Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

DESY.

Differential ratio measurements of $t\bar{t}\gamma/t\bar{t}$ and a search for exclusive $t\bar{t}$ with the CMS experiment

Dissertation

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

vorgelegt von

Beatriz Ribeiro Lopes

aus

Cascais, Portugal

Hamburg
2023

CERN-THESIS-2023-256
26/10/2023

Eidesstattliche Versicherung / Declaration on oath

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen benutzt habe. Die eingereichte schriftliche Fassung entspricht der auf dem elektronischen Speichermedium. Die Dissertation wurde in der vorgelegten oder einer ähnlichen Form nicht schon einmal in einem früheren Promotionsverfahren angenommen oder als ungenügend beurteilt.

I hereby declare in lieu of oath that I have written this dissertation myself and that I have not used any auxiliary materials or sources other than those indicated. The written version submitted corresponds to the one stored electronically. The dissertation presented in this form, has not already been accepted in an earlier doctoral procedure or assessed as unsatisfactory.

Hamburg, November 27, 2023

Beatriz Ribeiro Lopes

Gutachterinnen der Dissertation:	Dr. Abideh Jafari
	Prof. Dr. Elisabetta Gallo
Zusammensetzung der Prüfungskommission:	Prof. Dr. Sven-Olaf Moch
	Prof. Dr. Elisabetta Gallo
	Prof. Dr. Christian Schwanenberger
	Dr. Abideh Jafari
	Dr. Maria Aldaya Martin
Vorsitzender der Prüfungskommission:	Prof. Dr. Sven-Olaf Moch
Datum der Disputation:	26.10.2023
Vorsitzender des	
Fach-Promotionsausschusses Physik:	Prof. Dr. Günter H. W. Sigl
Leiter des Fachbereichs Physik:	Prof. Dr. Wolfgang J. Parak
Dekan der Fakultät MIN:	Prof. Dr.-Ing. Norbert Ritter

Abstract

In this thesis, measurements performed using proton-proton collision data at $\sqrt{s} = 13\,\text{TeV}$ collected by the CMS experiment at the CERN LHC, in the years 2016, 2017, and 2018, are presented. The data correspond to an integrated luminosity of $138\,\text{fb}^{-1}$.

New results on top quark pair production in association with a photon ($\text{t}\bar{\text{t}}\gamma$) are presented. Differential cross sections are presented, measuring five different observables that are sensitive to modelling aspects and possible effects of new physics. In particular, top quark observables are measured for the first time, using a top quark pair ($\text{t}\bar{\text{t}}$) reconstruction algorithm. Experiment independent distributions are obtained at the parton level.

The ratio between the cross sections for $\text{t}\bar{\text{t}}\gamma$ and $\text{t}\bar{\text{t}}$ is measured for the first time at the LHC, inclusively and differentially. Additionally, the top quark charge asymmetry is measured in $\text{t}\bar{\text{t}}\gamma$ events. The results are obtained using maximum likelihood fits which directly account for and constrain backgrounds and systematic uncertainties. The results are compared to theory predictions. No significant deviations from standard model predictions are observed.

Using a fraction of the same data, collected in 2017, a search for central exclusive production of $\text{t}\bar{\text{t}}$, where the protons remain intact and are measured in the Precision Proton Spectrometer, is also presented. An upper limit for the cross section of this process is obtained.

Zusammenfassung

In dieser Arbeit werden Messungen vorgestellt, die mit Daten aus Proton-Proton Kollisionen bei $\sqrt{s} = 13\,\text{TeV}$ durchgeführt wurden, die vom CMS-Experiment am LHC des CERN in den Jahren 2016, 2017 und 2018 gesammelt wurden. Diese Daten entsprechen einer integrierten Luminosität von $138\,\text{fb}^{-1}$.

Es werden neue Ergebnisse zur Top-Quark-Paarproduktion in Verbindung mit einem Photon ($\text{t}\bar{\text{t}}\boldsymbol{\gamma}$) vorgestellt. Differentielle Wirkungsquerschnitte werden als Funktion von fünf verschiedenen kinematischen Observablen gemessen. Die Verteilungen liefern Aufschluss über Modellierungsaspekte und Sensitivität auf mögliche Effekte von Neuer Physik. Insbesondere die Top-Quark-Observablen werden zum ersten Mal mit einem Top-Quark-Paar ($\text{t}\bar{\text{t}}$) Rekonstruktionsalgorithmus gemessen. Experimentunabhängige Verteilungen werden für Definitionen auf Parton-Level erhalten.

Das Verhältnis zwischen den Wirkungsquerschnitten von $\text{t}\bar{\text{t}}\boldsymbol{\gamma}$ und $\text{t}\bar{\text{t}}$ wird zum ersten Mal am LHC gemessen, sowohl inklusiv als auch differentiell. Zusätzlich wird die Ladungs-Asymmetrie des Top-Quarks in $\text{t}\bar{\text{t}}\boldsymbol{\gamma}$-Ereignissen gemessen. Die Ergebnisse werden mit Hilfe von Maximum-Likelihood-Anpassungen erzielt, die Untergründe und systematische Unsicherheiten direkt berücksichtigen. Die Ergebnisse werden mit Theorievorhersagen verglichen. Es werden keine signifikanten Abweichungen von den Vorhersagen des Standardmodells beobachtet.

Unter Verwendung eines Teils der gleichen Daten, die 2017 gesammelt wurden, wird auch eine Suche nach der zentralen exklusiven Produktion von $\text{t}\bar{\text{t}}$ mit intakten Protonen vorgestellt, die im Precision Proton Spectrometer gemessen werden. Mit Hilfe einer Maximum-Likelihood-Anpassung wird eine Obergrenze für den Wirkungsquerschnitt dieses Prozesses ermittelt.

Contents

Introduction

The standard model (SM) of particle physics is the theoretical framework in which all the work in this thesis is conducted. Developed during the latter half of the twentieth century, it stands as an exceptionally successful theory that comprehensively describes three out of the four fundamental interactions in nature, as well as all the elementary particles that have been observed in experiments so far. Despite its achievements, the SM falls short of being a complete theory of nature at its most fundamental level. It leaves several questions unanswered, such as the nature of dark matter, the origin of the matter–antimatter asymmetry in the Universe, and the existence and hierarchy of the neutrino masses. The modern understanding is that the SM can be seen as an effective theory that provides a suitable description of matter up to the energy scales that are reachable with current experiments, but which leaves room for the emergence of new physical phenomena at higher energy scales.

In experimental particle physics, the quest to find such new effects is carried out following two main approaches. The first and perhaps more intuitive approach involves direct searches for new particles or phenomena as foreseen by theoretical models proposing extensions to the SM. Alternatively, a second strategy involves conducting rigorous tests of the model by finely measuring established processes and seeking deviations from predictions. Both approaches rely on the analysis of large amounts of data from high energy particle colliders. The Large Hadron Collider (LHC) at CERN, in Switzerland, is unrivalled in the energy frontier. This thesis uses data collected at the Compact Muon Solenoid (CMS), one of the particle detectors installed at this collider.

The LHC has been collecting proton-proton collision data since 2008, and has completed two successful runs. This thesis uses data collected during its second run, and focuses on the production of top quarks. Top quarks are the heaviest known fundamental particles, and a great test ground for precision tests of the SM. Due to their large mass, they could potentially be involved in new physical effects occurring at higher energies. In particular, new physics effects could modify the way that the top quark interacts with the other SM particles. This thesis explores the interaction between the top quark and the photon, which is very well understood theoretically. This means that if even the slightest deviation from the predictions is observed in the data, it is a strong hint for the presence of new physics.

The thesis is organised as follows. The first chapter introduces the SM. A particular emphasis is given to the top quark and its interaction with the photon. Two processes are introduced in detail: the central exclusive production of top quark pairs ($t\bar{t}$) and the associated production of $t\bar{t}$ with one photon ($t\bar{t}\gamma$).

The subsequent chapter presents the LHC and the CMS detector. The main features of the detector are described, and the way the data used for this thesis work are produced and reconstructed is explained. The algorithms used for particle reconstruction are presented, and the methods used for simulating collision events are briefly introduced. Moreover, the concept of luminosity and the methods to measure it in the CMS experiment are discussed. Some results on luminosity measurements which were obtained as part of this thesis project are also shown.

Chapter 3 describes a set of inclusive and differential measurements of $t\bar{t}\gamma$, using data collected by the CMS experiment in 2016–2018. This process is interesting due to its sensitivity to the coupling between the photon and the top quark, and its cross section is high enough to allow for precision measurements with the data currently available. Events containing two leptons, two quark-initiated jets and a photon fulfilling specific criteria are selected. A data-driven approach is developed to estimate the background arising from misidentified photons. A list of observables related to the photon, lepton and top quark kinematics is chosen to perform the differential measurements, using a likelihood-based approach. Additionally, the first measurements of the ratio between the production rates of $t\bar{t}\gamma$ and that of $t\bar{t}$ at the LHC are presented. Finally, a study of the top quark charge asymmetry in $t\bar{t}\gamma$ events is described. The results are discussed and compared to the SM predictions.

Chapter 4 presents a search for the central exclusive production of $t\bar{t}$. This process occurs mainly via the interaction between two highly energetic photons, and has the particularity that the interacting protons remain intact and do not dissociate in the collision. It is also sensitive to the coupling between the photon and the top quark, and thus provides a unique way to study top quark pair production in a very rare production channel. In order to measure it, data collected by the CMS experiment in 2017 are used, including data from the Precision Proton Spectrometer, a CMS subdetector which can detect protons that leave the collisions intact. A top quark reconstruction algorithm is employed to reconstruct the $t\bar{t}$ system, and compare its kinematics to that of the intact protons. This comparison is done using a multivariate analysis technique with Boosted Decision Trees. Several sources of systematic uncertainty arising from theoretical aspects and from the experimental setup are included and described in detail. This process has a very low cross section and the current data collected by the CMS experiment do not provide enough sensitivity to allow for its observation, however upper limits are set on the cross section, using a likelihood-based fit.

CHAPTER 1

Theoretical framework

The standard model of particle physics (SM) is a well-established physics theory that describes all known fundamental particles and their interactions. It is the product of decades of theoretical and experimental work, starting in the latter half of the twentieth century. Section 1.1 provides a description of the theory: the elementary particles are introduced, and the strong, weak, and electromagnetic interactions are briefly explained. More in-depth derivations of the theory can be found, e.g., in Refs. [1, 2].

Although extremely successful in reproducing experimental data thus far, the SM is known to be incomplete. It does not include gravitation, nor does it provide suitable explanations for dark matter and dark energy. This implies the existence of physical effects which are beyond the SM (BSM). High-energy particle collision experiments are being used to perform stringent tests of the SM and search for such new physics. The analyses presented in this thesis make use of data collected in proton-proton (pp) collisions at the LHC, the most powerful particle collider to date. Proton collisions are introduced in section 1.2. In particular, this thesis focuses on the production of the heaviest SM particle, the top quark, through rare processes involving the electromagnetic interaction. The top quark and its interactions are explored in section 1.3.

1.1 The standard model of particle physics

The SM is a quantum field theory (QFT). This means it relies on the concept of quantum fields, which are fundamental and defined at all points in spacetime - particles emerge as excitations of these fields. The fundamental interactions or forces described by the SM are the strong, weak, and electromagnetic interactions. The fourth fundamental force, gravity, is not included in the theory. Based on the principle of least action, the SM can be mathematically represented using the Lagrangian formalism [2]. The Lagrangian density, $\mathcal{L}$, contains the fields and their interactions, incorporating all information about the

system. From $\mathcal{L}$, the equations of motion ruling the particle trajectories and interactions can be obtained. The SM is a gauge theory, and the fundamental interactions reflect local internal symmetries of $\mathcal{L}$. According to Noether's theorem, to each local symmetry corresponds a conserved quantity, or a quantum number [3].

The Lagrangian density of the SM is built by combining the individual Lagrangian densities of the aforementioned interactions. Quantum chromodynamics (QCD), which has a $\mathrm{SU}(3)_C$ symmetry, is used to describe the strong force, as detailed in section 1.1.2. The C index represents the colour quantum number. Electromagnetism and the weak force are generally unified using the electroweak (EW) theory, as detailed in section 1.1.3. The EW theory obeys a $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ symmetry, and the quantum numbers associated with it are the third component of the weak isospin T_3 and the hypercharge Y. The L index represents the left-chiral nature of the weak interaction, which is introduced in section 1.1.3. In terms of gauge symmetries, the SM can therefore be written as

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

The fundamental fields of the SM are organised as fermion fields, with spin 1/2, and vector boson fields, with spin 1. Fermion fields are represented by spinors, which are two-component fields that satisfy the Dirac equation [4]. Depending on how they behave under the parity transformation, spinors can be classified in two states, left-chiral, χ_L, with spin orientation $+1/2$, or right-chiral, ξ_R, with spin orientation $-1/2$. Parity is a transformation that inverts the spatial axes, i.e., transforms a right-handed coordinate system into left-handed, and thus left-chiral spinors into right-chiral ones. Considering chirality, we can write the SM fermion fields as bispinors, or Dirac fields:

$$\psi = \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix}, \tag{1.1}$$

which are four-component fields. Fermion fields are divided into quark and lepton families, each containing three generations. The isospin doublets of quarks contain each an up-type quark, with electric charge $+2/3$, and a down-type quark, with electric charge $-1/3$. The six kinds of quarks (two types and three generations) are often designated *flavours*. In the lepton family, every charged lepton (electric charge -1) has a corresponding neutrino (electrically neutral) as its isospin partner. The EW theory requires the introduction of an additional scalar (spin 0) field in the theory, known as the Higgs field, as described in section 1.1.3.

1.1.1 The Feynman diagram representation

Richard Feynman introduced a graphical representation for the interactions between fundamental fields, known as the Feynman diagram representation [5]. This representation is useful to visualise the different QFT processes intuitively, as well as to facilitate the computation of the probability amplitudes for those processes.

In Feynman diagrams, time is read from left to right. Straight lines represent fermions and antifermions. Arrows pointing left to right represent fermions, while those pointing right to left represent antifermions. Curly lines represent the bosons of the strong interaction (gluons, introduced in section 1.1.2) and wavy lines represent the electroweak bosons, introduced in section 1.1.3. The Higgs boson, also introduced in section 1.1.3, is normally represented as a dashed line. One diagram represents the sum over all possible time orderings in which a given process can occur.

For a given initial and final state, one can in principle draw an infinite number of Feynman diagrams, with different numbers of interaction vertices. A set of defined rules, the Feynman rules, are used to translate each diagram into a formula, and the sum of all possible diagrams gives us the probability amplitude of a process. However, an infinite number of diagrams leads to an infinite number of terms in the amplitude calculation.

In many cases, a perturbative approach can be used as an approximation, to reduce this sum to a finite number of terms. Each vertex in a diagram introduces in the calculation a multiplicative term involving the coupling strength of the corresponding interaction. If the coupling strength of the interaction is smaller than 1, the more vertices the diagram has, the less likely the process is to occur. That is the case where a perturbative approach is justified: the probability amplitude is first computed considering only the diagrams with the minimum number of vertices, also known as the leading order (LO); diagrams with higher number of vertices can then be added as corrections to the LO amplitude. Diagrams with one extra vertex, compared to the LO, are called next-to-LO (NLO), those with two extra vertices are called next-to-NLO (NNLO), and so on.

Renormalisation

In order to describe the propagation of fields in an accurate way, an infinite set of higher-order corrections needs to be taken into account. In particular, diagrams with internal "loops" need to be considered. The particles involved in these loops can have arbitrarily large momenta, which introduces diverging integrals in the amplitude calculations and leads to non-physical infinite results. This is circumvented by a mathematical procedure called regularisation, which introduces an arbitrary cut-off scale Λ, such that corrections involving momenta much larger than Λ are ignored. As a consequence of this, the formulas need to be rewritten in terms of renormalised quantities, with an energy scale dependence. Renormalised quantities are the ones we measure experimentally, and they need to be expressed at some arbitrary scale, called the renormalisation scale μ_R. It is common to

choose μ_R to be of the order of the momentum scale of the process of interest, to avoid large higher-order corrections in the amplitude calculation.

