [91] pdf sets along with either the CUETP8M1 [92] or CUETP8M2T4 [93] tunes to model the pdfs and underlying events.

The $t\bar{t}W$ signal samples are generated with NLO accuracy using the MADGRAPH5_aMC@NLO v2.6.0 event generator. The simulations account for several kinematic observables of the $t\bar{t}W$ process, which are sensitive to terms involving the strong coupling constant (α_S) and the fine-structure constant (α). The simulation takes into account LO and NLO terms of order α^3 and $\alpha_S\alpha^3$ as well as $\alpha_S^2\alpha$ and $\alpha_S^3\alpha$, whereas the former term is "EWK" and the term is QCD [37, 42, 94, 95]. The samples are produced separately for the QCD and EWK terms. The theoretical prediction of the cross section is calculated to be 529^{+155}_{-96} fb, this calculation includes the QCD and EWK effects at NLO accuracy and QCD corrections at NNLL accuracy [34]. The simulations are specifically generated for the phase spaces with at least two leptons in the final states, where one lepton originates from the associated W boson and the other lepton arises from the top quark decay via $t \to Wb \to \ell\nu b$.

Similarly, the MADGRAPH5_aMC@NLO programme is used for event generation at NLO accuracy in QCD for other processes. These include top quark-antiquark pairs with a vector boson ($t\bar{t}Z/\gamma^*$ and $t\bar{t}\gamma$), two or three vector bosons (WZ, WWW, WWZ, WZZ, ZZZ), two top quark-antiquark pairs ($t\bar{t}t\bar{t}$), Drell-Yan (DY) process ($Z/\gamma^* \to \ell^+\ell^-$), and a single top quark with a boson (tZq and tZ/γ). However, for the production of a single top quark in association with a Higgs boson (tHq and tHW), and a top quark-antiquark pair in association with two vector bosons ($t\bar{t}WW$, $t\bar{t}WZ$, and $t\bar{t}ZZ$), the simulations are generated at LO accuracy with the MADGRAPH5_aMC@NLO program.

The production of W or Z boson pairs (WW and ZZ), a top quark-antiquark pair ($t\bar{t}$), and a top quark-antiquark pair with a Higgs boson ($t\bar{t}H$) employ the POWHEG program instead of MADGRAPH5_aMC@NLO. These events are generated at NLO accuracy. For the QCD multijet samples used in studying the nonprompt lepton background, events are generated at LO accuracy using PYTHIA8 with either CUETP8M1 [92] tuning for 2016 or CP5 [89, 90] for the other years.

3.1.3 Event calibrations and corrections

L1 prefiring

During the data-taking years of 2016 and 2017, a gradual shift in the timing input of the ECAL L1 was observed [96]. This effect is caused by the progressive darkening of ECAL crystals which is radiation-induced, and is primarily affecting the endcap region ($|\eta| > 2$). The accumulation of this effect eventually prompts the L1 electronics to mistakenly assign a fraction of events to the preceding bunch crossing. Since the L1

rules do not allow two consecutive bunch crossings to fire the trigger, it leads to the rejection of events by the L1. At the beginning of the 2018 data-taking year, this effect was identified and corrected. Retroactively, correction factors were computed from the data and applied to simulations to correct for the shape discrepancy.

Trigger scale factors

Both data and simulated events are subject to a set of leptonic HLTs. The HLTs impose a requirement on the number of leptons in the events, either one, two, or three leptons, and the leptons are required to satisfy some specific p_{T} thresholds. The HLTs are applied to the events using the logic "OR" to maximise the selection efficiency. However, due to inherent differences between data and simulations, the trigger efficiencies in simulation may slightly deviate from those observed in data. To account for this discrepancy, a scale factor is applied to the simulation, calculated as the ratio of the efficiency measured in data to that in simulation. In dilepton channel, single-lepton and dilepton HLTs are utilised, whereas in the trilepton channel and four lepton control regions, single-lepton, dilepton and trilepton triggers are all used.

In data, the trigger efficiencies are measured using the JetMET and JetHT data streams, which apply specific requirements related to missing transverse momentum ($p_{\text{T}}^{\text{miss}}$) and the scalar sum of jets' transverse momentum (p_{T}), known as H_{T}. Using these data streams helps minimise biases introduced by lepton selection compared to multi-lepton data streams. For the measurement of trigger efficiencies, events must satisfy a set of reference $p_{\text{T}}^{\text{miss}}$ and H_{T}-related HLTs, and contain either two or three leptons passing the tight selection criteria described in Sec. 3.2. For dilepton final states, additional requirements on the invariant mass of the lepton pair ($\text{m}(\ell\ell) > 20$ GeV) and $p_{\text{T}}^{\text{miss}}$ ($p_{\text{T}}^{\text{miss}} > 40$ GeV) are applied. In the case of trilepton final states, the p_{T} of the leptons are required to be at least 25, 15, and 15 GeV, respectively. These selection criteria help ensure the inclusion of events that exhibit the desired characteristics for the $t\bar{t}W$ analysis.

Moreover, a thorough examination has been conducted to assess the correlation between reference HLTs and the leptonic HLTs use in this measurement are uncorrelated. A correlation factor, denoted as α, is computed using the formula $\alpha = \frac{N_{\text{ref}} \times N_{\text{HLT}}}{N_{\text{both}} \times N_{\text{none}}}$, where N_{ref}, N_{HLT}, N_{both}, and N_{none} represent the number of events that solely pass the reference HLTs, solely pass the main leptonic HLTs, pass both triggers, and pass neither of the triggers, respectively. To ascertain the absence of correlation between the two sets of HLTs, the correlation factor α is anticipated to equal unity if complete uncorrelation is present. Since the correlation factors have been calculated for dilepton triggers in another CMS analysis using the same set of HLTs [21], hence they are only calculated for trilepton triggers in this analysis. The correlation factors are calculated for each data-taking year, and in each case, a value of 1.0 is consistently obtained. This observation confirms the absence of correlation between the reference HLTs and the leptonic HLTs, solidifying the independence of these trigger selections in the analysis. The measurement of trigger efficiencies, the scale factors and the trigger correlation factors in the trilepton

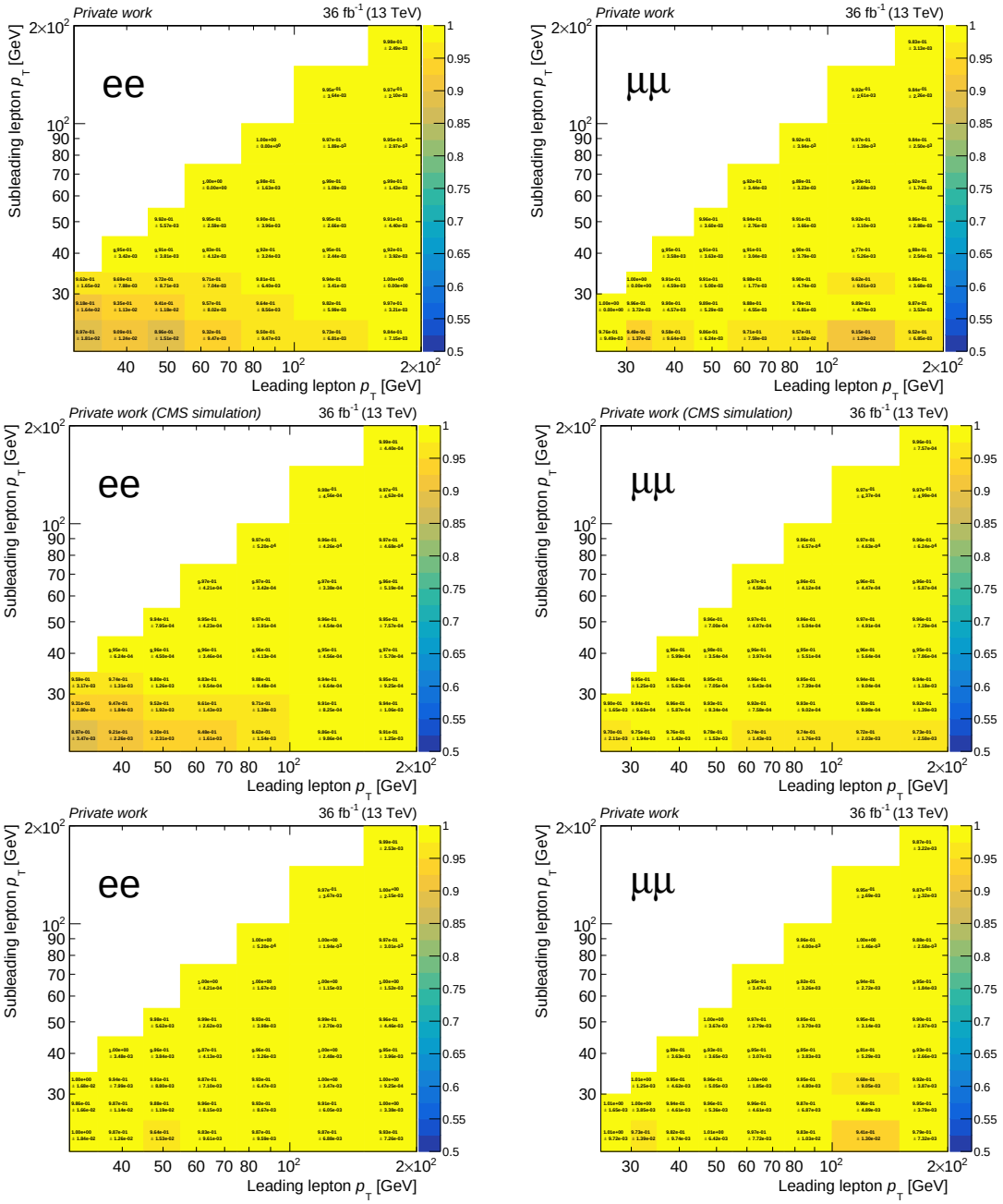

Figure 3.1: Trigger efficiencies measured in 2016 JetMET+JetHT data sets (top row) and $t\bar{t}$ simulations (middle row), categorised by di-electron (left column) and di-muon (right column) channels. The associated scale factors for these efficiencies are presented in the bottom row.

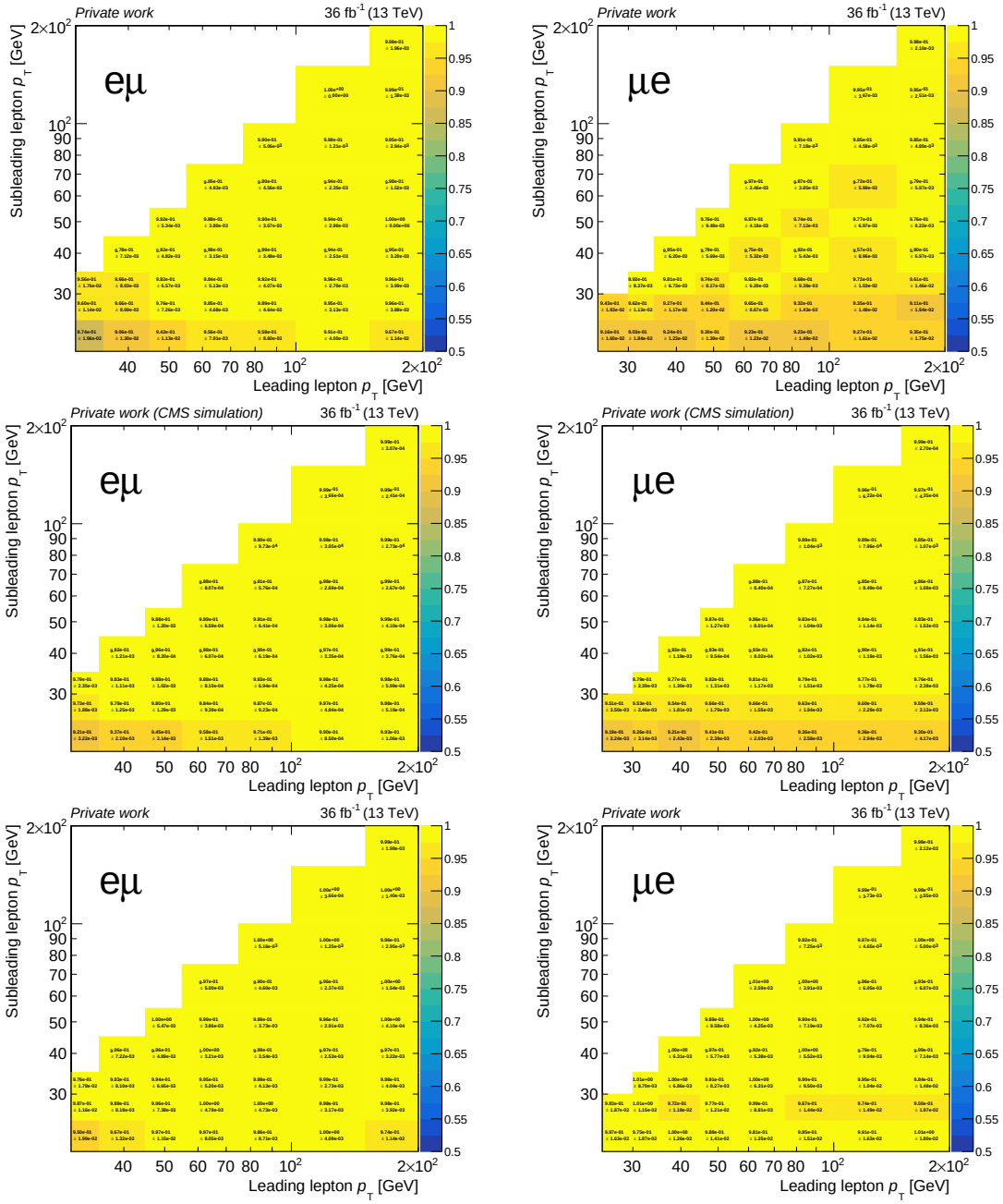

Figure 3.2: Trigger efficiencies measured in the 2016 JetMET+JetHT data sets (top row) and $t\bar{t}$ simulations (middle row), categorised by eμ (left column) and μe (right column) channels. The associated scale factors for these efficiencies are presented in the bottom row.

channels are performed by my collaborators at the University of Oviedo.

Similarly, the efficiencies of the same set of HLTs are measured in the $t\bar{t}$ simulation samples with the identical event and lepton selection criteria. However, due to the predominantly nonprompt lepton composition in the trilepton final states of the $t\bar{t}$ sample, the trigger efficiencies for trilepton events are instead measured using the $t\bar{t}W$ samples. This approach ensures that the trigger efficiencies accurately reflect the behavior observed in the signal process of interest and appropriately account for the characteristics of nonprompt leptons in the $t\bar{t}$ simulation sample.

The efficiencies and corresponding scale factors are parametrised as functions of the leading and subleading lepton p_T in the dilepton channels. In the trilepton channels, they are calculated based on the leading lepton p_T. These calculations are performed separately for each data-taking year and each dilepton final state - ee, eμ, μe and $\mu\mu$, with the electron acting as the leading lepton in eμ, and the muon taking the lead in μe. The triggers efficiencies and their corresponding scale factors measured using 2016 data sets and simulation sample are shown in Fig. 3.1 and Fig. 3.2. In certain bins with low lepton p_T, there is a noticeable drop in efficiency. This is attributed to the p_T thresholds of various triggers being set above these lower p_T values.

However, in the case of trilepton final states, limited statistics prevent the measurement of efficiencies and scale factors for each individual final state. As a result, a single set of efficiencies and corresponding scale factors is determined for all trilepton final states within each data-taking year. This approach ensures that the trigger efficiencies and scale factors are appropriately accounted for and applied consistently across the various lepton final states and data-taking years.

Pileup reweighting

At the LHC, the average number of collisions varies depending on the accelerator conditions, such as the number of proton bunches in each beam and the distance between the bunches. Producing simulations that accurately mirror the observed pileup profile in data is a complex task. This difficulty arises as simulation samples are typically generated in advance, before the actual data collection, thus, the pileup distributions in simulations have to be estimated [97, 98]. As a result, the pileup profile in simulation might not match the observed pileup in data and to address this discrepancy, pileup reweighting is employed.

In simulations, the pileup profile is represented as a histogram of the average number of interactions for each bunch crossing. A separate histogram is generated for each simulation sample, reflecting the known number of interactions in each event. On the other hand, in data, the pileup profile is derived from the instantaneous luminosity and the cross section of inelastic scattering of 69.2 mb [97, 98]. Pileup reweighting involves computing a ratio between the pileup profile observed in data and the profile obtained from simulation. This ratio is determined for each bin of the pileup distribution and is

subsequently applied to the simulated events that correspond to the specific number of interactions.

The goal of pileup reweighting is to correct for any difference between the observed pileup distribution in the data and the simulated pileup distribution. This correction ensures that the simulated events more accurately reflect the actual pileup conditions present in the recorded data. The correction is achieved by applying a weight to each event in the simulation, based on the number of collisions in that event. This weight is determined such that the overall distribution of pileup in the simulated events matches that observed in the data.

Lepton selection efficiency scale factor

The identification and selection of leptons in the CMS experiment involve applying a set of criteria based on their kinematic properties and their interactions with various sub-detectors. It is a complex task to accurately identify and select leptons from each event that come from interesting physics. This is where the concept of "lepton selection efficiency" becomes crucial.

The lepton selection efficiency quantifies how effectively the selection criteria can identify leptons from the total number of leptons produced in an event, and it hugely depends on the selection criteria. In this analysis, since a Multivariate Analysis (MVA) algorithm is used to identify the leptons of interest, e.g. leptons coming from W, either the one associated with $t\bar{t}$ production or the one coming from top decay. The lepton selection efficiency is measured based on this selection criteria. The efficiency is measured separately for both data and simulation, using the DY events where a Z boson decays into a lepton pair ($Z \to \ell\ell$), and is parametrised as a function of lepton p_T and η [70, 99]. The details on the MVA algorithm is elaborated in Ref 3.2.2.

A set of scale factors is derived for lepton selection efficiency to account for any discrepancies between data and simulation in terms of lepton selection efficiency. These scale factors are determined by comparing the efficiencies observed in data and simulation and are subsequently applied to each lepton in the simulated events based on p_T and η of the lepton.

Jet energy scale correction

Since the energy of the reconstructed jets does not represent the true energy of the partons (quarks or gluons). This is due to the convolution of various effects: the unavoidable imperfection of detector response, the impossibility to define a one-to-one correspondence between a parton and the hadrons from its fragmentation, pileup, and others. As such, reconstructed energy of jets is calibrated in both data and simulations to align with theirs true energy. The jet energy scale corrections are a series of corrections, starting with the removal of pileup contribution to the energy cluster of the jets. Then, in simulation, the jets at generator level are matched with those at reconstructed level,

hence, a correction factor is calculated from energy of the generated and reconstructed jets. As it is not possible to know the true energy of jets in data, jet energy scale correction relies on simulations, therefore, the correction factor obtained in simulation is applied. To account for nonlinear response of the detectors, the correction factors are calculated as functions of jets p_{T} and η. These corrections are cross-checked with data-driven approach by studying dijets events and Z + jets or γ + jets events. In dijet events, the jets are in opposite direction and they must have the same energy. If their energy is imbalanced, a correction is applied to correct the mismeasurement of energy in this case. On the other hand, in Z + jets and γ + jets events, the bosons can undergo decay through leptonic channels, allowing for precise measurements. The occurrence of $p_{\mathrm{T}}^{\mathrm{miss}}$ in these events is primarily attributed to the inaccurate measurement of jet recoil energy. Consequently, a correction is applied to account for this discrepancy.

3.2 Object and event selections

3.2.1 Trigger requirements

As this study focuses on final states involving two or three leptons, the selection of HLT paths is based solely on the number of leptons present in the events and their p_T. The HLT requirements mandate that each event must contain at least one, two, or three isolated muons or electrons. Additionally, lepton p_T thresholds are applied, with the lowest threshold set at 5 GeV for the leading lepton. It should be noted that the p_T thresholds may vary slightly across different data-taking years.

Since an event can potentially satisfy multiple HLT paths simultaneously, all the triggered paths are taken into account. Consequently, an event is selected if it passes any of the HLT paths, thereby optimising the selection process and maximising the number of relevant events. The single-lepton and dilepton HLTs are applied to events with exactly two leptons. Meanwhile, in the trilepton signal region as well as the trilepton and four-lepton control regions, the single-lepton, dilepton and trilepton HLTs are all applied.

3.2.2 Object selections

Once events pass the HLT trigger requirements, objects in the events are identified and selected for further analysis. Objects reconstructed by the PF algorithm are then required to pass a series of selection criteria. For leptons, three sets of selection requirements, also known as working points, are defined, namely "baseline", "loose", and "tight", respectively.

Lepton baseline selection

Leptons identified by the PF algorithm undergo a baseline selection to ensure their quality and reliability for further analysis. The algorithm provides additional information about the identified leptons, including their flavour, p_T, charge, associated tracks, and isolation values of the leptons. These variables play a crucial role in the baseline selection criteria for leptons.

The variable of **isolation** is defined as the sum of transverse momentum of candidates within a cone of size ΔR of 0.3 around the direction axis of the lepton, excluding the energy of the lepton itself. The purpose of isolation is to quantify the amount of additional energy deposited in the detector surrounding the lepton's track. The amount of energy due to pileup is subtracted from this sum of energy to mitigate the effect of pileup. In the case of a prompt lepton, the energy within the isolation cone is typically small compared to the lepton's energy.On the other hand, a nonprompt lepton, arising from the decay of a hadron, tends to have a larger amount of additional activity around it. To account for this, the concept of **relative isolation** is used, which is the ratio of the energy within the isolation cone to the energy of the lepton itself.

However, when a lepton has a high energy, its energy deposits tend to spread out over

a larger area, making the isolation calculation less precise. To address this issue, a modified version of isolation known as **relative mini-isolation** is used instead. This variable is defined similarly to relative isolation, but with a variation in the size of the cone depending on the lepton's energy. The size of the cone is computed from the following formula:

$$\Delta \mathrm{R} = \frac{10}{min(max(p_{\mathrm{T}}^{\ell}, 50), 200)} \tag{3.1}$$

As the lepton's energy increases, the cone size becomes smaller, taking into account the collimation of hadronic showers and improving the isolation measurement in high-energy scenarios. By utilising relative mini-isolation, the analyses can effectively characterise the isolation of leptons and enhance the discrimination between prompt and nonprompt leptons in a more precise manner.

The baseline selection criteria use kinematics information and PF variables of the leptons, and they are defined as follows:

- **Lepton** p_{T}: Leptons must have a transverse momentum greater than 10 GeV;
- **Lepton** $|\eta|$: The absolute value of pseudorapidity $|\eta| < 2.4$ for muons, while it must be less than 2.5 for electrons;
- **Compatibility with the PV**: The lepton's track must be consistent with originating from the PV. This is determined by the requirements that the transverse impact parameter ($|d_{xy}|$) is less than 0.05 cm and the longitudinal impact parameter ($|d_z|$) is less than 0.1 cm;
- **Relative mini-isolation**: The relative mini-isolation (I_{mini}^{rel}) is required to be less than 0.4. This measures the isolation of the lepton by summing the transverse momenta of nearby particles and normalising it by the lepton's p_{T};
- **Electron missing hits**: The tracks of electrons are allowed at most one missing hit through the inner tracker.

Moreover, to avoid double counting of leptons, electrons originating from muon decays are removed from the events. A cone of $\Delta \mathrm{R} = 0.4$ is considered around the direction of each electron, and if a "baseline" muon is found within this cone, the electron is then excluded from the analysis.

Machine learning in lepton selection

Nonprompt leptons, which are leptons not originating from the decays of W or Z bosons, contribute significantly to the background in this analysis as nonprompt lepton background. To distinguish between prompt and nonprompt leptons, a sophisticated MVA technique, which is based on the BDT algorithm, is employed. The lepton MVA is

specifically used to select "tight" leptons, and it has been employed in a number of CMS measurements [100,101] and is based on the earlier developments in Refs. [102,103].

The BDT is trained using a combination of prompt leptons from simulated $t\bar{t}W$, $t\bar{t}Z$, and tZq samples, as well as nonprompt leptons from the semileptonic decays of $t\bar{t}$ samples. Leptons that pass the aforementioned baseline selection criteria are selected for training and validation purposes.

Training of the BDT takes various kinematic variables of the leptons as input features. These include the lepton's p_T, pseudorapidity, relative isolation, relative mini-isolation, impact parameters and so on. For electrons, the electron MVA provided centrally by CMS [70] is used as an input variable, while for muons, compatibility with track segments in the muon system is considered. Additionally, properties of nearby jets within a cone of $\Delta R < 0.4$ around the lepton are also incorporated in the training.

The output of the BDT is computed individually for each lepton and ranges from -1 to 1. A lepton with a BDT output close to 1 is considered more likely to be a prompt lepton, while a value closer to -1 indicates a higher probability of being a nonprompt lepton. This lepton MVA serves as a discriminant to distinguish between prompt and nonprompt leptons, improving in the identification of leptons in the analysis.

Fig. 3.3 shows the efficiency of the lepton MVA in selecting prompt electrons (left) and muons (right). The efficiencies are calculated for lepton with $10 < p_T < 25\,\text{GeV}$ and $p_T \leq 25\,\text{GeV}$ separately. When compared with the cut-based selection used in the previous $t\bar{t}W$ measurement by CMS [19], the lepton MVA emerges as a markedly superior technique, offering notable advancements in selection efficiency. This improvement underscores the advantages of employing machine learning algorithms in lepton selection processes over traditional cut-based methods. At the chosen value of lepton MVA (see Sec. 3.2.2), the efficiency of selecting prompt leptons show a good agreement between data and simulation. In most p_T-η bins, the discrepancy is less than 10% for electrons and 5% for muons.

The lepton MVA clearly differentiates between prompt and nonprompt leptons. This distinction is shown in Fig. 3.4, which is based on simulated $t\bar{t}$ events decaying into final states with a single lepton. For an event to be selected, it must contain at least two leptons that satisfy the baseline selection criteria, and at least one of these leptons should be nonprompt.

Loose and tight lepton identifications

Leptons that satisfy the baseline selection criteria undergo further identification requirements, specifically the "loose" or "tight" identification criteria. The "tight" lepton identification criteria are designed to select leptons that are most likely originating from the decay of the W or Z bosons, effectively rejecting nonprompt leptons from decays of heavy- and light-flavour particles. On the other hand, the "loose" identification criteria

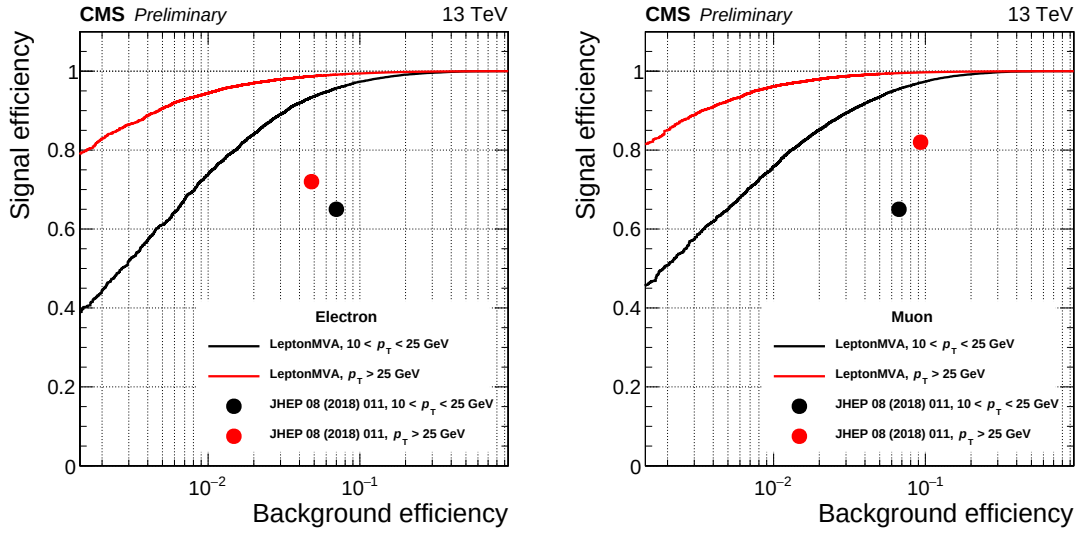

Figure 3.3: The signal efficiency and background efficiency of selection of electron (left) and muon (right) using the lepton MVA. The efficiencies of cut-based selection used in the previous $t\bar{t}W$ measurement by CMS [19] are also shown for comparison. Courtesy of Dr. Kirill Skovpen, Ghent University.

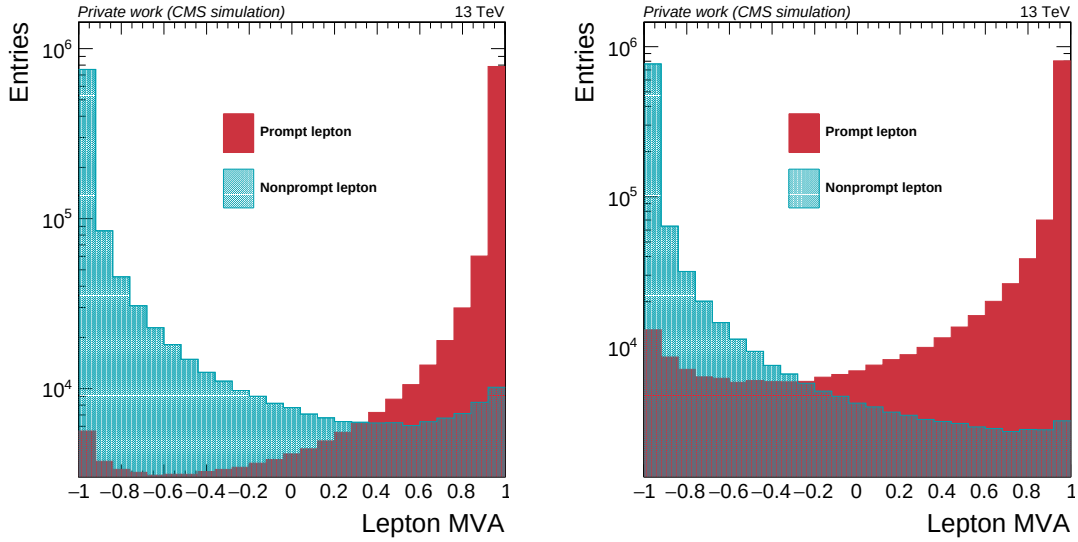

Figure 3.4: Lepton MVA distributions: electron (left) and muon (right) for both prompt and nonprompt leptons. These distributions are derived from a simulated $t\bar{t}$ event sample with single lepton final states.

are tailored for the estimation of the nonprompt lepton background using the **"Fake-rate"** method, as detailed in Sec. 3.3.2.

For electrons, additional requirements [70, 104] are applied to identify them as "tight"

or "loose" electrons:

- Electrons must not be associated with a Secondary Vertex (SV) that is consistent with a photon conversion;
- $\sigma_{i\eta i\eta} < 0.011$ for electron detected within the barrel region of ECAL or $\sigma_{i\eta i\eta} < 0.03$ for electron detected by the endcap ECAL, this variable is a lateral shower-shape variable, providing a measure of the spread of an electromagnetic shower in η direction. It is defined as the logarithm of the energy-weighted Root Mean Squared (RMS) of the electromagnetic shower in units of crystals;
- H/E < 0.1, the ratio of energy deposit in the HCAL to that in the ECAL system of the CMS detector, this variable is a powerful discriminant for distinguishing electrons/photons from hadrons;
- 1/E−1/p > -0.04, where E is energy measured in ECAL and p is track momentum determined from the curvature of the electrons at the point of closest approach to PV.

An electron is classified as "tight" if its lepton MVA score is above 0.4. If it fails the lepton MVA requirement but satisfies the following criteria, it is considered "loose":

- lepton p_{T}^{ratio} of at least 0.5, which is the ratio of p_{T} of the lepton to p_{T} of the parton;
- Loosely identified electron using MVA discriminant centrally provided by the CMS [70];
- DeepJet score (as explained in Sec. 2.3.4) of the closest jet passes an additional requirement.