1.1.2 The strong interaction

The strong interaction, described by QCD, acts between spinors and gauge fields with colour quantum numbers, i.e., quarks, antiquarks, and gluons. The colour quantum numbers are red (r), green (g), and blue (b), in analogy with the three fundamental colours in nature. Quarks carry one of three possible colour charges, so are represented as colour triplets, $Q = [\mathrm{q_r}, \mathrm{q_g}, \mathrm{q_b}]^T$. Antiquarks carry one of the three "anti-colour" charges and can be represented by $\bar{Q}$. Gluons carry both a colour and an anti-colour charge, forming a colour octet. Being colourless, leptons do not take part in the strong interaction. In other words, they are singlets under SU(3) transformations. The Lagrangian density of QCD [2] can be written as:

$$\mathcal{L}_{\mathrm{QCD}} = \bar{Q}_i \left(i\gamma^\mu (D_\mu)_{ij} - m\,\delta_{ij} \right) Q_j - \frac{1}{4} G^a_{\mu\nu} G^{\mu\nu}_a \tag{1.2}$$

with

$$D_\mu = \partial_\mu - i g_s T_a G^a_\mu, \tag{1.3}$$

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s f^{abc} G^b_\mu G^c_\nu. \tag{1.4}$$

The Greek indices μ, ν represent the four spacetime dimensions. The indices i, j represent a sum over the six quark flavours. The first term of the covariant derivative D_μ in equation 1.3 gives the propagation of free quark fields. The second term gives the interaction between the quark and gluon fields G_μ; the strength of this coupling is g_s. The generators of the SU(3) group are denoted T_a, with $a = 1...8$. A common representation of T_a are the Gell-Mann matrices, λ_a, with $T_a = \lambda_a/2$ [6]. The mass of the quark fields is m. The gluon field strength tensor $G^a_{\mu\nu}$ gives the self-interactions between the gluon fields. In equation 1.4, f^{abc} are the structure constants of SU(3), defined as $[T^a, T^b] = i f^{abc} T^c$.

The allowed interactions in QCD can be read directly from the Lagrangian. From the covariant derivative, appear the terms $g_s \bar{Q}_i \gamma^\mu T_a G^a_\mu Q_i$ that describe the gluon emission from quarks. Expanding the Lagrangian term with the field strength tensor, terms involving three and four gluons appear, which represent three and four gluon interactions. These interactions can be illustrated using the Feynman diagrams of figure 1.1.

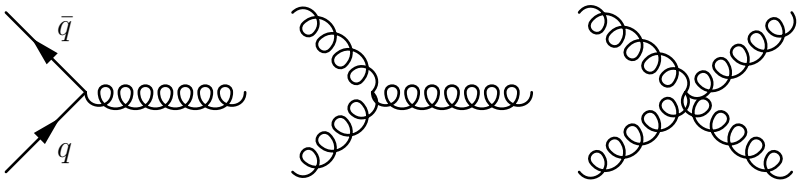

Figure 1.1: Feynman diagrams showing the interaction vertices of QCD.

Due to the renormalisation procedure introduced in section 1.1.1, the QCD coupling strength g_s depends on the energy scale k. The coupling constant α_S is defined as:

$$\alpha_S(k^2) = \frac{g_s^2(k^2)}{4\pi} \tag{1.5}$$

and its energy dependence is known as the "running" of the coupling. Towards higher energies, α_S decreases logarithmically. In this context, higher energies are equivalent to smaller distances between the interacting particles, thus at small distances quarks and gluons can move in a practically free way. This is called "asymptotic freedom". Conversely, lower energies correspond to an increase of α_S, leading to a phenomenon called "confinement", that causes quarks and gluons to be bound together forming hadrons. Bare quarks and gluons cannot be observed directly, because hadrons are formed almost immediately, in a time scale of about 10^{-24} s [7].

1.1.3 The electroweak interaction

The electromagnetic interaction can be explained using Quantum Electrodynamics (QED). This theory describes electrically charged fermion fields (leptons and quarks), the photon vector field, and their interactions. It is analogous to QCD, but instead of colour, it is the electric charge that rules the interaction. Conceptually, it is a very simple theory, yet it allows for extremely accurate predictions, which have been extensively verified experimentally. This led Richard Feynman to call it "the jewel of physics".

The weak interaction can be described by a QFT called Quantum Flavour Dynamics. It is the only flavour-changing interaction: it allows quarks to swap flavour, and leptons to be converted to their corresponding neutrino (or vice-versa). It has the very peculiar property that it is not invariant under parity transformations. This was discovered in the Wu experiment in 1956 [8] and was not a prediction from the theory. Later, it was found that the mediators of the weak interaction couple only to left-chiral particles, and consequently right-chiral particles do not participate in the weak interaction, thus the L index in $\mathrm{SU}(2)_L$.

As mentioned earlier, the electromagnetic interaction and the weak force are unified in one single mathematical description, using the EW theory. This theory was developed around 1968 by Glashow, Salam, and Weinberg, who received the Nobel Prize in Physics

for their work [9–11]. It treats the EW interaction as one unified interaction at high energy scales, where the Lagrangian is symmetric under $\mathrm{SU(2)}_L \times \mathrm{U(1)}_Y$. Below the energy of about 246 GeV, there is a spontaneous symmetry breaking, which breaks the SU(2) symmetry and leads to the separate electromagnetic and weak interactions. The symmetry breaking is explained using the Higgs mechanism, which is briefly introduced at the end of this section.

The charges of the electroweak interaction are the hypercharge Y and the third component of the weak isospin T_3. A combination of these two charges gives the electric charge from QED, Q:

$$Q = \frac{Y}{2} + T_3. \tag{1.6}$$

This identity is known as the Gell-Mann-Nishijima equation [12].

As mentioned before, right-chiral fermion fields are not affected by SU(2) transformations. This is reflected in the fact that they have $T_3 = 0$, and are considered as weak isospin singlets. Let us denote the weak isospin singlets for up- and down-type quark fields u_R and d_R, respectively, and for charged lepton fields e_R. Assuming that neutrinos are massless, there are no right-handed neutrinos in the SM[1].

Conversely, left-chiral fermions have weak isospin 1/2. This means that T_3 can have values $\pm 1/2$, with the positive (negative) value belonging to up- (down-) type quarks. Similarly, left-handed neutrinos have $T_3 = +1/2$ and left-handed charged leptons have $T_3 = -1/2$. The up- and down-type quarks in the same generation form a weak isospin doublet, denoted Q_L, and the neutrinos and corresponding charged leptons form three more, denoted L_L. Left-chiral fermions can transform into their weak isospin partner, within the same doublet. Let us encode all that into the Lagrangian density of the EW theory. The Lagrangian can be separated in terms describing the dynamics of the fermion fields, $\mathcal{L}_{EW,f}$ and terms concerning the gauge boson fields, $\mathcal{L}_{EW,g}$ [2]:

$$\mathcal{L}_{EW} = \mathcal{L}_{EW,f} + \mathcal{L}_{EW,g}. \tag{1.7}$$

Starting with the fermion fields, and using the notation introduced above for the weak isospin singlets and doublets (the L and R subscripts are omitted for readability):

$$\mathcal{L}_{EW,f} = \bar{Q}_i (D_\mu)_{ij} Q_j + \bar{L}_i (D_\mu)_{ij} L_j + \bar{\mathrm{u}}_i (D_\mu)_{ij} \mathrm{u}_j + \bar{\mathrm{d}}_i (D_\mu)_{ij} \mathrm{d}_j + \bar{\mathrm{e}}_i (D_\mu)_{ij} \mathrm{e}_j, \tag{1.8}$$

where the covariant derivative is now:

[1] Observations of neutrino oscillations contradict this assumption [13]. Such oscillations can be explained by the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix [14, 15], in an analogous way to the quark mixing with the CKM matrix introduced at the end of this section. This mechanism requires neutrinos to have non-zero mass, which can be realised in a straightforward way in the SM by introducing right-handed neutrinos. Neutrino masses have not been measured yet, but experiments like KATRIN report increasingly more stringent upper limits for their values [16].

$$D_\mu = i\gamma^\mu(\partial_\mu - i\frac{g'}{2}\hat{Y}_W B_\mu - ig\hat{T}_a W_\mu^a). \tag{1.9}$$

The indices i, j run over the three fermion generations. The covariant derivative, like in the case of QCD, gives the propagation of the fermion fields and their interaction with the gauge boson fields. These fields act in a similar way as the gluon fields in QCD, the difference being that instead of SU(3), here there are SU(2) and U(1) transformations. In case of SU(2), there are three generators, and therefore there are three W_μ^a bosons. The U(1) transformation, in turn, has only one generator, the B_μ gauge boson. The coupling constant associated with the interaction with the W bosons is g, and the one associated with the interaction with the B field is g'. The generator of U(1) is denoted as $\hat{Y}_W$ and the three generators of SU(2) as $\hat{T}_a$, with $a = 1...3$.

The gauge boson field term is written as:

$$\mathcal{L}_{EW,g} = -\frac{1}{4}W_{\mu\nu}^a W_a^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}, \tag{1.10}$$

where the field strength tensors are defined in a similar way as those of QCD:

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \tag{1.11}$$

$$W_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a + g f^{abc} W_\mu^b W_\nu^c, \tag{1.12}$$

with f^{abc} being the structure constant of SU(2). From the expressions above we can see that, unlike W_μ^a, B_μ has no self-interactions. We can also see from the Lagrangian that no mass term for vector boson fields is present. In fact, such a term would break SU(2) symmetry. Nevertheless, massive bosons have been experimentally observed [17], which implies that the symmetry has to be broken at low energies.

The Brout-Englert-Higgs mechanism

The electroweak symmetry breaking (EWSB) is explained by the Brout-Englert-Higgs mechanism [18, 19]. Introducing an isospin doublet of scalar fields ϕ, with $\phi = (\phi^+, \phi^0)$, is possible without explicitly breaking the SU(2) symmetry. It is done by adding the following term to the Lagrangian density [2]:

$$\mathcal{L}_{\text{Higgs}} = (D_\mu\phi)^\dagger(D_\mu\phi) - V(\phi), \tag{1.13}$$

where the Higgs potential $V(\phi)$ is given by:

$$V(\phi) = -\rho^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2. \tag{1.14}$$

The minimum of the potential can be computed in the usual way by imposing $\partial V(\phi)/\partial\phi = 0$. If $\lambda < 0$, there is no minimum, which leads to an unreasonable theory. Imposing $\lambda > 0$

brings two possibilities: $\rho < 0$ creates a potential with one trivial minimum at $\phi_{min} = 0$, while $\rho > 0$ yields

$$\phi_{min} = \sqrt{\frac{\rho^2}{2\lambda}} e^{i\varphi}. \tag{1.15}$$

In this case, there is a minimum for every value of the phase φ. This corresponds to a circle in the complex ϕ plane, with radius $\sqrt{\frac{\rho^2}{2\lambda}}$. This can be visualised in 3 dimensions, by putting the value of the potential $V(\phi)$ on the vertical axis, and gives rise to the famous *sombrero* shape, as shown in figure 1.2. The symmetry breaking can be seen as a marble rolling down the hat and having to choose spontaneously one specific vacuum value, among infinite possibilities.

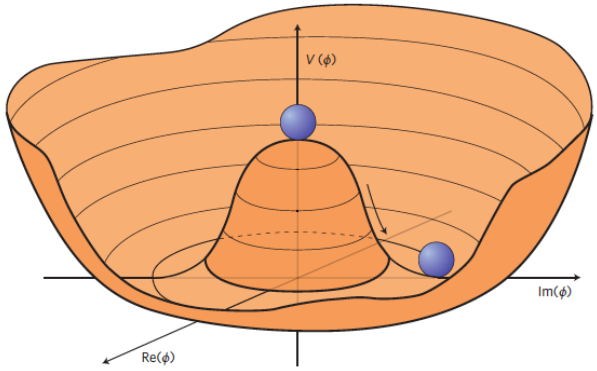

Figure 1.2: Illustration of the Higgs potential in the case of $\rho < 0$. The purple marble illustrates the choice of a point at the minimum of the potential, which breaks spontaneously the symmetry. Image taken from Ref. [20].

One can choose a minimum such that

$$\phi_{min} = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix}, \tag{1.16}$$

with $v = \sqrt{\rho^2/\lambda}$ being the Higgs vacuum expectation value. Expanding ϕ around the minimum and inserting it back into equation 1.13, gives rise to a non-diagonal mass matrix for the vector boson fields. This means that the original fields from the Lagrangian do not correspond directly to the vector bosons observed in nature. Instead, the observed boson fields, which correspond to the mass eigenstates, are a mix of the original ones. The fields W^1_μ and W^2_μ mix to give the mass eigenstates W^+ and W^-, while the B_μ and W^3_μ give the electromagnetically neutral photon and Z boson fields, denoted respectively as A_μ and Z_μ. The mixing can be represented as:

$$\begin{pmatrix} \mathrm{W}^+_\mu \\ \mathrm{W}^-_\mu \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} W^1_\mu \\ W^2_\mu \end{pmatrix}, \tag{1.17}$$

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W^3_\mu \end{pmatrix}. \tag{1.18}$$

The angle θ_W is known as the Weinberg angle or the weak mixing angle, and its value is $\approx 15^\circ$ [21]. Introducing these new fields in the Lagrangian produces mass terms for the W and Z bosons, while the photon remains massless. The measured values for these masses are about 80.4 GeV for the W bosons and 91.2 GeV for the Z boson [21].

From the new Lagrangian, after symmetry breaking, one can read the interactions that are predicted by the EW theory. The allowed vertices can be represented by the Feynman diagrams of figure 1.3 (excluding those involving the Higgs boson).

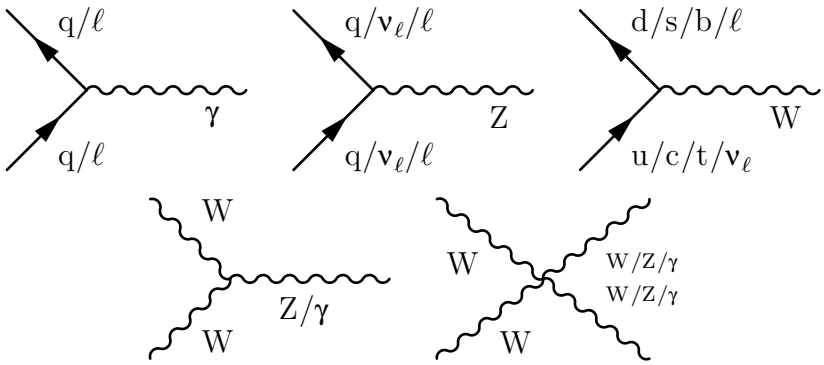

Figure 1.3: Feynman diagrams showing the interaction vertices in the EW theory, excluding those involving the Higgs boson.

The Yukawa interaction

It is possible, without breaking the SU(2) symmetry, to add yet another term to the Lagrangian: an interaction term between the Higgs field ϕ and the fermions. This is described by the Yukawa interaction [11], which was originally developed to explain the nuclear force between hadrons, mediated by pions,

$$\mathcal{L}_{\text{Yukawa}}, f = -\lambda_f(\bar{\mathbf{\Psi}}_L\ \phi\ \psi_R + \bar{\psi}_R\ \phi\ \mathbf{\Psi}_L). \tag{1.19}$$

Here λ_f is the Yukawa coupling constant, different for each fermion, $\mathbf{\Psi}_L$ is the corresponding left-chiral weak isospin doublet (L_i or Q_i), and ψ_R is the right-chiral singlet (u_i,

d_i or e_i). This term does not only describe the interactions between the fermions and the Higgs field, but due to spontaneous symmetry breaking, it also causes fermions to acquire a mass proportional to the vacuum expectation value of the Higgs field, v.