In recorded data, where parton information is not available, p_{T}^{ratio} is the ratio of the lepton p_{T} to the sum of p_{T} of jets within a cone of $\Delta\text{R} = 0.4$ around the lepton.

For muons, a "tight" muon simply refers to a muon that has a lepton MVA score greater than 0.4. A muon is considered "loose" if it fails the lepton MVA requirement but satisfies the following conditions:

- The lepton p_{T}^{ratio} must be at least 0.45;
- DeepJet score of the closest jet passes an additional requirement.

The DeepJet value of the closest jet requirement is linearly interpolated between two working points as a function of lepton p_{T}, and is defined as:

$$\texttt{DeepJet cut} = \begin{cases} y_1 & p_{\text{T}} < p_{\text{T}}^1 \\ y_1 - \frac{y_1 - y_2}{p_{\text{T}}^2 - p_{\text{T}}^1} \times (p_{\text{T}} - p_{\text{T}}^1) & p_{\text{T}}^1 \leq p_{\text{T}} \leq p_{\text{T}}^2 \\ y_2 & p_{\text{T}} > p_{\text{T}}^2 \end{cases} \tag{3.2}$$

where y_1 and y_2 are DeepJet score of the closest jet to the lepton at lepton p_{T}^1 and p_{T}^2, respectively. p_{T}^1 and p_{T}^2 are two selected values of lepton p_{T} between which the

interpolated value of DeepJet is used. A lepton is selected if its closest jet has a DeepJet score lower than the interpolated DeepJet cut. This selection criterion ensures an equal probability for the lepton to originate from either light- or heavy-flavour quarks.

Lepton cone-corrected p_{T}

For leptons that fail tight lepton selection criteria, an alternative p_{T} variable known as cone-corrected p_{T} is used instead. This is because those leptons are mostly nomprompt and come from a jet or is a jet that is incorrectly identified as a lepton. The cone-corrected p_{T} represents the p_{T} of the parton which the lepton originates from. In the recorded data, this variable reflects the p_{T} of jets associated with the lepton, confined within a cone of $\Delta\mathrm{R} = 0.4$ around the lepton. The cone-corrected p_{T} is calculated using the following equation:

$$p_{\mathrm{T}}^{cone} = C\frac{p_{\mathrm{T}}}{p_{\mathrm{T}}^{ratio}} \tag{3.3}$$

Consider the distribution of lepton cone-corrected p_{T} as a function of lepton MVA. At the value of MVA cut, the distribution shows a discontinuity, which can be corrected by applying the coefficient C, the value of C is found to be 0.69 for both electron and muon in all three years. The coefficient C is calculated using semileptonic $\mathrm{t\bar{t}}$ simulated events that include a minimum of two loose leptons. This selection criterion is applied because such events have a higher likelihood of containing nonprompt leptons. This correction is illustrated in Fig. 3.5.

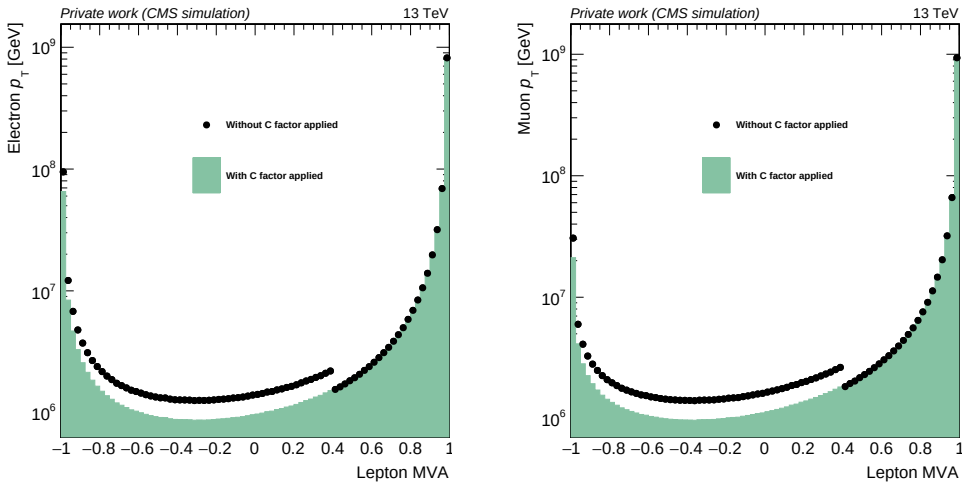

Figure 3.5: Distributions of electron (left) and muon (right) cone-corrected p_{T} as a function of lepton MVA, before and after setting the empirical C factor to remove the discontinuity at the value of the lepton MVA cut of 0.4.

This variable is also crucial in the estimation of nonprompt lepton background using the Fake-rate (FR) method. Since the FR depends on the p_{T} of mother parton instead of

the lepton, cone-corrected p_T is utilised in the measured measurement of FR as well as a default value for p_T of leptons that fail the tight selection criteria.

Jet selection and flavour tagging

Jets are selected if they have a p_T of at least 25 GeV and an absolute pseudorapidity of $|\eta| < 2.4$. Consider a cone of $\Delta R < 0.4$ around a jet, if there is at least one loose lepton found within the cone, the jet is not selected. This procedure is necessary to avoid double counting of energy and momentum.

In order to identify jets coming hadronisation of b quark, an algorithm known as DeepJet [76–78] is used. For each jet, a weight is calculated which present the chance of the jet being a b-jet, the weight ranges from 0 to 1 where one is the most likely to be b-jet. Two working points, donated as "loose" and "medium" are defined, corresponding to the chance of correctly selecting a b-jet (selection efficiencies) of 90 and 85%, respectively.

Finally, missing transverse momentum, which is a negative vector sum of all the transverse momentum of all the jets and lepton identified by the PF algorithm in an event.

3.2.3 Event selection

Events are selected based on the number of specific objects present, as well as other kinematic variables associated with them. The selection criteria are optimised to enhance the signal of $t\bar{t}W$ events and reduce contamination from background processes.

The first criterion is the number of loose leptons in the event. Events with exactly two, three, or four loose leptons are selected for further analysis. Events with additional loose or baseline leptons are rejected to ensure that orthogonal events samples are used. Events with two or three leptons appear in both signal regions and control regions, while events with four leptons are used in control regions only.

The definitions of the dilepton and trilepton signal regions are discussed in detail below, while the control regions are defined in Sec. 3.3. These selection criteria help isolate the desired $t\bar{t}W$ signal and provide a framework for distinguishing it from the background processes.

Same-sign dilepton channel

In the same-sign dilepton channel, events are required to have exactly two loose leptons with the same charge, and these leptons must also satisfy the tight lepton selection criteria. This same-sign requirement is implemented to reduce background contributions from processes such as $t\bar{t}$, DY, and WZ.

The leading lepton, whether it is an electron or a muon, is required to have a p_T greater than 30 GeV or 25 GeV, respectively. The subleading lepton must have a p_T of at least 20 GeV, regardless of its flavour. These requirements are dictated by the transverse

momentum requirement in the triggers used for this analysis, they enhance the suppression of background from low mass resonances. Additionally, the two leptons must be separated by a distance $\Delta R(\ell\ell)$ greater than 0.4, and their combined invariant mass ($m(\ell\ell)$) must be at least 30 GeV. These requirements are imposed to reduce background contributions from low-mass resonances and photon conversions.

In events with two electrons, an additional requirement is applied to further remove contributions from the DY process, specifically $Z \to e^+e^-$, where one of the electron charges is misidentified. The mass of the electron pair is required to be outside a 15 GeV window centred around the mass of the Z boson.

In addition to the criteria related to leptons, further requirements are applied to the number of jets and b-tagged jets. Events are selected if they contain at least two jets, out of which either at least two jets satisfy the loose b-tagged jet selection criteria or at least one jet passes the medium b-tagged jet selection criteria. If the missing transverse momentum, p_T^{miss}, is lower than 30 GeV, the events are rejected. These additional criteria help enhance the signal of $t\bar{t}W$ events and reduce background contamination.

In Fig. 3.6, the observed and predicted events in the same-sign dilepton signal region are compared for several observables, namely p_T of the leading and subleading leptons, the number of jets, the number of loose b-tagged jets, and p_T^{miss}. These are the most important input variables for the training of the BDT and NN algorithms, as they provide the best discrimination power between signal and background events. Overall, a reasonably good agreement between the observed and predicted distributions. At the tail of the distributions, there are some discrepancies due to imperfect modelling. However, they are reduced once the results from the fit are incorporated, resulting in an improved agreement between data and prediction.

Trilepton channel

Events in trilepton channel must consist of exactly three tight leptons with no additional loose leptons. The sum of the lepton charges should be either +1 or −1. Regardless of their flavour, the three leptons must have p_T values of at least 25 GeV, 15 GeV, and 15 GeV, respectively.

For any pair of leptons in the event, their invariant mass ($m(\ell\ell)$) must be at least 12 GeV. In the case where the pair consists of opposite-charge, same-flavour leptons, their invariant mass must not be within a 10 GeV window centred around the mass of the Z boson, i.e., $|m(\ell\ell) - m_Z| > 10\,\text{GeV}$. These requirements on the invariant mass help reduce contributions from low-mass resonances and the DY background.

Similar to the same-sign dilepton channel, events in the trilepton channel are selected if they have at least two jets, out of which at least one jet is identified as a medium b-tagged jet.

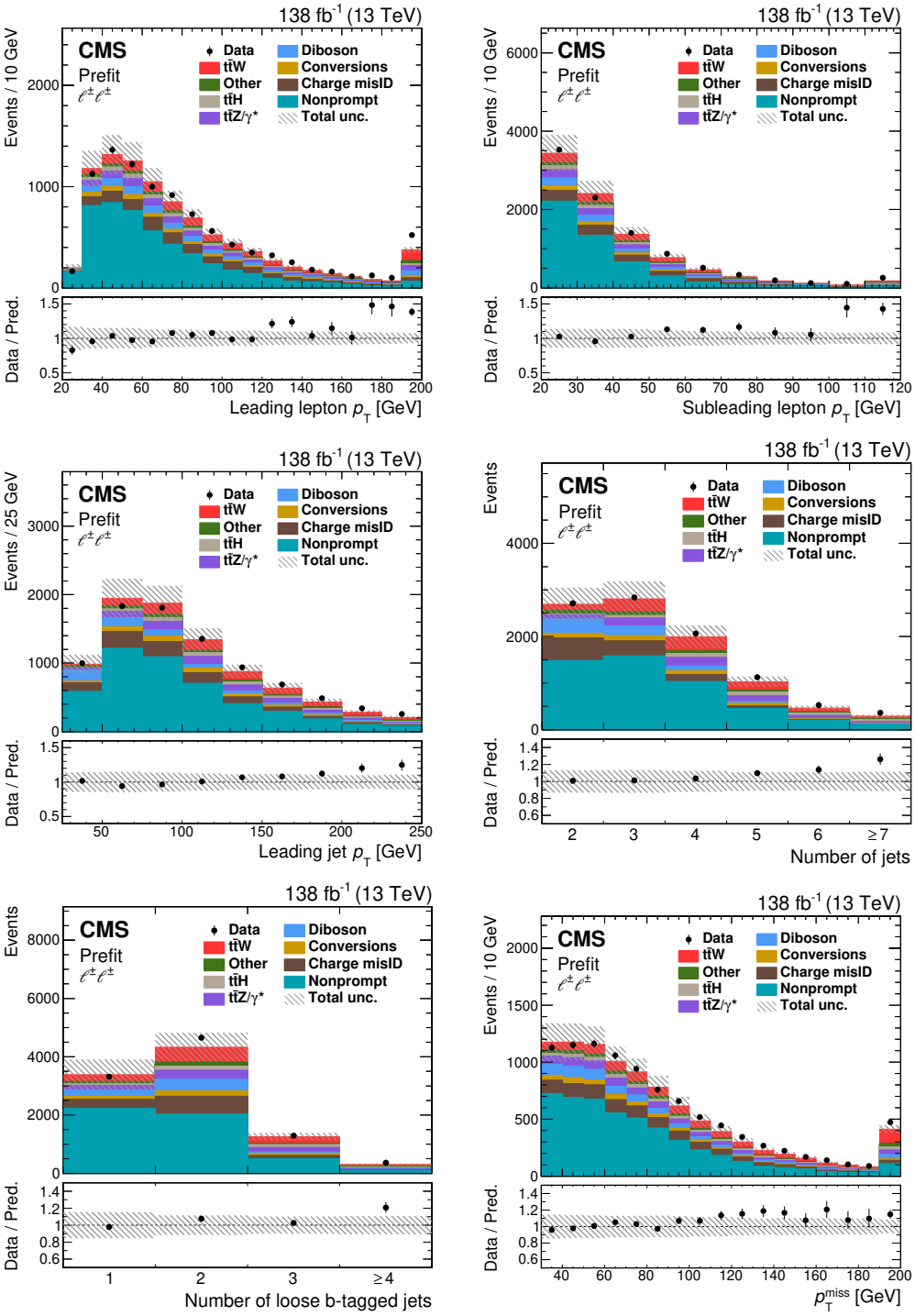

Figure 3.6: Comparison of observed (points) and predicted (coloured histograms) events in the same-sign dilepton signal region. Distributions are shown for the leading and subleading lepton p_{T}, the leading jet p_{T}, the number of jets and loose b-tagged jets, and $p_{\mathrm{T}}^{\mathrm{miss}}$. The predictions are shown before the fit to data ("Prefit"). The vertical bars represent the statistical uncertainties in the data, while the hatched bands indicate the systematic uncertainty in the predictions. The last bins include the overflow contributions. In the lower panels, the ratio of the event yields in data to the overall sum of the predictions is presented.

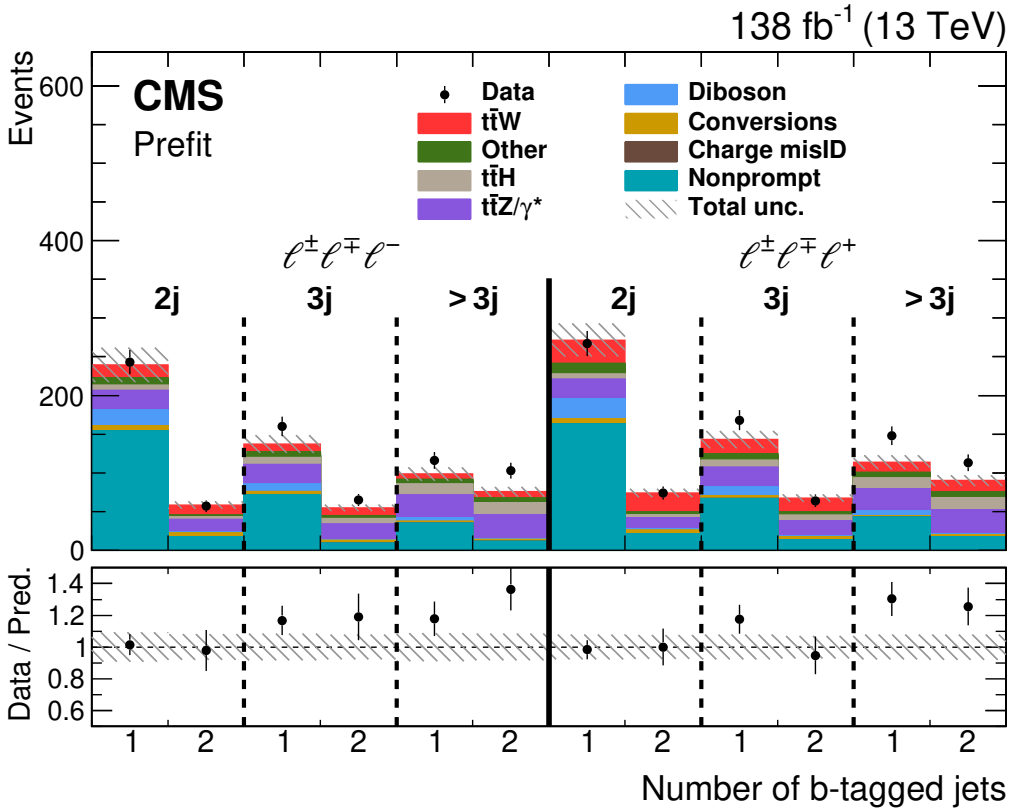

Figure 3.7: Distribution of the number of medium b-tagged jets per event from data (points) and prediction (coloured histograms) for events in the trilepton signal region with different numbers of jets (j). Events with total trilepton charge -1 and $+1$ are shown in the left and right halves of the plot, respectively. The predictions are shown before the fit to data ("Prefit"). The vertical bars on the points represent the statistical uncertainties in the data, and the hatched bands the systematic uncertainty in the predictions. The lower panel shows the ratio of the event yields in data to the overall sum of the predictions. Courtesy of Clara Ramón Álvarez and Carlos Vico Villalba, University of Oviedo.

3.3 Background modelling

The nature of this analysis is to precisely measure the inclusive cross section of $t\bar{t}W$, necessitating a comprehensive understanding of the background processes. This section provides a detailed explanation of various background processes and their estimation techniques.

The nonprompt lepton background is the most dominant and crucial background in this analysis. It mostly originates from the decay of light- and heavy-flavour hadrons. This background is relevant to both same-sign dilepton and trilepton channels, and it is estimated using a data-driven technique called the "FR" method.

In dilepton final states with at least one electron, events with opposite-sign electrons may be misidentified as a same-sign event due to the incorrect reconstruction of the charge of one electron. These events contribute to the charge misID background and are estimated from data.

All other background processes are estimated from simulations and are SM processes that decay to multilepton final states. These processes include $t\bar{t}H$, $t\bar{t}Z/\gamma^*$, diboson (WZ and ZZ), and other rare processes (e.g. $t\bar{t}$ $t\bar{t}$, $t\bar{t}VV$, tHq, tHW, tZq, and VVV).

3.3.1 Charge misidentification background

In same-sign dilepton channels, the charge misidentification (charge misID) background mainly arises from the Bremsstrahlung effect of electrons, and therefore, it is only observed in events with at least one electron. As electrons traverse the CMS detector, they emit photons through the Bremsstrahlung effect, and these photons subsequently convert into electron-positron pairs. The conversions make it more difficult to determine the curvature of the track of the original electron. As a result, the track of one of the electrons from the conversions is identified as the original electron, leading to the electron charge being misidentified if its charge is opposite to that of the converted electron.

The charge misID background primarily occurs in final states involving electrons (ee and eμ), while its contribution to final states with only muons is minimal. This is attributed to the substantial mass of muons, which results in them emitting minimal energy that is insufficient for the production of $\mu^+\mu^-$ pairs. As a result, the charge misID background is largely insignificant in final states with muons. Furthermore, the charge misID background is absent in trilepton final states

To estimate this background, a charge misID rate is first estimated using a pool of events from DY, $t\bar{t}$, and WZ simulations. The charge misID rate is the probability of a simulated lepton having its charge incorrectly reconstructed. Events are selected if they contain two reconstructed leptons that pass the tight lepton selection criteria as defined in Sec. 3.2.2. Each lepton is also required to match with a prompt simulated lepton, and its reconstructed charge is then cross-referenced its generated charge. The

charge misID rate is then parametrised as a function of lepton p_{T} and absolute value of pseudo-rapidity, $|\eta|$.

The charge misID rates for electrons and muons are shown in Fig. 3.8, revealing a notable difference in the charge misID rate between muons and electrons, with muons exhibiting significantly lower misidentification rates compared to electrons. The charge misID rate increases with the increase in $|\eta|$. At higher $|\eta|$, electrons travel a longer distance inside the CMS detector, they have a higher chance to emit Bremsstrahlung photons, hence, higher chance of their charge being wrongly reconstructed.

In order to validate the measured charge misID rate, events with two well identified electrons are selected from DY ($\mathrm{Z} \rightarrow \mathrm{ee}$) control regions from data. Furthermore, the invariant mass of the electrons must fall within a 15 GeV window centred around of the mass of Z boson, $|m(\mathrm{Z}) - m(\mathrm{ee})| < 15$ GeV. Events with two same sign electrons are classified as "Observed" while events with two opposite electrons are known as "Predicted". To obtain the yield of "Predicted" events, an event weight which is computed from the charge misID rate is applied. The charge misID weight is defined as following:

$$w = P_{\mathrm{CM}}(\mathrm{e}^+)(1 - P_{\mathrm{CM}}(\mathrm{e}^-)) + P_{\mathrm{CM}}(\mathrm{e}^-)(1 - P_{\mathrm{CM}}(\mathrm{e}^+)) \approx P_{\mathrm{CM}}(\mathrm{e}^+) + P_{\mathrm{CM}}(\mathrm{e}^-) \quad (3.4)$$

As seen in Fig. 3.8, the charge misID probability is predominantly in the range of $\mathcal{O}(10^{-3} - 10^{-4})$. Consequently, the term $P_{\mathrm{CM}}(\mathrm{e}^+) \times P_{\mathrm{CM}}(\mathrm{e}^-)$ is on the order of $\mathcal{O}(10^{-6})$ or smaller, rendering it negligible. However, in the highly forward region with $|\eta| > 2.2$, for electrons with $p_{\mathrm{T}} > 100$ GeV, the charge misID probability can be on the order of $\mathcal{O}(10^{-2})$. It is important to note that such high-energy electrons are not commonly encountered, hence the contribution of two such electrons to the charge misID background is not significant.

When comparing the number of "Observed" events with the "Predicted" events, discrepancies are observed in 2017 and 2018 data sets. The ratio between the "Observed" and "Predicted" event yields in this charge misID dominated sample are shown in Tab. 3.1. This indicates that the charge misID rate is underestimated in 2017 and 2018 simulations. The reason for this underestimation comes from the fact that the Pixel tracker of the CMS detector was upgraded after 2016 data-taking period, however, simulation samples were not yet tuned thoroughly to reflect the changes in the detector.

To account for the underestimation of the charge misID rate in the simulation, the rates presented in Fig. 3.8 are adjusted using the scale factors from Tab. 3.1.

With the corrected charge misID rate, several kinematic variables are used to further validate it, the "Predicted" and "Observed" events are compared using those variables and good closures are obtained, Figs. 3.9, 3.10, and 3.11. This further confirms that the binning used to measure the charge misID rate is optimal, and the assumption that the rate is independent for each electron is justified.

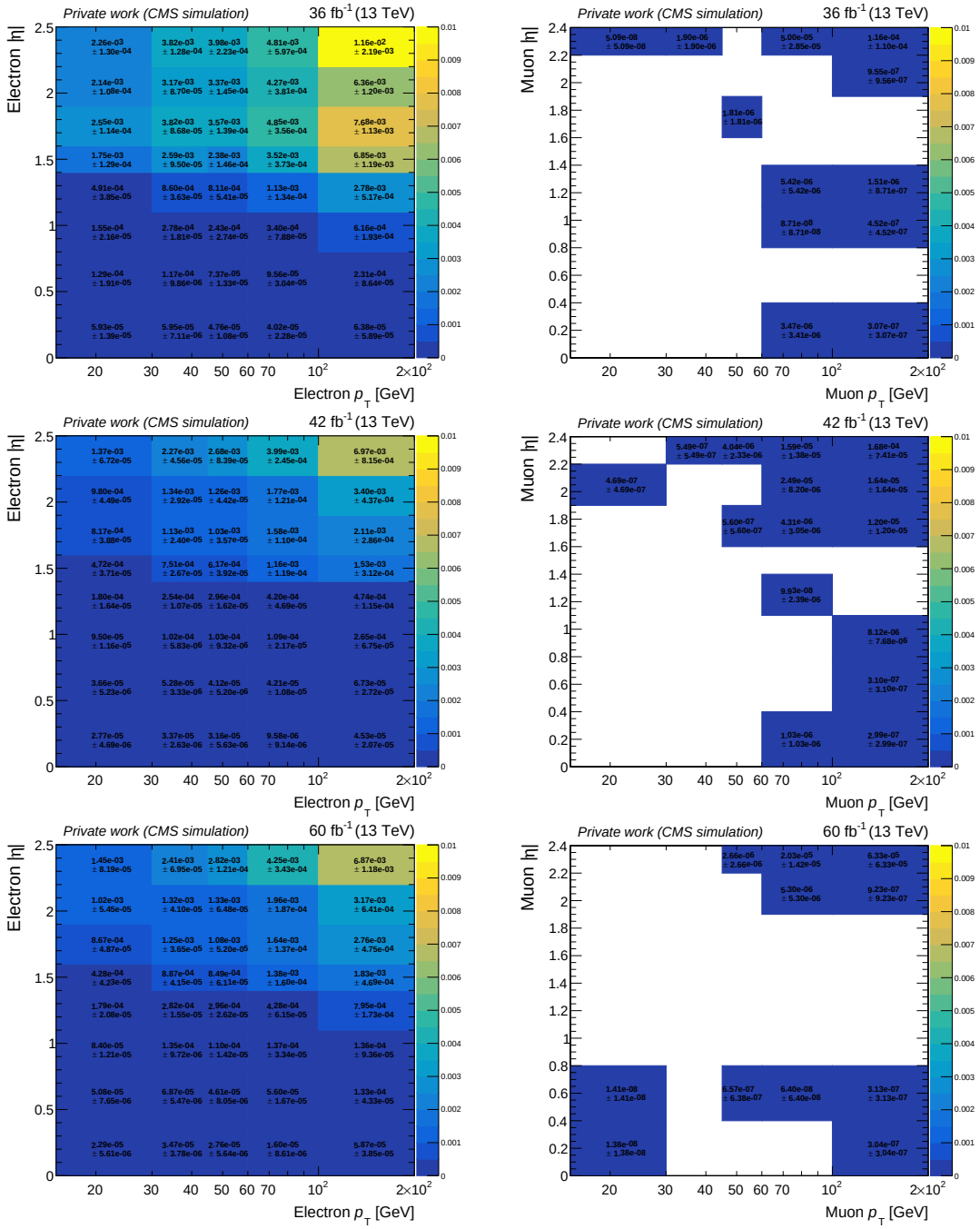

Figure 3.8: The charge misID rate of electron (left) and muon (right) measured in DY, $t\bar{t}$, and WZ simulation samples by comparing the reconstructed charge to the generated charge of the leptons. The figures are arranged in descending order: 2016 (top), 2017 (middle), and 2018 (bottom) simulations.

Year	Obs/Pred
2016	1.03 ± 0.01
2017	1.48 ± 0.01
2018	1.34 ± 0.01

Table 3.1: Ratio between observed same-sign Z $\rightarrow$ ee in data and predicted from opposite-sign in data and simulation samples using the charge misID rates. The uncertainty quoted here is statistical, and it is included in the 20% systematic uncertainty on charge misID background.

However, while the p_{T} and η distributions of both leading and subleading lepton show good agreements between the observed and predicted events, it is not the case for the di-electron invariant mass distribution. This could be due to the fact that electrons with a mis-reconstructed charge are likely to have lost more energy due to Bremsstrahlung, which results in a shift and larger radiative tail for the reconstructed Z resonance mass. Furthermore, there could be some correlation between the p_{T} of leading and subleading electrons, which was not investigated in this analysis. This correlation could be unfolded in a 2D distribution of leading and subleading p_{T}.

Based on the closure test performed as well as the fact that the charge misID rate from simulation is used together with calibrations obtained from Z $\rightarrow$ $\mathrm{e}^{+}\mathrm{e}^{-}$ events in data, a 20% systematic uncertainty is concluded for this prediction. As demonstrated in Figs. 3.9, 3.10, and 3.11 the 20% is slightly conservative, but it is in-line with other measurements in CMS.

To estimate the contribution of charge misID background to the signal region, events from data are selected with almost identical selection criteria as the same-sign dilepton signal region, except the charge of the leptons are required to be opposite. An event weight calculated from the charge misID rate using Equation 3.4 is then applied to those opposite-sign dilepton events, this procedure is similar to the closure test describe earlier in this section. Note that, an opposite-sign di-electron event in fact can contribute to both positive and negative di-electron categories, while an opposite-sign e$-\mu$ can contribute to the category with the same charge as the muon since charge misID rate of muon is negligible.

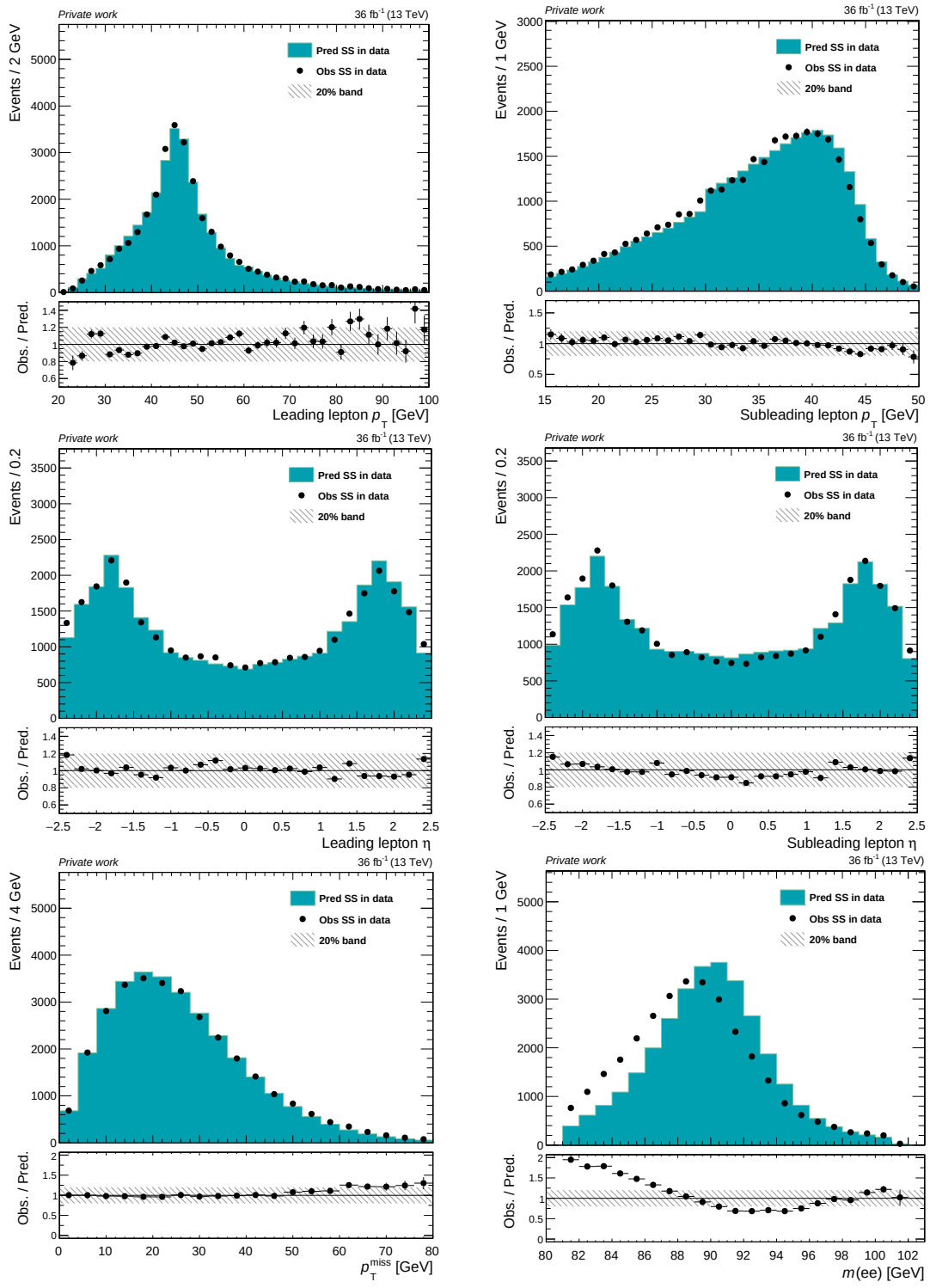

Figure 3.9: Observed events in same-sign Z $\rightarrow$ ee in 2016 data, and predicted events from opposite-sign Z in data and simulation samples, for observables: p_T and η of the leading and the subleading leptons, p_T^{miss}, and m(ee). A 20% band is added to the ratio plot to guide the eye.