Quark mixing through the CKM matrix

In the EW theory, interactions via W bosons lead to a change of quark flavour. According to the explanations given so far, quarks are allowed to transform into their isospin partners, but not to change generation. Nevertheless, it is experimentally well known that such transitions between generations occur. This can be explained if the quark mass eigenstates q are different from the EW eigenstates that appear in the Lagrangian, q'. Assuming q eigenstates are a linear combination of their EW counterparts, one can write:

$$\begin{pmatrix} \mathrm{d}' \\ \mathrm{s}' \\ \mathrm{b}' \end{pmatrix} = V_{\mathrm{CKM}} \begin{pmatrix} \mathrm{d} \\ \mathrm{s} \\ \mathrm{b} \end{pmatrix} . \tag{1.20}$$

The mixing matrix V_{CKM} is well known as the Cabibbo–Kobayashi–Maskawa (CKM) matrix [22,23]. This matrix, whose values are experimentally measured with uncertainties of the order of 10^{-3} or smaller, can be written in the following way [21]:

$$V_{\mathrm{CKM}} = \begin{pmatrix} V_{\mathrm{ud}} & V_{\mathrm{us}} & V_{\mathrm{ub}} \\ V_{\mathrm{cd}} & V_{\mathrm{cs}} & V_{\mathrm{cb}} \\ V_{\mathrm{td}} & V_{\mathrm{ts}} & V_{\mathrm{tb}} \end{pmatrix} \approx \begin{pmatrix} 0.974 & 0.226 & 0.004 \\ 0.226 & 0.973 & 0.040 \\ 0.009 & 0.040 & 0.999 \end{pmatrix} . \tag{1.21}$$

In practice, this means that any up-type quark can transform into any down-type quark, with probabilities given by the square of the corresponding entry of the CKM matrix. However, since the non-diagonal entries of this matrix are small, some of these processes are heavily suppressed. In particular, it is important to note that $|V_{tb}| \approx 1$, which means that the top quark decays almost exclusively to a b quark and a W boson. Additionally, it is worth noting the small non-diagonal entries related to the bottom quark. These cause b hadrons to have their decay suppressed and therefore have a long lifetime. This is relevant experimentally because it facilitates their identification, as explained in section 2.3.2.

1.1.4 Particle content of the SM

After the EWSB, the SM contains 12 fermions and 12 vector bosons, known as "matter particles" and "force carriers", respectively. As shown in figure 1.4, the vector bosons in the SM are the 8 gluons (g), the photon (γ), and the Z and W $^{\pm}$ bosons. The up-type quarks are the up (u), charm (c), and top (t) quarks. The corresponding down-type

quarks are the down (d), strange (s), and bottom (b) quarks. In the lepton family, there are electrons (e) and the corresponding electron neutrinos (ν_e), muons (μ) and the muon neutrinos (ν_μ), and finally the tau lepton (τ) and its neutrino (ν_τ). Particles with the same isospin in different generations share all quantum numbers, except their mass which increases from the first to the third generation. Though there is no fundamental reason why there should be exactly three generations, their existence is well established experimentally [24]. The possibility of there being a fourth generation is strongly constrained by experiments, namely through measuring the decay modes of the Z boson, and through experiments that measure neutrino production [25].

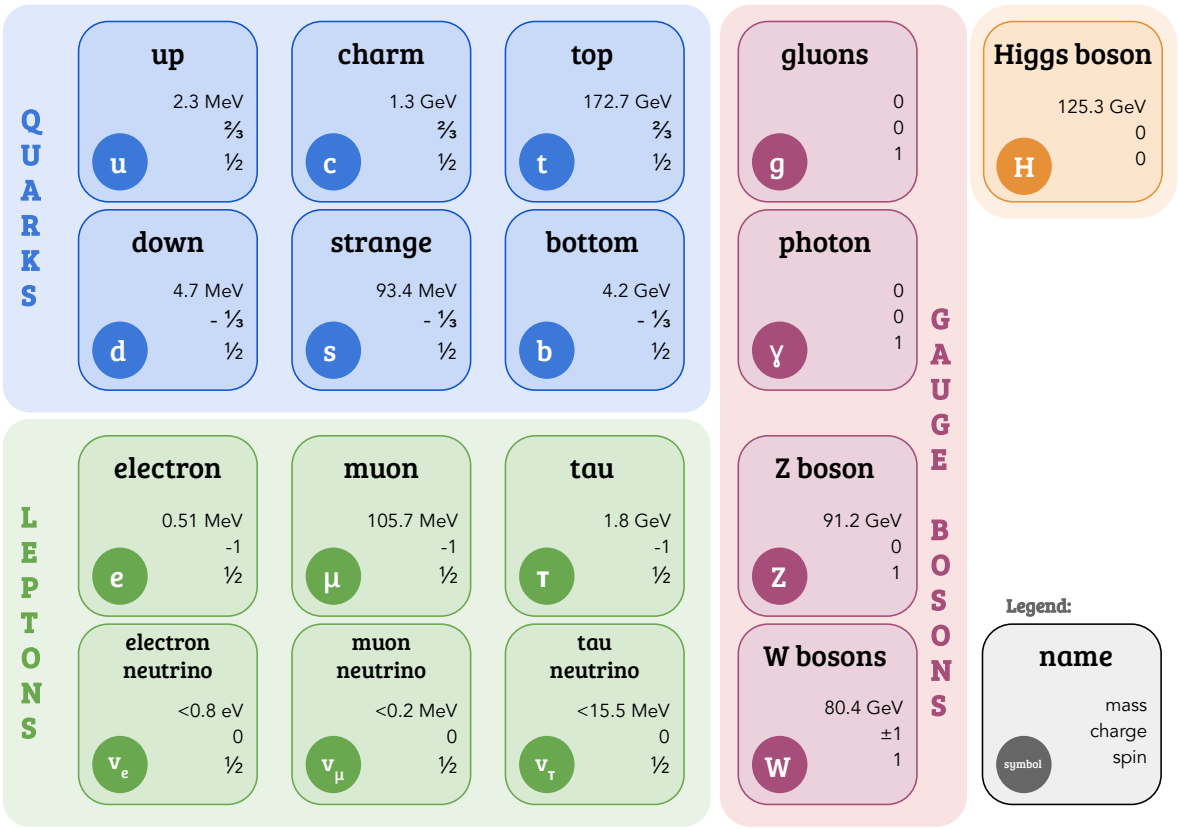

Figure 1.4: The particle content of the SM. For each particle, a measured value for its mass is shown [21], as well as the electric charge and spin quantum numbers.

Each fermion in the table of figure 1.4 has a corresponding antifermion, with the same mass but all opposite sign quantum numbers, or "charges". Similarly, there are two W bosons, W^+ and W^-, with opposite electric charges. From the Higgs field, an extra scalar boson arises, the Higgs boson (H), which was experimentally observed by the CMS and ATLAS collaborations in 2012 [26, 27], making it the last SM particle to be observed.

1.2 Proton collisions at high energies

Protons are bound states of three quarks, two u and one d, with a total electric charge of +1. These are the so-called *valence quarks*. Inside the proton, apart from these, there are also quark-antiquark pairs, as well as gluons, which are constantly being formed and dissolved, and constitute the *sea*. All constituents of the proton are collectively denoted as *partons*.

In a pp collider such as the LHC, described in the next chapter, two beams of protons are accelerated in opposite directions until they reach energies at the TeV scale, and are subsequently allowed to collide in dedicated interaction points. In a typical high energy collision, the main interaction, known as the hard scattering, is carried out between two partons. Protons are normally dissociated as a result of collisions so, in addition to the hard scattering, other particles arise from the interaction of the spectator partons, or proton remnants. This is not the case in elastic collisions, as described in section 1.2.3.

The probability that a pp collision results in a specific set of final state particles with certain kinematics depends on the proton structure and the partonic cross section, as described in section 1.2.1. The hard scattering may result in one or more short-lived resonances, such as top quarks, Z, W or H bosons, which will almost immediately decay into more stable particles. All these steps are simulated using Monte Carlo (MC) event generators, as introduced in section 1.2.2. The concepts of additional radiation and hadronization are also discussed in this section.

1.2.1 QCD factorization and proton structure

In pp collisions at the LHC, a wide range of processes is governed by the strong interaction. The hard scattering happens in the high-energy regime, where α_S is sufficiently small and direct computations with perturbation theory are possible. However, computing cross sections at the LHC also requires knowing the probability of each parton inside the protons contributing to the hard scattering. This corresponds to a lower energy scale (larger distance between interacting particles), where α_S becomes large and perturbative calculations are not possible.

The QCD factorization theorem [28] describes an approximation that makes it possible to separate long-distance and short-distance effects by introducing a *factorization scale* μ_F. Long-distance effects, which cannot be computed perturbatively, are described by parton distribution functions (PDFs). These distributions give the probability to find each type of parton in the proton carrying a fraction x of the proton's momentum (*Bjorken* x [29]). Different PDF parametrisations are available whose parameters are extracted through fits to the data. The short distance cross section can be computed using perturbation theory, from the scattering amplitudes. This is normally referred to as *matrix element* (ME) calculation. The cross section for a certain process $\mathrm{pp} \to ab + \mathrm{X}$ can therefore be written as:

$$\sigma_{\mathrm{pp}\to ab+\mathrm{X}} = \sum_{i,j} f_i(x_1, \mu_F^2) f_j(x_2, \mu_F^2) \hat{\sigma}_{ij\to ab}(\hat{s}, \mu_F^2, \mu_R^2, \alpha_S)\, \mathrm{d}x_1\, \mathrm{d}x_2, \tag{1.22}$$

where $f_{i(j)}(x_{1(2)}, \mu_F^2)$ stands for the PDF of each of the incoming protons, for a factorization scale μ_F, describing the probability for a parton with flavour i (j) to carry a fraction x_1 (x_2) of the proton's momentum. The term $\hat{\sigma}_{ij\to ab}$ is the ME calculation for the partonic cross section. It is a function of the centre-of-mass energy of the collision, $\hat{s}$, as well as of the scales μ_F, μ_R, and α_S. The cross section has to be independent of the choice of scale when all terms of the perturbation series are considered. In practice, fixed-order calculations are performed, where only the first few terms of the series are included. This introduces a scale dependence in the cross section prediction, which is considered as a theoretical uncertainty in measurements.

Parton distribution functions

A precise knowledge of the PDFs is crucial to describe the initial state of the system in pp collisions. As mentioned earlier, they cannot be computed perturbatively but are extracted from fits to the experimental data. The set of PDFs used in the analyses presented in this thesis is known as the NNPDF3.1 NNLO PDFs [30]. They are extracted using data from deep inelastic scattering events in electron-proton collisions, fixed-target Drell-Yan events, and $\mathrm{p\overline{p}}$ and pp collision measurements from the Tevatron and the LHC, respectively. The Dokshitzer–Gribov–Lipatov–Altarelli–Parisi (DGLAP) equation [31–33] provides the dependence of the PDFs on the chosen μ_F. The data are fitted using a machine learning-based methodology, and the resulting PDFs are shown in figure 1.5. The figure shows the value of the PDF as a function of the Bjorken x, for two different choices of scale μ (μ_F and μ_R are both set to the same value μ). As shown in the figure, gluons typically carry a lower fraction of the proton momentum, while the valence quarks carry the highest fractions of the momentum. As the energy increases, the fraction of momentum carried by the valence quarks becomes smaller, and the frequency of sea quarks and gluons increases.

1.2.2 Simulating pp collisions

Measurements of SM processes are usually performed by comparing the experimental data to theory predictions. This requires two main ingredients: a very reliable modelling of the physical process, and an accurate simulation of the detector behaviour. If these are achieved, the information recorded in the detectors can be translated back to the original particles which were produced in the hard interaction.

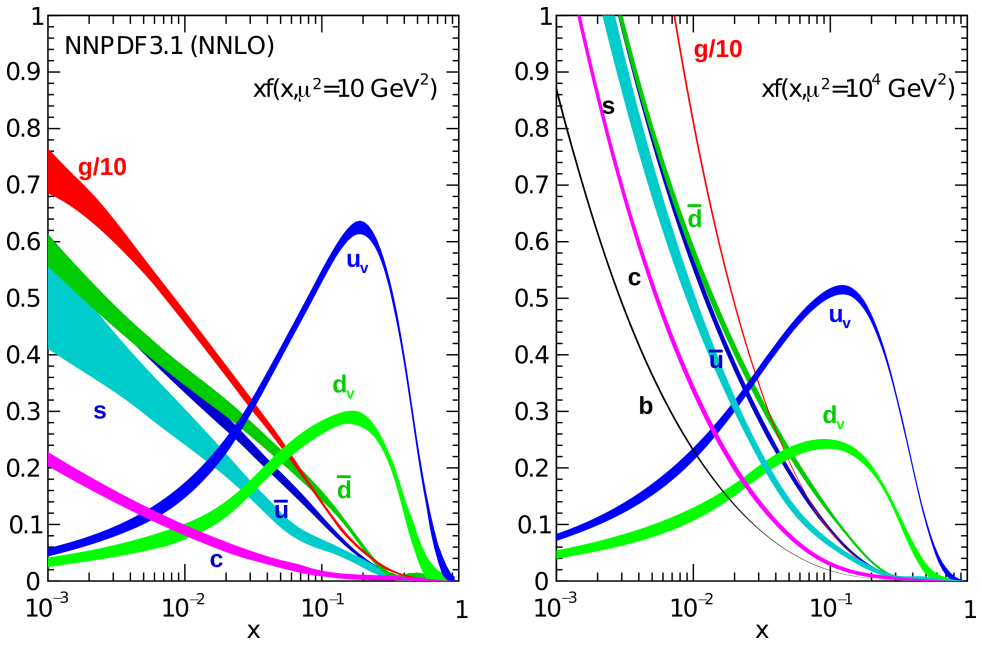

Figure 1.5: The NNPDF3.1 NNLO PDFs, evaluated at $\mu^2 = 10\,\mathrm{GeV}^2$ (left) and $\mu^2 = 10^4\,\mathrm{GeV}^2$ (right) [30]. The quarks with subscript v represent the valence quarks, while the remaining ones represent sea quarks. The curves corresponding to gluons, in red, are divided by a factor of 10 to make them visible in the figure.

MC modelling

To simulate pp collisions, not only a simulation of the hard interaction is necessary, but also an accurate description of the way partons evolve into stable particles, including decay and hadronization, is crucial. Generators of MC events were developed to provide this description. They are numerical algorithms capable of producing (or "generating") random sequences of simulated events based on a modelled probability distribution. Such event generators combine complex theoretical inputs with phenomenological models in order to achieve the best possible description of real collisions. Figure 1.6 shows a simplified scheme of an example pp collision event, where a $t\bar{t}$ pair is produced. This scheme is based on the event generator PYTHIA [34], but can be used to illustrate the general structure of MC generation in particle physics. The scheme should be read from the centre to the periphery, where the central white circle represents the hard interaction, and the particles on the outside edges represent the stable final-state particles.

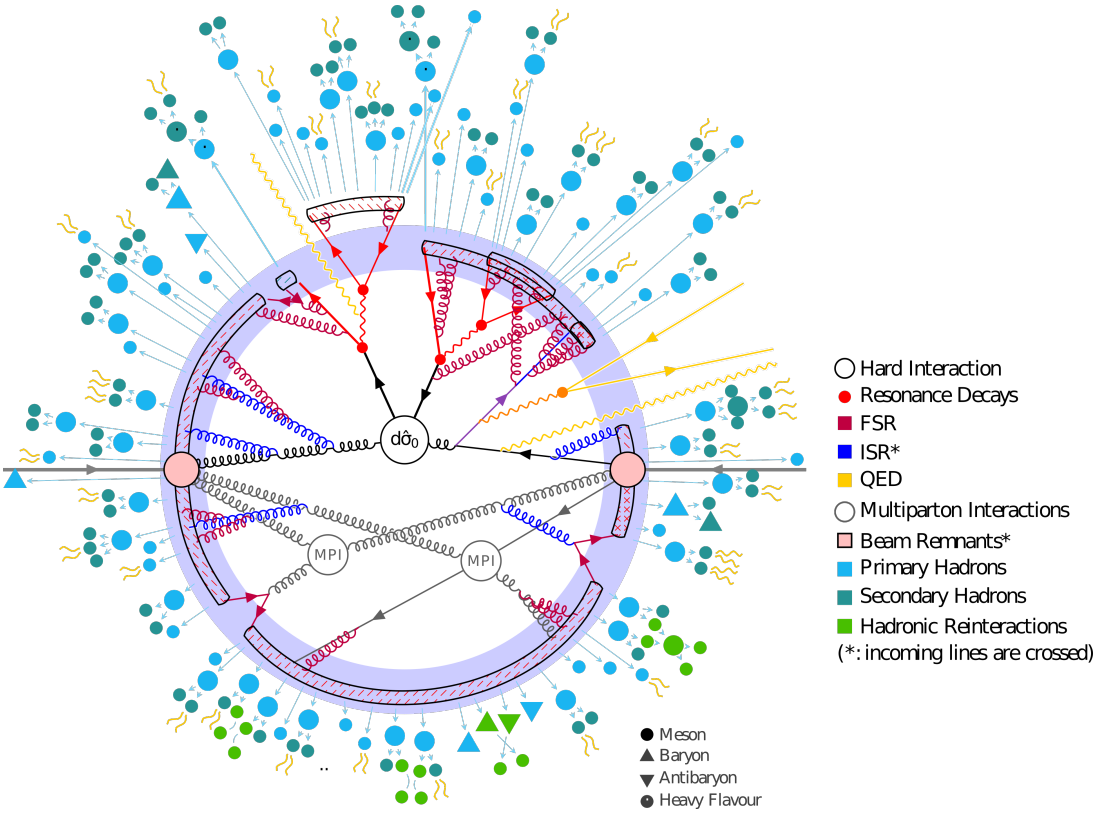

Figure 1.6: Schematic of the structure of a pp $\rightarrow t\bar{t}$ event, as modelled by PYTHIA, from Ref. [34] (modified).