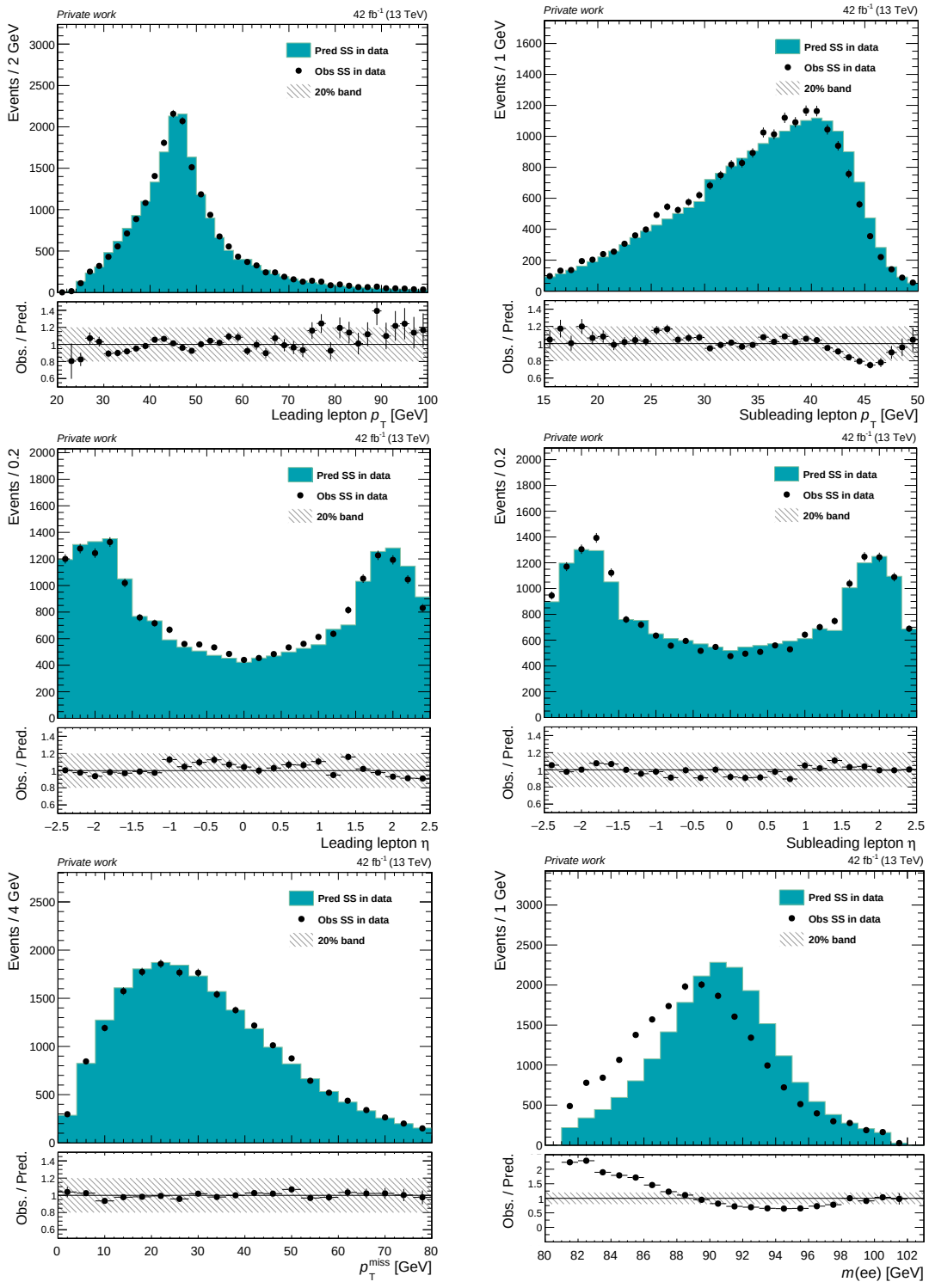

Figure 3.10: Observed events in same-sign Z → ee in 2017 data, and predicted events from opposite-sign Z in data and simulation samples, for observables: p_{T} and η of leading and subleading leptons, $p_{\mathrm{T}}^{\mathrm{miss}}$, and m(ee). A 20% band is added to the ratio plot to guide the eye.

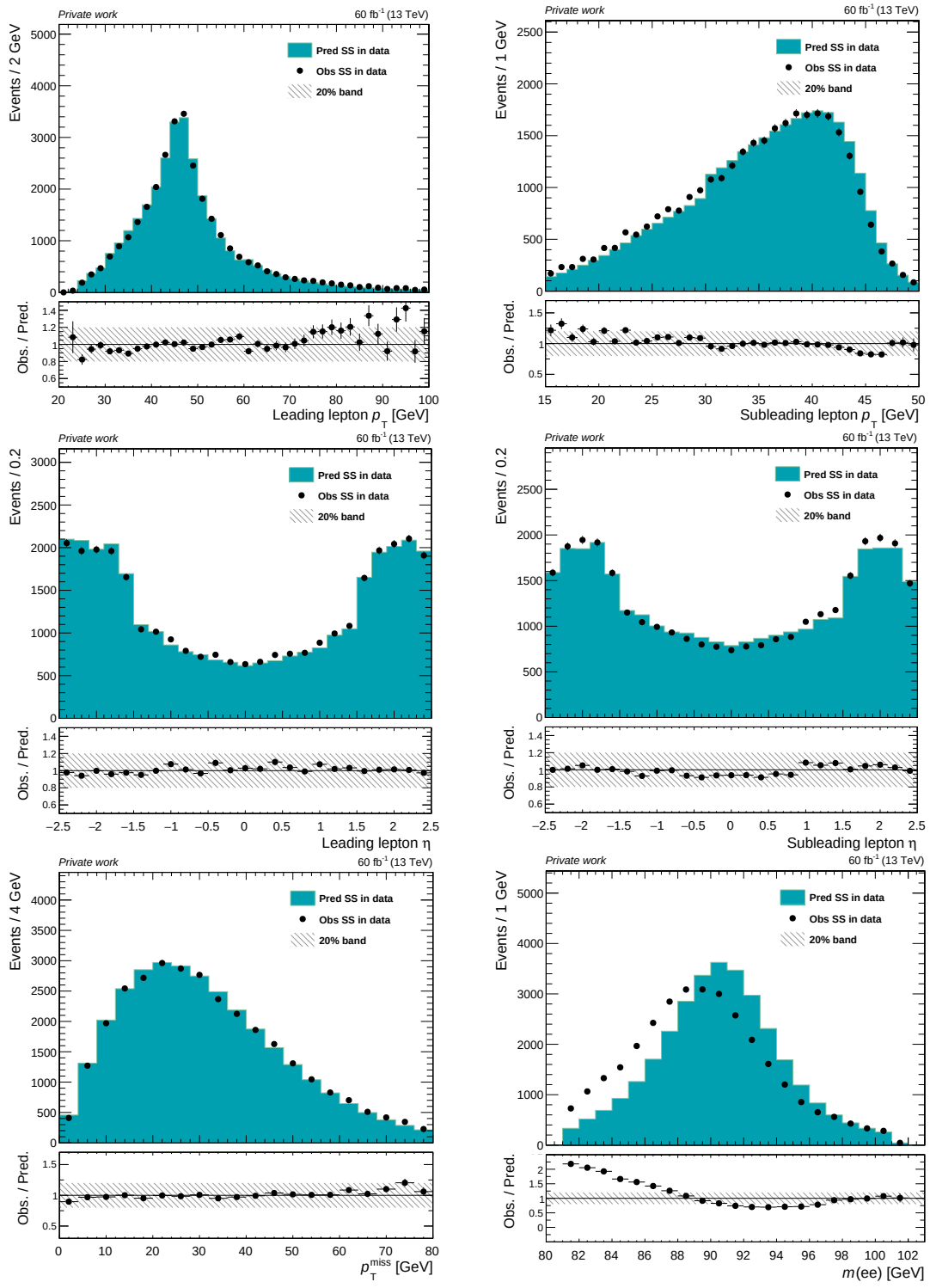

Figure 3.11: Observed events in same-sign Z $\rightarrow$ ee in 2018 data, and predicted events from opposite-sign Z in data and simulation samples, for observables: p_T and η of leading and subleading leptons, p_T^{miss}, and m(ee). A 20% band is added to the ratio plot to guide the eye.

3.3.2 Nonprompt lepton background

Nonprompt leptons refer to leptons that are not produced directly from the decays of heavy bosons. Instead, they arise from other sources, such as the decays of hadrons (e.g. $K^+ \to \ell^+\nu$, $D^+ \to \mu^+\nu_\mu$) or hadrons that are mistakenly identified as leptons. In multileptonic final states, particularly in the case of same-sign dilepton events, the contribution of the nonprompt lepton background is substantial and cannot be disregarded. This background source poses a challenge in accurately identifying and distinguishing between prompt and nonprompt leptons. It is worth noting that the decays of SM processes leading to multilepton final states, wherein only two same-sign leptons fall within the detector acceptance while the others do not, are exceedingly rare and can be neglected. Additionally, SM processes featuring two opposite-sign leptons in their final states, with the misidentification of their charge, are classified as charge misID background.

Nonprompt lepton background is estimated using a data-driven method where a Fake-rate (FR), a probability of a loosely identified lepton being wrong identified as a tight lepton, is measured. Thus, this method is commonly known as "fake-rate" method. In this method, FR is measured in a multijet-enriched region, then a weight, known as tight-to-loose ratio is computed from FR, and applied to the sideband regions which are similar to the dilepton and trilepton signal regions, except that at least one of the leptons fails the tight selection criteria. The multijet-enriched region is defined by the following criteria:

- Exactly one lepton that passes loose requirements;
- At least one recoiling jet which is separated from the lepton by $\Delta R \leq 0.7$;
- Missing transverse momentum less than 20 GeV.

The tight-to-loose ratio is computed using the following equation:

$$w = (-1)^{n+1} \prod_{n}^{i=1} \frac{f_i}{1 - f_i} \tag{3.5}$$

where n is the number of loose leptons in the event, f_i is the FR of loose lepton i. In this equation, the term $(-1)^{n+1}$ is used to ensure that contributions from events with different numbers of nonprompt leptons are properly taken into account. Considering a dilepton event, a background event with two nonprompt leptons would contribute twice to the region with one tight and one loose lepton, and thus, it leads to an overestimation in the predicted background of events with two tight leptons. This is corrected by subtracting events with two loose-not-tight leptons with the appropriate scale factor applied [105]. The diagram in Fig. 3.12 illustrates a simplified summary of the estimation of nonprompt lepton background using FR method. For the simplicity, it only shows the case of dilepton channel. Events in the sideband region (beige) pass almost the same

selection criteria (Sec.3.2.3) as those in the signal region (blue), except that at least one of the leptons fails the tight selection criteria.

The study of nonprompt lepton backgrounds employs two distinct methodologies: a simulation-based approach and a data-driven approach. In the simulation-based approach, the FR is determined using QCD multijet-enriched simulation samples. Its validity is subsequently confirmed through closure tests on simulated $t\bar{t}$ events. This process not only tests the reliability of the FR method but also aids in refining the criteria for selecting loose leptons, which is further elaborated on later in this section. Conversely, the data-driven approach measures FR directly from data in multijet-enriched regions. In this scenario, templates for both prompt and nonprompt leptons are derived from simulations, and used in a fit for the number of nonprompt leptons. The measured FR is then validated in a region designated for nonprompt validation.

Fake-rate measurement

In simulation, FR is measured in QCD multijet simulation samples, these samples are generated for the electron and muon enriched decays from heavy- and light-flavour hadrons. Events are required to pass the multijet region selection criteria as described above, and the lepton are required to be nonprompt lepton using Generator information from the simulations. The FR is measured in $p_{\text{T}} - |\eta|$ bins and separately for electron and muon. In each $p_{\text{T}} - |\eta|$ bin, FR is calculated as:

$$f = \frac{N^T}{N^L} \tag{3.6}$$

where f is FR, N^T is the number of leptons that pass the tight selection and N^L is the number of leptons that pass loose selection but fail to pass tight selection.

Closure test in simulation

The FR measured in the QCD multijet dominated samples is validated in $t\bar{t}$ simulation samples. Events with exactly two same-sign leptons that pass the loose lepton selection criteria (Sec. 3.2) are chosen. Since the events are from simulations, it is possible to require that at least one lepton is nonprompt, e.g. the selected constructed lepton is not matched with any of the prompt leptons at generator level.

Events are classified as "Observed" when both leptons meet the "tight" selection criteria. Conversely, events with at least one lepton failing this criteria are then applied a "tight-to-loose" weight, derived from the lepton's fake rate based on its p_{T} and η. These events are categorised them as "Predicted". This classification mirrors the closure test methodology used for the charge misID background.

Figs. 3.14, 3.16 and 3.16 present closure test of the FR as measured in QCD simulations, specifically applied to simulated $t\bar{t}$ events. In this closure test, several observables are considered, namely: p_{T} of both the leading and subleading leptons, the numbers of jets

and loose b-tagged jets, $p_{\mathrm{T}}^{\mathrm{miss}}$, and the event Neural-Network (NN) discriminant. The distribution of the event NN discriminant plays a crucial role in distinguishing signal events from background events in the same-sign dilepton channel, and is used to extract the cross section of $\mathrm{t\bar{t}W}$. A detailed discussion of its application and significance will be presented later in Section 3.4. In the majority of bins for these observables—particularly in the most populated ones—there's a notable agreement between observed and predicted values. This concurrence further attests to the validity of the fake rate method.

Loose lepton selection tuning

The selection criteria for loose leptons are carefully defined to ensure that almost the same number of nonprompt leptons come from both light- and heavy-flavour hadrons. Tuning of the selection is achieved by adjusting the cuts on two variables: the p_{T}^{ratio} and the DeepJet score of the closest jet to the lepton, which are closely related to the origin of the leptons. For electrons, the electron MVA score is also utilised in the selection process. The selection process involves exploring a large number of combinations of cut values from these variables. In each case, a FR is estimated using QCD simulations, and then it is validated in a closure test on $\mathrm{t\bar{t}}$ simulated events. The final working point is chosen based on a comparison of the "observed" and "predicted" nonprompt events, with a focus on achieving a good agreement and low FR values.

Data-driven FR measurement

In data, the measurement of FR is performed in the multijet-enriched region, as previously described. In this region, prompt leptons mainly come from W+jets, $\mathrm{t\bar{t}}$ and DY, hence, simulations are used as temples for these processes. For nonprompt leptons, a template is taken from QCD simulations. The events selected for this measurement are additionally required to satisfy a set of non-isolated single lepton triggers, which have low p_{T} thresholds and do not impose any isolation requirement on the leptons. However, This leads to significantly high bandwidths. As the result, the triggers are prescaled in order for the CMS data acquisition system to handle. On the other hand, these triggers are not prescaled in simulations, therefore, the prescale value of the triggers need to be identified from data and applied to simulations.

The prescale values are measured in a region which is enriched in electroweak events from W+jets simulation. The selection requirements of this region are very similar to the multijet-enriched region, except that the missing transverse momentum is required to be above 40 GeV. In this electroweak enriched region, the distribution of transverse mass, m_{T}, peaks around the mass of W boson, and the contributions from $\mathrm{t\bar{t}}$ and DY are significantly lower than from W+jets. To calculate the prescale value for each trigger, the trigger is applied to both data and simulations, and then the event yields are obtained from the m_{T} range between 80 GeV and 120 GeV from both data and simulations. The ratio of yield from simulation to that from data is the prescale value. Events from simulations are subsequently scaled using a weight, which is the multiplicative inverse of the prescale value associated with the applied trigger. Fig. 3.17 presents the $m_{\mathrm{T}}^{\mathrm{fixed}}$

distribution obtained from 2018 data and simulation before (top row) and after (bottom row) the prescale is applied to simulation samples. In this example, a non-isolated single muon trigger with a p_T threshold of 20 GeV is considered.

In the multijet-enriched region, the m_T^{fixed} observable is used to extract the number of nonprompt events, this observable is defined as:

$$m_T^{\text{fixed}} = \sqrt{2p_T^{\text{fixed}} p_T^{\text{miss}} (1 - \cos(\Delta\phi))} \tag{3.7}$$

where $\Delta\phi$ is the azimuthal angle separation between the lepton and missing transverse momentum, and p_T^{fixed} is set to 35 GeV. m_T^{fixed} is specifically chosen in this case due to the fact that in multijet region, p_T^{miss} is reconstructed due to resolution effect and is typically small.

The choice of m_T^{fixed} is due to the fact that prompt-lepton events with only one lepton mainly come from W+jets which generally have high p_T and their m_T peaks around the mass of W boson, 80 GeV. On the other hand, nonprompt events tend to have lower p_T^{miss} and m_T due to the lack of high p_T neutrinos. As shown in Equation 3.7, m_T is correlated with the lepton p_T, hence the lepton FR. By setting lepton p_T to a fixed value, this correlation can be avoided, this lead to m_T depending on the relative direction of lepton p_T and p_T^{miss} only.

The estimation of nonprompt event numbers is derived directly from the data, achieved through the subtraction of prompt leptons. For this subtraction, the simulations of W+jets, $t\bar{t}$, and DY serve as templates for prompt leptons, while the template for nonprompt leptons is taken from QCD simulations. This analysis employs various methods for prompt lepton subtraction, each of which is detailed comprehensively later in this section. After subtracting prompt lepton from data, the number of nonprompt leptons is obtained in each p_T-$|\eta|$ bin, for each of the cases where the leptons pass or fail the tight lepton criteria. Similar to the simulation approach, the FR is calculated using for each data-taking year and lepton flavour, using Eq. 3.6.

Prompt lepton subtraction

Three approaches have been investigated for the subtraction of the yield of prompt lepton events from data, thereby determining the number of nonprompt events. Each approach yields a specific set of FRs. The values from the three methods are combined to obtain the nominal FR and its uncertainty. The average of the maximum and the minimum value in each bin is taken as the nominal FR, while the two extremes are considered as an envelope uncertainty. Fig. 3.19 shows FRs measured in data as a result of the three prompt lepton subtraction methods discussed in this section.

The first method involves conducting an independent binned profile likelihood fit for the m_T^{fixed} distribution within each lepton p_T-$|\eta|$ bin. This approach is based on the assumption that there is no correlation among the bins. In each bin, the fit returns a signal strength, which show how much the yield of nonprompt lepton taken from

simulation need to scale to match the data. In the second method, events from all lepton p_{T}-$|\eta|$ bins are correlated, therefore, a simultaneous fit is performed for all bins, as a result, only one signal strength for the nonprompt lepton yield. Each p_{T}-$|\eta|$ is scaled by the same signal strength.

The fits in the two methods described above are performed independently for each lepton flavour, data-taking year, and lepton identification (e.g, "loose but fails tight" or "tight" lepton) category. In the fits, statistical uncertainty of the simulations are taken as a source of systematic uncertainty. An example of $m_{\mathrm{T}}^{\mathrm{fixed}}$ distribution of muon using 2017 data and simulation is shown in Fig. 3.18. The muons are in a specific p_{T}-$|\eta|$ bin with p_{T} between 20 and 25 GeV, and $|\eta|$ less than 1.2.

In the third method, the FR is measured using the two $p_{\mathrm{T}}^{\mathrm{miss}}$ regions: "small" with $p_{\mathrm{T}}^{\mathrm{miss}} < 20\,\mathrm{GeV}$ and "large" with $45 < p_{\mathrm{T}}^{\mathrm{miss}} < 80$. Since FR does not depend on $p_{\mathrm{T}}^{\mathrm{miss}}$ but on the relative fraction of prompt and nonprompt leptons. This method relies on data and simulation to estimate the relative yields of prompt events in low and high $p_{\mathrm{T}}^{\mathrm{miss}}$ regions. It then extracts the FR for nonprompt leptons base on this fraction.

Let f_{N} and f_{P} be the FRs for nonprompt and prompt leptons, and $f^{\mathrm{S}} = N^{\mathrm{S}}_{\mathrm{tight}}/N^{\mathrm{S}}_{\mathrm{loose}}$, $f^{\mathrm{L}} = N^{\mathrm{L}}_{\mathrm{tight}}/N^{\mathrm{L}}_{\mathrm{loose}}$, the overall FRs in the small and large $p_{\mathrm{T}}^{\mathrm{miss}}$ regions. The overall FRs are defined as:

$$\begin{aligned} f^{\mathrm{S,L}} &= \frac{N^{\mathrm{S,L}}_{\mathrm{tight}}}{N^{\mathrm{S,L}}_{\mathrm{loose}}} = \frac{N^{\mathrm{S,L}}_{\mathrm{N,tight}} + N^{\mathrm{S,L}}_{\mathrm{P,tight}}}{N^{\mathrm{S,L}}_{\mathrm{loose}} + N^{\mathrm{S,L}}_{\mathrm{P,loose}}} = f_{\mathrm{N}} \cdot \frac{N^{\mathrm{S,L}}_{\mathrm{N,loose}}}{N^{\mathrm{S,L}}_{\mathrm{loose}}} + f_{\mathrm{P}} \cdot \frac{N^{\mathrm{S,L}}_{\mathrm{P,loose}}}{N^{\mathrm{S,L}}_{\mathrm{loose}}} \\ &= f_{\mathrm{N}} \cdot \left(1 - \frac{N^{\mathrm{S,L}}_{\mathrm{P,loose}}}{N^{\mathrm{S,L}}_{\mathrm{loose}}}\right) + f_{\mathrm{P}} \cdot \frac{N^{\mathrm{S,L}}_{\mathrm{P,loose}}}{N^{\mathrm{S,L}}_{\mathrm{loose}}} = f_{\mathrm{N}} \cdot \left(1 - x^{\mathrm{S,L}}_{\mathrm{P}}\right) + f_{\mathrm{P}} \cdot x^{\mathrm{S,L}}_{\mathrm{P}} \end{aligned} \tag{3.8}$$

where $x^{\mathrm{S,L}}_{\mathrm{P}} := N^{\mathrm{S,L}}_{\mathrm{P,loose}}/N^{\mathrm{S,L}}_{\mathrm{loose}}$ is the fraction of prompt leptons in the two regions, these are loose leptons as defined in Sec. 3.2.

The ratio of the contaminations in the two $p_{\mathrm{T}}^{\mathrm{miss}}$ regions can be factorised and written as:

$$r^{\mathrm{SL}}_{\mathrm{P}} := \frac{x^{\mathrm{S}}_{\mathrm{P}}}{x^{\mathrm{L}}_{\mathrm{P}}} = \frac{N^{\mathrm{S}}_{\mathrm{P,loose}}/N^{\mathrm{S}}_{\mathrm{loose}}}{N^{\mathrm{L}}_{\mathrm{P,loose}}/N^{\mathrm{L}}_{\mathrm{loose}}} = \left(\frac{N^{\mathrm{S}}_{\mathrm{P,loose}}}{N^{\mathrm{L}}_{\mathrm{P,loose}}}\right)\left(\frac{N^{\mathrm{L}}_{\mathrm{loose}}}{N^{\mathrm{S}}_{\mathrm{loose}}}\right) \tag{3.9}$$

where the first term is estimated from simulated events, while the second is a ratio of overall observed event yields in data. This leads to a set of two linear equations:

$$f^{\mathrm{L}} = f_{\mathrm{N}} \cdot \left(1 - x^{\mathrm{L}}_{\mathrm{P}}\right) + f_{\mathrm{P}} \cdot x^{\mathrm{L}}_{\mathrm{P}} = f_{\mathrm{N}} + (f_{\mathrm{P}} - f_{\mathrm{N}}) \cdot x^{\mathrm{L}}_{\mathrm{P}} \tag{3.10}$$

$$f^{\mathrm{S}} = f_{\mathrm{N}} + (f_{\mathrm{P}} - f_{\mathrm{N}}) \cdot x^{\mathrm{S}}_{\mathrm{P}} = f_{\mathrm{N}} + (f_{\mathrm{P}} - f_{\mathrm{N}}) \cdot r^{\mathrm{SL}}_{\mathrm{P}}\, x^{\mathrm{L}}_{\mathrm{P}} \tag{3.11}$$

with two unknowns f_{N} and $(f_{\mathrm{P}} - f_{\mathrm{N}}) \cdot x^{\mathrm{S}}_{\mathrm{P}}$, that can be solved directly:

$$f_{\mathrm{N}} = \frac{f_{\mathrm{S}} - r^{\mathrm{SL}}_{\mathrm{P}} f_{\mathrm{L}}}{1 - r^{\mathrm{SL}}_{\mathrm{P}}} \tag{3.12}$$

This approach is viable as long as r_{P}^{SL} is significantly smaller than unity. As shows in Equation 3.12, the FR of nonprompt leptons depends on the overall FRs and number of prompt leptons in both small and large $p_{\text{T}}^{\text{miss}}$ regions.

Nonprompt lepton validation region

A validation region in data is designed to validate the data-driven FR. This region has similar selection as the same-sign dilepton signal region, except the $p_{\text{T}}^{\text{miss}}$ cut is inverted. By inverted the $p_{\text{T}}^{\text{miss}}$ cut, most of the leptonic decays W is supressed, thus it was significant reduces the number of $t\bar{t}W$ signal. Since nonprompt leptons are mostly the products of light- and heavy-flavour hadron decays, their yield is not affected as much in this low $p_{\text{T}}^{\text{miss}}$ region.

As demonstrated in Fig. 3.20, the data-driven technique used for estimating the nonprompt lepton background demonstrates a good closure with observed events across various observables in the validated region. In the majority of bins for these observables, the predicted and observed events are in agreement within the associated uncertainties. However, in a few bins, a noticeable discrepancy is observed, which is likely attributed to the mis-modelling of simulations.

3.3.3 Other background contributions

Background arising from the conversions of photons is known as "Conversions". It appears when an off-shell photon converts into an e^+e^- pair where of the electrons carries most of the energy and pass the lepton selection criteria while the other does not. This background is mostly reduced by the electron selection criteria (Sec.3.2.2) and by applying the $m(\ell\ell)$ and $\Delta R(\ell\ell)$ cuts (Sec.3.2.3). However, there is still non-negligible number of conversions event enter the signal regions. Conversions are estimated using $t\bar{t}\gamma$ simulation, where at least of the electron is required to match to a photon at generator level.

Other backgrounds come from SM processes with at least one top quark, namely $t\bar{t}H$, $t\bar{t}Z/\gamma^*$, tZq, tHq, tHW, $t\bar{t}t\bar{t}$, $t\bar{t}VV$, and as well as processes with two or three vector bosons V, where V is either W, Z or γ. These backgrounds are estimated using simulation, with the number of events being scaled corresponding to their cross section obtain from previous measurements or SM prediction. The study of the control regions outlined in this section was undertaken by Clara Ramón Álvarez and Carlos Vico Villalba, University of Oviedo.

Among those processes $t\bar{t}H$, $t\bar{t}Z/\gamma^*$, and WZ are the most dominant. With $t\bar{t}H$ and $t\bar{t}Z/\gamma^*$ have very similar topology to $t\bar{t}W$, especially in events with large number of jets, b-tagged jets. On the other hand, WZ contribute more to regions with lower number of jets, b-tagged jets. These backgrounds are grouped into 5 groups, namely Conversions, $t\bar{t}H$, $t\bar{t}Z/\gamma^*$, diboson, which includes WZ and ZZ, and "Other" (e.g. tZq, tHq, tHW, $t\bar{t}t\bar{t}$, $t\bar{t}VV$, and VVV).

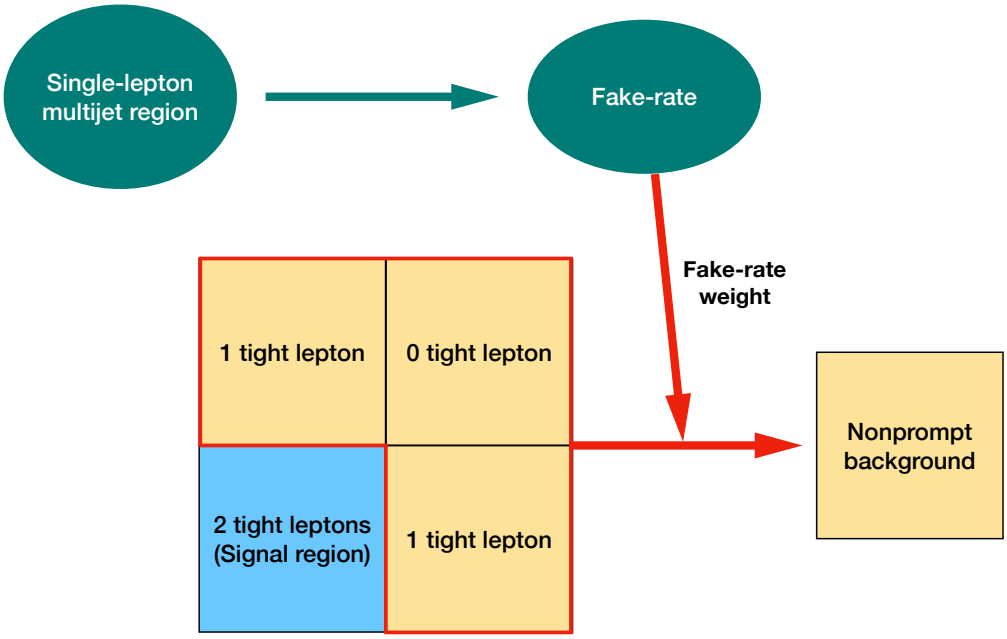

Figure 3.12: This diagram illustrates the estimation of nonprompt lepton background using FR method. The FR is measured in a single lepton multijet region, and applied to a sideband region where at least one lepton fails the tight selection criteria.

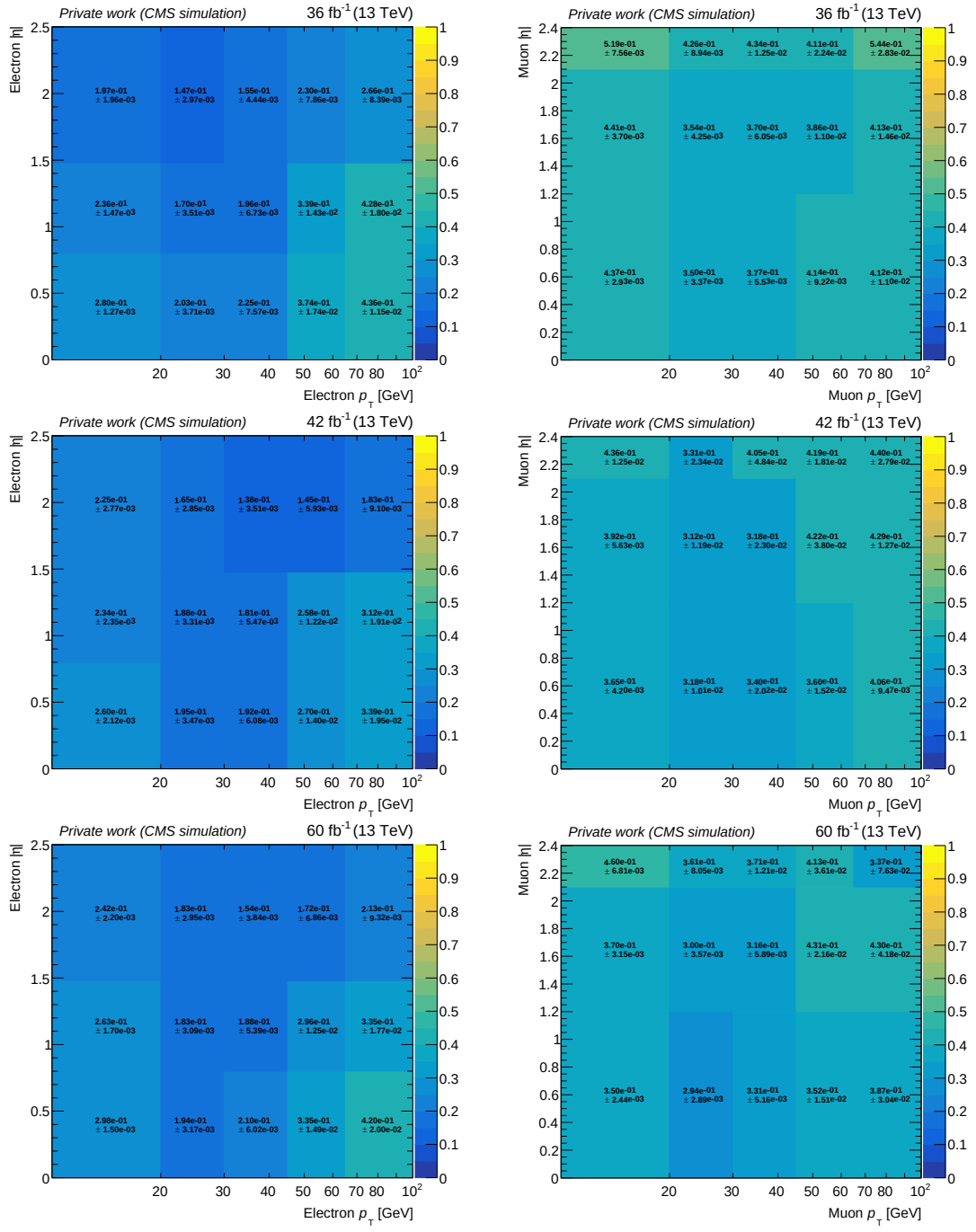

Figure 3.13: Electron (left) and muon (right) FRs measured in 2016, 2017 and 2018 QCD simulation samples.