The initial partons to interact are selected in a probabilistic way using the PDFs, and the type and kinematics of the outgoing particles are determined by the ME calculations,

as described in the previous section. The hard scattering may result in a short-lived resonance - in the case of the figure, a pair of top quarks - represented by solid black lines, which promptly decay into more stable particles.

Parton shower (PS) models are used to describe the emission of soft and collinear partons and photons, to account for additional orders in perturbation theory. Such emissions from initial state particles are known as initial-state radiation (ISR) and are represented in the figure in dark blue. Similar emissions but arising from the products of the hard scattering or their decay are called final-state radiation (FSR) and are represented in dark red. Fixed-order perturbative calculations describe well-separated emissions of particles, while the PS simulates soft and collinear emissions. Combining the two processes can result in double counting of emissions. This is avoided by using matching and merging schemes, such as MLM [35] at LO and FxFx [36] at NLO. Simultaneously, the proton remnants give rise to several other particles (denoted as *underlying event*) through Multiple Parton interactions (MPIs), as shown in grey.

Subsequently, partons from the hard scattering, PS, and MPIs are recombined according to colour potentials; this is described by colour reconnection models. The strong interaction finally causes these groups of partons to be confined into colour-singlet states of hadrons. Hadronization models are used to describe the decay of unstable hadrons until stable hadrons are formed.

Detector simulation

At the CMS experiment, a full simulation of the detector is performed using the GEANT4 package [37]. The particles emerging from the MC simulation, called "generator-level" particles, are propagated through the simulation, and modified according to the different detector effects, such as the magnetic field, scatterings in the detector materials, and bremsstrahlung. This simulation is tuned on real collision data, to make sure it provides an adequate description of the detector behaviour. The objects obtained after the simulation are called "detector-level" objects, and can be treated in the same way as collision data, i.e. be used as inputs to the particle reconstruction and identification algorithms described in section 2.3.2.

1.2.3 Elastic collisions - central exclusive production

The central exclusive production (CEP) of a system of particles X, i.e. $\mathrm{pp} \to \mathrm{p} + X + \mathrm{p}$, occurs when X is created in an elastic pp collision. A colour-singlet exchange of photons, gluons, or Z bosons may occur between the interacting protons and X is created, while the protons are not disrupted and leave the collision intact. The only collision products are the system X, e.g. ee, $\mu\mu$, $\gamma\gamma$, WW, ZZ, $\mathrm{Z}\gamma$ or $\mathrm{t\bar{t}}$, and where relevant, their decays products, and there are no proton remnants.

CEP processes can be purely photon-induced (QED) or have both photon-induced and gluon-induced (QCD) components, as shown schematically in figure 1.7. In the QCD case, a so-called "pomeron" is exchanged. Pomerons can be described at the lowest order in perturbation theory as colour-singlet two-gluon exchanges [38, 39].

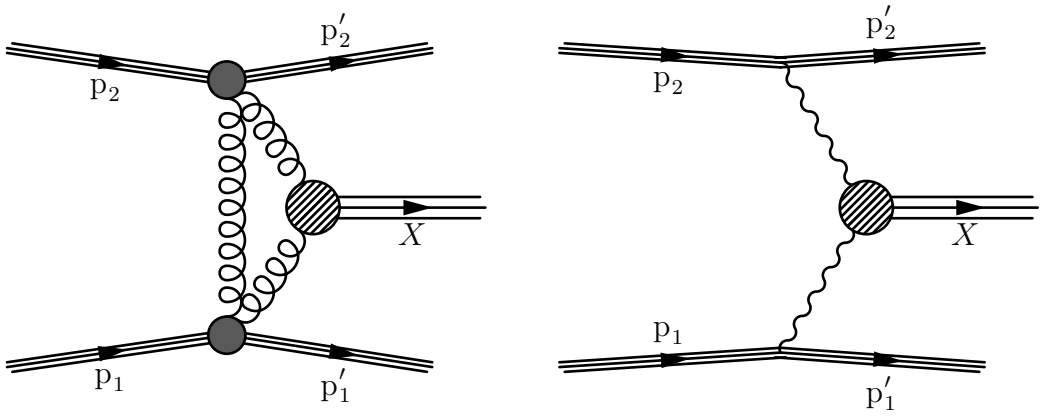

Figure 1.7: Diagrams representing CEP via pomeron exchange (QCD, left) and via photon exchange (QED, right).

A critical element in CEP, in particular in the case of strong interaction processes, is the proton survival probability. This quantifies the probability that no additional soft interactions between the spectator partons of the colliding protons take place, that is, it quantifies the probability that the process is truly exclusive and there is no underlying event. It is also named "rapidity gap survival probability" since these additional particles would be produced with a wide angular distribution and would fill the gap in the detectors that is characteristic of CEP.

CEP processes are extremely rare at high energy hadron colliders. However, they offer a unique playground for measuring physics processes at high energy in a relatively background-free environment, due to the absence of proton remnants. Moreover, if the outgoing intact protons can be tagged and their momentum measured, the full kinematics of the process can be reconstructed with very high resolution. Such a measurement was recently performed by the CMS collaboration in Ref. [40], where the CEP of electron and muon pairs (e^+e^- and $\mu^+\mu^-$) is measured by combining data from the CMS detector with tagged protons from PPS. Fig. 1.8 (left) shows the observed events in data (black) compared to the estimated background from non-CEP processes (in red), and the fit to the sum of the background and the $pp \to p\mu^+\mu^-p$ signal. In Fig. 1.8 (right) only the signal after subtracting the estimated contribution from the background is shown.

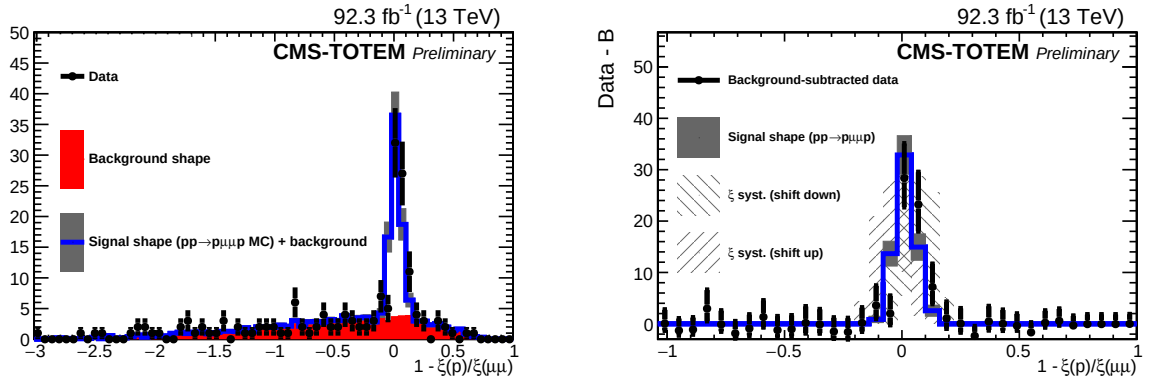

Figure 1.8: Correlation between the kinematics of the intact protons and the dimuon system, for data collected by the CMS experiment in 2017 and 2018, showing a clear peak corresponding to the observation of CEP of muon pairs [40]. The figure on the left shows the events observed in data after event selection, as well as the estimated background (in red) and signal+background (in blue) contributions. The figure on the right shows these data after subtraction of the background contribution.

1.3 The top quark

A third generation of quarks was first postulated in 1973, by Makoto Kobayashi and Toshihide Maskawa, to explain the observed CP violation in kaon decays [23]. The name "top" quark was later introduced in 1975, by Haim Harari [41], to designate the up-type quark of the third generation. Early searches for this particle, for example, at the Large Electron Positron (LEP) collider, were assuming the top quark to have a mass smaller than that of the W boson, so that the top quark could be a decay product of the W boson. The LEP results did not reveal any observation of the top quark, and the final combination results suggested a very high top mass, which led to a change of paradigm. Finally in 1995, the CDF and D0 collaborations [42] reported the observation of top quark production in proton-antiproton collisions. Indeed, the top quark has a considerably large mass: with a mass of $172.76 \pm 0.30\,\text{GeV}$ [21], it is by far the heaviest particle of the SM. It is also the only elementary particle with a mass close to the electroweak symmetry breaking energy scale, 246 GeV.

The large mass of the top quark makes it an interesting object of study and a potential window to BSM physics. On one hand, a large mass implies a large coupling to the Higgs boson and, therefore, studying the top quark can offer insight into the EW sector. On the other hand, the top quark has a very short lifetime, about $5 \times 10^{-25}\,\text{s}$ [21], shorter than the hadronisation time mentioned in section 1.1.2, making it the only quark that decays before hadronisation. Hence, its properties, such as the spin and electric charge, are passed on to its decay products. This presents a unique opportunity to study a bare

quark and measure its properties.

Top quark pair production

At hadron colliders, the dominant production mode of top quarks is via the strong interaction, resulting in the production of top quark-antiquark pairs ($t\bar{t}$), either via gluon fusion or quark-antiquark annihilation, as shown in the diagrams of figure 1.9.

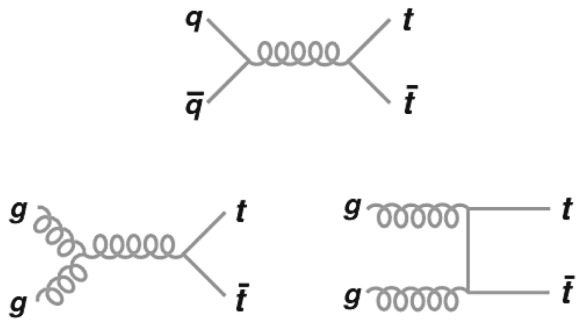

Figure 1.9: Feynman diagrams for top quark pair production at hadron colliders.

At the LHC, $t\bar{t}$ production is dominated by gluon fusion, which accounts for about 90% of the cross section at 13 TeV. Each top quark decays almost exclusively into a W boson and a b quark. A $t\bar{t}$ event is characterized by the decay of the W bosons. If both W bosons decay into a charged lepton and a neutrino, the event is labelled as dileptonic; if both W bosons decay into quark pairs, it is labelled as fully-hadronic; and in the case where one W boson decays into lepton and neutrino, while the other one decays into quarks, the event is labelled as lepton+jets (ℓ + jets). The fraction of events decaying through each channel is summarised in table 1.10. It is obtained by applying combinatorics to the branching ratios of the top quark found in Ref. [21]. The dileptonic channel has the smallest branching fraction, but it has the advantage of having the lowest background contribution. Additionally, it contains only two jets in the final state, leading to lower uncertainties, as jets are typically measured less precisely than leptons. The disadvantage of the dileptonic channel is the difficult reconstruction of the $t\bar{t}$ system, due to the presence of two neutrinos. This thesis focuses on the dileptonic channel. In particular, the focus is on events with electrons or muons in the final state. Tau leptons are not explicitly considered but are still included in case they decay into electrons or muons.

The LHC experiments have measured the inclusive $t\bar{t}$ production cross section in pp interactions at various centre-of-mass energies, using different top quark decay channels [43–53]. The dependence of the $t\bar{t}$ cross section on the energy has been studied at the LHC and compared to NNLO theoretical predictions, as shown in figure 1.11. Top

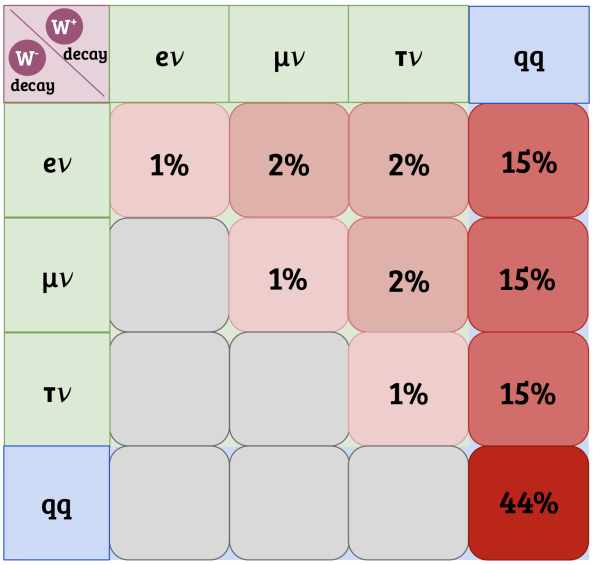

Figure 1.10: Branching ratios for the different top quark pair decay channels.

quarks can also be produced singly in electroweak processes, in three different modes known as t channel, s channel, and tW production. The ATLAS and CMS collaborations have observed or reported evidence for single top quark production in all three modes at different centre-of-mass energies [54–57].

Taking advantage of the abundant production of top quark pairs at the LHC, more exclusive final states than $t\bar{t}$, such as $t\bar{t}$+V, where $V = \gamma, Z, H, W^{\pm}$, become accessible and can offer great additional insight on the structure of the SM, despite having cross sections orders of magnitude smaller than that of $t\bar{t}$. At the CMS and ATLAS experiments, all these processes have been measured, and a summary of cross section measurements for these processes by both experiments at a centre of mass energy of 13 TeV is shown in figure 1.12, together with the most accurate available theoretical predictions. Production of single top quarks in association with a photon or a Z-boson have also been measured by CMS and ATLAS [59–64]. Additionally, CMS recently reported the first evidence for the production of tWZ [65].

1.3.1 Top quark coupling with the photon

The tbW vertex has been extensively studied using top quark decays in $t\bar{t}$ and single top production. Experimentally, it is accessible through precision measurements of W helicities and top quark spin correlations [66–68]. The electroweak couplings, i.e., the coupling with the Z, γ, and the Higgs boson, have been much less explored. Understanding precisely the strength and structure of the electroweak $t\gamma$ coupling constitutes a powerful test of the predictions of the SM, and may therefore be a gateway to new physics. Presently,

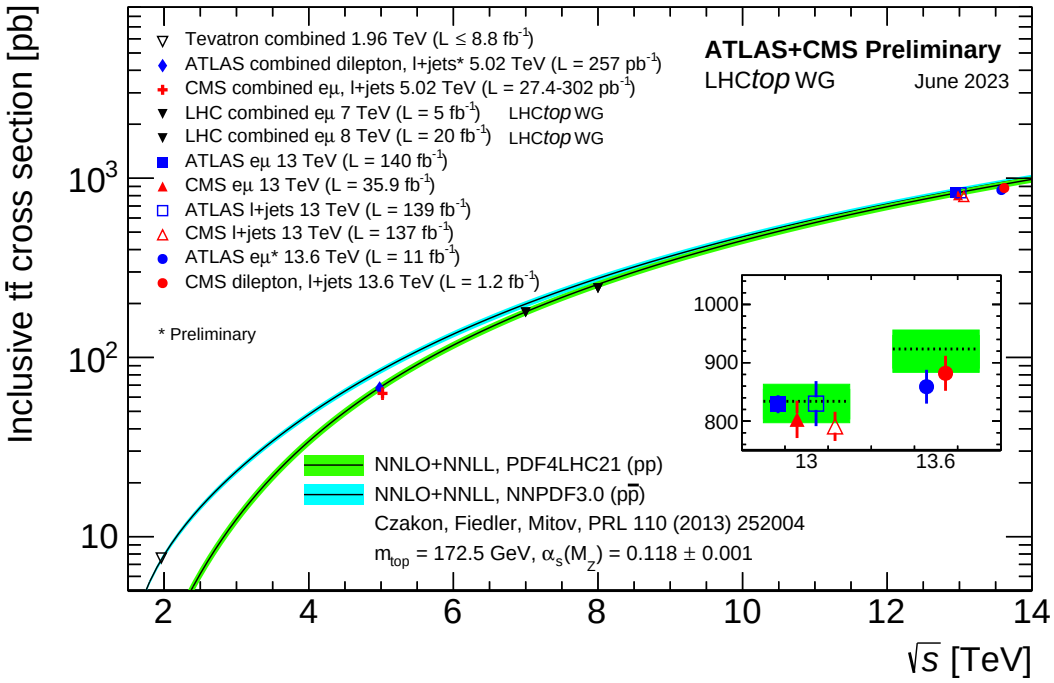

Figure 1.11: Top quark pair production cross section summary of ATLAS and CMS measurements, in comparison with the theory calculation at NNLO+NNLL accuracy. The Tevatron combined measurement is also shown [58].

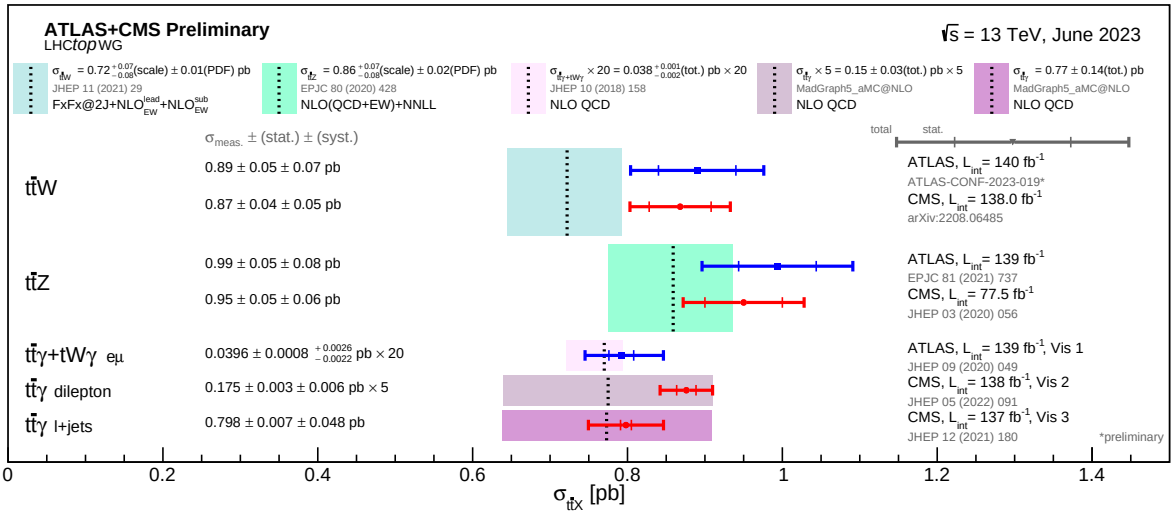

Figure 1.12: Summary of CMS and ATLAS $t\bar{t}$ + vector boson inclusive cross section measurements and comparison with the SM predictions [58].

the LHC works as a top factory, due to its high centre-of-mass energy and large amounts of data recorded. This means that top quarks are copiously produced, and processes involving the $t\gamma$ coupling, so far considered very rare, occur at sufficient rates as to allow precision measurements of this coupling.

Section 1.3.2 focuses on the potential of measuring the $t\gamma$ interaction when it comes to looking for new physics. A model-independent way to probe such effects that is currently widely used in high energy physics is Effective Field Theory (EFT).

The most relevant process when it comes to exploring this coupling is the associated production of $t\bar{t}$ with a photon ($t\bar{t}\gamma$). Measuring this process precisely has been possible only recently, and it is now crucial to study its modelling and sensitivity to new interactions. In this thesis, we perform a variety of differential measurements, and then go one step forward: looking at the ratio between $t\bar{t}\gamma$ and the $t\bar{t}$ production, we can reduce systematic uncertainties and achieve higher sensitivity to BSM effects. Section 1.3.3 provides a detailed explanation of the $t\bar{t}\gamma$ process, as well as an introduction to the differential and ratio measurements that are performed in this thesis. There, it is explicit that new structures in the top EW couplings would be shared between the $t\gamma$ and the tZ couplings [69]. Hence, in a bigger programme, one can combine $t\bar{t}\gamma$ measurements with other top quark associated production results to achieve an even higher sensitivity.

A more unexplored territory of processes involving $t\gamma$ couplings is the photon-induced $t\bar{t}$ production, occurring in the CEP of $t\bar{t}$. Details about CEP of top quark pairs are given in section 1.3.6, and a search for this process is also conducted in the context of this thesis.

1.3.2 Top quark and Effective Field Theory

Many theories that predict new physics postulate the existence of new particles and processes at an energy scale Λ, way above the TeV scale reached at the LHC. Such new physics, though not directly reachable, can still produce observable deviations in the lower energy distributions that we can presently measure with high precision. In the standard model EFT (SMEFT) approach, such deviations are parameterised in a model-independent way by extending the SM Lagrangian to include effective interactions between the SM fields. These interactions can be written as a sum of higher dimension operators $\mathcal{O}_i$:

$$\mathcal{L} = \mathcal{L}_{SM} + \sum_i C_i/\Lambda^2 \mathcal{O}_i. \tag{1.23}$$

The coefficients C_i are called Wilson coefficients. Here, the parametrisation of Ref. [69], with 59 dimension-6 operators, is followed. From these, 15 are associated with modifications of top quark interactions. The measurement of $t\bar{t}\gamma$ production, introduced in the next section, provides sensitivity to the electroweak dipole moments of the top quark. These are quantified by the coefficients $C_{uW}^{(33)}$ and $C_{uB}^{(33)}$.

One can write other coefficients which directly describe the modifications in the $t\bar{t}\gamma$ ($c_{t\gamma}$ and $c^{I}_{t\gamma}$) and $t\bar{t}$ Z (c_{tZ} and c^{I}_{tZ}) vertices, as a function of $c^{(33)}_{uW}$ and $c^{(33)}_{uB}$:

$$
\begin{aligned}
c_{tZ} &= \mathrm{Re}\left(-\sin\theta_W c^{(33)}_{uB} + \cos\theta_W c^{(33)}_{uW}\right), \\
c^{I}_{tZ} &= \mathrm{Im}\left(-\sin\theta_W c^{(33)}_{uB} + \cos\theta_W c^{(33)}_{uW}\right), \\
c_{t\gamma} &= \mathrm{Re}\left(\cos\theta_W c^{(33)}_{uB} - \sin\theta_W c^{(33)}_{uW}\right), \\
c^{I}_{t\gamma} &= \mathrm{Im}\left(\cos\theta_W c^{(33)}_{uB} - \sin\theta_W c^{(33)}_{uW}\right).
\end{aligned}
$$

Measuring the production of $t\bar{t}\gamma$ allows setting bounds on $c_{t\gamma}$ and $c^{I}_{t\gamma}$, or alternatively on c_{tZ} and c^{I}_{tZ}. Such a measurement by the CMS collaboration is described at the end of section 1.3.3.

1.3.3 Production of $t\bar{t}$ in association with a photon

When compared to other $t\bar{t} + V$ processes, $t\bar{t}\gamma$ has the highest production rate at the LHC, making it an excellent candidate for precision measurements. We define $t\bar{t}\gamma$ as the production of $t\bar{t}$ with an additional hard (high transverse momentum) isolated photon in the final state. The electroweak t-$\bar{t}$-γ vertex is present if the photon is emitted from one of the top quarks, which makes it the most interesting physics case, but this is not the only possibility. The $t\bar{t}\gamma$ process includes cases where the photon is radiated from the initial state quarks, as well as the decay products of the top quarks. This thesis focuses on the dileptonic $t\bar{t}$ decays, so the decay products that can radiate photons are in this case W bosons, b quarks, and charged leptons. Examples of Feynman diagrams for all these cases - photon from a top quark, from an initial state quark, and from a decay product at LO - are shown in the left, centre, and right of Fig. 1.13, respectively.

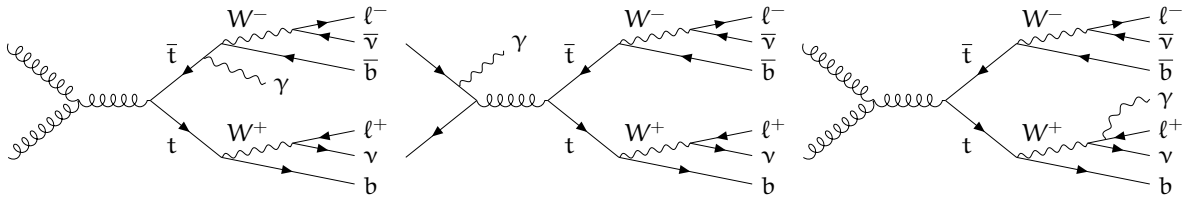

Figure 1.13: Examples of Feynman diagrams at LO for the production of $t\bar{t}\gamma$, decaying dileptonically.

Photons from different origins are not experimentally distinguishable. Moreover, diagrams with two resonant top quarks, as the ones in figure 1.13, cannot be experimentally distinguished from those with only one or even no resonant top quarks.

From a theoretical perspective, this means that a complete description can only be achieved by considering the full $2 \to 7$ process $pp \to \ell^+\nu_\ell\ell'^-\bar{\nu}_{\ell'}b\bar{b}\gamma$. The first full computation of $t\bar{t}\gamma$ at NLO in QCD, including all resonant and non-resonant diagrams, interferences, and off-shell effects of the top quarks and W bosons was performed by Bevilacqua and others in Ref. [70]. Based on their full calculation, the same authors demonstrated in Ref. [71] that approximately 57% of the photons are emitted either from the initial state quarks or from the off-shell top quarks (that go on-shell after emission). Conversely, 43% of the photons are emitted in the decay (from on-shell top quarks or their decay products). Hence, contributions from initial and final state emissions are almost equally relevant and considering only one of the cases cannot provide a suitable description of the experimental data. These fractions are altered when applying selection cuts to the physics objects, however sizeable contributions from the different photon origins will still be present. The simulation of $t\bar{t}\gamma$ used for the work of this thesis accounts for these effects, as detailed in section 3.2.1.

Measuring the inclusive cross section of this process is a direct probe of the electroweak $t\gamma$ coupling, and consequently of the top quark electric charge [72]. The top quark charge is known to be $Q_t = +2/3$ with a confidence level higher than $5\,\sigma$, consistent with the SM, and theoretical scenarios with $Q_t = -4/3$ are excluded. This has been measured in $t\bar{t}$ events, using the electric charges of the W boson and the b jet [73–76]. However, these are indirect measurements: the top quark charge has not been measured directly in processes where the fundamental $t\bar{t}\gamma$ interaction is potentially present. The inclusive measurement of $t\bar{t}\gamma$ is a "diluted" probe of the top quark electric charge. This is because only a fraction of the $t\bar{t}\gamma$ cross section is related to this coupling. Moreover, inclusive measurements do not provide information on the photon origin or on the relative contribution from different kinds of diagrams.

Performing differential measurements of $t\bar{t}\gamma$, that is, measuring its production cross section as a function of several kinematic observables, can bring additional sensitivity to the $t\gamma$ coupling and to all the theoretical aspects involved in the modelling of the process. For example, kinematic observables such as the p_T of the leptons or the angular distance ΔR between the photon and other particles, have a strong dependence on the origin of the radiated photon [71], as shown in figure 1.14. At the tail of such distributions (high p_T and large ΔR), the contribution from photons emitted by the top quarks is expected to dominate, which makes these observables very sensitive to BSM effects that potentially modify the strength and/or the structure of the t-$\bar{t}$-γ vertex.

1.3.4 Ratio between $t\bar{t}\gamma$ and $t\bar{t}$

Accurate theoretical predictions for $t\bar{t}\gamma$ are essential so that precision measurements can be compared to the theory and thus be sensitive to BSM effects. For $t\bar{t}$, calculations exist up to NNLO in QCD [77], but for $t\bar{t}\gamma$ going beyond NLO in QCD is currently not feasible. In Ref. [78], the authors present the possibility of achieving higher precision

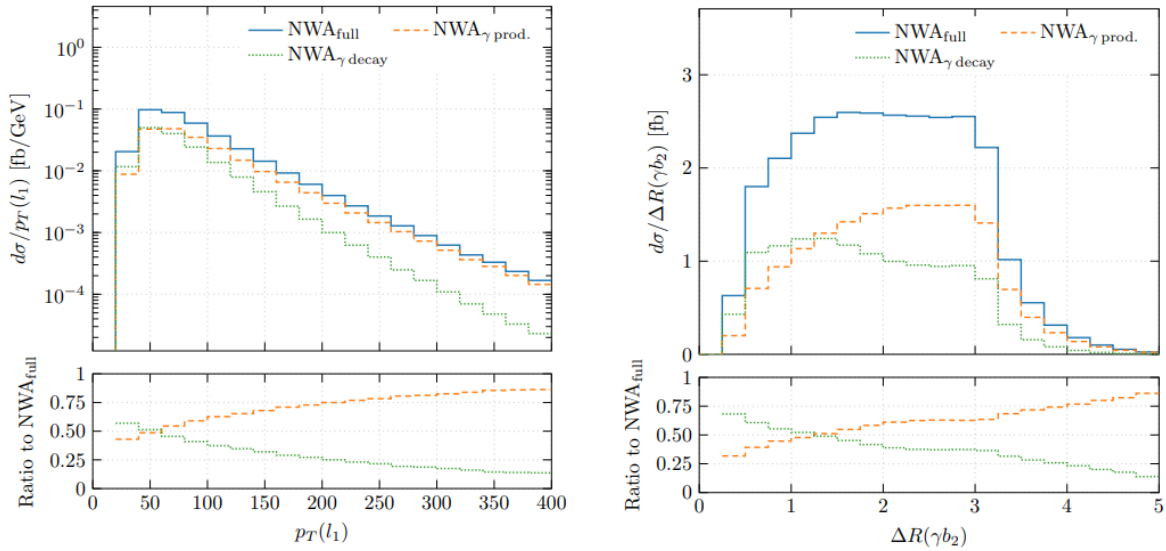

Figure 1.14: Theoretical predictions from Ref. [71] for the differential cross section of $pp \to e^+\nu_e\mu^-\bar{\nu}_\mu b\bar{b}\gamma$ as a function of the p_T of the leading (highest p_T) lepton (left) and the ΔR between the photon and the second leading b jet. The blue line shows the total contribution, while the orange and green lines show the contributions from photons emitted in the production and in the decay of $t\bar{t}$, respectively. The acronym "NWA" refers to the Narrow Width Approximation, a mathematical approximation that is used to derive these results, which relies on the fact that the top quark decay width is much smaller than its mass.

in $t\bar{t}\gamma$ predictions without the need to go beyond NLO, by taking the ratio between $t\bar{t}\gamma$ and $t\bar{t}$. When taking the ratio, many theoretical uncertainties cancel out and a higher precision can be achieved, when compared to predictions for $t\bar{t}\gamma$ alone. Moreover, ratios are more stable against radiative corrections and have reduced scale dependence [79].

From an experimental perspective, measuring the ratio also allows us to achieve a higher precision. Since $t\bar{t}$ and $t\bar{t}\gamma$ have very similar final states, many of the uncertainties associated with the experimental conditions and the reconstruction of physics objects are correlated and cancel out, leaving mainly statistical uncertainties and some remaining experimental uncertainties arising, for example, from the reconstruction of photons.