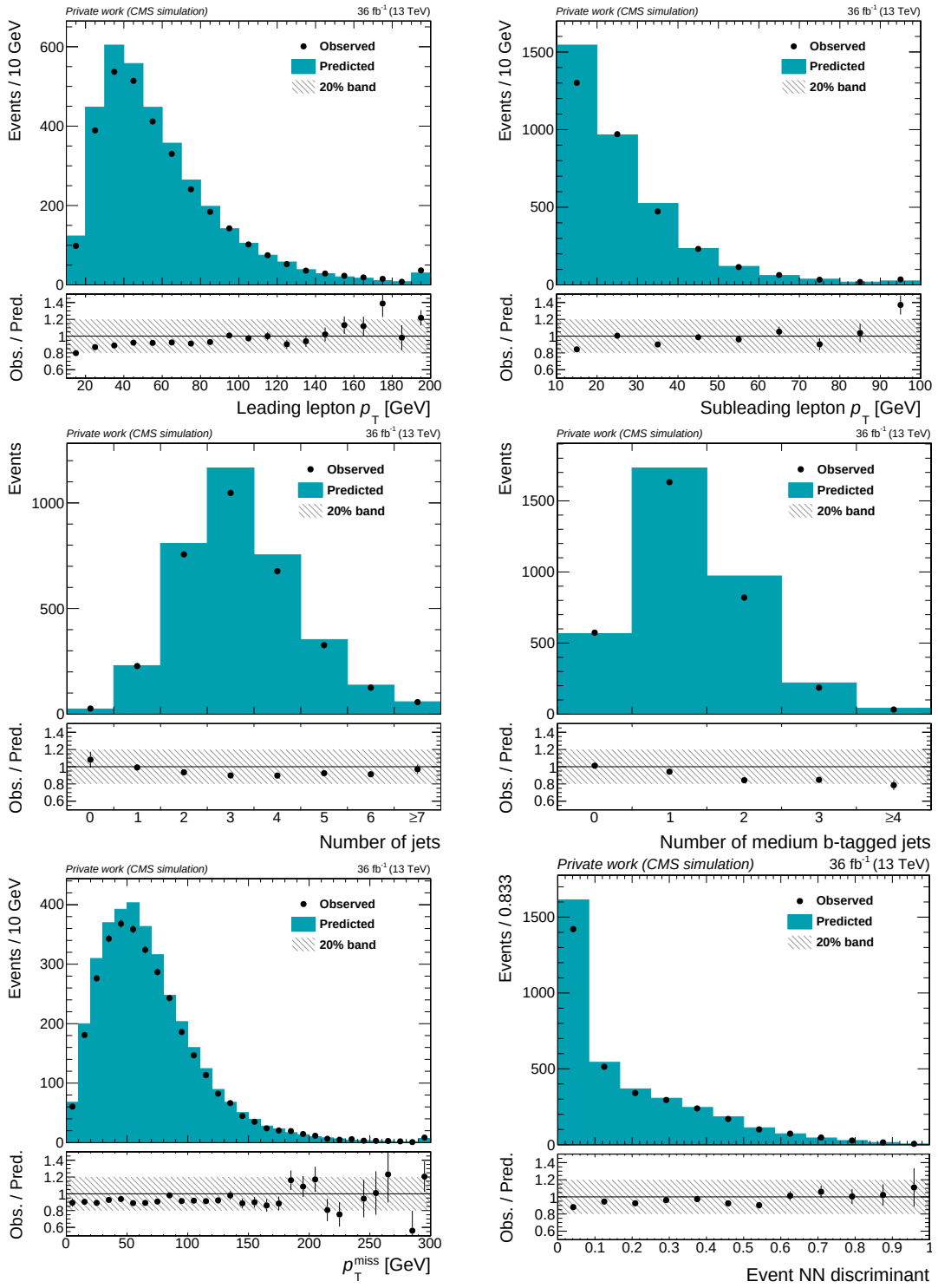

Figure 3.14: Comparison of observed and predicted nonprompt events using $t\bar{t}$ 2016 simulation samples, for p_T of the leading and the subleading leptons, numbers of jets and loose b-tagged jets, p_T^{miss}, and the event NN discriminant. A 20% band is added to the ratio plot to guide the eye.

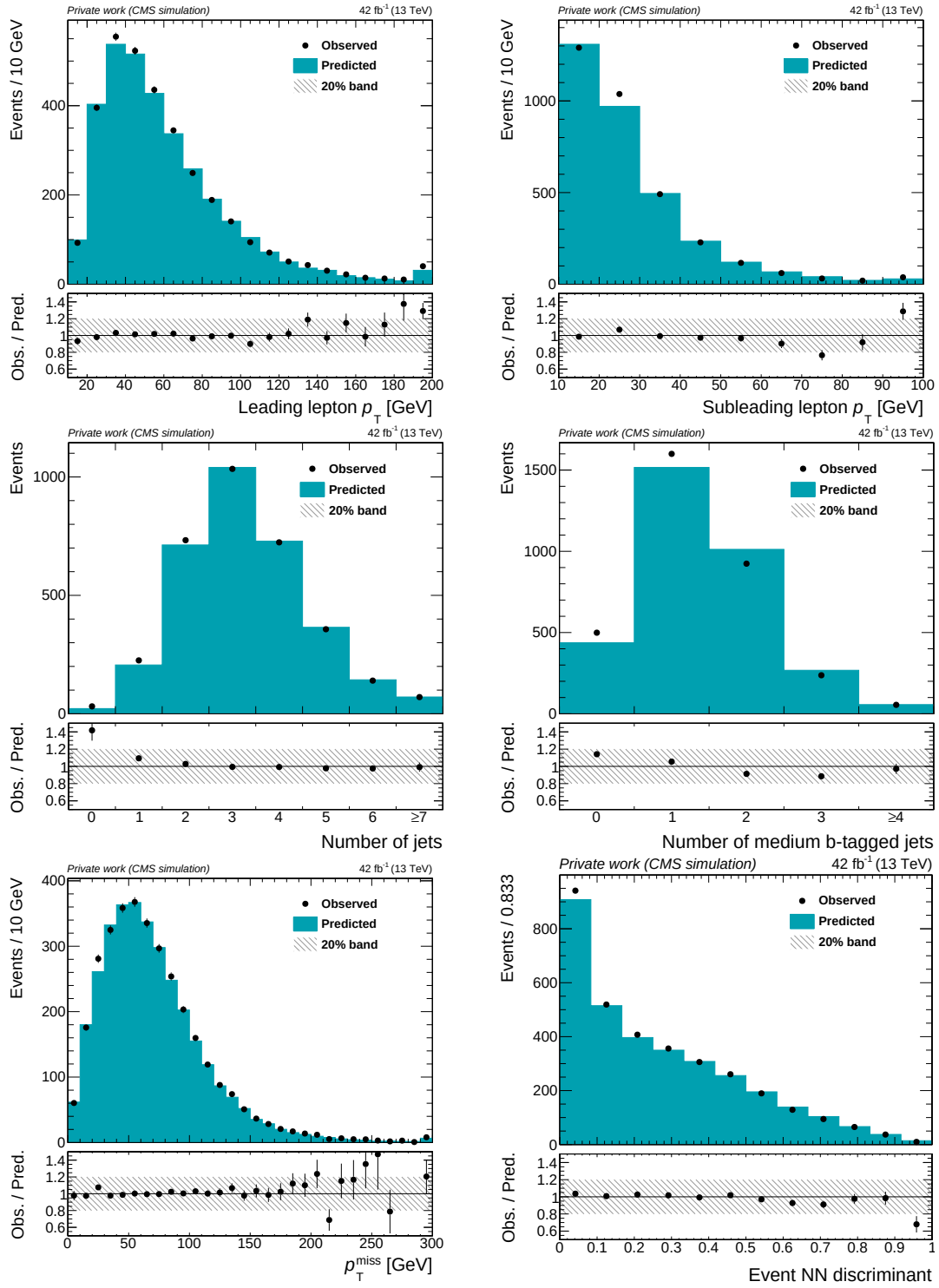

Figure 3.15: Comparison of observed and predicted nonprompt events using $t\bar{t}$ 2018 simulation samples, for p_T of the leading and the subleading leptons, numbers of jets and loose b-tagged jets, p_T^{miss}, and the event NN discriminant. A 20% band is added to the ratio plot to guide the eye.

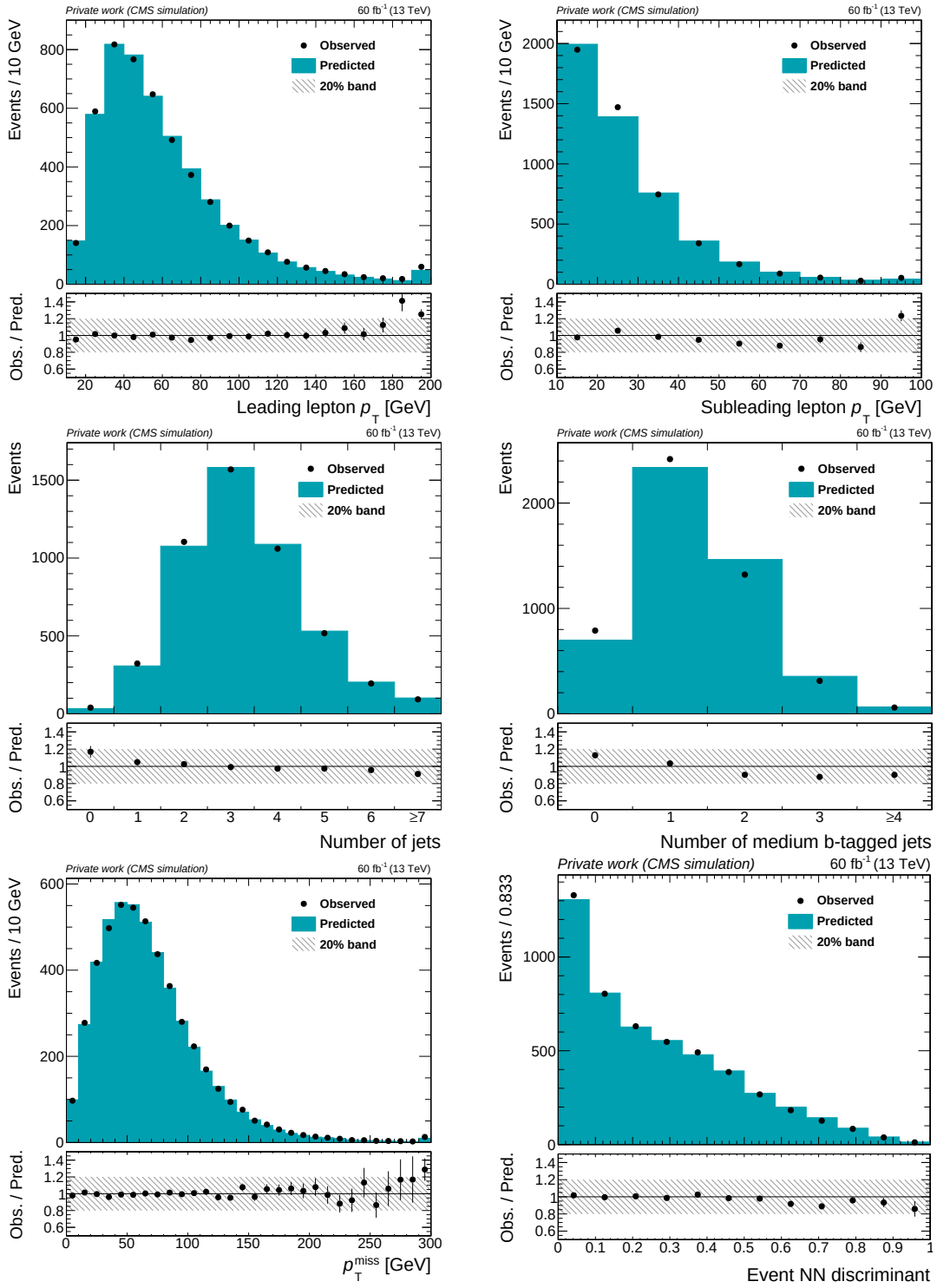

Figure 3.16: Comparison of observed and predicted nonprompt events using $t\bar{t}$ 2018 simulation samples, for p_T of the leading and the subleading leptons, numbers of jets and loose b-tagged jets, p_T^{miss}, and the event NN discriminant. A 20% band is added to the ratio plot to guide the eye.

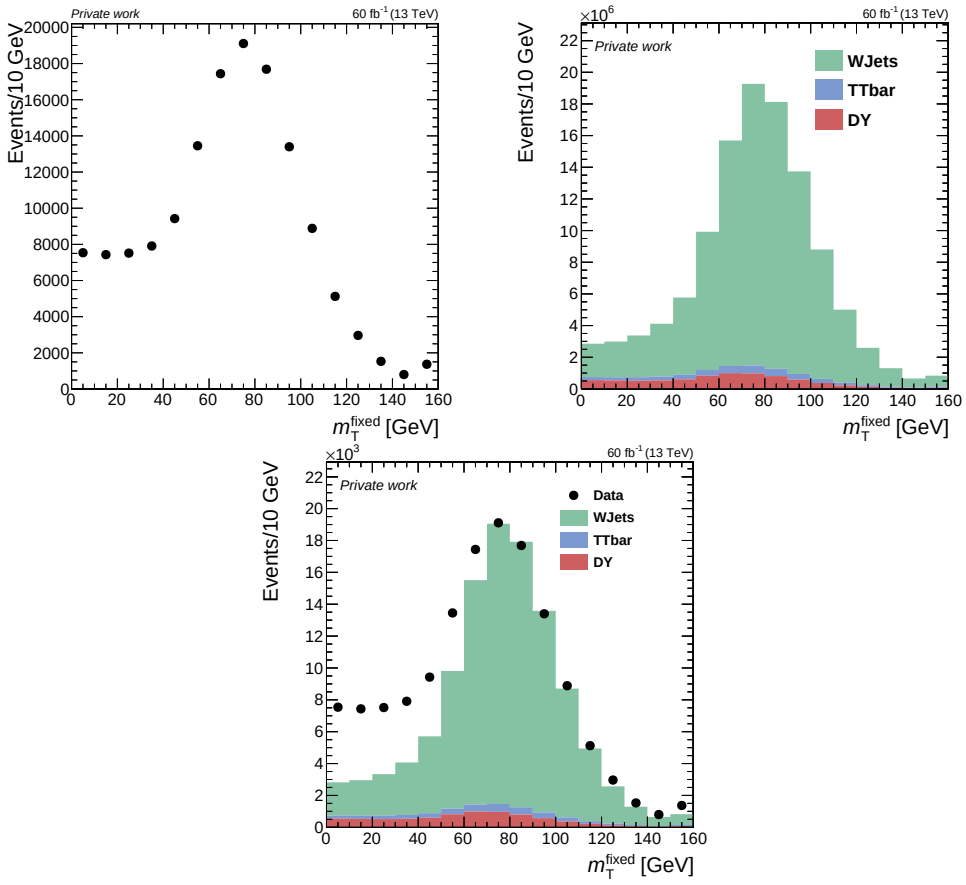

Figure 3.17: The $m_{\mathrm{T}}^{\mathrm{fixed}}$ distributions that are used in the measurement of the prescale of a non-isolated single muon trigger with a p_{T} threshold of 20 GeV. Owing to prescaling, the event yield from data (top left) is significantly lower than that obtained from simulations (top right). Consequently, the simulations are scaled down for alignment with the data, as illustrated in the bottom plot.

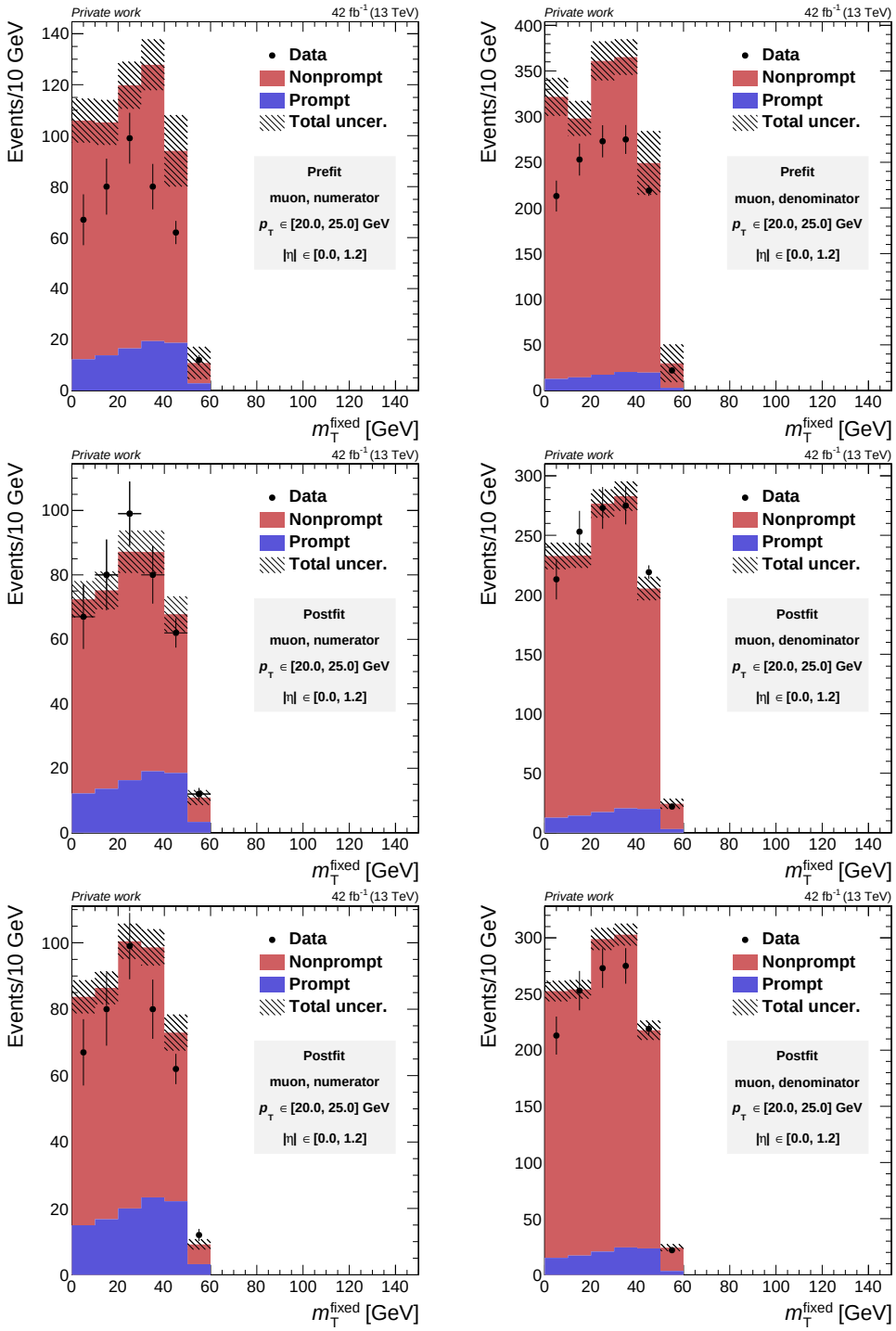

Figure 3.18: The $m_{\mathrm{T}}^{\mathrm{fixed}}$ distributions of muon in a p_{T}-$|\eta|$ bin before (top row) and after the fit, with the fit results from independent fit in each bin (middle row) and simultaneous (bottom row). On the left are distributions of muons that pass the tight selection criteria, while on the right, muons pass the loose but fail the tight selection criteria.

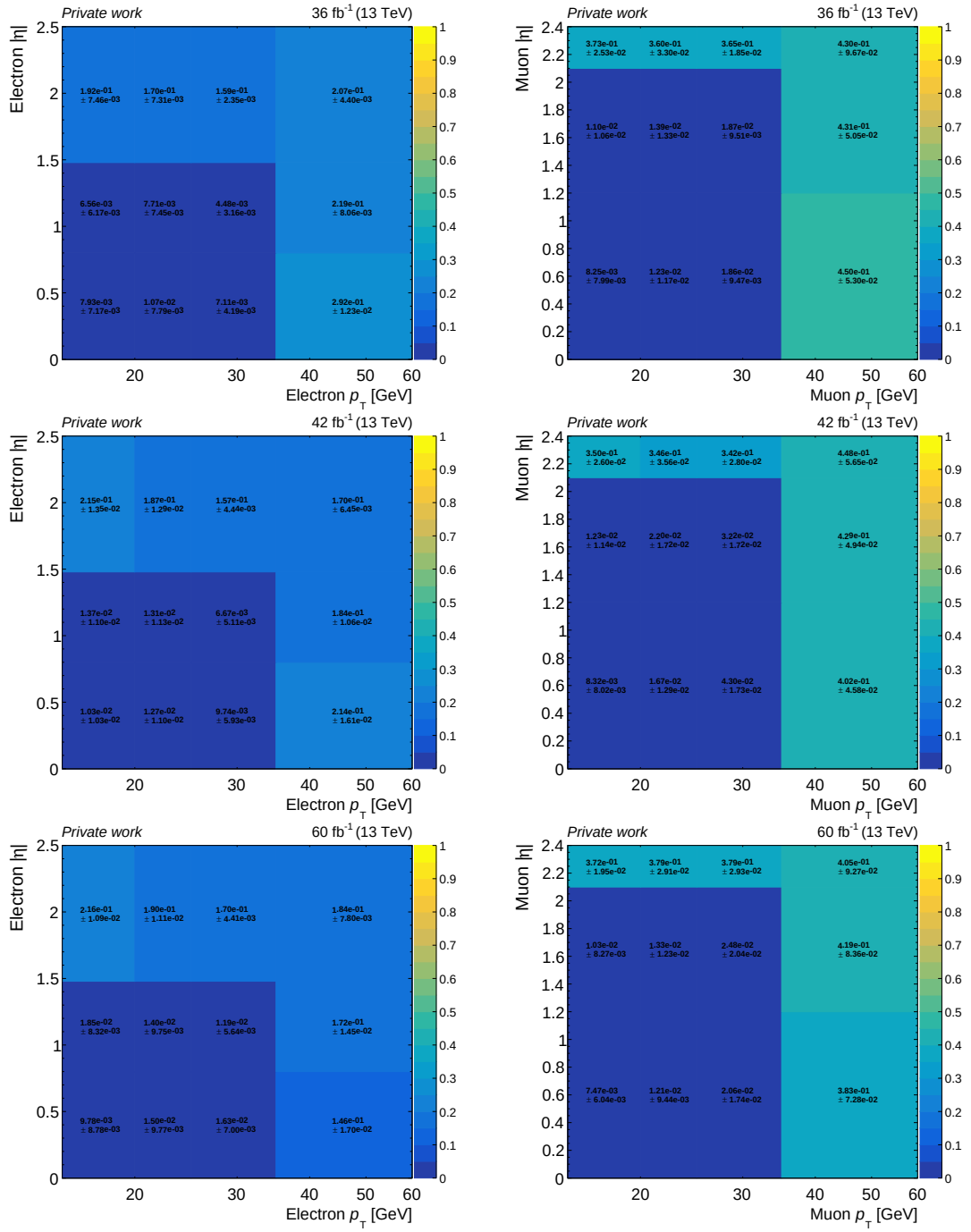

Figure 3.19: Electron (left) and muon (right) FRs measured in 2016, 2017 and 2018 using data-driven technique.

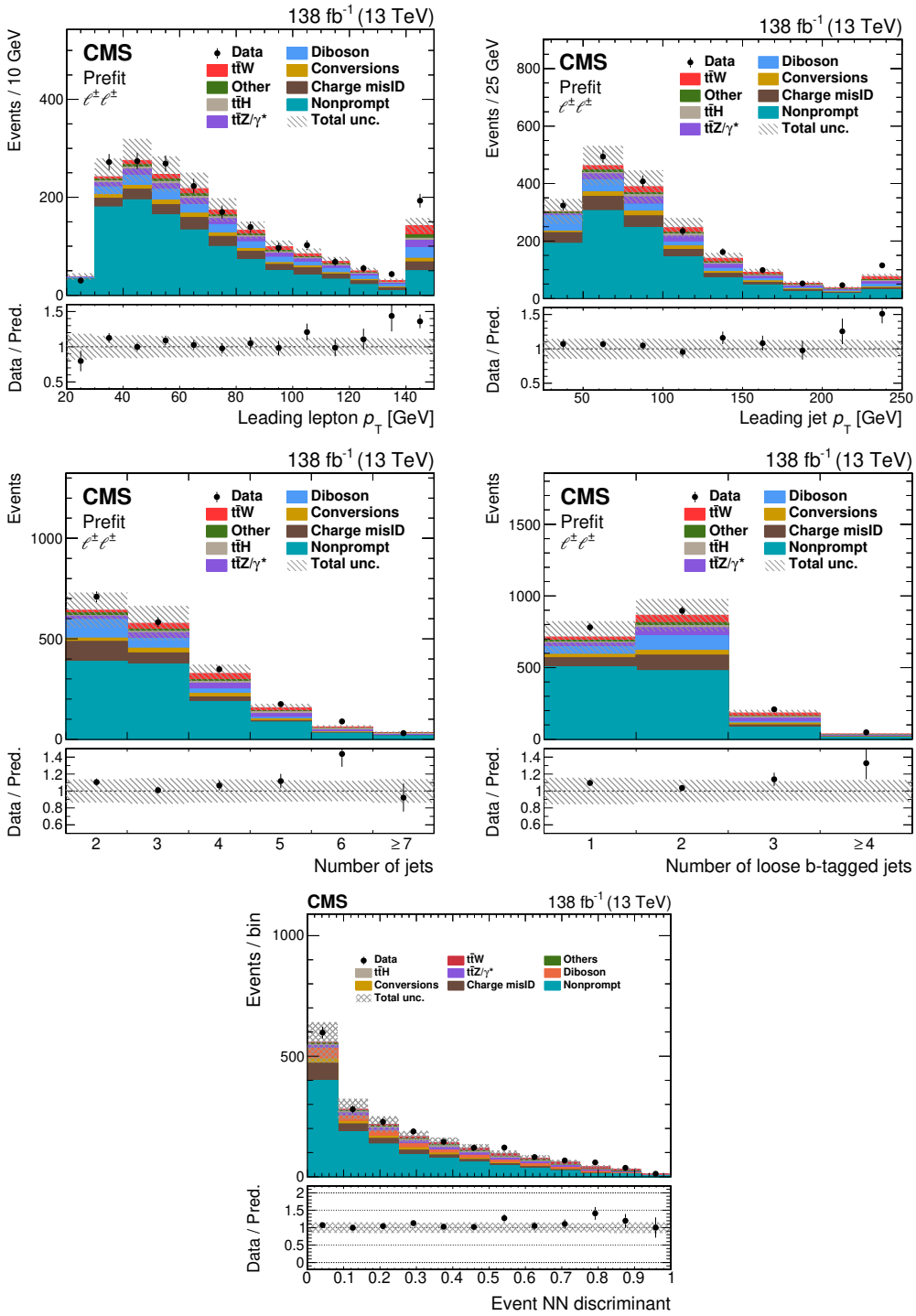

Figure 3.20: Comparison of the number of observed (points) and predicted (coloured histograms) events in the validation region for nonprompt leptons. The distributions of the number of jets (upper left) and loose b-tagged jets (upper right), and of the p_{T} of the leading jet (middle left) and of the leading lepton (middle right), and the NN discriminant are displayed. The predictions are shown before the fit to data ("Prefit"). The vertical bars on the points represent the statistical uncertainties in the data, and the hatched bands the systematic uncertainty in the predictions. The last bins include the overflow contributions. In the lower panels, the ratio of the event yields in data to the overall sum of the predictions is presented.

3.3.4 Backgrounds control regions

The estimations and modellings of WZ, ZZ, and $t\bar{t}Z/\gamma^*$ backgrounds are validated in two control regions, namely trilepton and four-lepton final states. These regions also enter the simultaneous fit for the yield of $t\bar{t}W$ to further constrain those backgrounds.

Trilepton control region

The trilepton control region is defined by the following requirements:

- Contain exactly three leptons that pass tight selection;
- The leading, sub-leading and trailing leptons have a p_T of at least 25, 15 and 15 GeV, respectively;
- At least one same-flavour opposite-sign lepton pair with $|m(\ell\ell) - m_Z| < 10\,\text{GeV}$;
- Invariant mass of any lepton pair must be at least 12 GeV;
- Sum of lepton charges is either 1 or -1.

The events in trilepton control region are categorised based on numbers of jets and b-tagged jets. The event yield of each category is shown in Fig. 3.21.

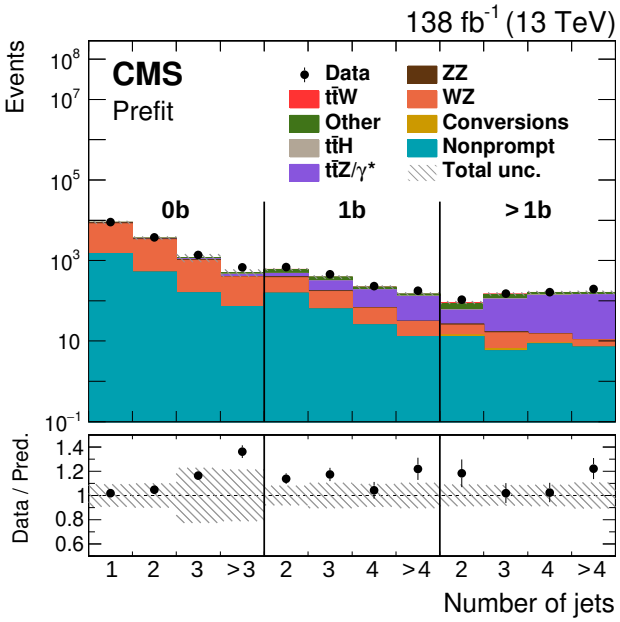

Figure 3.21: Distributions of the subcategories used for the 3ℓ channel control region for all lepton flavour channels, considering the full Run II data set. The uncertainty band accounts for the full uncertainty, including both statistical and systematic sources of uncertainty. Courtesy of Clara Ramón Álvarez and Carlos Vico Villalba, University of Oviedo.

Four-lepton control region

For the four-lepton control region (Fig. 3.22), the following requirements are applied:

- Exactly four tight leptons with $p_{\mathrm{T}} > 25$, 15, 15, 10 GeV, respectively
- At least one opposite-sign same-flavour (OSSF) lepton pair with $|m(\ell\ell) - m_Z| < 10\,\mathrm{GeV}$

The events are categorised based on numbers of jet, b-tagged jets and OSSF lepton pairs on Z, namely:

- 2 OSSF lepton pairs on Z;
- 1 OSSF lepton pairs on Z with 0 jets;
- 1 OSSF lepton pairs on Z with 1 b-tagged jet;
- 1 OSSF lepton pairs on Z with more than 1 b-tagged jet.

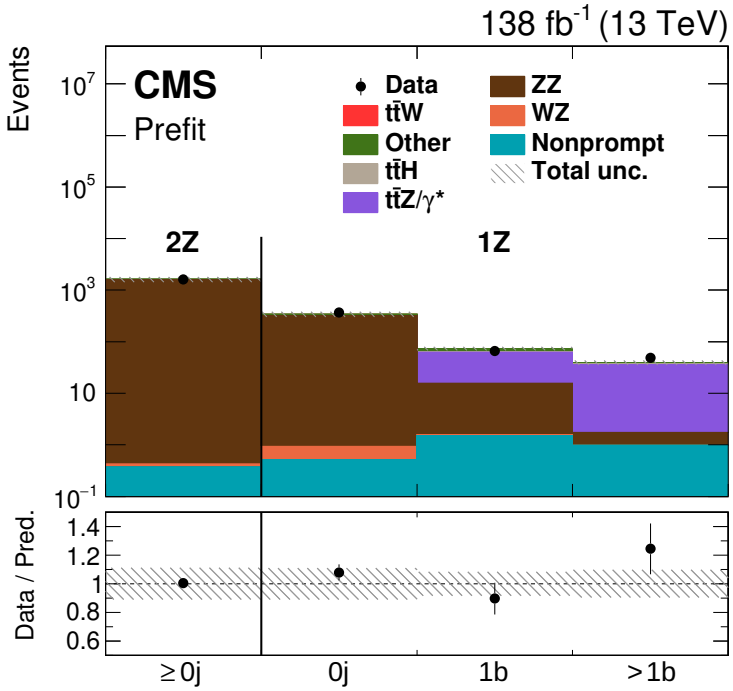

Figure 3.22: Distributions of the subcategories used for the 4ℓ control region for all lepton flavour channels, considering the full Run II data set. The uncertainty band accounts for the full uncertainty, including both statistical and systematic sources of uncertainty. Courtesy of Clara Ramón Álvarez and Carlos Vico Villalba, University of Oviedo.

3.4 Signal discrimination

Despite the careful selection criteria for objects and events to minimise background contributions, there can still be a considerable number of background events present in the signal regions. Therefore, it is essential to employ additional methods to effectively differentiate between signal and background events. Different approaches are thoroughly investigated and compared to select the most suitable ones that optimise the selection of signal events. In the same-sign dilepton and trilepton channels, different approaches are employed. The former uses Multivariate Analysis (MVA)-based techniques, while the latter relies on single-variable discriminants.

3.4.1 Multivariate analysis in same-sign dilepton channel

In the same-sign dilepton channel, a Multivariate Analysis (MVA) approach is based on two algorithms, Boosted Decision Tree (BDT) [106] and Neural-Network (NN). The two algorithms are thoroughly studied and compared to assess their performance in discriminating between signal and background events. They take advantage of the correlations between different variables and combines them to archive optimal discrimination power.

In both algorithms, events from $t\bar{t}W$, $t\bar{t}$, $t\bar{t}H$, $t\bar{t}Z$, and $t\bar{t}\gamma$ simulation samples are used for training and validation purposes. To minimise any biases during the application of the models, only a fraction of events is used: 20% from the $t\bar{t}W$ signal samples and 15% from the background samples.

In order to prevent the overrepresentation of certain processes during training, events from each simulation sample are assigned weights corresponding to the respective cross section of the process. This weighting scheme ensures that the training process is fair and unbiased.

While the BDT and NN algorithms employ slightly different sets of input variables, they both exploit the kinematic properties of the leptons and jets in the events. Variables such as lepton flavour, charge, and the multiplicities of jets and b-tagged jets are also included in the input sets. Within the two methods explored in this section, I exclusively undertook the study of the BDT algorithm, whereas Dr. Pieter David from the Catholic University of Louvain focused on the NN approach.

Boosted decision tree

The BDT [106] is trained using a binary approach, where events are divided into two categories: “signal” and “background”. The signal category exclusively consists of $t\bar{t}W$ simulated events, while the background category includes events from $t\bar{t}$, $t\bar{t}H$, $t\bar{t}Z$, and $t\bar{t}\gamma$ simulations.

The BDT algorithm employs gradient boosting [107] techniques to optimise its performance. Gradient boosting is an ensemble method that combines multiple weak classifiers,

known as decision trees, to create a strong classifier. During the training process, the BDT algorithm iteratively constructs decision trees by focusing on the misclassified events from previous iterations. It assigns higher weights to the misclassified events, enabling the algorithm to prioritise and improve the classification of those events in subsequent iterations. By iteratively combining multiple decision trees, the BDT algorithm builds a powerful classifier that can effectively discriminate between signal and background events. In this analysis, the BDT algorithm is trained using TMVA package [108], with the following parameters:

- Number of trees: 250
- Minimum node size: 5%
- Number of cuts: 25
- Maximum depth: 3
- Bagged sample fraction: 0.8
- Shrinkage: 0.25

These parameters are optimised by performing a numerous number of trainings with different combinations of parameters' values. The chosen setting has the best separation between signal and background, hence, the highest Area Under the Curve (AUC). The meaning of AUC is explained later on in this section.

The input variables that are used in the training include:

- data taking year;
- flavour, charge, p_{T}, η, and m_{T} of each lepton;
- invariant mass of the two leptons $m(\ell\ell)$;
- separation between the leptons;
- minimum and maximum separations between each lepton and jets;
- number of jets;
- numbers loose and medium b-tagged jets;
- average separation between the jets;
- missing transverse momentum $p_{\mathrm{T}}^{\mathrm{miss}}$;
- scalar sum of jet p_{T};
- p_{T}, η, and mass of each of the two jets with highest p_{T}.

The output of the BDT algorithm is a continuous spectrum ranging from -1 to 1. A higher output value indicates a higher likelihood of the event belonging to the $\mathrm{t\bar{t}W}$ signal

category, while a lower value indicates a higher likelihood of it belonging to one of the background processes.

To improve the interpretability and enhance the discrimination between signal and background events, the histograms of BDT distributions has a variable binning, with the bin width gradually increases from the first bin (closest -1) to the final bin. This ensures that the total yield of expected background in each bin in the BDT output distribution follows an approximately linearly decreasing function. Therefore, the discrimination power between signal and background events can be enhanced, allowing for better separation and identification of the $t\bar{t}W$ signal from the background processes. Then, a monotonic transformation is applied to the BDT output distribution to give an equidistant binning, making it easier to visually interpret the histograms. The distributions of BDT output are shown in Fig. 3.23 and 3.24 for each channel separately.

Multiclass neural-network

The training of the NN model is performed using the TensorFlow package [109] with the Keras [110] interface. An ADAM optimiser [111] is used to minimise a cross-entropy loss function [112] during the training process. In this approach, the events are divided into four different classes: $t\bar{t}W$ signal, $t\bar{t}H$ and $t\bar{t}Z/\gamma^*$ backgrounds, nonprompt lepton background from $t\bar{t}$, and $t\bar{t}\gamma$ background. The NN architecture consists of two fully connected hidden layers with 128 and 64 nodes [113], respectively, and an output layer with four nodes representing the four event classes.

To prevent overfitting, a dropout layer [114] with a dropout rate of 0.5 is added after the hidden layers. The dropout rate of 0.5 means that during training, each input variable has a 50% chance of being randomly set to zero. This technique helps reduce the dependence of the model on specific input variables and encourages the model to learn more robust and generalised features.

Two important parameters in training a neural network are the learning rate and the number of epochs. The learning rate determines the step size taken by the optimisation algorithm to update the neural network's weights at each iteration. In this case, a learning rate of 0.0015 is chosen. An epoch represents a complete iteration through the entire data set during training. Within each epoch, the model makes predictions, calculates the loss, and updates the parameters based on the optimisation algorithm. For this training, the number of epochs is set to 20.

Compared to BDT, the input variables used in NN are slightly different. The following inputs are passed to the neural network:

- data-taking year;
- lepton flavour category (dielectron, dimuon, or mixed; the mixed category is split depending on the flavour of the leading lepton);
- charge of the leptons ($+1$ or -1);

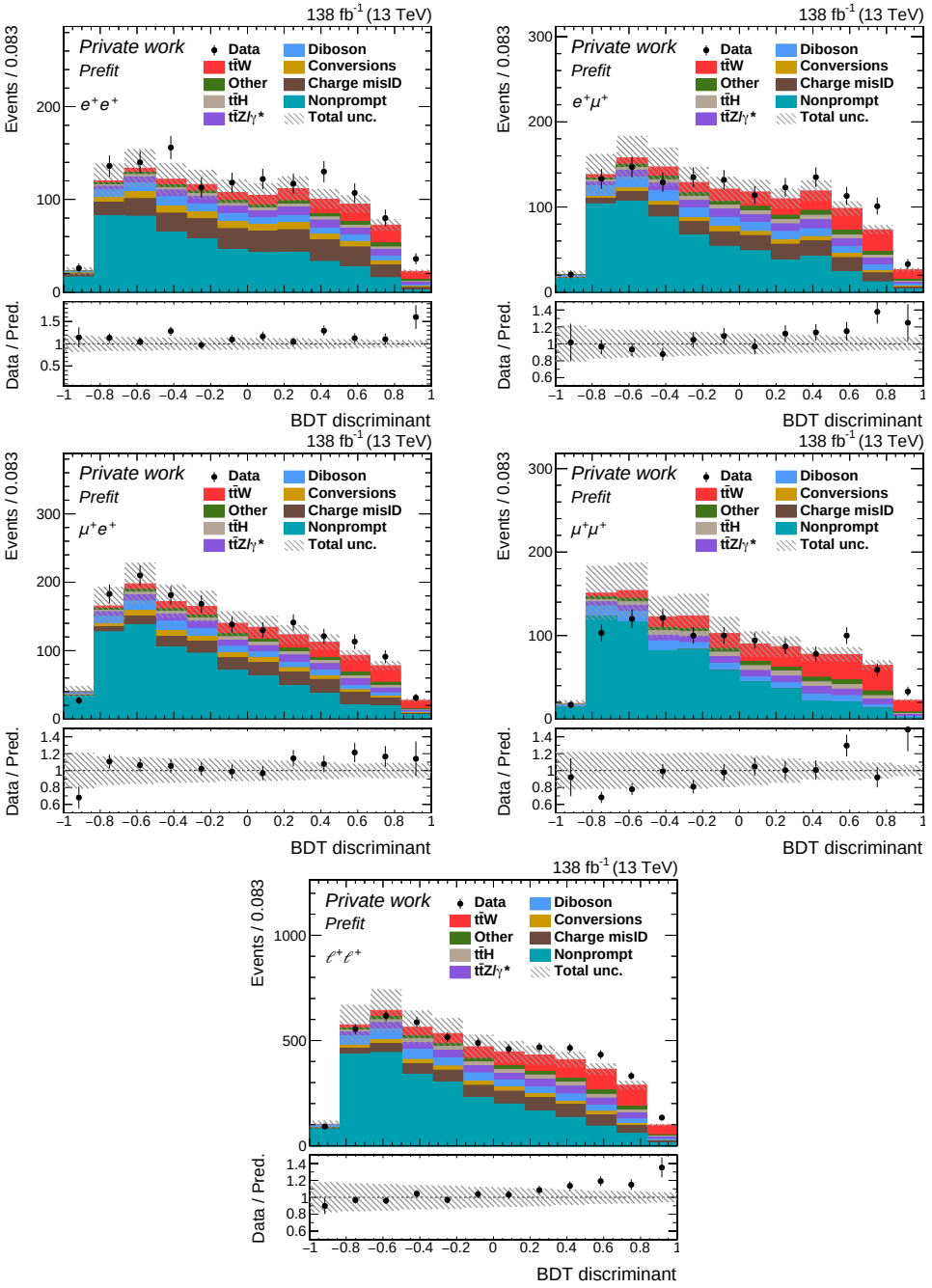

Figure 3.23: The distribution of BDT discriminant in e^+e^+ (top left), $e^+\mu^+$ (top right), μ^+e^+ (middle left), $\mu^+\mu^+$ (middle right), and all channels combined (bottom). The vertical bars on the points represent the statistical uncertainties in the data, and the hatched bands the systematic uncertainty in the predictions. The lower panel shows the ratio of the event yields in data to the overall sum of the predictions.

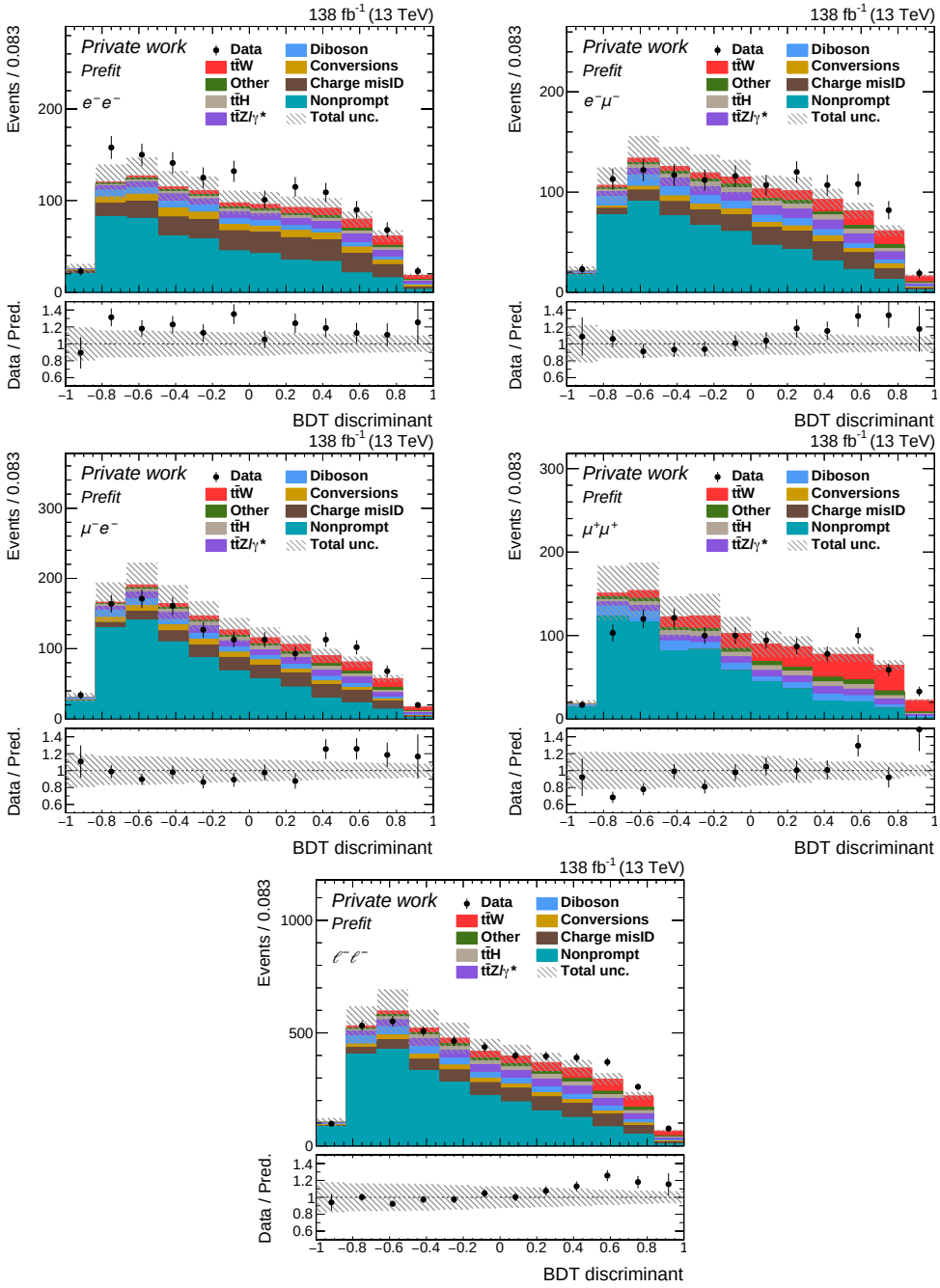

Figure 3.24: The distribution of BDT discriminant in e^-e^- (top left), $e^-\mu^-$ (top right), μ^-e^- (middle left), $\mu^-\mu^-$ (middle right), and all channels combined (bottom). The vertical bars on the points represent the statistical uncertainties in the data, and the hatched bands the systematic uncertainty in the predictions. The lower panel shows the ratio of the event yields in data to the overall sum of the predictions.

- transverse momentum and pseudorapidity of the leptons;
- invariant mass, and difference in azimuthal angle and pseudorapidity between the leptons;
- transverse mass of the leptons;
- missing transverse momentum, and the azimuthal angle difference between the missing transverse momentum and each of the leptons;
- number of cleaned jets passing the loose b-tagging working point, at least one and at most two are considered as b-tagged jets (if there are more than two b-tagged jets they are part of the other jets collection);
- number of cleaned jets that are not among the b-tagged;
- transverse momentum and pseudorapidity for up to 2 b-tagged and up to 4 non-b-tagged jets;
- azimuthal angle difference and invariant mass for any combination of a lepton and jet;
- azimuthal angle difference and ΔR between any pair of jets;
- for b-tagged jets, a flag indicating whether they also pass the medium b-tagging working point;
- for other jets, a flag indicating whether they also pass the loose b-tagging working point (this can be the case for events with more than two jets that pass the loose b-tagging working point).

Unlike the BDT model, in this model, the data-taking year, lepton flavour category, charge, and jet multiplicities are one-hot encoded, this means that there is an input node for every possible value, which is set to 0 or 1, instead of a single integer with all possible values. This should help the network to adjust weights for the different categories.

The NN model outputs four values for each event, representing the probabilities of the event belonging to each of the four classes. These values range from 0 to 1 and can be interpreted as the probability for the event to belong to the corresponding class. Similarly to the BDT, the output distribution of the $t\bar{t}W$ node in the NN is subjected to variable histogram binning to achieve a linearly decreasing distribution for the expected background and monotonic transformation is also applied. The distributions of the $t\bar{t}W$ node of the NN output are shown in Fig. 3.25 and 3.26 for each channel separately.

The performance of NN and BDT are assessed using the Receiver Operating Characteristic (ROC) curves and the AUC. The ROC curves represent the true positive rate of selecting $t\bar{t}W$ signal as a function of the false positive at various point along the output distribution of the discriminant. It provides an insight into the trade-off between the sensitivity and specificity. The ROC curved are computed with simulation events that

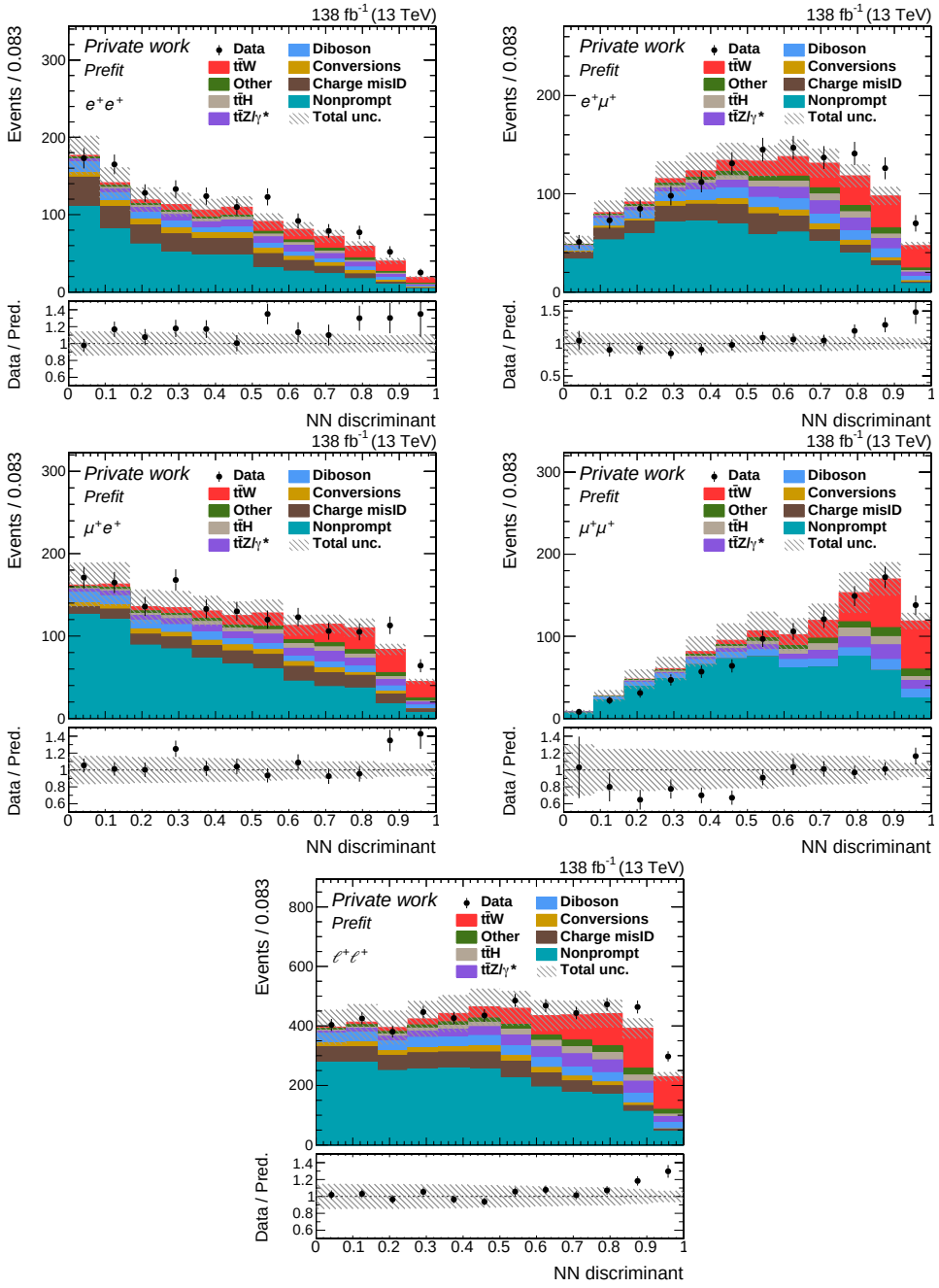

Figure 3.25: The distribution of NN discriminant in e^+e^+ (top left), $e^+\mu^+$ (top right), μ^+e^+ (middle left), $\mu^+\mu^+$ (middle right), and all channels combined (bottom). The vertical bars on the points represent the statistical uncertainties in the data, and the hatched bands the systematic uncertainty in the predictions. The lower panel shows the ratio of the event yields in data to the overall sum of the predictions.

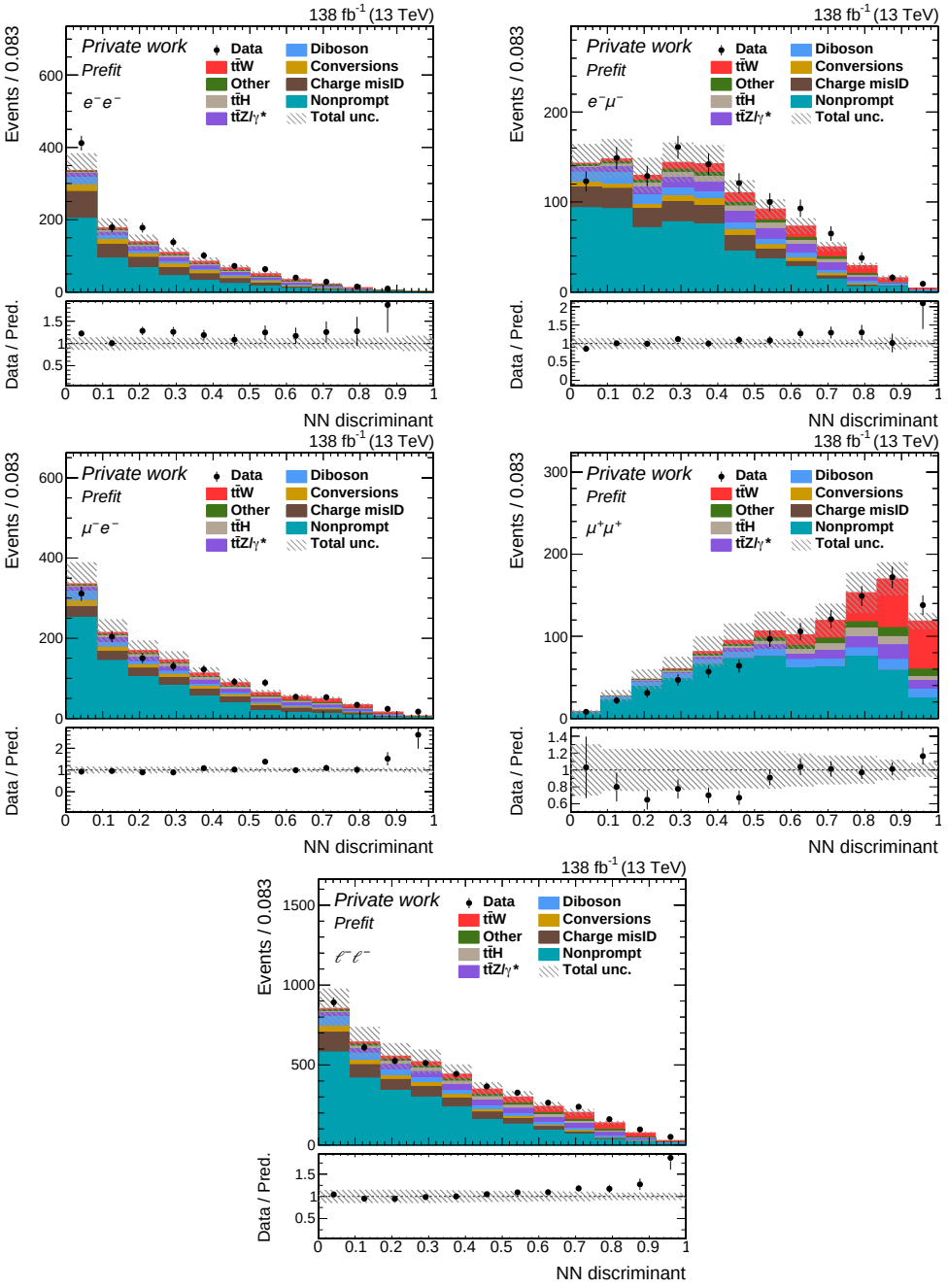

Figure 3.26: The distribution of NN discriminant in e^-e^- (top left), $e^-\mu^-$ (top right), μ^-e^- (middle left), $\mu^-\mu^-$ (middle right), and all channels combined (bottom). The vertical bars on the points represent the statistical uncertainties in the data, and the hatched bands the systematic uncertainty in the predictions. The lower panel shows the ratio of the event yields in data to the overall sum of the predictions.

are not used during the training to retain the independency of the data and avoid bias. For NN, the $t\bar{t}W$ output node is used to compute the ROC curve. Fig. 3.27 represent the ROC curves computed from both NN and BDT discriminants.

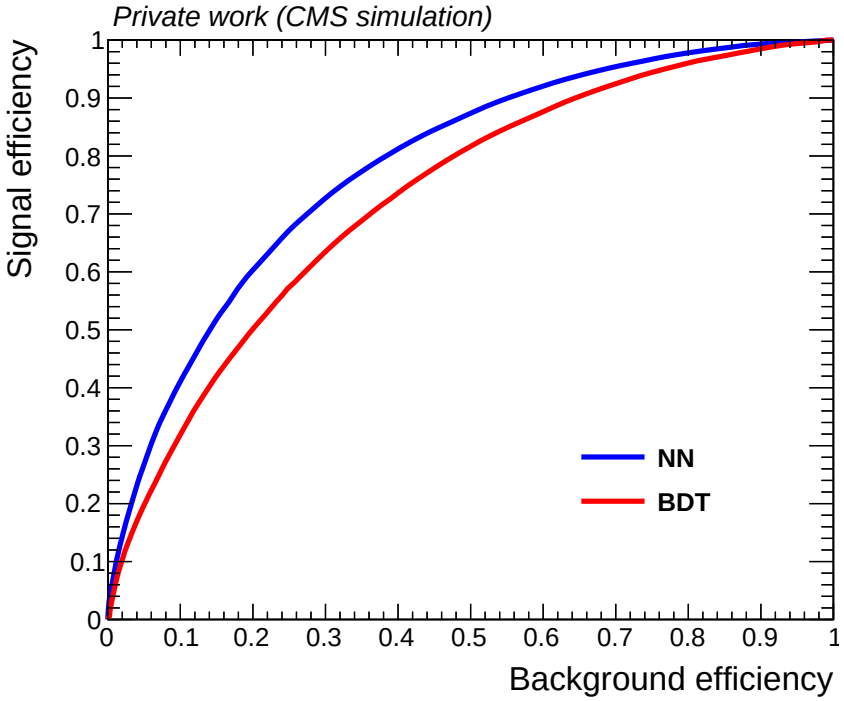

Figure 3.27: The figure compares the ROC curves of both NN and BDT. The x-axis represents the signal efficiency, while the y-axis depicts the background efficiency.

On the other hand, the AUC give an overview on the performance of a discriminant. At an AUC of 1.0, the discriminant would have a perfect distribution; while with an AUC of 0.5, the algorithm has no discriminating power. Thus, the objective is to archive a ROC curve that closely approaches the top-left corner of the figure, thereby maximising the AUC. Upon comparison, the NN shows a better performance than the BDT, as seen in Fig. 3.27, it has an AUC of 0.79 compared to 0.73 from the BDT algorithm. Furthermore, fits were conducted to Asimov data set [115] - a data set comprised solely of simulation events representing both "observed data" and "predicted data" - the fit using the NN discriminant outperformed the one using the BDT distribution. Notably, it archived a better precision of $\pm 10\%$ compared to $\pm 9\%$ from the fit to the BDT distribution. As a result, the NN discriminant is chosen for the signal extraction and the cross section measurement in the same-sign dilepton channel.

3.4.2 Single variable discriminant in trilepton channel

In the trilepton channel where statistics are limited, an events categorisation strategy is adopted instead of taking the MVA approach. Events are categorised based on the

number of jets and medium b-tagged jets present, as well as the charge combination of the leptons. This results in a total of twelve categories.

To further discriminate between signal and background events within each category, the invariant mass of the three leptons, denoted as $m(3\ell)$, is chosen as the discriminant variable. It has been found to provide the best discriminating power between signal and background processes in this channel. The $m(3\ell)$ distributions are binned in a way that is optimised to enhance the signal-to-background ratio and mitigate the effects of statistical uncertainties.

By categorising events based on different kinematic properties and charge combinations, the sensitivity to the $t\bar{t}W$ signal can be enhanced, even with limited statistics. This strategy allows for a more targeted analysis of the trilepton channel and improves the overall sensitivity of the measurement.

3.5 Measurements of the cross sections

In the final stage of the analysis, the cross section of $t\bar{t}W$ is measured using a binned profile likelihood fit [115, 116] to the predicted distribution of the NN $t\bar{t}W$ node output in dilepton channel and $m(3\ell)$ in trilepton channel. In addition to the signal regions, the trilepton and four-lepton control regions, that have been introduced in Sec. 3.3.4, are also used to control the $t\bar{t}Z/\gamma^*$, WZ, and ZZ backgrounds. The fit is performed simultaneously for all signal regions and control regions.