The SMEFT approach can also be used to constrain BSM scenarios in ratio measurements. In Ref. [79], both $t\bar{t}\gamma/t\bar{t}$ and $t\bar{t}Z/t\bar{t}$ ratios are analysed, and found to have complementary sensitivity to the $c_{uB}^{(33)}$ and $c_{uW}^{(33)}$ coefficients, as shown in figure 1.15. The plot on the left refers to $\mathcal{R}_\gamma$, the $t\bar{t}\gamma/t\bar{t}$ cross section ratio, while the plot at the centre shows $\mathcal{R}_Z$, the $t\bar{t}Z/t\bar{t}$ cross section ratio. The plot on the right is the result of a simple χ^2 combination between the two others. The two ratios bring complementary, almost orthogonal, sensitivity to the presence of the weak mixing angle in the $c_{uB}^{(33)}$ and $c_{uW}^{(33)}$ formulas. In other words, between R_γ and R_Z, there is a rotation between $c_{uB}^{(33)}$ and $c_{uW}^{(33)}$. This motivates future measurements of the $t\bar{t}Z/t\bar{t}$ cross section ratio as complementary results to the measurements presented in this thesis.

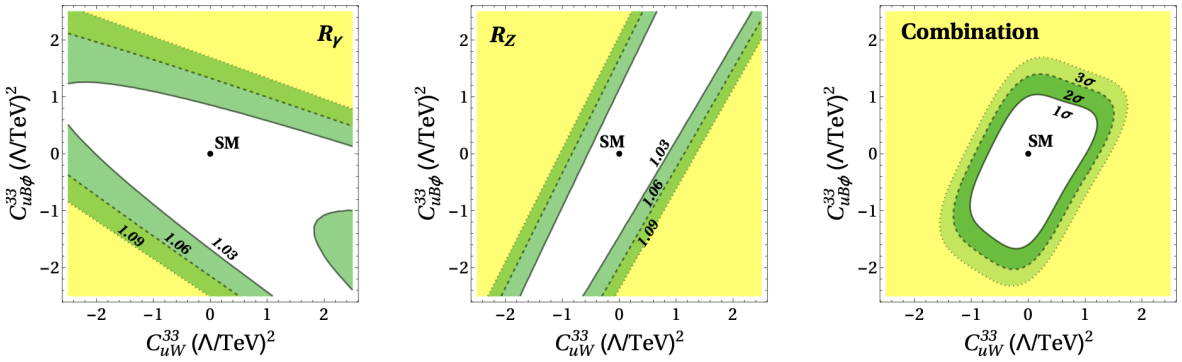

Figure 1.15: Cross section ratios R_γ (left) and R_Z (middle) normalized to their SM values (R_{SM}) as a function of the γ/Z anomalous dipole operator couplings. The contours show the deviation from the SM value in steps of 3, 6, and 9 per cent. On the right, we show the 1, 2, 3σ contours from combining R_γ and R_Z with an assumed uncertainty of 3%. From Ref. [79].

Differential cross section ratios

While the inclusive ratio measurement between $t\bar{t}\gamma$ and $t\bar{t}$ can already result in higher precision and new physics sensitivity, more interesting conclusions can be achieved in differential measurements. For example, correlations between the $t\bar{t}$ and $t\bar{t}\gamma$ processes can depend on the region of the phase space, and this can be explored in differential ratio measurements using strategically chosen observables. In Ref. [78], the authors define differential cross section ratios,

$$\mathcal{R}_X = \left(\frac{d\sigma_{t\bar{t}\gamma}}{dX}\right)\left(\frac{d\sigma_{t\bar{t}}}{dX}\right)^{-1}, \tag{1.24}$$

where X is any observable that can be measured in $t\bar{t}$ and $t\bar{t}\gamma$ events. Not all observables offer the same amount of information and sensitivity to BSM effects and the $t\bar{t}\gamma$ modelling. Dimensionless observables such as rapidities or angular distances do not reveal large shape differences between $t\bar{t}\gamma$ and $t\bar{t}$. Dimensionful observables, on the other hand, especially those related to the top quark, like the p_{T} of the top quarks or the invariant mass of the $t\bar{t}$ system, $m_{t\bar{t}}$, show clearly a harder spectrum for $t\bar{t}\gamma$ than for $t\bar{t}$, thus making them very promising for differential ratio measurements [78]. Similarly, the invariant mass of the dilepton system, $m_{\ell\ell}$, presents a different spectrum between $t\bar{t}\gamma$ and $t\bar{t}$, as shown in figure 1.16. The advantage of the latter is that it does not require the reconstruction of top quarks, which is difficult in dileptonic $t\bar{t}$ events.

1.3.5 Top quark charge asymmetry

In $t\bar{t}$ production, charge asymmetry refers to an anisotropy in the angular distributions of the final-state top quark and antiquark. Despite its name, it does not refer to an asymmetry in the charge conjugation operator C, but to an asymmetry when exchanging the top quark t with the anti-quark $\bar{t}$ in the final state, i.e., $t \leftrightarrow \bar{t}$.

If we denote the difference between the rapidity of the top quark and antiquark as $\Delta y = y_t - y_{\bar{t}}$, we can define an observable to quantify this asymmetry, called the forward-backward asymmetry, A_{FB}, as follows:

$$A_{FB} = \frac{\sigma(\Delta y > 0) - \sigma(\Delta y < 0)}{\sigma(\Delta y > 0) + \sigma(\Delta y < 0)}. \tag{1.25}$$

As mentioned before, in hadron colliders $t\bar{t}$ is produced mainly in gluon fusion and $q\bar{q}$ annihilation. In the SM, gluon fusion is symmetric at all orders. The $q\bar{q}$ annihilation is also symmetric at leading order, but higher order effects bring multiple sources of asymmetry. At NLO in QCD, there is an interference between terms in the cross section that are odd under the interchange $t \leftrightarrow \bar{t}$ when keeping the initial quarks fixed, namely from diagrams with ISR and FSR gluon emission (Fig. 1.17, top), and from the box and lower-order diagrams (Fig. 1.17, bottom) [80]. The first of these interference effects

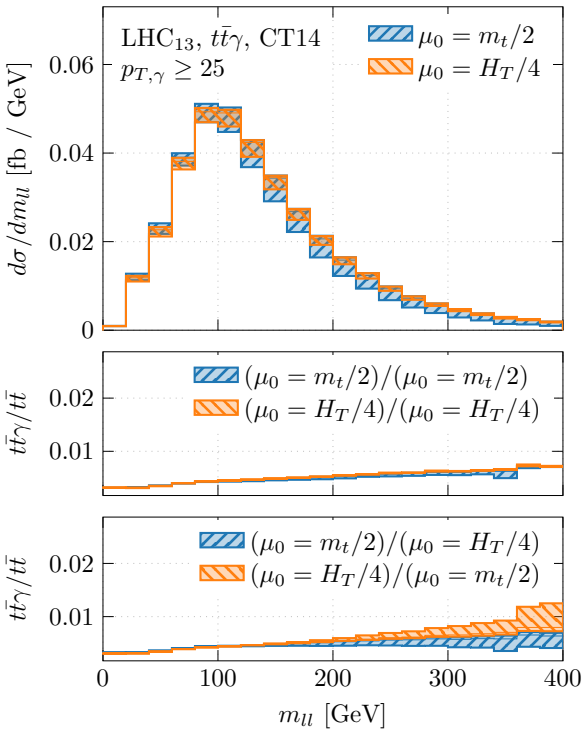

Figure 1.16: Differential cross section predictions as a function of the invariant mass f the two charged leptons for the pp $\to e^+\nu_e\mu^-\bar{\nu}_\mu b\bar{b}\gamma$ process at the LHC run II with $\sqrt{s} = 13\,\text{TeV}$. The upper panel shows absolute NLO predictions for $p_T(\gamma) > 25\,\text{GeV}$ together with the uncertainty band for renormalization and factorization scale variations. The lower panels show the differential cross section ratio together with their uncertainty bands, for different scale choices. From Ref. [78].

introduces a positive asymmetry, while the other introduces a negative asymmetry. The overall consequence of these asymmetries is that the top quark is preferentially produced in the direction of the incoming quark. Measuring this effect and comparing it to the existing SM predictions at NLO in QCD constitutes an interesting test of the SM, as there are several new physics scenarios which predict a different value for the asymmetry [81–85].

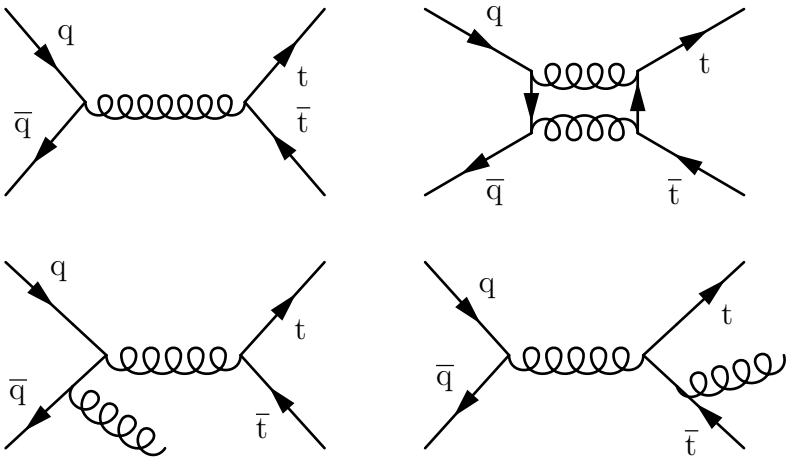

Figure 1.17: Top: $t\bar{t}$ production via $q\overline{q}$ annihilation, at the lowest order in QCD (left) and in a box diagram (right). Bottom: $t\bar{t}$ +g production via $q\overline{q}$ annihilation (NLO in QCD), with the gluon emitted from ISR (left) and FSR (right).

The top quark forward-backward asymmetry was measured for the first time at the Tevatron $p\overline{p}$ collider, by the D0 and CDF collaborations [86, 87].

At the LHC pp collider, $t\bar{t}$ production is dominated (≈89%) by gluon fusion. Therefore, the asymmetry is heavily diluted. Moreover, since the initial state is symmetric (two protons), the effect that is present in the Tevatron vanishes, and there is no A_{FB}. Nevertheless, since the valence quarks in the proton carry a larger average momentum than the antiquarks from the sea, a forward top quark (with respect to the incoming quark direction) will tend to scatter with larger absolute rapidity $|y|$ than the antiquark. With this in mind, one can define a forward-central asymmetry, A_C (from now on referred to simply as charge asymmetry), which considers the absolute values of the top quark and antiquark rapidities:

$$A_C = \frac{\sigma(\Delta|y| > 0) - \sigma(\Delta|y| < 0)}{\sigma(\Delta|y| > 0) + \sigma(\Delta|y| < 0)}, \tag{1.26}$$

with $\Delta|y| = |y_t| - |y_{\bar{t}}|$. The NLO prediction for this asymmetry at the LHC at a centre-of-mass energy of 13 TeV is about 0.6% [88].

Charge asymmetry in $t\bar{t}\gamma$

The production of $t\bar{t}\gamma$, when compared to $t\bar{t}$, is enhanced in the $q\bar{q}$ production mechanism, due to the middle diagram in figure 1.13. Therefore, the charge asymmetry effects are less washed out by the gg contribution than in the $t\bar{t}$ case. For this reason, even though the cross section, and thus the amount of data available, is much smaller for $t\bar{t}\gamma$ than for $t\bar{t}$, it is still worth studying $t\bar{t}\gamma$ events in this context.

Furthermore, the charge asymmetry in $t\bar{t}\gamma$, unlike in $t\bar{t}$, appears already at tree level in $q\bar{q} \rightarrow t\bar{t}\gamma$. This is caused by the interference between diagrams with ISR and FSR photons. The explanation is the same as for the gluon emission in $t\bar{t}$, except that the latter is an NLO effect. For this reason, the asymmetry is expected to be larger for $t\bar{t}\gamma$ than it is for $t\bar{t}$ [89].

1.3.6 Central exclusive production of $t\bar{t}$

The CEP mechanism described in section 1.2.3 can lead to the production of top quark-antiquark pairs in pp scattering. The process $pp \rightarrow pt\bar{t}p$ receives contributions from QED and QCD diagrams, as sketched in Fig. 1.18.

Figure 1.18: Leading diagrams for CEP of $t\bar{t}$ via $\gamma\gamma$ fusion (left) and pomeron exchange (right).

The diagram with photon exchange dominates at the energy scales achieved in current hadron colliders, while the pomeron exchange is heavily suppressed. Predictions for CEP of $t\bar{t}$ in the framework of the SM are available, including both QED and QCD contributions [90–95]. The predicted cross section amounts to $0.22 \pm 0.05\,\text{fb}$, for pp collisions at $\sqrt{s} = 13\,\text{TeV}$, including NLO perturbative QCD corrections [96]. This small cross section could be enhanced in case there are BSM contributions. In particular, the process is sensitive to the $t\gamma$ vertex (present twice in the QED diagram), which makes it suitable for interpretations in the context of EFT [69] or anomalous couplings [92, 97]. The process was also shown to be sensitive to BSM models including extra dimensions [98]. Therefore, studying CEP of $t\bar{t}$ can, despite the small cross section in the SM, offer complementary information to precision measurements in $t\bar{t}\gamma$.

CHAPTER 2

The LHC and the CMS experiment

The collision data used to perform the work of this thesis were produced at the Large Hadron Collider (LHC) at the Conseil Européen pour la Recherche Nucléaire (CERN) in Switzerland, the largest and most powerful particle accelerator to date.

The LHC accelerator complex, where protons are produced and accelerated, is briefly described in section 2.1. Protons are subsequently collided at several interaction points (IPs), one of them housing the CMS detector, a general-purpose particle detector which is used to collect the data studied in this thesis. Different components of the CMS detector are detailed in section 2.2. Section 2.2.6 describes the Precision Proton Spectrometer (PPS), a subdetector of CMS which is located near the LHC beam pipe, at about 200 m from each side of CMS, and is able to detect protons that leave the collisions intact.

A critical concept when talking about particle accelerators is the luminosity, which quantifies the number of collisions and consequently the amount of data produced. The concept of luminosity is discussed in section 2.4, and the main method to measure it is explained. Additionally, some results are shown for the analysis of emittance scans at the LHC, which are used to correct the luminosity measurements as well as estimate their uncertainties.

Finally, section 2.3 describes how the information provided by the various layers of the CMS detector is used to reconstruct the different physics objects.

2.1 The Large Hadron Collider

In order to probe the predictions of the SM and push its boundaries towards new physics, particle colliders need to reach high centre-of-mass energies, $\sqrt{s}$, and produce large amounts of data. For that purpose, the LHC was built at CERN, crossing the border between Switzerland and France. With a circumference length of about 27 km, it is designed to collide protons at $\sqrt{s}$ up to 14 TeV, and heavy ions at about 5 TeV per

nucleon. The name chosen for the LHC contains the three main ingredients that allow it to achieve such high energies. The first is its size (*Large*), because in circular accelerators the radius is directly related to the maximum energy that can be obtained. The second is the fact that it is a hadronic collider (*Hadron*), since accelerating hadrons in a circular trajectory (as opposed to, for example, electrons) results in a much smaller energy loss through radiation. The last word is *Collider*, and indeed collider experiments with counter-circulating beams can achieve higher energies than fixed-target ones, since the $\sqrt{s}$ in the former is given by the sum of the energies of the colliding particles.

Colliding hadrons at high energies, however, comes with its challenges: as explained in section 1.2, in a pp collision there are typically many final state particles due to the presence of proton remnants. This does not happen in an e^+e^- machine such as the Large Electron-Positron Collider (LEP, 1989–2000) [99], that produced much "cleaner" events, i.e., events with fewer particles in the final state. Therefore, lepton colliders are often regarded as "precision machines", while hadron colliders are seen as "discovery machines". In other words, hadron colliders are capable of producing new high-mass resonances, but not of performing precision measurements. Interestingly, the LHC has proven its capability of not only discovering new particles, such as the Higgs boson, but also performing a wide variety of unprecedentedly precise measurements. The machine is currently in its third data-taking run, known as Run 3, colliding protons at $\sqrt{s}$ =13.6 TeV. It has successfully completed two runs before: Run 1 between 2010 and 2012, at $\sqrt{s}$ =7 and 8 TeV, and Run 2, between 2016 and 2018, at $\sqrt{s}$ =13 TeV. This thesis focuses on data collected during Run 2.