Various systematic uncertainties are considered in the cross section measurement and introduced to the fit as nuisance parameters [117]. These uncertainties arise from a variety of sources, such as the modelling of the signal and background processes, the calibration and resolution of detector effects, the estimation of background contributions, and the theoretical uncertainties in the cross section calculations. Each source of systematic uncertainties is evaluated and assigned a corresponding value that can affect the measured cross section.

The results of the cross section measurements are compared to theoretical predictions and used to test the SM. The impact of systematic uncertainties on the measurements are discussed.

3.5.1 Profile likelihood fit

The profile likelihood fit is a statistical technique commonly used in particle physics measurements at the CMS experiment. In this analysis, the observed and predicted events are arranged in binned histograms, thus, a variation known as binned profile likelihood is employed due to its ability to account for the discrete nature of the data. This algorithm provides parameter estimates by maximising the likelihood function while taking into consideration the bin counts.

In the fit presented in this analysis, the fit result does not yield a direct measurement of the cross section itself. Instead, it estimates a parameter called "signal strength", denoted by r. This parameter quantifies the deviation of the observed value from the theoretical prediction. In another word, it is the ratio of the observed yield of $t\bar{t}W$, s^{obs}, to the predicted yield, s^{pred}, $r = s^{obs}/s^{pred}$.

Consider a signal region that contains a set of n events observed from data spreading over N bins. In each bin i, there are n_i events observed from data, and the expectation value of n_i is $E[n_i] = rs_i + b_i$, where r is the signal strength, s_i and b_i are the expectation values for the signal and the background in bin i, respectively. The number of events in each bin i follows a Poisson distribution:

$$\mathbb{P}(n_i|r, s_i, b_i) = \frac{(rs_i + b_i)^{n_i} e^{-(rs_i+b_i)}}{n_i!} \tag{3.13}$$

The expectation values of signal and background in each bin i can be written as:

$$s_i = s^{pred} \int_i f_s(x, \Theta_s)\, dx, \tag{3.14}$$

$$b_i = b^{pred} \int_i f_b(x, \Theta_b)\, dx \tag{3.15}$$

where s^{pred} and b^{pred} are predicted number signal and background events. The terms f_s and f_b are probability density functions of observable x for signal and background events, respectively. These functions also depend on nuisance parameters Θ_s and Θ_b which can affect the yields or the shape of the observable x of the signal and background [115, 118], respectively. These nuisance parameters, can be collectively written as $\Theta = (\Theta_s, \Theta_b)$.

In additional to the signal region, a control region which is enriched in background events is used to constrain background. In each bin j, $(j = 1, \ldots, M)$, m_j events are observed from data and the expected number of events in each bin is $E[m_i] = u_j(\Theta)$, which is some calculable quantities depending on the set of nuisance parameters Θ. As in the signal regions, events in the control region also follows a Poisson distribution:

$$\mathbb{P}(m_j|u_j) = \frac{(u_j)^{m_j} e^{-(u_j)}}{m_j!} \tag{3.16}$$

A likelihood function $\mathcal{L}(r, \Theta)$ can be constructed, this function is a product of the probability density functions of the signal region and the control region. The function $\mathcal{L}(r, \Theta)$ is written as:

$$\mathcal{L}(r, \Theta) = \prod_i^N \mathbb{P}(n_i|r, s_i, b_i, \Theta) \prod_j^M \mathbb{P}(m_j|u_j) \tag{3.17}$$

In CMS, it is a common approach to perform a hypothesis test using profile likelihood ratio [115], which is defined as:

$$\lambda(r) = \frac{\mathcal{L}(r, \hat{\hat{\Theta}})}{\mathcal{L}(\hat{r}, \hat{\Theta})} \tag{3.18}$$

The numerator is represented by a profile likelihood function, which is a conditional maximum likelihood function. For every fixed value of r, there is an optimal value of Θ, $\Theta = \hat{\hat{\Theta}}$, which gives the likelihood function the maximum value. Contrastingly, the denominator is an unconditional likelihood function, in which both r and Θ are optimised such that the likelihood function reaches its maximum. At the global maximum, r and

Θ take the best-fit values of $r = \hat{r}$ and $\Theta = \hat{\Theta}$. Hence, the value of $\lambda(r)$ is between 0 and 1.

However, as most optimisation algorithms are designed to minimise functions, the test statistic $t = -2ln[\lambda(r)]$, is used instead.

$$t(r) = 2ln[\mathcal{L}(\hat{r}, \hat{\Theta})] - 2ln[\mathcal{L}(r, \hat{\hat{\Theta}})] \tag{3.19}$$

At its optimal fit, the test statistic t reaches its minimum value of zero.

There are several reasons to make the case for using binned profile likelihood. Firstly, it is a computationally efficient method, especially when dealing with a huge amount of data, thus binning the data can reduce the complexity of the data set and simplify the analysis. This method also is robust to fluctuations in the data, as data often contains a lot of noise or random fluctuations, binning the data means grouping multiple data points into discrete intervals. This effectively averaging out the fluctuation within each bin. Each bin is considered independent of each other, as shown in Eq. 3.17. During the fit procedure, while focusing on the parameters of interest, difference sources of systematic uncertainties, referred to as nuisance parameters, are taken into account, allowing for estimation of these nuisance parameters. When performing the fit for the parameter of interest, the nuisance parameters are profiled over, it adjusts these parameters to their best-fit values for each value of the parameter of interest, reducing their impact on the final result. This makes the analysis less sensitive to the specifics of these less important parameters. The impacts of those nuisance parameters on the likelihood function are directly incorporated into the fit, this results in a reduction of bias in the parameter of interest.

3.5.2 Sources of systematic uncertainties

In experimental physics, not only statistics of data can influence measurements' result, but systematic uncertainties can also exert a significant impact. This is true in many particle physics measurements, where systematic uncertainties often emerge as the dominant factor. These uncertainties can originate from a number of sources such as the measurement and calibration of physical quantities, our limited understanding of the experiment apparatus, imperfect simulations, estimation of background processes, as well as theoretical modelling. Unlike statistical uncertainty, which reduces as more data are collected, systematic uncertainties remained resilient. Their effect may only diminish with more comprehensive understanding of the sources of uncertainties.

At the CMS, most of the systematic uncertainties are often considered as "shape" uncertainties, which have the potential to affect the shape or distribution of the parameter of interest. Whereas some other systematic uncertainties are "normalisation", which scales the yield of each bin by a fixed factor. In the likelihood fit, the shape uncertainties are represented by Gaussian distributions and normalisation uncertainties follow log-normal functions. Based on their nature, systematic uncertainties are classified as experimental uncertainties, background uncertainties or modelling uncertainties.

Experimental uncertainties

Experimental uncertainties consist of corrections that are implemented to account for the detector's performance. In this analysis, the most relevant experimental uncertainties include factors such as luminosity, lepton identification efficiency, jet flavour tagging, jet energy scale and resolution, triggers, and pileup.

- **Luminosity**
  The systemic uncertainty associate with the luminosity measurements consist of two components: correlated and uncorrelated. The uncorrelated components are the uncertainties associated with the measurement of integrated luminosity of each data-taking year: 2016, 2017, and 2018, which are estimated to be 1.2%, 2.3%, and 2.5%, respectively. They are applied to all simulation events in the corresponding data-taking year. While the correlated component represents an overall uncertainty of 1.6% across all three data-taking years and is consistently applied to all simulation samples. All values are incorporated into the likelihood function as normalisation uncertainties.

- **Pileup reweighting**
  As discussed earlier in Sec. 3.1.3, the distributions of the number of additional p-p interactions per event in the simulation do not perfectly match those observed in the data. Consequently, the profile of the true number of interactions is adjusted through reweighting to align with the inferred distribution derived from the instantaneous luminosity in the data. To compute the expected pileup distribution for the data, a minimum-bias inelastic cross section of 69.2 mb is used. This cross section was measure with a total uncertainty of 4.6% [97,98]. As a result, an uncertainty of 4.6% is assigned to the reweighting of pileup and is applied consistently fully correlated throughout the data-taking years for all signal and background simulated events.

- **Trigger efficiency**
  The application of the trigger introduces certain inefficiencies in event selection for both data and simulation samples. However, the efficiencies measured in data and in simulations do not completely align. Hence, a scale factor is applied to each event in the simulation to correct for the discrepancies in efficiencies. In the dilepton channel, the scale factor is parametrised as a function of the leading and subleading leptons' p_{T}. On the other hand, it is parametrised as a function of the p_{T} of the leading lepton in the trilepton channel. The uncertainty introduced in the measurement of the scale factor consists of a statistical uncertainty and a systematic uncertainty of 2%. The statistical uncertainties are less than 1% in most bins. The systematic uncertainty arises from the variation in efficiency when measuring using the MET and JetHT data sets separately compared to using the combined data set. This uncertainty is treated as uncorrelated across the data-taking years and lepton final states in the dilepton channel. However, in the trilepton final state, the uncertainty is correlated among the different final states

and uncorrelated among the data-taking years.

- **L1 prefiring**
  As described in Sec. 3.1.3, the L1 prefiring effect causes efficiency loss in L1 triggers during 2016 and 2017 data-taking periods. This efficiency loss is estimated to be around 10-20%, depending on p_T, η, and timing of the particles [96]. Therefore, a scale factor is applied to 2016 and 2017 simulation samples in order to correct the effect of L1 prefiring. Associate with this scale factor is a shape uncertainty which is correlated between the two years.

- **Lepton identification**
  In the identification of tight leptons, there is an inherent efficiency associated with the selection criteria. The efficiencies for tight lepton selection are measured in p_T-$|\eta|$ bins, using $Z \to \ell^+\ell^-$ events from data and simulations. These measurements are performed separately for electrons and muon. To rectify any discrepancies between the lepton selection efficiencies in data and simulations, a scale factor is applied to each selected lepton in the simulations. The uncertainties tied to these scale factors are typically on the order of a few percent in each bin. These uncertainties are considered shape uncertainties, and treated as correlated between the data-taking years and measurement channels.

- **Jet energy scale and resolution**
  There are numerous sources of uncertainty that can impact the Jet Energy Scale (JEC) and Jet Energy Resolution (JER) [119]. These uncertainty sources are organised into 17 groups in order to simplify the implementation and reduce the number of nuisance parameters. Among these groups, 11 are associated with JEC, they account for variations in different detector regions, jet flavours, and discrepancies between data and simulation, and between different event generators. The remaining groups pertain to JER and cover variations in jet p_T and pseudorapidities. Each variation within these groups can influence the p_T of jets and the missing transverse momentum p_T^{miss} in each event. Therefore, it becomes necessary to recompute the jet p_T and p_T^{miss} at each iteration of these variations. Regarding the JEC sources, some are partially correlated among the data-taking years, meaning that certain sources are treated as uncorrelated while others are treated as correlated. On the other hand, the JER sources are fully uncorrelated between the years. This correlation treatment helps account for shared effects among the data-taking years, enabling a more accurate consideration of the uncertainties associated with JEC and JER.

- **b-tagging**
  Similar to the lepton identification and trigger efficiencies, the shape of the jet flavour tagging algorithm, DeepJet, in simulations is adjusted to match that observed in data. This is accomplished by applying scale factors that depend on the jet p_T and η. The scale factors are determined using the "iterative fit" method [76, 120], and they are varied to assess the associated uncertainties. The scale factors for

light- and heavy-flavoured quarks are varied independently. There are two sets of uncertainties for the flavour tagging scale factors. One set is uncorrelated across the data-taking years as it is due to statistical uncertainties of the measurements, while the other set is fully correlated.

Background uncertainties

Depending on the nature of the background and how it is estimated, uncertainties associate with it can be either shape or normalisation in nature. For most background sources, a normalisation is applied to all bins, some of these normalisations come from SM predictions, it reflects our current understanding of the background processes, and some other normalisations are obtained from the latest experimental results.

- **Nonprompt lepton background**
  The prediction of nonprompt lepton background relies heavily on the modelling of the multijets events and the subtraction of electroweak events from the FR measurement region. Since the electroweak subtraction is conducted using three different approaches, it results in three distinct sets of FRs. To accommodate this, an envelope systematic uncertainty is introduced to account for the variability in the nonprompt lepton background. The up and down variations of this envelope uncertainty is the maximum and the minimum of the three FR values in each bin. Moreover, as the FRs are dependent on the lepton p_{T} and η, two additional uncertainties are included to account for the p_{T} and η dependencies of the nonprompt leptons and the observed mismodelling effects. These three uncertainties also encompass the statistics of the measurements. They are considered fully correlated across the years and uncorrelated between different lepton flavours. Finally, in the validation region discussed in Sec. 3.3.2, an additional uncertainty of 20% is assigned to cover the discrepancy between the observed and predicted yields. This uncertainty is fully correlated among the data-taking years and lepton final states.

- **Charge misidentification background**
  In Sec. 3.3.1, the charge misID background estimation method is validated using $\mathrm{Z} \rightarrow \mathrm{ee}$ events from data and simulations, revealing discrepancies between the observed and predicted charge misID. To account for this difference, a flat uncertainty of 20% is applied to dilepton events with at least one electron. This uncertainty is fully correlated between the years and final states. It does not apply to dimuon events or trilepton channels.

- **WZ and ZZ**
  The yields of WZ and ZZ are obtained through a fit from the trilepton and four-lepton control regions described in Sec. 3.3.4. However, events with a large number of b-tagged jets contribute to the yields due to either misidentification of the jet's flavour or phase space mismodelling in simulations. Consequently, an additional uncertainty of 40% is applied to events with at least two b-tagged jets, while an uncertainty of 10% is applied to events with fewer than two b-tagged jets [121].

- **$t\bar{t}H$, $t\bar{t}Z$, and $t\bar{t}\gamma$**
  A 20% uncertainty is applied to the normalisation of $t\bar{t}H$ to account for the experimental uncertainty of the $t\bar{t}H$ cross section [30] measurement. While the yield of $t\bar{t}Z$ is obtained from the control regions during the fit, it is constrained within theoretical uncertainties of $^{+8.6\%}_{-9.5\%}$. For the prompt contribution from $t\bar{t}\gamma$, a normalisation uncertainty of 8% [122] is considered, and for events from photon conversions, a 30% [123] uncertainty is applied to address mismodelling of photon conversions in simulations.

- **Other background processes**
  Background contribution from tZq and $t\bar{t}VV$ each is constrained with uncertainties of 10% and 50%, respectively. These uncertainties base on the latest measurement by CMS [100, 124]. The normalisations of tHq and tHW backgrounds are constraints within an uncertainty of 50% [125]. A 50% uncertainty also applied to the VVV production based on a study of WWW, which is the main contribution to VVV, as well as to other rare processes. While some of the rare processes have been measured at a higher precision than 50%, the assigned uncertainty takes a conservative approach to account for the extrapolation of their predictions to relevant the kinematic region of this analysis.

Modelling uncertainties

The modelling of signal and background processes is subject to certain limitations arising from theoretical models and the production of simulation samples. The associated uncertainties pertaining to these modelling aspects can potentially impact the acceptance and selection efficiencies, yields, and the shape of the fitted distribution. Therefore, they are carefully considered in order to mitigate their effect on the measurement. As these modellings are consistent across the data-taking years, their uncertainties are treated as fully correlated across all years and final states.

- **Factorisation and renormalisation scales**
  To assess the uncertainties in the acceptance of both signal and background processes resulting from missing higher-order diagrams in the ME calculations, variations are applied to the factorisation scale, μ_F, and the renormalisation scale, μ_R. Six different variations are considered: four involve fixing either μ_F or μ_R while varying the other parameter upwards or downwards by a factor of 2. In the remaining two cases, both μ_F and μ_R are simultaneously varied upwards or downwards by the same factor. The uncertainty in each bin is evaluated by constructing the envelope from these six variations. This type of uncertainty is relevant for the $t\bar{t}W$ signal and all background processes estimated from simulations, and it is applied in correlation among different processes.

- **Parton distribution functions**
  The impact of the pdfs on $t\bar{t}W$ signal is considered using the Hessian NNPDF3.1 set of 100 pdf variations [88]. These variations are generated for the central

value $\alpha_s(m_Z^2) = 0.118$. Additionally, two variations with $\alpha_s(m_Z^2) = 0.1165$ and $\alpha_s(m_Z^2) = 0.1195$ are also provided. The difference between each pdf variation and the nominal value is combined quadratically, and then added in quadrature to the mean of the two α_s variations to obtain the final uncertainty variations for the pdf effects. This uncertainty is applied as a shape uncertainty on the $t\bar{t}W$ signal. Furthermore, a 2.3% flat uncertainty is applied to $t\bar{t}Z$ to account for its pdf uncertainty.

- **Parton shower**
  The Initial State Radiation (ISR) and Final State Radiation (FSR) energy scales have the potential to affect the shape of the $t\bar{t}W$ signal. Therefore, their effects are considered as two shape uncertainties. These uncertainties are obtained from the alternative PS simulation samples. The impact of colour reconnection on the $t\bar{t}W$ signal is investigated using two alternative simulations samples: the "QCD-inspired" model and the "Gluon-move" model [126]. In the QCD-inspired model, all QCD dipole pairs are allowed to undergo reconnection according to QCD colour rules that determine the compatibility of colours between the two strings. In the Gluon-move model, quarks remain in their original positions without any colour exchange, while gluons are moved from one system of two connected partons to another. To assess the systematic uncertainty related to colour reconnection, the differences in the distributions (such as NN output) between the two models and the nominal $t\bar{t}W$ samples are compared. It shows that the two models agree with the nominal sample within statistical uncertainty, indicating that this is not a significant physical effect. Therefore, a flat uncertainty of 1% is applied to $t\bar{t}W$ events. to account for the observed difference

- **Statistical uncertainty on simulation**
  Finally, in the prediction of signal and background processes, events are drawn from simulations or selected from sideband regions in the data. Since the number of events used for the estimation is limited, it is important to consider the associated statistical uncertainties. To tackle this, the Barlow-Beeston approach [127] is adopted, treating the statistical uncertainty on predicted events as a unified nuisance parameter for all processes within each bin of the histograms utilised in the fitting process. This ensures a comprehensive treatment of statistical uncertainties across the various processes in each histogram bin.

3.5.3 Result

A binned profile likelihood fit to is performed on data using the predicted distribution of NN output from dilepton signal region, $m(3\ell)$ from trilepton signal regions, and the event yields for each category in the trilepton and four-lepton control regions. In the dilepton signal region, events are split into 24 groups based on lepton flavour combination (ee, eμ, μe, and $\mu\mu$), charge of the lepton pairs (++ and $--$), and data taking year (2016, 2017, and 2018). Events from trilepton signal region categorised based on the sum of charge of the leptons, jet and medium b-tagged jet multiplicities as well as data taking years, this comes to a total of 36 categories. There are 12 categories in trilepton control region which are grouped by lepton flavour (eee, eeμ, e$\mu\mu$, and $\mu\mu\mu$) and data taking years; and the four-lepton control regions are split by data taking year only. In total, there are 75 separate categories to enter the fit.

The fit incorporates all the aforementioned systematic uncertainties, where the focus lies on the signal strength of $t\bar{t}W$ cross section as the primary parameters of interest, which is allowed to vary freely during the fit and based on the theory prediction from Ref. [34]. Additionally, the signal strength of $t\bar{t}Z$ cross section serves as the secondary parameter of interest, also subject to variation in the fit but within the bounds of its theoretical uncertainties. Conversely, the yields of WZ and ZZ are constrained by the trilepton and four-lepton control regions.

The observed signal strength obtained fit results is:

$$r_{t\bar{t}W} = 1.47 \pm 0.07\,(\text{stat})^{+0.09}_{-0.08}\,(\text{syst}), \tag{3.20}$$

which gives an inclusive cross section of

$$\sigma_{t\bar{t}W} = 868 \pm 40\,(\text{stat}) \pm 51\,(\text{syst})\,\text{fb}. \tag{3.21}$$

The impacts of systematic uncertainties in the measured inclusive cross section are provided in Tab. 3.2. Among the dominant experimental uncertainties are those associated with the measurement of integrated luminosity, the background estimation for electron charge misidentification, and the identification efficiency of b jets. The normalisation uncertainties in the predictions of the $t\bar{t}H$, VVV, and $t\bar{t}VV$ processes also make significant contributions to the total uncertainty in the measurement. Despite the low number of expected events for these rare processes, their similarities to the decay products of the $t\bar{t}W$ signal process result in a substantial impact on the overall uncertainty. Other important sources of uncertainty arise from modelling uncertainties in the simulation of $t\bar{t}W$ events and from the statistical uncertainty in the predicted number of signal and background events. To give a better insight to the impact of nuisance parameters on the fit result, an impact plot is shown in Fig 3.28.

In this analysis, a tension between the observed and predicted cross section normalisations of $t\bar{t}W$ and $t\bar{t}H$ is observed, and this is evident in the impact plot shown in Fig. 3.28. Specifically, associated with the cross section normalisation of $t\bar{t}H$, there is an

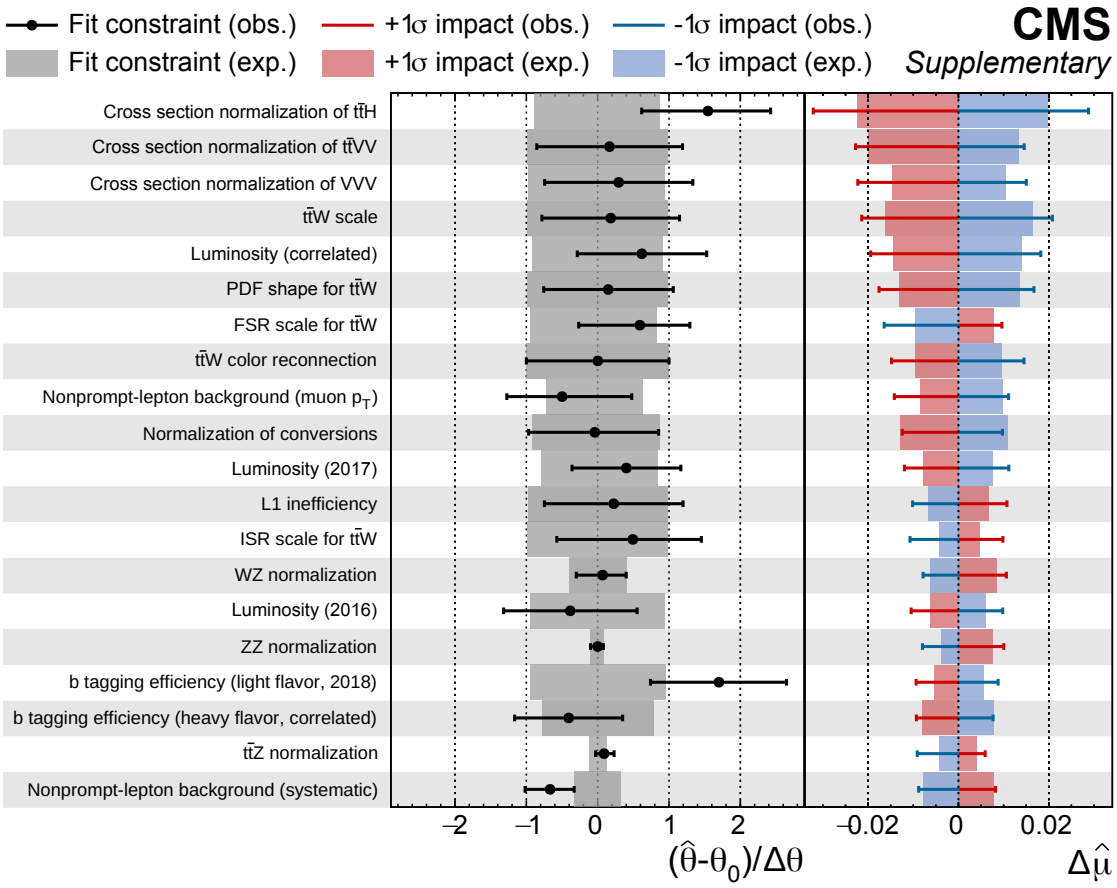

Figure 3.28: For the nuisance parameters listed in the left column, the pulls $(\hat{\theta}-\theta_0)/\Delta\theta$ (middle column) and impacts $\Delta\hat{\mu}$ (right column) are displayed. The 20 nuisance parameters with the largest impacts in the fit used to determine the $\mathrm{t\bar{t}W}$ cross section are shown. The impact $\Delta\hat{\mu}$ is defined as the shift induced in the signal strength μ when the nuisance parameter θ is varied by ±1 standard deviation (σ). The pull $(\hat{\theta}-\theta_0)/\Delta\theta$ is calculated from the values $\hat{\theta}$ and θ_0 after and before the fit of θ and from its uncertainty $\Delta\theta$ before the fit. The label "correlated" and the per-year labels indicate nuisance parameters associated with the correlated and uncorrelated parts of a systematic uncertainty.

Table 3.2: Sources of systematic uncertainty in the predicted signal and background event yields with their impact on the measured $t\bar{t}W$ production cross section, estimated after the fit to the data. Only systematic uncertainty sources with values greater than 0.1% are included in the table. The production cross sections of the $t\bar{t}W$, WZ, ZZ, and $t\bar{t}Z$ processes are simultaneously constrained in the fit. The second-to-last row refers to the statistical uncertainty in the simulated event samples.

Source	Uncertainty [%]
Experimental uncertainties	
Integrated luminosity	1.9
b tagging efficiency	1.6
Trigger efficiency	1.2
Pileup reweighting	1.0
L1 inefficiency	0.7
Jet energy scale	0.6
Jet energy resolution	0.4
Lepton selection efficiency	0.4
Background uncertainties	
$t\bar{t}H$ normalisation	2.6
Charge misidentification	1.6
Nonprompt leptons	1.3
VVV normalisation	1.2
$t\bar{t}VV$ normalisation	1.2
Conversions normalisation	0.7
$t\bar{t}\gamma$ normalisation	0.6
ZZ normalisation	0.6
Other normalisations	0.5
$t\bar{t}Z$ normalisation	0.3
WZ normalisation	0.2
tZq normalisation	0.2
tHq normalisation	0.2
Modelling uncertainties	
$t\bar{t}W$ scale	1.8
$t\bar{t}W$ colour reconnection	1.0
ISR & FSR scale for $t\bar{t}W$	0.8
$t\bar{t}\gamma$ scale	0.4
VVV scale	0.3
$t\bar{t}H$ scale	0.2
Conversions	0.2
Simulation statistical uncertainty	1.8
Total systematic uncertainty	5.8

upward deviation of 30% in the post-fit value compared to the pre-fit value. This tension is also reported in the $t\bar{t}H$ measurement [30]. A comparison can be done with the $t\bar{t}H$ measurement [30] which also utilises the same-sign dilepton and trilepton final states. In the measurement using only the same-sign dilepton channel, the signal strength of $t\bar{t}H$ is $1.01^{+0.32}_{-0.29}$, whereas in the triplepton channel, the signal strength being reported at $1.53^{+0.43}_{-0.38}$. These two signal strengths are higher than the signal strength of $0.92^{+0.26}_{-0.23}$ measured using all channels combined, and they are compatible with the upward pull of 30% observed in this analysis. As a cross check, the fit for $t\bar{t}W$ signal strength is modified, now allowing the $t\bar{t}H$ cross section normalisation to vary freely. The result indicates a normalisation of $2.1^{+0.6}_{-0.5}$ for $t\bar{t}H$ and the signal strength of $t\bar{t}W$ is observed to be 7% lower, corresponding to a value of $1.36^{+0.12}_{-0.11}$. However, the absence of a control region for $t\bar{t}H$ means that allowing the normalisation of $t\bar{t}H$ to float in the fit leads to unreliable results.

Fig. 3.29 show the observed and predicted NN discriminant distributions for events in same-sign dilepton channel, categorised by the charge of the leptons, with positive charged leptons on the left and negative charged leptons on the right. The upper panel shows the distributions before the fit to the data and the lower panel shows the distributions after the fit. Since the NN incorporates the lepton charges as input variable, it is enabled to separately optimise the separation of signal and background for events with positive and negative lepton charges. As a result, the shapes of the NN discriminant distributions are different for the two classes of events.

In Figs. 3.30 and 3.31, the distributions of $m(3\ell)$ (invariant mass of the three leptons) are presented for events in the trilepton signal region with a total trilepton charge of $+1$ and -1, respectively. The upper panels in each figure display the distributions before the fit, while the lower panels show the distributions after the fit has been performed. On the left side of both Figs. 3.30 and 3.31, the distributions represent events with two jets, of which exactly two are medium b-tagged jets. On the right side of both figures, the distributions represent events with three jets, of which exactly two are medium b-tagged jets.

The impact of these uncertainties on the measurement can be observed in the reduction of uncertainties between the prefit and postfit values, as illustrated in Figs. 3.29–3.31. This reduction is primarily attributed to the categorisation of signal regions and the inclusion of control regions in the fit, which helps constrain certain sources of uncertainty.

Furthermore, the discrepancies between data and prediction seen in Fig. 3.6 are reduced as the fit results are implemented. Fig. 3.32 present the postfit version of Fig. 3.6.

In Fig. 3.29, 3.30, and 3.31, the yield of $t\bar{t}W$ is notably lower in channels with negative leptons ($\ell^-\ell^-$ and $\ell^\pm\ell^\pm\ell^-$). To evaluate the influence of these negative channels on the cross section, a fit excluding these channels was performed. The resulting cross section is $845^{+46}_{-45}\,(\text{stat})^{+54}_{-54}\,(\text{syst})\,\text{pb}$, indicating a decrease of 23 fb, or 2.6%, compared to the result presented in 3.21. Concurrently, there is an approximate increase of $\approx 1\%$ in the total uncertainty.

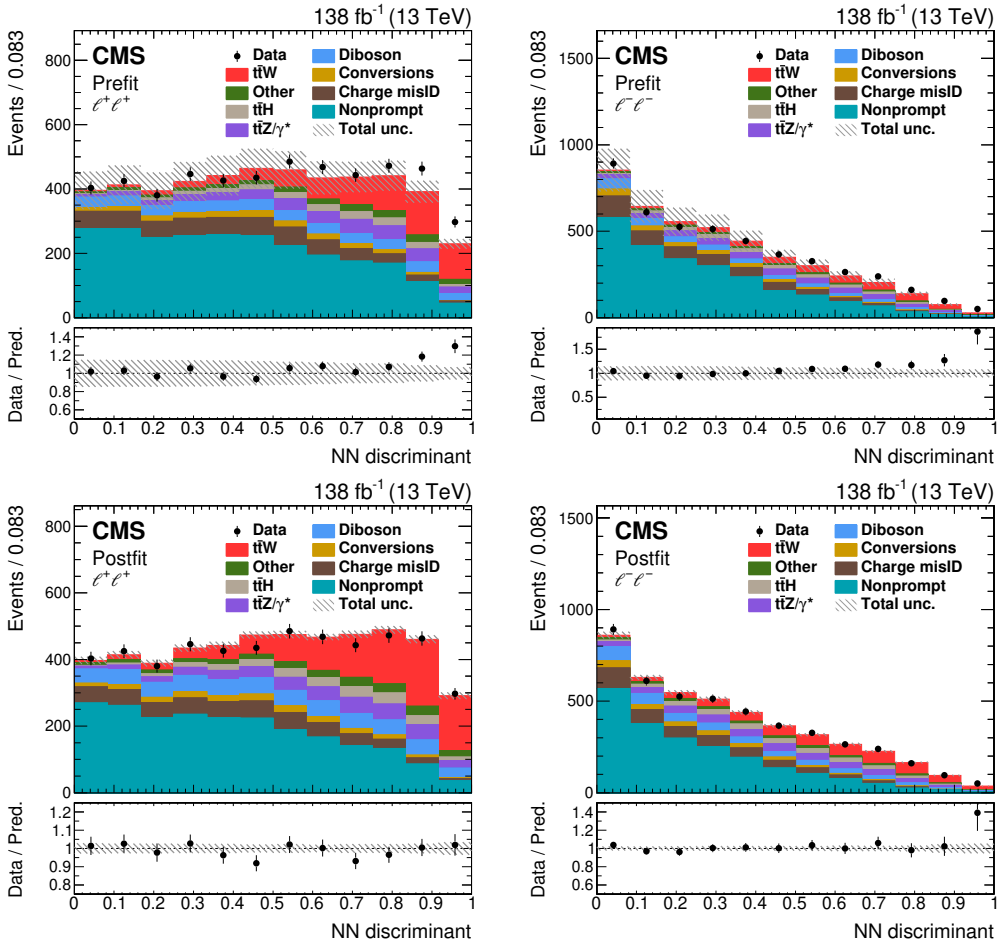

Figure 3.29: Comparison of the number of observed (points) and predicted (coloured histograms) events in the same-sign dilepton signal region. The distributions of the NN output score are displayed in final states with two leptons of positive (left) and negative (right) charges. The predictions are shown "prefit" as in Fig. 3.6 (upper row), and with the values of the normalisations and nuisance parameters obtained in the fit to the data applied ("postfit", lower row). The vertical bars on the points represent the statistical uncertainties in the data, and the hatched bands the systematic uncertainty in the predictions. In the lower panels, the ratio of the event yields in data to the overall sum of the predictions is presented.

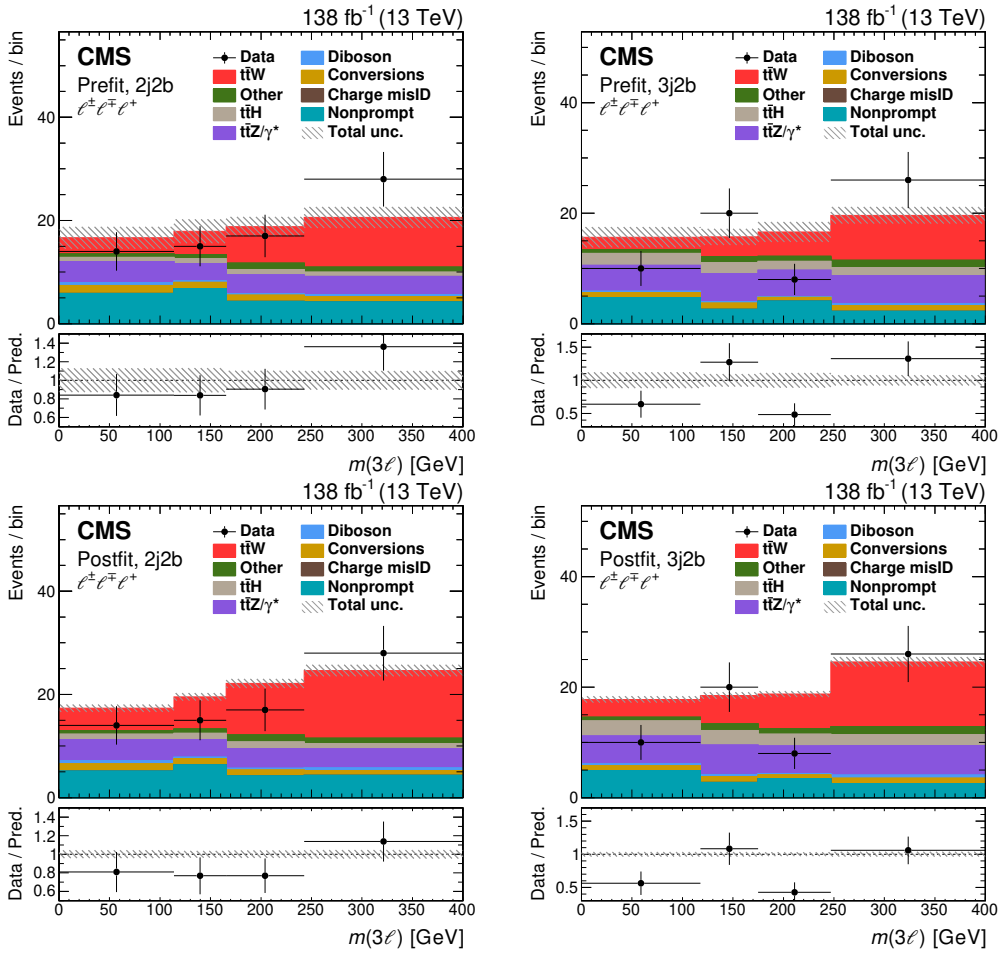

Figure 3.30: Comparison of the number of observed (points) and predicted (coloured histograms) events in the trilepton signal region with positive sum of lepton charges. The distributions of the three-lepton invariant mass are displayed in final states with two (left) and three (right) jets, of which exactly two pass the medium b tagging requirements. The predictions are shown "prefit" as in Fig. 3.6 (upper row), and with the values of the normalisations and nuisance parameters obtained in the fit to the data applied ("postfit", lower row). The vertical bars on the points represent the statistical uncertainties in the data, the horizontal bars the bin widths, and the hatched bands the systematic uncertainty in the predictions. The last bins include the overflow contributions. In the lower panels, the ratio of the event yields in data to the overall sum of the predictions is presented. Courtesy of Clara Ramón Álvarez and Carlos Vico Villalba, University of Oviedo.

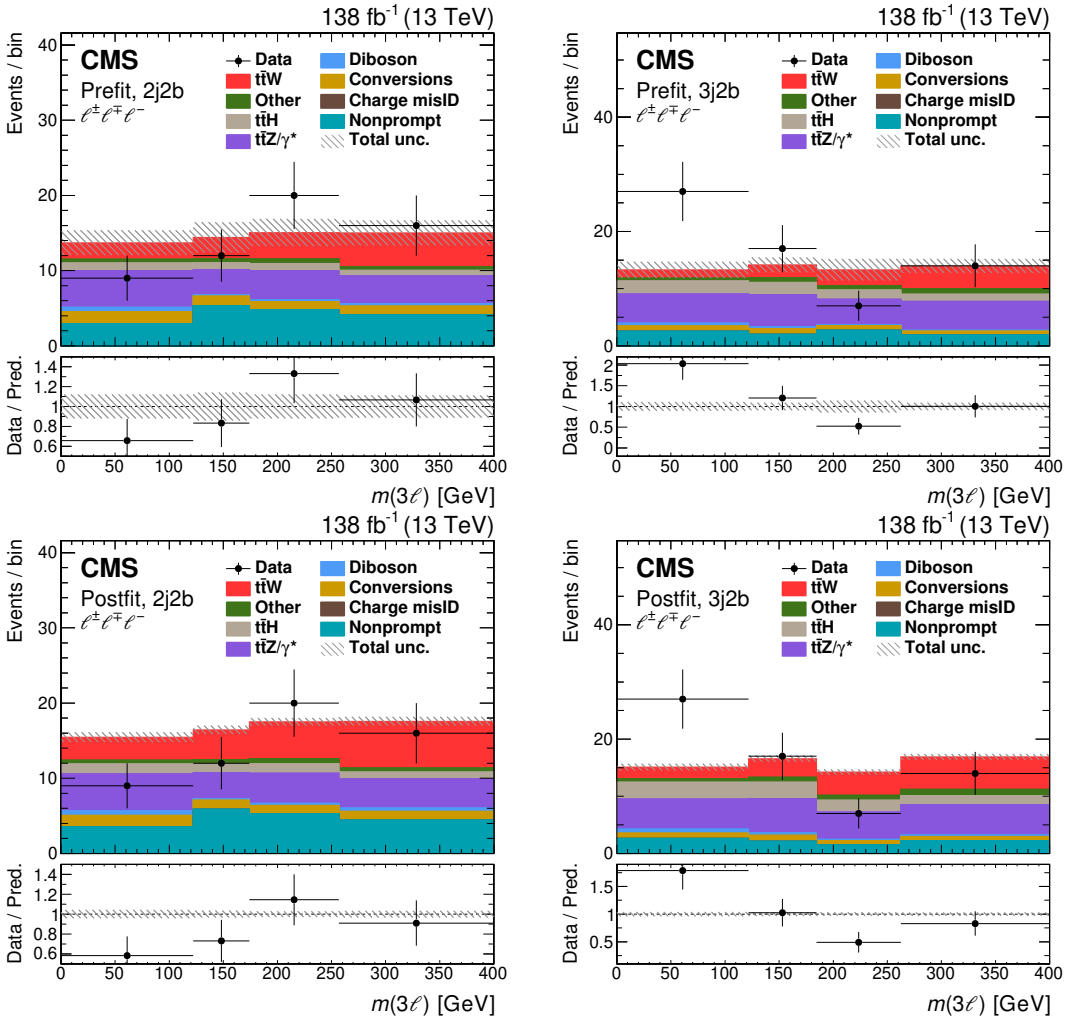

Figure 3.31: Comparison of the number of observed (points) and predicted (coloured histograms) events in the trilepton signal region with negative sum of lepton charges. The distributions of the three-lepton invariant mass are displayed in final states with two (left) and three (right) jets, of which exactly two pass the medium b tagging requirements. The predictions are shown "prefit" as in Fig. 3.6 (upper row), and with the values of the normalisations and nuisance parameters obtained in the fit to the data applied ("postfit", lower row). The vertical bars on the points represent the statistical uncertainties in the data, the horizontal bars the bin widths, and the hatched bands the systematic uncertainty in the predictions. The last bins include the overflow contributions. In the lower panels, the ratio of the event yields in data to the overall sum of the predictions is presented. Courtesy of Clara Ramón Álvarez and Carlos Vico Villalba, University of Oviedo.

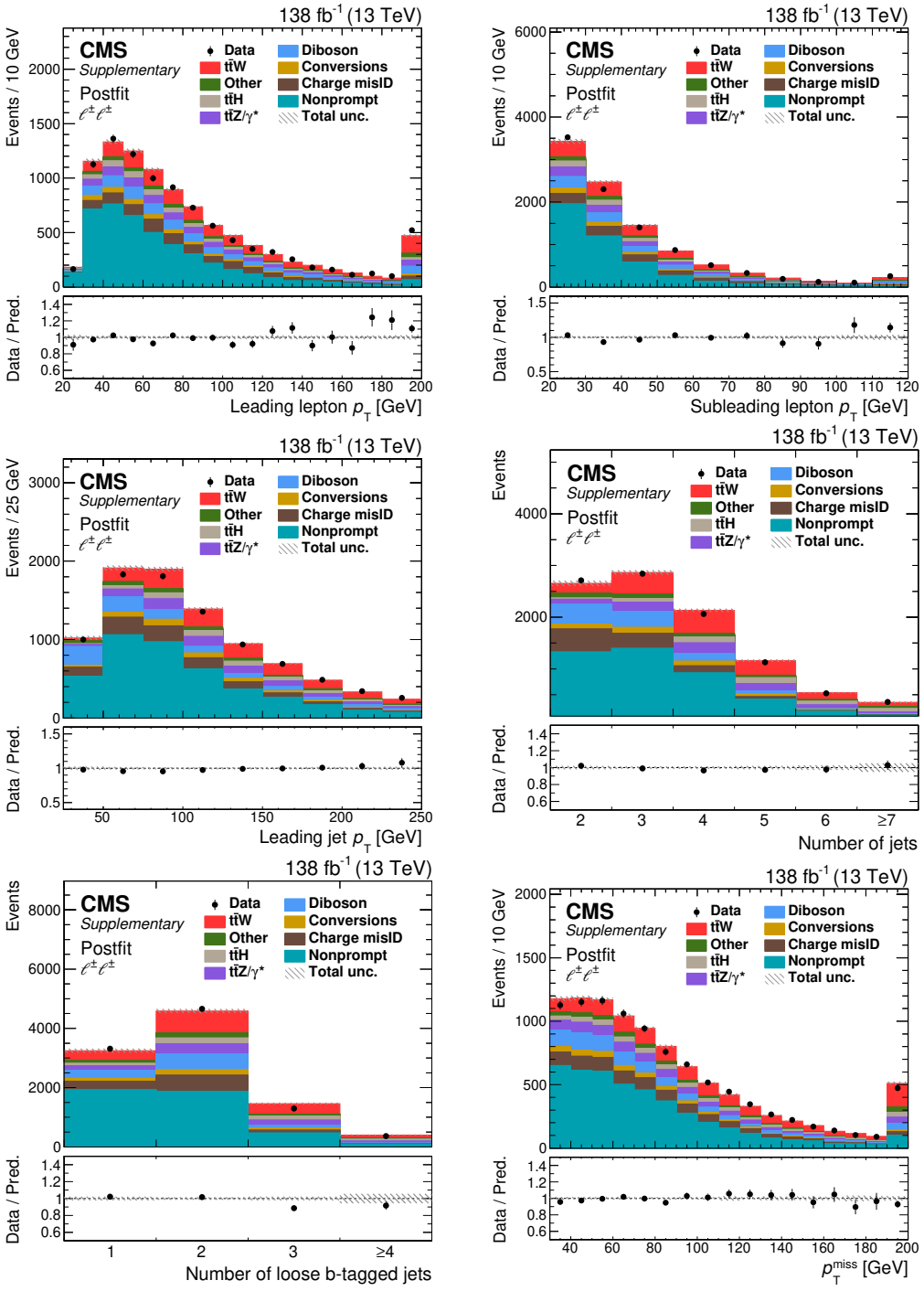

Figure 3.32: Comparison of observed (points) and predicted (coloured histograms) events in the same-sign dilepton signal region. Distributions are shown for the leading and subleading lepton p_{T}, the leading jet p_{T}, the number of jets and loose b-tagged jets, and $p_{\mathrm{T}}^{\mathrm{miss}}$. The predictions are shown with the values of the normalisations and nuisance parameters obtained in the fit to the data applied ("Postfit").

The fit robustness is validated by performing a fit to different subsets of the data, namely trilepton, dilepton channels as well as ee, eμ and μe, $\mu\mu$ channels separately, the fit is also performed on each data taking year. The comparison of the fits are shown in Fig. 3.33.

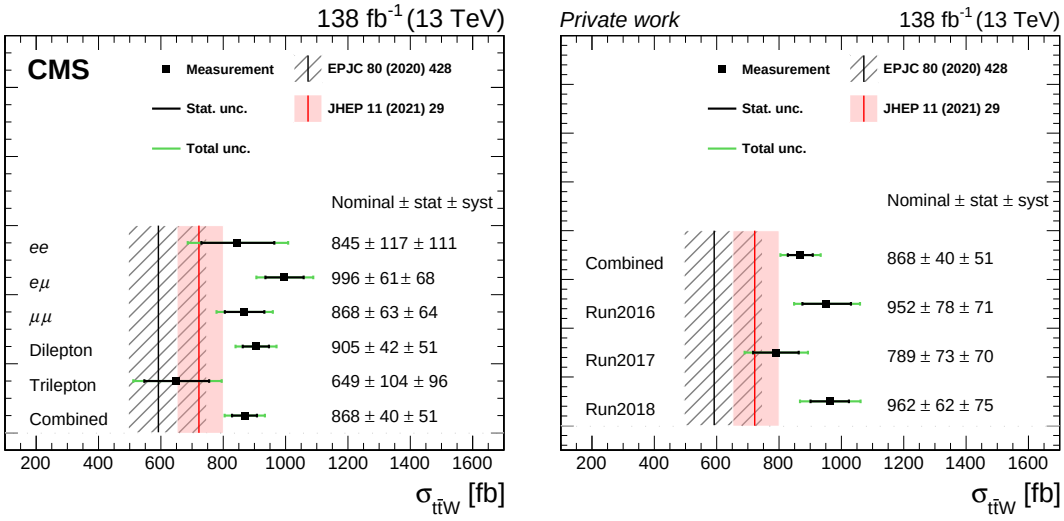

Figure 3.33: The measured cross sections in channels (left) and data-taking years (right) compared to the theoretical predictions by authors in of Refs. [34, 39].

Compared to previous measurements of $t\bar{t}W$ cross section from CMS [19], which was $0.77^{+0.12}_{-0.11}\,(\text{stat})^{+0.13}_{-0.12}\,(\text{syst})\,\text{pb}$, this measurement has archived a significant improvement in precision with a reduction of systematic uncertainty by a factor of two. At the time of writing this dissertation, this measurement is the most precise measurement of inclusive cross section of $t\bar{t}W$. The improvement is attributed to the improvement in lepton selection efficiency by using lepton MVA as well as a better understanding of background processes leading to a better estimation of backgrounds. The obtained results for the inclusive $t\bar{t}W$ cross section, along with their complete statistical and systematic uncertainties, are presented in Tab. 3.3. Furthermore, two SM predictions are provided, both in Tab. 3.3 and Fig. 3.33.

When comparing the measured cross section to the SM predictions, including the SM at NLO + NNLL calculation [34] and the SM prediction using NLO with an improved FxFx merging procedure [39], it is observed that the measured cross section is larger. However, it remains consistent within a two standard deviation range. Moreover, the constraints on $t\bar{t}W$ production obtained from the $t\bar{t}H$ and $t\bar{t}t\bar{t}$ studies, as referenced in Refs. [30, 31], are consistent with our measured cross section within one standard deviation. These findings further validate the under-prediction of the $t\bar{t}W$ cross section by the SM predictions.

In addition to the inclusive cross section of $t\bar{t}W$, the inclusive cross section of $t\bar{t}W^+$ and $t\bar{t}W^-$ are also measured separately. This measurement is performed by slightly modify

Table 3.3: Summary of measured and predicted production cross sections of $t\bar{t}W$, $t\bar{t}W^+$, and $t\bar{t}W^-$ production, as well as of the $\sigma_{t\bar{t}W^+}/\sigma_{t\bar{t}W^-}$ ratio. The SM predictions quoted at NLO + NNLL accuracy are taken from Refs. [33,34]. The SM predictions quoted at NLO accuracy and including corrections from an improved FxFx merging procedure (NLO + FxFx) have been provided by the authors of Ref. [39]. The theoretical uncertainties include scale variations and PDF uncertainties.

Observable	Measurement	SM prediction	
		NLO+NNLL	NLO+FxFx
$\sigma_{t\bar{t}W}$	$868 \pm 40\,(\text{stat}) \pm 51\,(\text{syst})$ fb	$592\,^{+155}_{-97}\,(\text{theo})$ fb	$722\,^{+71}_{-78}\,(\text{theo})$fb
$\sigma_{t\bar{t}W^+}$	$553 \pm 30\,(\text{stat}) \pm 30\,(\text{syst})$ fb	$384\,^{+53}_{-33}\,(\text{theo})$fb	$475\,^{+46}_{-52}\,(\text{theo})$fb
$\sigma_{t\bar{t}W^-}$	$343 \pm 26\,(\text{stat}) \pm 25\,(\text{syst})$ fb	$198\,^{+26}_{-17}\,(\text{theo})$fb	$247\,^{+24}_{-27}\,(\text{theo})$fb
$\sigma_{t\bar{t}W^+}/\sigma_{t\bar{t}W^-}$	$1.61 \pm 0.15\,(\text{stat})\,^{+0.07}_{-0.05}\,(\text{syst})$	$1.94\,^{+0.37}_{-0.24}\,(\text{theo})$	$1.92\,^{+0.27}_{-0.29}\,(\text{theo})$

the fit model, in which now the main parameter of interest, the signal strength of $t\bar{t}W$, is replaced by two parameters: the signal strength of $t\bar{t}W^+$ and $t\bar{t}W^-$. This procedure is very straight forward given that the data are already categorised by the sum of lepton charges, which is also corresponding to the charge of the W boson. The fit model is then modified further to extract the ratio of $t\bar{t}W^+$ cross section to $t\bar{t}W^-$ cross section, $\sigma_{t\bar{t}W^+}/\sigma_{t\bar{t}W^-}$ and it is performed independently to the fit for the cross sections. In this model, the parameter of interest is the "signal strength" of the ratio, which shows how much the observation deviates from theory.

The prediction for $t\bar{t}W$ is normalised to the cross section provided by the authors in Ref. [34]. However, the individual cross sections of $t\bar{t}W^+$ and $t\bar{t}W^-$ are not provided in the same study. To address this, the work of the authors in Ref. [33] is taken into consideration for the cross sections of $t\bar{t}W^+$ and $t\bar{t}W^-$. These cross sections are then scaled to align with the prediction from Ref. [34].

The measured values of the $t\bar{t}W^+$ and $t\bar{t}W^-$ cross sections are $553\pm30\,(\text{stat})\pm30\,(\text{syst})$ fb and $343\pm26\,(\text{stat})\pm25\,(\text{syst})$ fb, respectively. Both of the cross section of $t\bar{t}W^+$ and $t\bar{t}W^-$ are measured to be higher than the reference theoretical prediction at about 44% higher for $t\bar{t}W^+$ and 70% for $t\bar{t}W^-$. As shown in Fig. 3.34, there is some correlation between the two measured cross sections, which is originated from the use of the same control regions and common nuisance parameters, e.g. uncertainties on theoretical prediction of the signal and backgrounds processes. They are consistent with the measurement of the inclusive cross section of $t\bar{t}W$ within statistical uncertainties.

While the cross sections are higher than predictions, the ratio $\sigma_{t\bar{t}W^+}/\sigma_{t\bar{t}W^-}$ is measured to be lower than prediction at $1.61 \pm 0.15\,(\text{stat})\,^{+0.07}_{-0.05}\,(\text{syst})$, compared to $1.92^{+0.27}_{-0.29}$ predicted by the author of Ref. [39]. A scan of the negative log-likelihood function with the best fit value, 68 and 95% confidence levels, and theoretical prediction value are shown

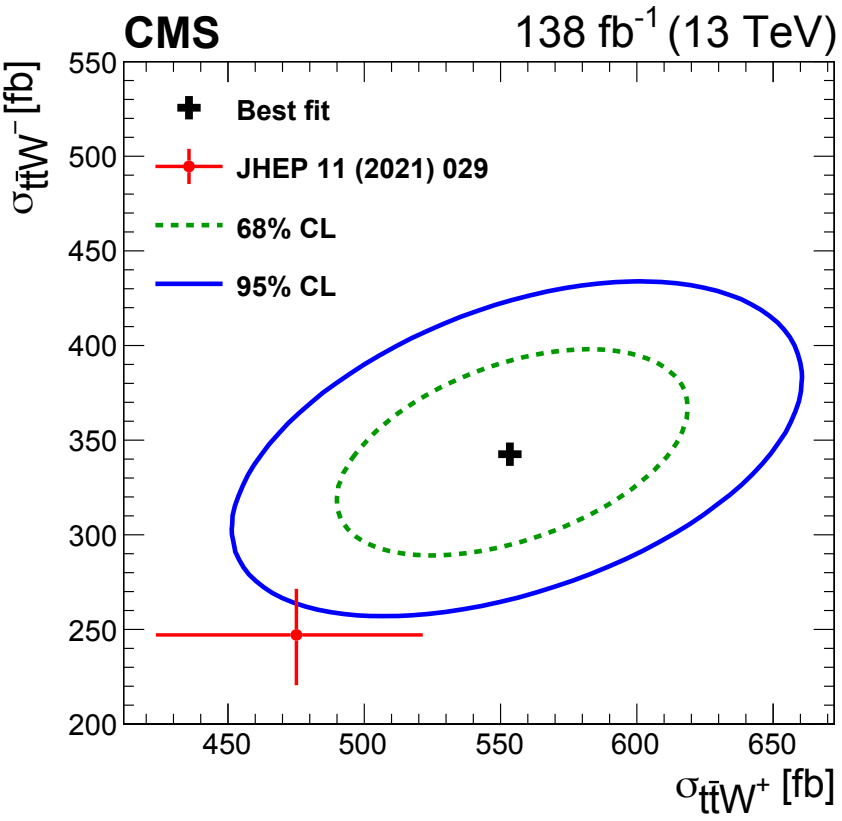

Figure 3.34: The measurement of the $t\bar{t}W^+$ vs. $t\bar{t}W^-$ cross sections (black cross), along with the 68 (green dashed) and 95% (blue solid) CL intervals. The SM prediction provided by the authors of Ref. [39] is shown by the red dot, with the horizontal and vertical bars corresponding to the total uncertainties.

in Fig. 3.35. Compared to the cross section measurements, the ratio has much lower systematic uncertainty of about 3%. This reduction in systematic uncertainty is mainly due to the correlation between some uncertainty sources get cancelled out in the ratio. Due to the unavailability of the theoretical prediction of the ratio, the ratio is computed from the cross sections of $t\bar{t}W^+$ and $t\bar{t}W^-$ with the assumption that the two cross sections are completely independent, hence, it results in a large theoretical uncertainty on the ratio.

The ratio $\sigma_{t\bar{t}W^+}/\sigma_{t\bar{t}W^-}$ is particularly interesting as it can reveal the effects of spin-correlation [37], which influences the value of the ratio. These effects can be investigated by measuring the ratio $\sigma_{t\bar{t}W^+}/\sigma_{t\bar{t}W^-}$ in different final states.

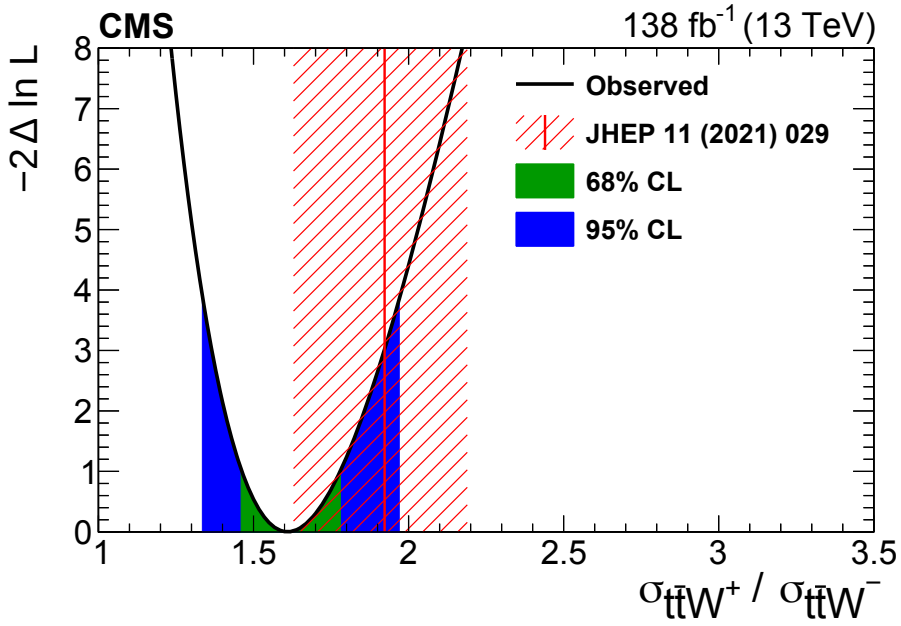

Figure 3.35: Likelihood scan for the ratio of the $t\bar{t}W^+$ and $t\bar{t}W^-$ signal strengths, using all signal and control regions and the full data sample.

Chapter 4

Conclusion and outlook

The considerable volume of data collected by the CMS detector during the second run of the LHC enable the $t\bar{t}W$ process to be measured with utmost precision. This dissertation presents the most accurate measurement to date, boasting a statistical uncertainty of only 4.6% and a systematic uncertainty of 5.8%. The statistical uncertainty achieved in this study is more than three times lower the CMS' last result, presented in Ref. [19]. This dramatic improvement stems from a near quadruple increase in integrated luminosity, along with enhancements in lepton and event selections. The systematic uncertainty has improved by a factor of 2.8 relative to the same result from the CMS [19]. This progress is thanks to a deeper understanding of background processes, achieved through meticulous studies, tighter constraints on background processes, more precise experimental results, and refined theoretical models. Between these two analyses, significant efforts have been invested by the CMS to improve the B-tagging algorithm, the measurement of integrated luminosity, pileup modelling, and the reconstructions of jets and leptons—all of which significantly impact systematic uncertainty. Furthermore, the employment of the multiclass NN vastly improves the separation between the signal and background processes. With an unprecedented precision, this result provides a stringent constraint of the $t\bar{t}W$ as a background in the measurements of $t\bar{t}H$ and $t\bar{t}t\bar{t}$. This is proven to be the case for $t\bar{t}t\bar{t}$, with its existence confirmed by the CMS [128].

The result presented in this dissertation aligns well with recent measurements, where the $t\bar{t}W$ cross section is assessed as a background process [30, 31]. Additionally, a more recent finding from the ATLAS collaboration [129] measures the inclusive cross section of the $t\bar{t}W$ process in the same final states as this result. Ref. [129], using the complete data from Run II, reports an inclusive cross section of $890 \pm 50\,(\text{stat}) \pm 70\,(\text{syst})$. This ATLAS result is in good agreement with the measurement presented in the dissertation, further validating this finding.

In recent years, the theoretical community has invested significant effort into refining the modelling of the $t\bar{t}W$ process, from the incorporation of QCD and EWK corrections to the inclusion of multi-jet merging. Recently, a calculation of $t\bar{t}W$ has been conducted

at NNLO in QCD [130]. Hence, simulations need to be updated to align with the most current theoretical precision, necessitating an updated measurement to bolster confidence in these predictions. This enhancement could potentially reconcile the discrepancy observed between the predicted and actual $t\bar{t}W$ cross section.

Future analyses should consider conducting the measurements of $\sigma_{t\bar{t}W^+}/\sigma_{t\bar{t}W^-}$ in different phase spaces to study the effect of spin-correlations. The effect of spin-correlation on the ratio varies depending on the choice of phase space. In same-sign dilepton channel, the $\bar{t}$-quark tends to decay hadronically more often than the t-quark, this is because $t\bar{t}W^+$ having higher cross section and the lepton that comes from the t-quark is more central due to spin-correlation. This results in a stronger effect on the ratio in same-sign dilepton channel than in trilepton channels [37]. Furthermore, at the LHC, the production of $t\bar{t}W^+$ ($t\bar{t}W^-$) is dominated by the scattering of u-$\bar{d}$ (d-$\bar{u}$). Hence, $\sigma_{t\bar{t}W^+}/\sigma_{t\bar{t}W^-}$ measurements can also be an alternative approach to measurements of pdfs, specifically, the ratio of the parton distribution of the u-quark to that of the d-quark.

Conversely, with the full Run II data set, differential measurement of $t\bar{t}W$ has now become feasible. The first differential measurement has been conducted and reported by the ATLAS collaboration, as reported in the same Ref. [129]. Although this report indicates a discrepancy in the absolute cross section measurements concerning overall normalisation, it is consistent with the inclusive measurement. In the short term, a differential measurement should be conducted by the CMS to offer a complementary perspective to the ATLAS result.

ATLAS has also reported on a measurement of the leptonic charge asymmetry in $t\bar{t}W$ [131], in which the charge asymmetry is measured based on pseudorapidities of the leptons from the top-quark and top-antiquark decays. The measured asymmetry variable, $A_c^l(t\bar{t}W) = -0.123 \pm 0.136\,(\text{stat}) \pm 0.051\,(\text{syst})$, concurs with the SM expectation. However, it does bear the significant drawback of insufficient statistics. With the LHC Run III projected to yield 300 fb^{-1} of data, the statistical robustness of the asymmetry variables measurements in $t\bar{t}W$ is set to improve, yielding more dependable results.

The high energy physics community has been constantly developing sophisticated analysis techniques as the LHC embarks on its third data collection phase and prepares for the HL-LHC. The HL-LHC will be equipped with superior detectors, triggers, and a greater acceptance coverage. This era is anticipated to yield an abundance of insightful results as well as challenges, potentially bringing us closer to unveiling the true nature of our universe.