In the future, the LHC is expected to undergo a major upgrade, into what is known as the High Luminosity LHC (HL-LHC), which is planned to be ready for data taking by 2028. It is planned that it will run at a centre-of-mass energy of 14 TeV, but its main novelty will be the unprecedented amounts of data (integrated luminosity) that it is expected to collect [100]. It will allow us to pin down rare SM processes such as the production of two Higgs bosons (HH), and thus improve our current understanding of the SM.

The proton journey: from a small bottle to the LHC

The LHC itself is the last element of a complex accelerator chain, as shown in figure 2.1. Protons are first produced at the linear accelerator Linac4 [101]. This accelerator is filled with negative hydrogen atoms, H^-, coming from a small gas bottle, and uses radiofrequency (RF) cavities to accelerate them to 160 MeV. The ions are then stripped of their electrons, and the resulting protons are injected into the Proton Synchrotron Booster (PSB). There, they are accelerated to 2 GeV and continue to the Proton Synchrotron (PS), which brings their energy up to 26 GeV. The next step is the Super Proton Synchrotron (SPS), which accelerates the protons to 450 GeV. At this point, they are ready to enter the LHC and start the last phase of acceleration, all the way to 6.5 TeV (in Run 2). In

a proton beam, there are up to 2808 bunches, with about 1.2×10^{11} protons each; it takes about 4 minutes to fill the LHC ring and 20 minutes for the protons to reach their maximum energy.

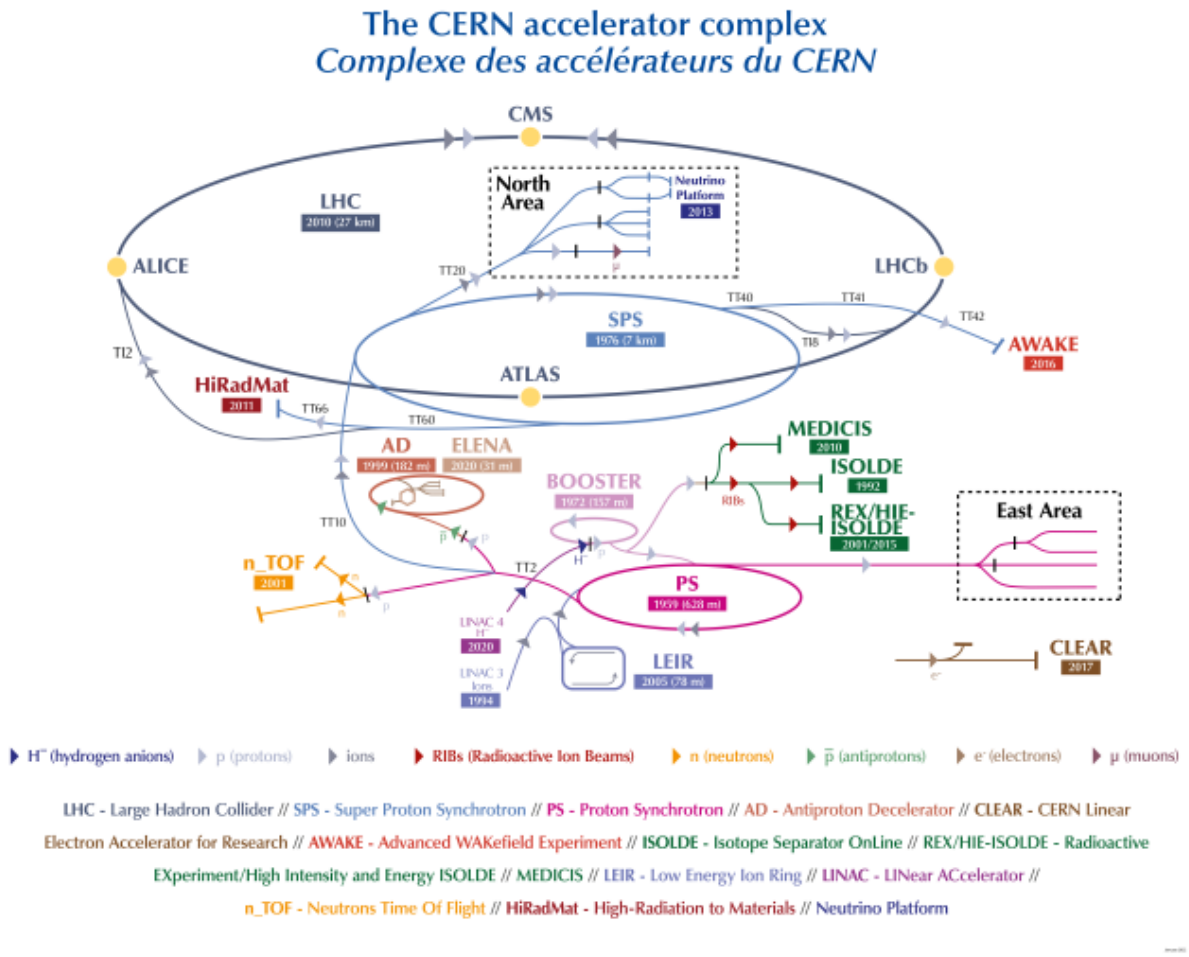

Figure 2.1: CERN accelerator complex, from Ref. [102].

The LHC machine

The LHC is composed of two beam pipes, which are tubes kept at ultralight vacuum. It is not a perfect circle, being composed of eight straight sections, called "insertions", connected by curved arcs. The curved arcs are called "sectors" and contain the bending dipole magnets, 154 per arc, for a total of 1232. These magnets are superconducting dipoles which operate at a temperature of 1.9 K [102,103]. The straight insertions contain

the IPs where the particle detectors are located, as well as other auxiliary components, for example for injection, beam dumping and beam cleaning.

Apart from the main dipole magnets that bend the beam, the LHC is equipped with quadrupole magnets that keep the beam focused, by squeezing it either vertically or horizontally. The dipoles are also equipped with sextupole, octupole and decapole magnets, which correct for small imperfections in the magnetic field near the ends of the dipoles.

To accelerate the protons from 450 GeV up to their final energy, eight RF cavities per beam are installed in the insertions of the machine. They are superconducting and operate at 4.5 K.

The accelerated and focused proton beams are allowed to interact at four different IPs along the tunnel. The bunches collide every 25 ns, which translates to about 40 million collisions per second. Around each of these collision points are installed detectors of the main LHC experiments. The largest are the two general-purpose experiments: A Toroidal LHC Apparatus (ATLAS) [104] and the Compact Muon Solenoid (CMS). There is also LHCb [105], which focuses on the study of processes involving the bottom quark and ALICE [106], which mainly studies heavy-ion collisions.

2.2 The Compact Muon Solenoid

The CMS detector is a general-purpose detector, located at Point 5, one of the LHC caverns, about 100 m underground, near the village of Cessy, in France. It has a cylindrical shape, with a length of 21.6 m and a diameter of 14.6 m, and weighs a total of 14500 tons. It is built in an onion-like structure, with different layers designed to detect different particle properties. The general layout and main features of CMS are shown in figure 2.2. Its most distinctive component is a 3.8 T solenoidal magnet which provides the magnetic field to bend the particles produced in the collision. From the centre outwards, it contains a tracker system, to register the trajectory of charged particles, a calorimeter system, where particles can interact and deposit their energy, and a set of muon chambers, which track the passage of muons. More details about each subdetector are given in sections 2.2.2 to 2.2.5. Section 2.2.6 describes PPS, a CMS subdetector which is not part of its main cylindrical structure, but is located a few hundred metres away in the LHC tunnel, and is used to detect protons which leave the CMS detector intact after each collision.

2.2.1 CMS coordinate system and definitions

Figure 2.3 shows two transverse views of the CMS detector, with the conventional coordinate system marked with green arrows. The polar angles θ and ϕ, defined with respect to the particle direction, are also shown.

The transverse momentum $\vec{p_{\mathrm{T}}}$ is defined as the momentum of a particle in the transverse $(x - y)$ plane. Its magnitude is given by:

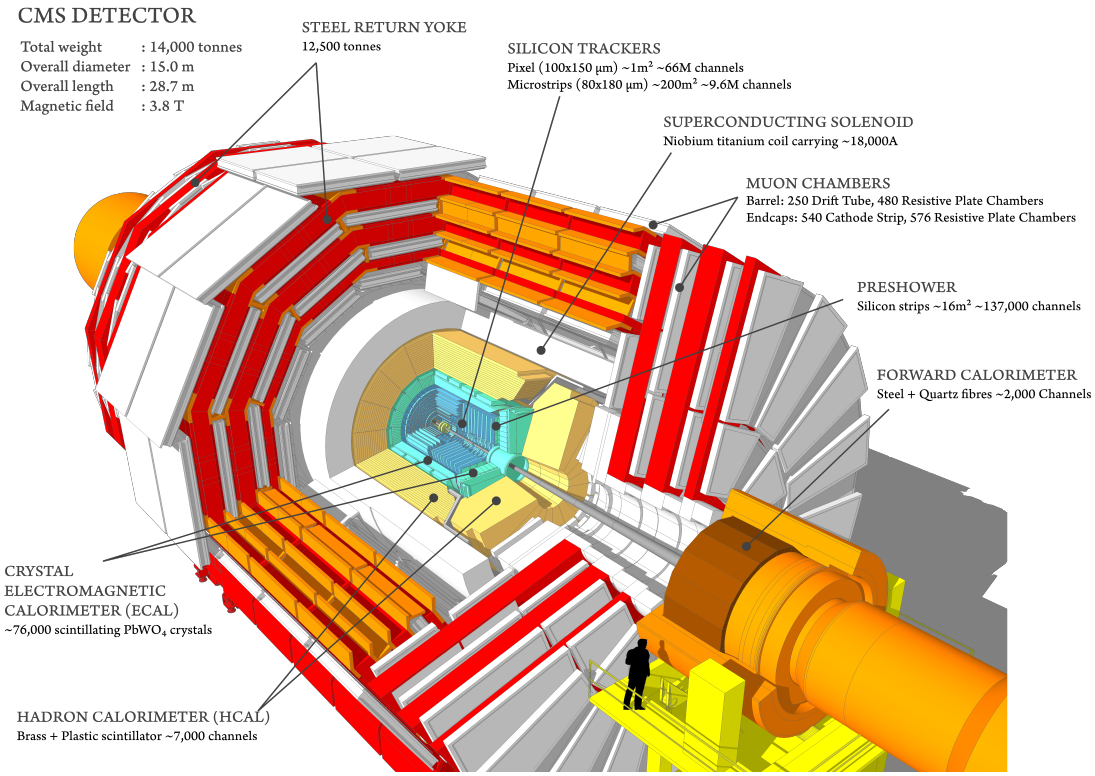

Figure 2.2: Schematic view of the CMS detector, showing the location of its various subdetectors. A person is shown in black for scale. Picture taken from Ref. [107].

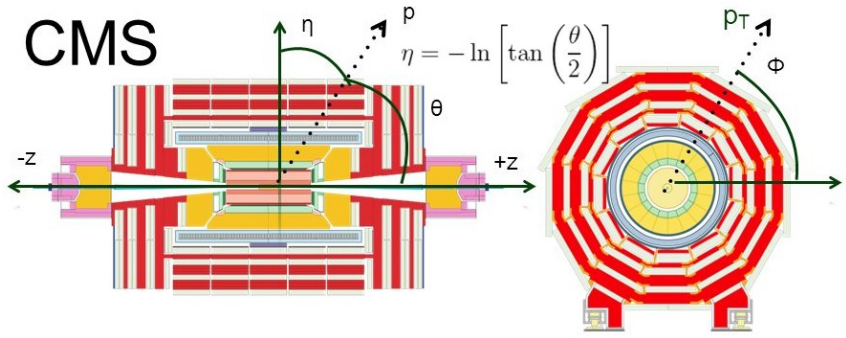

Figure 2.3: Schematic description of the geometry and coordinate system at the CMS experiment. The z-axis corresponds to the beam axis, the y axis is vertical and pointing upwards, and the x axis is horizontal, perpendicular to the beam and pointing towards the centre of the LHC. Adapted from Ref. [108].

$$p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2} \tag{2.1}$$

The p_{T} sum of all particles, including those escaping the detector without interaction, should be equal to zero. This is because the beams do not travel in the transverse plane. Therefore, a measured value that differs from zero is generally attributed to the presence of undetectable neutrinos in the event (or other unknown particles which would also be undetectable). The missing transverse momentum $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ is defined as the negative vector sum of the transverse momenta of all measured particles. Its magnitude $E_{\mathrm{T}}^{\mathrm{miss}}$ is commonly used to reconstruct the neutrino momenta and/or search for new "invisible" particles.

Apart from the energy and momentum observables, it is useful to take advantage of the angular information provided by the detectors. The pseudorapidity, η, defined as $-\ln\left[\tan\left(\theta/2\right)\right]$, is typically used to quantify how "centrally" a particle is produced. It ranges from 0 (for $\theta = 90^\circ$, a particle aligned with the transverse plane) to infinity (for a particle with $\theta = 0^\circ$).

An alternative variable to the pseudorapidity is the rapidity, y, calculated as

$$y \equiv \frac{1}{2} \ln\left(\frac{E + p_{\mathrm{z}}}{E - p_{\mathrm{z}}}\right) \quad . \tag{2.2}$$

In the relativistic limit, where the momentum of the particle is much larger than its mass ($E \approx |\vec{p}|$), η and y become approximately the same.

The angular separation between two particles is defined as the sum of the squared differences of their η and the angle ϕ. This is called ΔR, and defined as:

$$\Delta R \equiv \sqrt{\left(\Delta\eta\right)^2 + \left(\Delta\phi\right)^2} \quad . \tag{2.3}$$

2.2.2 The solenoid

The strong magnetic field surrounding the CMS IP is used to measure the electric charge and momenta of charged particles that originate in the collisions. Due to the Lorentz force $\vec{F}_L = q(\vec{v} \times \vec{B})$, charged particles are bent in the presence of a magnetic field. The direction of bending depends on the electric charge. The curvature radius indicates the magnitude of p_{T}. The solenoidal magnet of CMS has a length of 12.5 m and a diameter of 6.3 m, making it the largest device of this kind ever built. It is composed of four layers of niobium-titanium alloy and provides a magnetic field of 3.8 T, mostly uniform inside the solenoid. Inside the solenoid, the tracker and calorimeter systems are hosted. Similarly to the LHC dipoles, the niobium-titanium layers are cooled by a cryostat system operating with liquid helium at 4.7 K.

The steel return yoke is responsible for the return of the magnetic flux. It is composed of five barrel wheels and six endcap disks, weighing in total 10000 tonnes and accounting for most of the weight of the experiment. Apart from returning the magnetic flux, they

provide the mechanical structure for housing the muon detectors. They also act as a filter for these same detectors, i.e., by absorbing hadronic particles which may not be absorbed in the calorimeters, they ensure cleaner muon signals in the muon chambers.

2.2.3 The silicon tracker

The CMS tracker is the innermost sub-detector of CMS, immediately surrounding the IP. It is responsible for the precise determination of the trajectory of charged particles originating from the collision. Due to its location close to the IP, it has to withstand a very intense particle flux. The requirements on granularity and radiation hardness lead to the choice of silicon-based technologies. The tracker is composed of two main sub-detectors: the Pixel Tracker and the Silicon Strip Tracker, which together provide full coverage up to $|\eta| = 2.5$. Figure 2.4 shows a schematic cross section of the tracker, where the Pixel Tracker is shaded in red in the centre, and the remaining parts constitute the Silicon Strip Tracker.