FR Fake-rate

charge misID charge misidentification

OSSF opposite-sign same-flavour

BDT Boosted Decision Tree

NN Neural-Network

DNN Deep Neural-Network

MVA Multivariate Analysis

SM Standard Model

BSM Beyond Standard Model

SLAC Stanford Linear Accelerator Center

D$\emptyset$ DZero

CDF Collider Detector at Fermilab

ROC Receiver Operating Characteristic

AUC Area Under the Curve

CHS Charged-Hadron Subtraction

LS1 Long Shutdown 1

LS2 Long Shutdown 2

RMS Root Mean Squared

GUT Grand Unified Theory

SMEFT SM Effective Field Theory

SUSY Supersymmetry

MC Monte-Carlo

LHC Large Hadron Collider

HL-LHC High-Luminosity LHC

CERN the European Organisation for Nuclear Research

PV Primary Vertex

SV Secondary Vertex

HLT High-Level Trigger

CMS Compact Muon Solenoid

ATLAS A Toroidal LHC Apparatus

ALICE A Large Ion Collider Experiment

LHCb LHC Beauty

LHCf LHC-forward

TOTEM Total, elastic and diffractive cross-section measurement

SND@LHC Scattering and Neutrino Detector at the LHC

FASER Forward Search Experiment

MoEDAL Monopole and Exotics Detector at the LHC

LINAC 2 Linear Accelerator 2

PSB Proton Synchrotron Booster

SP Proton Synchrotron

SPS Super Proton Synchrotron

EWK Electroweak

HF Forward Hadron Calorimeter

HB Barrel Hadron Calorimeter

HE End-cap Hadron Calorimeter

HO Outer Hadron Calorimeter

ECAL Electromagnetic Calorimeter

HCAL Hadron Calorimeter

PF ParticleFlow

DT Drift Tube

CSC Cathode Strip Chamber

RPC Resistive Plate Chamber

GEM Gas Electron Multiplier

TIB Tracker Inner Barrel

TID Tracker Inner Disk

TOB Tracker Outer Barrel

TEC Tracker End-cap

APD Avalanche Photodiode

VPT Vacuum Photo-triode

TDC Time-to-Digital-Converter

L1 Level-1 trigger

TPG Trigger Primitive Generator

CTF Combinatorial Track Finder

KF Kalman Filter

GSF Gaussian-Sum filter

RF Radio-Frequency

JEC Jet Energy Scale

JER Jet Energy Resolution

LO Leading Order

NLO Next-to-Leading Order

NNLO Next-to-Next-to-Leading Order

NNLL Next-to-Next-to-Leading Logarithmic

pdf Parton Distribution Function

ISR Initial State Radiation

FSR Final State Radiation

PS Parton Shower

ME Matrix Element

QFT Quantum Field Theory

QCD Quantum Chromodynamics

QED Quantum Electrodynamics

SSB Spontaneous Symmetry Breaking

BEH Brout-Englert-Higgs

VEV Vacuum Expectation Value

DY Drell-Yan

NP New Physics

Bibliography

[1] A. Purcell, "Go on a particle quest at the first CERN webfest. Le premier webfest du CERN se lance à la conquête des particules", BUL-NA-2012-269. https://cds.cern.ch/record/1473657.

[2] CMS Collaboration, "Measurement of the top quark mass using events with a single reconstructed top quark in pp collisions at $\sqrt{s} = 13$ TeV", *JHEP* **12** (2021) 161, doi:10.1007/JHEP12(2021)161, arXiv:2108.10407.

[3] Particle Data Group, "Review of Particle Physics", *PTEP* **2022** (2022) 083C01, doi:10.1093/ptep/ptac097.

[4] M. Thomson, "Modern Particle Physics". Cambridge University Press, 2013. doi:10.1017/CBO9781139525367, ISBN 9781107034266.

[5] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys. Lett. B* **716** (2012) 30–61 CMS-HIG-12-028, CERN-PH-EP-2012-220, CMS-HIG-12-028, CERN-PH-EP-2012-220, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

[6] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett. B* **716** (2012) 1–29 CERN-PH-EP-2012-218, CERN-PH-EP-2012-218, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.

[7] ATLAS Collaboration, "Combined measurement of the Higgs boson mass from the $H \to \gamma\gamma$ and $H \to ZZ^* \to 4\ell$ decay channels with the ATLAS detector using $\sqrt{s} = 7$, 8 and 13 TeV pp collision data", arXiv:2308.04775.

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

[9] P. W. Higgs, "Broken symmetries, massless particles and gauge fields", *Phys. Lett.* **12** (1964) 132, doi:10.1016/0031-9163(64)91136-9.

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

[11] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global conservation laws and massless particles", *Phys. Rev. Lett.* **13** (1964) 585, doi:10.1103/PhysRevLett.13.585.

[12] P. W. Higgs, "Spontaneous symmetry breakdown without massless bosons", *Phys. Rev.* **145** (1966) 1156, doi:10.1103/PhysRev.145.1156.

[13] T. W. B. Kibble, "Symmetry breaking in non-Abelian gauge theories", *Phys. Rev.* **155** (1967) 1554, doi:10.1103/PhysRev.155.1554.

[14] F. Maltoni, L. Mantani, and K. Mimasu, "Top-quark electroweak interactions at high energy", *JHEP* **2019** (oct, 2019) doi:10.1007/jhep10(2019)004.

[15] N. P. Hartland et al., "A monte carlo global analysis of the standard model effective field theory: the top quark sector", *JHEP* **2019** (apr, 2019) doi:10.1007/jhep04(2019)100.

[16] I. Brivio et al., "O new physics, where art thou? a global search in the top sector", *JHEP* **2020** (feb, 2020) doi:10.1007/jhep02(2020)131.

[17] J. Ellis et al., "Top, higgs, diboson and electroweak fit to the standard model effective field theory", *JHEP* **2021** (apr, 2021) doi:10.1007/jhep04(2021)279.

[18] V. Miralles et al., "The top quark electro-weak couplings after LHC run 2", *JHEP* **2022** (feb, 2022) doi:10.1007/jhep02(2022)032.

[19] CMS Collaboration, "Measurement of the cross section for top quark pair production in association with a W or Z boson in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *JHEP* **08** (2018) 011, doi:10.1007/JHEP08(2018)011, arXiv:1711.02547.

[20] CMS Collaboration, "Probing effective field theory operators in the associated production of top quarks with a Z boson in multilepton final states at $\sqrt{s} = 13$ TeV", *JHEP* **2112** (2021) 083, doi:10.1007/JHEP12(2021)083, arXiv:2107.13896.

[21] CMS Collaboration, "Measurement of the inclusive and differential $\text{t}\bar{\text{t}}\gamma$ cross sections in the dilepton channel and effective field theory interpretation in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP* **2205** (2022) 091, doi:10.1007/JHEP05(2022)091, arXiv:2201.07301.

[22] CMS Collaboration, "Measurement of the inclusive and differential $\text{t}\bar{\text{t}}\gamma$ cross sections in the single-lepton channel and EFT interpretation at $\sqrt{s} = 13$ TeV", *JHEP* **2112** (2021) 180, doi:10.1007/JHEP12(2021)180, arXiv:2107.01508.

[23] ATLAS Collaboration, "Direct top quark mass measurements with the ATLAS and CMS detectors", arXiv:2211.11398. Proceedings for a talk at the 15th International Workshop on Top Quark Physics, Durham, UK, 4-9 September 2022.

[24] F. Maltoni, M. Mangano, I. Tsinikos, and M. Zaro, "Top-quark charge asymmetry and polarization in $t\bar{t}W^{\pm}$ production at the LHC", *Phys. Lett. B* **736** (2014) 252, doi:10.1016/j.physletb.2014.07.033, arXiv:1406.3262.

[25] CMS Collaboration, "Measurement of associated production of vector bosons and $t\bar{t}$ at $\sqrt{s}$= 7 TeV", *Phys. Rev. Lett.* **110** (2013) 172002, doi:10.1103/PhysRevLett.110.172002, arXiv:1303.3239.

[26] ATLAS Collaboration, "Measurement of the $t\bar{t}W$ and $t\bar{t}Z$ production cross sections in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector", *JHEP* **11** (2015) 172, doi:10.1007/JHEP11(2015)172, arXiv:1509.05276.

[27] CMS Collaboration, "Observation of top quark pairs produced in association with a vector boson in pp collisions at $\sqrt{s} = 8$ TeV", *JHEP* **01** (2016) 096, doi:10.1007/JHEP01(2016)096, arXiv:1510.01131.

[28] ATLAS Collaboration, "Measurement of the $t\bar{t}Z$ and $t\bar{t}W$ production cross sections in multilepton final states using 3.2 fb^{-1} of pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Eur. Phys. J. C* **77** (2017) 40, doi:10.1140/epjc/s10052-016-4574-y, arXiv:1609.01599.

[29] ATLAS Collaboration, "Measurement of the $t\bar{t}Z$ and $t\bar{t}W$ cross sections in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Phys. Rev. D* **99** (2019) 072009, doi:10.1103/PhysRevD.99.072009, arXiv:1901.03584.

[30] CMS Collaboration, "Measurement of the Higgs boson production rate in association with top quarks in final states with electrons, muons, and hadronically decaying tau leptons at $\sqrt{s} = 13$ TeV", *Eur. Phys. J. C* **81** (2021) 378, doi:10.1140/epjc/s10052-021-09014-x, arXiv:2011.03652.

[31] CMS Collaboration, "Search for production of four top quarks in final states with same-sign or multiple leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV", *Eur. Phys. J. C* **80** (2020) 75, doi:10.1140/epjc/s10052-019-7593-7, arXiv:1908.06463.

[32] ATLAS Collaboration, "Evidence for $t\bar{t}t\bar{t}$ production in the multilepton final state in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Eur. Phys. J. C* **80** (2020) 1085, doi:10.1140/epjc/s10052-020-08509-3, arXiv:2007.14858.

[33] A. Broggio et al., "Top-quark pair hadroproduction in association with a heavy boson at NLO+NNLL including EW corrections", *JHEP* **08** (2019) 039, doi:10.1007/JHEP08(2019)039, arXiv:1907.04343.

[34] A. Kulesza et al., "Associated top quark pair production with a heavy boson: differential cross sections at NLO+NNLL accuracy", *Eur. Phys. J. C* **80** (2020) 428, doi:10.1140/epjc/s10052-020-7987-6, arXiv:2001.03031.

[35] G. Bevilacqua et al., "NLO QCD corrections to off-shell $t\bar{t}W$ production at the LHC: correlations and asymmetries", *Eur. Phys. J. C* **81** (2021) 675, doi:10.1140/epjc/s10052-021-09478-x, arXiv:2012.01363.

[36] S. von Buddenbrock, R. Ruiz, and B. Mellado, "Anatomy of inclusive $t\bar{t}W$ production at hadron colliders", *Phys. Lett. B* **811** (2020) 135964, doi:10.1016/j.physletb.2020.135964, arXiv:2009.00032.

[37] R. Frederix and I. Tsinikos, "Subleading EW corrections and spin-correlation effects in $t\bar{t}W$ multi-lepton signatures", *Eur. Phys. J. C* **80** (2020) 803, doi:10.1140/epjc/s10052-020-8388-6, arXiv:2004.09552.

[38] F. Febres Cordero, M. Kraus, and L. Reina, "Top-quark pair production in association with a $W^{\pm}$ gauge boson in the POWHEG−BOX", *Phys. Rev. D* **103** (2021) 094014, doi:10.1103/PhysRevD.103.094014, arXiv:2101.11808.

[39] R. Frederix and I. Tsinikos, "On improving NLO merging for $t\bar{t}W$ production", *JHEP* **11** (2021) 29, doi:10.1007/JHEP11(2021)029, arXiv:2108.07826.

[40] G. Bevilacqua et al., "Modeling uncertainties of $t\bar{t}W^{\pm}$ multilepton signatures", *Phys. Rev. D* **105** (2022) 014018, doi:10.1103/PhysRevD.105.014018, arXiv:2109.15181.

[41] "Parton shower effects in $t\bar{t}W^{\pm}$ production at NLO QCD", The European Physical Society Conference on High Energy Physics (EPS-HEP2021). (2021). arXiv:2110.12512.

[42] R. Frederix, D. Pagani, and M. Zaro, "Large NLO corrections in $t\bar{t}W^{\pm}$ and $t\bar{t}t\bar{t}$ hadroproduction from supposedly subleading EW contributions", *JHEP* **02** (2018) 031, doi:10.1007/JHEP02(2018)031, arXiv:1711.02116.

[43] DO Collaboration, "Forward-backward asymmetry in top quark-antiquark production", *Physical Review D* **84** (dec, 2011) doi:10.1103/physrevd.84.112005.

[44] DO Collaboration, "Measurement of the forward-backward asymmetry in top quark-antiquark production in lepton+jets channel", *Physical Review D* **90** (oct, 2014) doi:10.1103/physrevd.90.072011.

[45] CDF Collaboration, "Evidence for a mass dependent forward-backward asymmetry in top quark pair production", *Physical Review D* **83** (jun, 2011) doi:10.1103/physrevd.83.112003.

[46] CDF Collaboration, "Measurement of the top quark forward-backward production asymmetry and its dependence on event kinematic properties", *Physical Review D* **87** (may, 2013) doi:10.1103/physrevd.87.092002.

[47] CDF Collaboration, "Measurement of the inclusive leptonic asymmetry in top-quark pairs that decay to two charged leptons at CDF", *Physical Review Letters* **113** (jul, 2014) doi:10.1103/physrevlett.113.042001.

[48] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004. Also published by CERN Geneva in 2010.

[49] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) S08003, doi:10.1088/1748-0221/3/08/S08003. Also published by CERN Geneva in 2010.

[50] ALICE Collaboration, "The ALICE experiment at the CERN LHC. A Large Ion Collider Experiment", *JINST* **3** (2008) S08002, doi:10.1088/1748-0221/3/08/S08002. Also published by CERN Geneva in 2010.

[51] LHCb Collaboration, "The LHCb Detector at the LHC", *JINST* **3** (2008) S08005 LHCb-DP-2008-001, doi:10.1088/1748-0221/3/08/S08005. Also published by CERN Geneva in 2010.

[52] FASER Collaboration, "The FASER Detector", CERN-FASER-2022-001, arXiv:2207.11427. https://cds.cern.ch/record/2816452.

[53] LHCf Collaboration, "LHCf experiment: Technical Design Report", CERN-LHCC-2006-004, LHCF-TDR-001. https://cds.cern.ch/record/926196.

[54] MoEDAL Collaboration, "Technical Design Report of the MoEDAL Experiment", CERN-LHCC-2009-006. https://cds.cern.ch/record/1181486.

[55] TOTEM Collaboration, "Total cross-section, elastic scattering and diffraction dissociation at the Large Hadron Collider at CERN: TOTEM Technical Design Report", CERN-LHCC-2004-002, TOTEM-TDR-001. https://cds.cern.ch/record/704349.

[56] SND@LHC Collaboration, "SND@LHC: The Scattering and Neutrino Detector at the LHC", CERN-EP-2022-203. https://cds.cern.ch/record/2834502.

[57] E. Mobs, "The CERN accelerator complex - August 2018. Complexe des accélérateurs du CERN - Août 2018",. General Photo. https://cds.cern.ch/record/2636343.

[58] CMS Collaboration, "Precise mapping of the magnetic field in the CMS barrel yoke using cosmic rays", *JINST* **5** (2010) T03021 CMS-CFT-09-015, doi:10.1088/1748-0221/5/03/T03021, arXiv:0910.5530.

[59] A. Dominguez et al., "CMS Technical Design Report for the Pixel Detector Upgrade", CERN-LHCC-2012-016, CMS-TDR-11. https://cds.cern.ch/record/1481838.

[60] CMS Collaboration, "Avalanche Photodiodes and Vacuum Phototriodes for the Electromagnetic Calorimeter of the CMS experiment at the Large Hadron Collider", CMS-CR-2008-083. https://cds.cern.ch/record/1143414.

[61] CMS Collaboration, "Overview of the CMS electromagnetic calorimeter", CMS-CR-1999-006, LYCEN-98-106, doi:10.1016/S0920-5632(99)00544-7.

[62] CMS Collaboration, "Calibration of the CMS hadron calorimeters using proton-proton collision data at $\sqrt{s} = 13$ TeV", *JINST* **15** (2020) P05002 CMS-PRF-18-001, CERN-EP-2019-179, CMS-PRF-18-001-003, doi:10.1088/1748-0221/15/05/P05002, arXiv:1910.00079. https://cds.cern.ch/record/2691403.

[63] CMS Collaboration, "ECAL 2016 refined calibration and Run2 summary plots",.

[64] CMS Collaboration, "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV", *JINST* **13** (2018), no. 06, P06015 CMS-MUO-16-001, CERN-EP-2018-058, CMS-MUO-16-001-003, doi:10.1088/1748-0221/13/06/P06015, arXiv:1804.04528.

[65] S. Cucciarelli, M. Konecki, D. Kotlinski, and T. Todorov, "Track reconstruction, primary vertex finding and seed generation with the Pixel Detector", CMS-NOTE-2006-026. https://cds.cern.ch/record/927384.

[66] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST* **9** (2014) P10009 CMS-TRK-11-001, CERN-PH-EP-2014-070, CERN-PH-EP-2014-070, CMS-TRK-11-001, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.

[67] P. Billoir and S.-J. Qian, "Simultaneous pattern recognition and track fitting by the kalman filtering method", *Nucl. Instrum. Methods. Phys. Res. B* **294** (1990), no. 1, 219–228, doi:10.1016/0168-9002(90)91835-Y.

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

[69] R. Frühwirth and A. Strandlie, "Secondary Vertex Reconstruction". Springer International Publishing, 2020. doi:10.1007/978-3-030-65771-0_9.

[70] CMS Collaboration, "Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC", *JINST* **16** (2021), no. 05, P05014 CMS-EGM-17-001, CERN-EP-2020-219, CMS-EGM-17-001-003, doi:10.1088/1748-0221/16/05/P05014, arXiv:2012.06888.

[71] W. Adam, R. Frühwirth, A. Strandlie, and T. Todor, "Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC", CMS-NOTE-2005-001. https://cds.cern.ch/record/815410.

[72] CMS Collaboration, "Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s}$ =8 TeV", *JINST* **10** (2015) P08010, doi:10.1088/1748-0221/10/08/P08010, arXiv:1502.02702.

[73] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm", *JHEP* **2008** (apr, 2008) 063–063, doi:10.1088/1126-6708/2008/04/063.

[74] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet user manual", *The European Physical Journal C* **72** (mar, 2012) doi:10.1140/epjc/s10052-012-1896-2.

[75] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector. Particle-flow reconstruction and global event description with the CMS detector", *JINST* **12** (2017), no. 10, P10003 CMS-PRF-14-001, CERN-EP-2017-110, CMS-PRF-14-001-004, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.

[76] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV", *JINST* **13** (2018), no. 05, P05011 CMS-BTV-16-002, CERN-EP-2017-326, CMS-BTV-16-002-004, doi:10.1088/1748-0221/13/05/P05011, arXiv:1712.07158.

[77] E. Bols et al., "Jet flavour classification using DeepJet", *JINST* **15** (dec, 2020) P12012–P12012, doi:10.1088/1748-0221/15/12/p12012.

[78] CMS Collaboration, "Performance of the DeepJet b tagging algorithm using 41.9/fb of data from proton-proton collisions at 13 TeV with Phase 1 CMS detector",. https://cds.cern.ch/record/2646773.

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

[80] CMS Collaboration, "Measurement of the cross section of top quark-antiquark pair production in association with a W boson in proton-proton collisions at $\sqrt{s}$ = 13 TeV", *JHEP* **2307** (2023) 219, doi:10.1007/JHEP07(2023)219, arXiv:2208.06485.

[81] S. Agostinelli et al., "GEANT4—a simulation toolkit", *Nucl. Instrum. Meth. A* **506** (2003) 250, doi:10.1016/S0168-9002(03)01368-8.

[82] J. Alwall et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *JHEP* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.

[83] R. Frederix and S. Frixione, "Merging meets matching in MC@NLO", *JHEP* **12** (2012) 061, doi:10.1007/JHEP12(2012)061, arXiv:1209.6215.

[84] P. Nason, "A new method for combining NLO QCD with shower Monte Carlo algorithms", *JHEP* **11** (2004) 040, doi:10.1088/1126-6708/2004/11/040, arXiv:hep-ph/0409146.

[85] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with parton shower simulations: the POWHEG method", *JHEP* **11** (2007) 070, doi:10.1088/1126-6708/2007/11/070, arXiv:0709.2092.

[86] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX", *JHEP* **06** (2010) 043, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.

[87] T. Sjöstrand et al., "An introduction to PYTHIA 8.2", *Comput. Phys. Commun.* **191** (2015) 159, doi:10.1016/j.cpc.2015.01.024, arXiv:1410.3012.

[88] NNPDF Collaboration, "Parton distributions from high-precision collider data", *Eur. Phys. J. C* **77** (2017) 663, doi:10.1140/epjc/s10052-017-5199-5, arXiv:1706.00428.

[89] P. Skands, S. Carrazza, and J. Rojo, "Tuning PYTHIA 8.1: the Monash 2013 tune", *Eur. Phys. J. C* **74** (2014) 3024, doi:10.1140/epjc/s10052-014-3024-y, arXiv:1404.5630.

[90] CMS Collaboration, "Extraction and validation of a new set of CMS PYTHIA 8 tunes from underlying-event measurements", *Eur. Phys. J. C* **80** (2020) 4, doi:10.1140/epjc/s10052-019-7499-4, arXiv:1903.12179.

[91] NNPDF Collaboration, "Parton distributions for the LHC run II", *JHEP* **04** (2015) 040, doi:10.1007/JHEP04(2015)040, arXiv:1410.8849.

[92] CMS Collaboration, "Event generator tunes obtained from underlying event and multiparton scattering measurements", *Eur. Phys. J. C* **76** (2016) 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.

[93] CMS Collaboration, "Investigations of the impact of the parton shower tuning in PYTHIA 8 in the modelling of $t\bar{t}$ at $\sqrt{s} = 8$ and 13 TeV",. https://cds.cern.ch/record/2235192.

[94] J. Dror, M. Farina, E. Salvioni, and J. Serra, "Strong tW scattering at the LHC", *JHEP* **01** (2016) 071, doi:10.1007/JHEP01(2016)071, arXiv:1511.03674.

[95] G. Bevilacqua et al., "The simplest of them all: $t\bar{t}W^{\pm}$ at NLO accuracy in QCD", *JHEP* **08** (2020) 043, doi:10.1007/JHEP08(2020)043, arXiv:2005.09427.

[96] CMS Collaboration, "Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *JINST* **15** (2020) P10017, doi:10.1088/1748-0221/15/10/P10017, arXiv:2006.10165.

[97] CMS Collaboration, "Measurement of the inclusive production cross sections for forward jets and for dijet events with one forward and one central jet in pp collisions at $\sqrt{s} = 7$ TeV", *JHEP* **06** (2012) 036, doi:10.1007/JHEP06(2012)036, arXiv:1202.0704.

[98] CMS Collaboration, "Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 7$ TeV", *Phys. Lett. B* **722** (2013) 5, doi:10.1016/j.physletb.2013.03.024, arXiv:1210.6718.

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

[100] CMS Collaboration, "Inclusive and differential cross section measurements of single top quark production in association with a Z boson in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *JHEP* **02** (2022) 107, doi:10.1007/JHEP02(2022)107, arXiv:2111.02860.

[101] CMS Collaboration, "Measurement of the inclusive and differential $t\bar{t}\gamma$ cross sections in the dilepton channel and effective field theory interpretation in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *JHEP* **05** (2022) 091, doi:10.1007/JHEP05(2022)091, arXiv:2201.07301.

[102] CMS Collaboration, "Search for new physics in same-sign dilepton events in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Eur. Phys. J. C* **76** (2016) 439, doi:10.1140/epjc/s10052-016-4261-z, arXiv:1605.03171.

[103] CMS Collaboration, "Observation of single top quark production in association with a Z boson in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Phys. Rev. Lett.* **122** (2019) 132003, doi:10.1103/PhysRevLett.122.132003, arXiv:1812.05900.

[104] CMS Collaboration, "Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8\,\text{TeV}$", *JINST* **10** (2015) P06005, doi:10.1088/1748-0221/10/06/P06005, arXiv:1502.02701.

[105] J. H. Hoß, "Search for Supersymmetry with Multiple Charged Leptons at *sqrts* = 13 TeV with CMS and Radiation Tolerance of the Readout Chip for the

Phase I Upgrade of the Pixel Detector". PhD thesis, ETH Zürich, 2017. doi:10.3929/ethz-b-000182698.

[106] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting", *Journal of Computer and System Sciences* **55** (1997), no. 1, 119–139, doi:https://doi.org/10.1006/jcss.1997.1504.

[107] J. H. Friedman, "Greedy function approximation: A gradient boosting machine.", *The Annals of Statistics* **29** (2001), no. 5, 1189 – 1232, doi:10.1214/aos/1013203451.

[108] A. Hoecker et al., "Tmva - toolkit for multivariate data analysis", arXiv:physics/0703039.

[109] M. Abadi et al., "TensorFlow: large-scale machine learning on heterogeneous distributed systems", 2016. arXiv:1603.04467.

[110] F. Chollet et al., "keras", 2015. https://keras.io, https://github.com/fchollet/keras.

[111] D. P. Kingma and J. Ba, "adam: a method for stochastic optimization", in *Proc. 3rd Int. Conf. on Learning Representations (ICLR 2015): San Diego, CA, USA, May 7–9, 2015*. 2015. arXiv:1412.6980.

[112] I. Goodfellow, Y. Bengio, and A. Courville, "Deep learning". MIT Press, Cambridge, MA, USA, 2016. http://www.deeplearningbook.org.

[113] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks", in *Proc. 14th Int. Conf. on Artifical Intelligence and Statistics (AISTATS 2011): Fort Lauderdale, FL, USA, April 11–13, 2011*. 2011. *Proc. Mach. Learn. Res.* **15** (2011) 315.

[114] N. Srivastavaet al., "Dropout: a simple way to prevent neural networks from overfitting", *J. Mach. Learn. Res.* **15** (2014) 1929.

[115] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J. C* **71** (2011) 1554, doi:10.1140/epjc/s10052-011-1554-0, arXiv:1007.1727. [Erratum: doi:10.1140/epjc/s10052-013-2501-z].

[116] The LHC Higgs Combination Group, "Procedure for the LHC Higgs boson search combination in Summer 2011",. https://cds.cern.ch/record/1379837.

[117] L.-G. Xia, "Study of constraint and impact of a nuisance parameter in maximum likelihood method", *Journal of Physics G: Nuclear and Particle Physics* **46** (Jul, 2019) 085004, doi:10.1088/1361-6471/ab02c0.

[118] J. Conway, "Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra", doi:10.5170/CERN-2011-006.115, arXiv:1103.0354.

[119] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", *JINST* **12** (2017) P02014, doi:10.1088/1748-0221/12/02/P02014, arXiv:1607.03663.

[120] CMS Collaboration, "Search for the associated production of the Higgs boson with a top-quark pair", *JHEP* **09** (2014) 087, doi:10.1007/JHEP09(2014)087, arXiv:1408.1682. Erratum: doi:10.1007/JHEP10(2014)106.

[121] CMS Collaboration, "Evidence for associated production of a Higgs boson with a top quark pair in final states with electrons, muons, and hadronically decaying τ leptons at $\sqrt{s} = 13$ TeV", *JHEP* **08** (2018) 066, doi:10.1007/JHEP08(2018)066, arXiv:1803.05485.

[122] CMS Collaboration, "Measurement of the inclusive and differential $\mathrm{t\bar{t}}\gamma$ cross sections in the single-lepton channel and EFT interpretation at $\sqrt{s} = 13\,\text{TeV}$", *JHEP* **12** (2021) 180, doi:10.1007/JHEP12(2021)180, arXiv:2107.01508.

[123] CMS Collaboration, "Search for electroweak production of charginos and neutralinos in multilepton final states in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP* **03** (2018) 166 CMS-SUS-16-039, CERN-EP-2017-121, doi:10.1007/JHEP03(2018)166, arXiv:1709.05406.

[124] CMS Collaboration, "Search for standard model production of four top quarks with same-sign and multilepton final states in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$", *Eur. Phys. J. C* **78** (2018) 140, doi:10.1140/epjc/s10052-018-5607-5, arXiv:1710.10614.

[125] CMS Collaboration, "A portrait of the Higgs boson by the CMS experiment ten years after the discovery", *Nature* (2022) CMS-HIG-22-001, CERN-EP-2022-039, CMS-HIG-22-001-003, doi:10.1038/s41586-022-04892-x, arXiv:2207.00043.

[126] CMS Collaboration, "A new set of CMS tunes for novel colour reconnection models in PYTHIA8 based on underlying-event data", CMS-PAS-GEN-17-002. https://cds.cern.ch/record/2780467.

[127] R. Barlow and C. Beeston, "Fitting using finite monte carlo samples", *Computer Physics Communications* **77** (1993), no. 2, 219–228, doi:https://doi.org/10.1016/0010-4655(93)90005-W.

[128] CMS Collaboration, "Observation of four top quark production in proton-proton collisions at $\sqrt{s} = 13$ TeV", CMS-TOP-22-013, CERN-EP-2023-090, CMS-TOP-22-013-004, arXiv:2305.13439. https://cds.cern.ch/record/2859484.

[129] ATLAS Collaboration, "Measurement of the total and differential cross-sections of $t\bar{t}W$ production in pp collisions at 13 TeV with the ATLAS detector", ATLAS-CONF-2023-019. https://cds.cern.ch/record/2855337.

[130] L. Buonocore et al., "Associated production of a W boson with a top-antitop quark pair: next-to-next-to-leading order QCD predictions for the LHC", arXiv:2306.16311.

[131] ATLAS Collaboration, "Search for leptonic charge asymmetry in $t\bar{t}W$ production in final states with three leptons at $\sqrt{s} = 13$ TeV", *JHEP* **07** (2023) 033 CERN-EP-2022-249, doi:10.1007/JHEP07(2023)033, arXiv:2301.04245.

Acknowledgements

My journey from starting my undergraduate to completing my PhD was not always a long, smooth journey, and it would have not been successful without the supports of many people around me.

First of all, I would like to express my gratitude to Andrea and Didar, without you offering me this PhD position, I would have not moved to Belgium in the first place. Thank you for guiding me through this journey, for giving me useful advice and for all the lovely discussions, whether they were about physics or about life. I also would like to extend my appreciation to every member of my thesis, for patiently reading through every chapter of my thesis and giving your insightful feedback. Pieter, you were a brilliant colleague, I was so lucky to work with you, to have someone to bother when I got stuck, and of course to drink beer with as well. Kirill, thank you for your effort and for being patient with me since my very early days up to the public defence. The $t\bar{t}W$ analysis would have not been completed without endless effort from Barbara, Clara, and Carlos. I am grateful that we had a chance to collaborate, to share our knowledge, even the pandemic was there most of the time. And of course, how can I forget the Flinke jongens. Gianny, Liam and Luka, I really enjoyed sharing the office with you, thanks for answering all my random silly questions from time to time, for all the good laughs, and for procrastinating together sometimes. To Illia, Tom and Willem, you had helped me so much to get started in my first day and I really appreciate that. Martina, Basile, Anna, Joscha, Sandhya, Khawla, Marek, Yanwen, Amrutha, and many other people at UGent and UCLouvain, even if we didn't work together much, it was such a pleasure to meet and get to know you.

To my friend, Anna, I can't believe it has been a decade since we met. All the lectures, chilling in JCMB, our wine nights in Edinburgh to beer nights in Belgium, the gossips, the ranting, the memes, thank you for them all. Thank you, Jana, for being there with me. Whether in Ghent or Brussels, we always have a good time, non-stop chats, and we definitely keep doing this. Chị Tú, chị Anh, and mini UN, many thanks to you for all the virtual support for more than 10 years.

Bá me, Thón đi được tới đây là nhờ bá me cho tất cả mọi thứ. Còn được sự ủng hộ, lo lắng hết mình từ nhà mình, cón, ngoại, Ngoán, tư, bảy, câu, út, ỵ ba, me hai, to cu,... And finally, Maxwell, I am thankful to have you. You moved to Belgium for me, you always be there for me, support me, make sure I am well fed. Thank you for everything. Love!