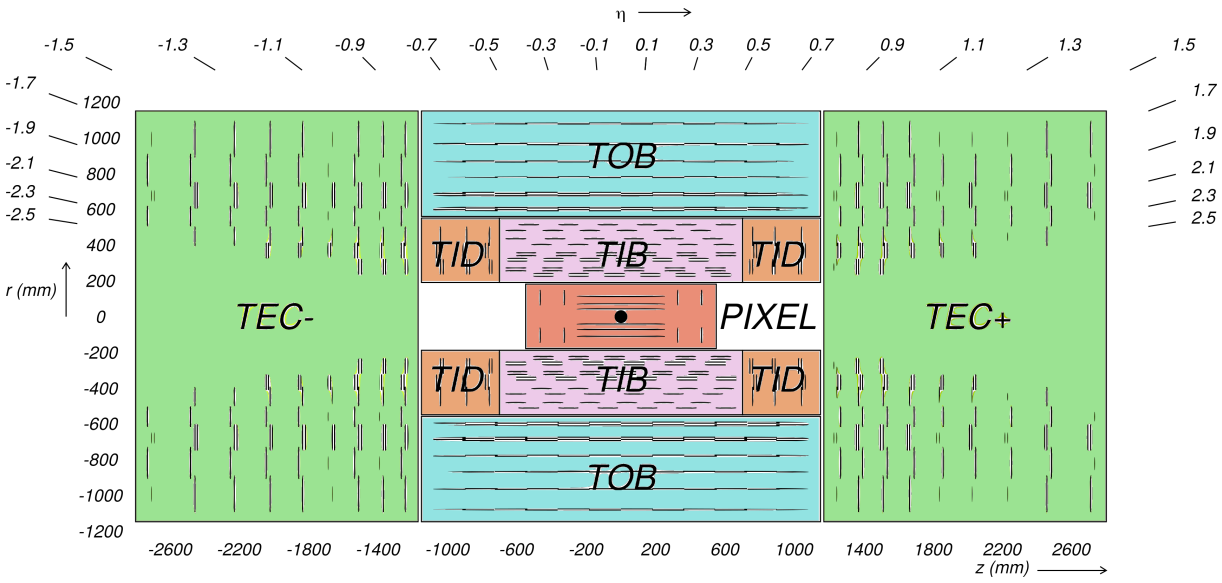

Figure 2.4: Schematic cross section of the CMS tracker. The lines represent the detector modules. Image adapted from Ref. [109].

The Pixel Tracker is the inner part of the tracker and is designed to provide precise measurements of charged particle trajectories and collision vertices. When it was built, it consisted of three cylindrical barrel layers and two endcap disks. Then, in 2016/2017, an upgrade took place and four barrel layers and three endcap disks were installed, to provide hit coverage up to an absolute value of $|\eta|$ of 2.5 [110]. The layers/disks are made up of pixel modules, each containing one or more pixel sensors, which are $100 \times 150\,\mu$m in size and $285\,\mu$m thick. The sensors are made of high-purity silicon and have a high spatial

resolution, providing precise measurements of the position of charged particles passing through them. The total active area of the Pixel Tracker is approximately $1\,\mathrm{m}^2$, and it contains about 66 million pixels.

The Silicon Strip Tracker is the outer part of the tracker and is designed to further measure the momenta of the charged particles and to provide information for particle identification. It is divided into the inner barrel part (TIB), the inner disks (TID), the outer barrel (TOB), and the outer endcaps (TEC), as shown in the figure. The TIB and TOB systems are composed, respectively, of four and six concentric layers. The TID system has three disk structures on each side, each divided into three concentric rings, while the TEC is made of nine disk structures on each side, each made of four to seven rings. Each of these structures contains the tracker modules, which are composed of silicon strip sensors, 10 cm long and with a thickness of $320\,\mu$m. The sensors have a lower spatial resolution than the pixel sensors, but cover a much larger area, thus providing crucial complementary information. The Silicon Strip Tracker has a total active area of about $200\,\mathrm{m}^2$ and contains about 15 million silicon strips. The strips are read out by electronics located at the end of each module, and the data are sent to the CMS data acquisition system for processing and analysis.

Together, the Pixel Tracker and Silicon Strip Tracker provide precise measurements of charged particle trajectories, momenta, and electric charge. The tracker is a crucial component of the CMS detector, and its high precision and reliability are essential for the success of the experiment.

2.2.4 The calorimeters

The calorimeters are located immediately after the tracker and are designed to precisely measure the energy of particles. While the electromagnetic calorimeter (ECAL) targets mainly electrons and photons, the hadron calorimeter (HCAL) aims to measure the energy of hadrons. Both calorimeters have a compact design so they could be installed inside the magnet structure.

Electromagnetic calorimeter

One of the physics goals of the CMS experiment, at the time of its design, was to search for the Higgs boson in its decays to two photons or two Z bosons. That goal was eventually achieved and led to the discovery of this particle [27]. A key reason for this success was the excellent photon and electron energy resolution provided by the ECAL. The ECAL detector is composed of a barrel (EB) structure and two endcaps (EE), each with a single layer of homogeneously distributed lead tungstate ($PbWO_4$) crystals. The crystals in the EB are organised in 36 supermodules, each with 1700 crystals, covering the region $|\eta| < 1.48$. In the EE, there are 4 "Dee" structures, each with 5382 crystals [111], covering the region $1.48 < |\eta| < 2.0$. One of the supermodules is indicated in figure 2.5 in orange

(and the crystals are represented in pink for another supermodule), while the crystals in the "Dees" are coloured in blue.

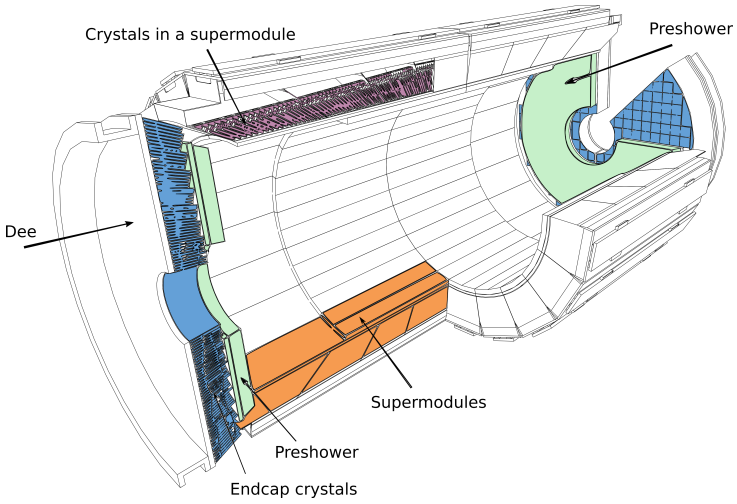

Figure 2.5: Scheme of the ECAL, highlighting its different components. The preshower detectors are coloured in light green. One of the supermodules that compose the barrel section is coloured in orange, while another one is shown without its cover so that the crystals, highlighted in pink, are visible. Finally, the crystals in the endcap "Dees" are shown in blue. Image adapted from Ref. [111].

In front of the endcaps, the preshower (ES) detector is installed, shown in light green in the figure. It contains 2 mm-long silicon strips, providing an improved spatial resolution in the region $1.65 < |\eta| < 2.6$. This is particularly important to improve the identification efficiency of low energy photons originating from decays of π^0 mesons.

The $PbWO_4$ crystals act simultaneously as absorbers and scintillators: when particles hit the ECAL, they interact with the material of the crystal and lose energy via bremsstrahlung. The emitted photons convert into electron and positron pairs, which in turn emit new photons. Those photons can undergo further conversion, continuing the chain. This process creates an electromagnetic shower. In the shower, electrons and photons lose almost all their energy (approximately 98% of the energy of electrons and photons below 1 TeV is absorbed). The ECAL has a length of approximately 1 nuclear interaction length (λ_I), therefore hadrons also lose a fraction of their energy in the ECAL.

Particles in the electromagnetic shower excite the atoms of the crystals, which subsequently relax by emitting an amount of blue light proportional to the deposited energy. This light is amplified by photomultipliers and collected in photodetectors, i.e., avalanche photodiodes in the EB and vacuum phototriodes in the EE.

The choice for $PbWO_4$ crystals is due to their short radiation length ($X_0 = 0.89\,\text{cm}$), low Molière radius (2.2 cm), quick scintillation response (80% of all the light is emitted within the 25 ns between two collisions), and radiation hardness.

Hadron calorimeter

The HCAL measures the energy of particles interacting mainly via the strong force. For neutral hadrons, it is the only subdetector that "sees" them, as they do not leave tracks in the tracker and do not deposit a significant fraction of their energy in the ECAL. For charged hadrons, the information from the HCAL complements the measurement from the tracker.

The HCAL is located outside the tracker and the ECAL and consists of four main parts: the hadron barrel (HB), the endcaps (HE), the outer subdetector (HO) and the forward calorimeter (HF). The layout of the HCAL is sketched in figure 2.6.

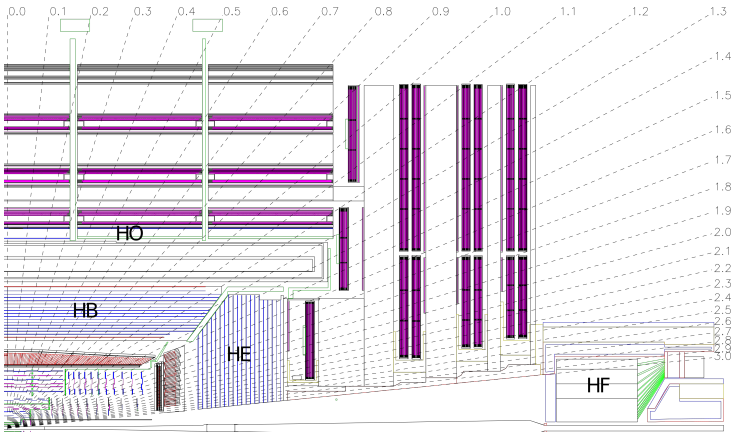

Figure 2.6: Layout of the HCAL, highlighting its different components. Image taken from Ref. [109].

The HB is divided into 36 azimuthal wedges and has a thickness of $5.8\lambda_I$. The HE are composed of two sections: inner and outer endcaps. The inner endcaps cover the range $1.3 < |\eta| < 3.0$, while the outer endcaps cover the region $3.0 < |\eta| < 5.0$. They are composed of 14 disks and have a thickness of $10\lambda_I$. Both HB and HE are made of brass absorber plates interleaved with plastic scintillator tiles.

The HO is located outside of the solenoidal magnet, extending the depth of the calorimeters to about $12\lambda_I$. It is composed of scintillators interleaved with the steel absorber plates of the magnet return yoke, and is designed to provide a complementary measurement of hadrons that escape the HCAL.

The HF is located at the very forward regions of the CMS detector, covering the region $2.9 < |\eta| < 5.2$, and is designed to measure the energy of hadrons produced at very small angles with respect to the beam axis. This region is especially challenging due to the high particle fluxes. Hence, the HF design requires very high radiation hardness. It is composed of steel absorber plates and a Cherenkov technology using quartz fibres placed behind plates of steel and borated polyethylene neutron shielding.

2.2.5 The muon chambers

Detecting muons with high precision was one of the core design specifications of the CMS detector, hence the name. One of the motivations was the search for the Higgs boson in its decay to two Z bosons that in turn decay to four muons. This was indeed one of the decay modes used for the discovery. In general, muons provide a very clear signature and are present in many interesting SM processes. Furthermore, they have a relatively long lifetime (2.2 μs [21]) and penetrate through the tracker, calorimeters, and solenoid, losing only a small fraction of their energy.

The muon chambers are the outermost subdetectors of CMS, located outside of the solenoidal magnet. The muon system is composed of three types of detectors: drift tubes (DTs), cathode strip chambers (CSCs), and resistive plate chambers (RPCs). A sketch of the layout of the muon chambers is shown in figure 2.7.

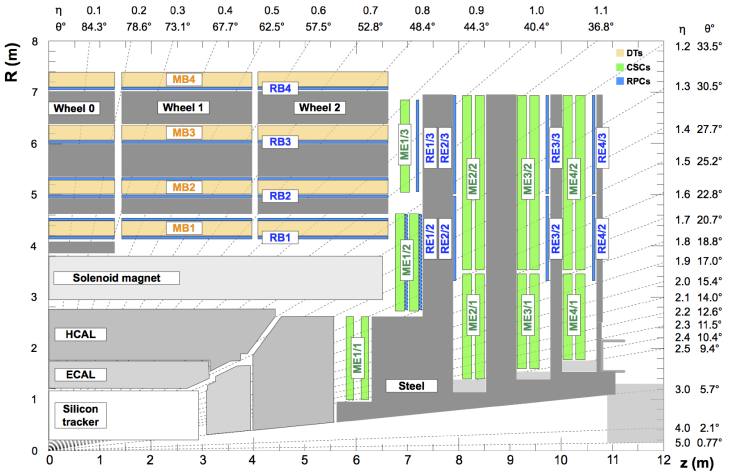

Figure 2.7: Layout of the CMS subdetectors as a function of the z and r axes. The muon systems are highlighted in different colours and the location of the DTs, CSCs and RPS is shown. Image taken from Ref. [112].

The DTs are used to detect muons in the barrel region of CMS, where the magnetic field is uniform and parallel to the beam axis. They consist of thin aluminium tubes filled with a gas mixture of argon and carbon dioxide, with a central tungsten wire running through the centre of each tube. When a muon passes through a DT, it ionises the gas, creating a trail of electrons that drift towards the wire. The resulting electrical signal is then read out by electronics at either end of the tube, determining the position of the muon. Additionally, from the curvature of the trajectory, the muon momentum is measured.

The CSCs, on the other hand, are used to detect muons in the endcap regions of CMS, where the magnetic field is stronger and non-uniform. They consist of arrays of

cathode strips crossed by arrays of anode wires, all contained within gas-tight chambers. When a muon passes through a CSC, it ionises the gas, creating a cloud of electrons and positively charged ions. The electric field in the chamber causes the electrons to drift towards the anode wires, while the positively charged ions drift towards the cathode strips. The resulting signals from the anode wires and cathode strips are then combined to give 3D measurements of the muon positions.

Finally, the RPCs are used to provide fast timing information for muons passing through the detector, both in the barrel and in the endcaps. They consist of two parallel plates made of a high-resistivity material, with a thin layer of gas in between. When a muon passes through the RPC, it ionises the gas, creating a large number of electron-ion pairs that produce an electrical signal. This signal can be read extremely fast (order of a few nanoseconds) and therefore the RPCs are used to trigger the readout of the other muon detectors.

2.2.6 The Precision Proton Spectrometer

The PPS is a detector which was built at the LHC as a joint project between the CMS and TOTEM (TOTal cross section, Elastic scattering and diffraction dissociation Measurement at the LHC) collaborations [113]. It has been integrated with the CMS experiment for Run 2 and operated during the full Run with 13 TeV proton-proton collisions. It is currently running with an upgraded setup, recording collisions at a centre-of-mass energy of 13.6 TeV.

The PPS is designed to measure the trajectory of protons that have left the collisions intact and lost a small fraction of their momentum. Thus, it is used to measure protons involved in CEP processes. These protons remain inside the beam pipe, displaced from the central beam orbit, and can be measured using detectors inserted in "Roman Pots" (RPs), which are movable beam pipe insertions that can approach the LHC beam up to a few mm ($\sim 15\sigma$ in standard runs). The RPs are located at 200-210 m from the IP 5, along the beamline in the LHC sector between the interaction points 5 and 6, referred to as sector 56, as well as between the interaction points 4 and 5, in sector 45. A scheme of one of the arms, from sector 56, is shown in figure 2.8.

Each arm consists of two RPs equipped with silicon tracking detectors and one RP instrumented with timing detectors. The tracking detectors are silicon devices that measure the transverse displacement of the protons with respect to the beam. A drawing of a tracking station is shown in figure 2.9. The two vertical pots are used only for alignment and the horizontal one for measurements in standard LHC runs. The tracking RP closer to the IP5 is referred to as "near", the other as "far". The timing detectors are used to measure the proton's time-of-flight. Timing stations have only one horizontal pot, and the alignment is done by propagating tracks from the tracking stations.

In 2016, each RP from the tracking stations contained 10 silicon strip sensor planes, positioned half at a +45° angle and half at a -45° angle with respect to the bottom of

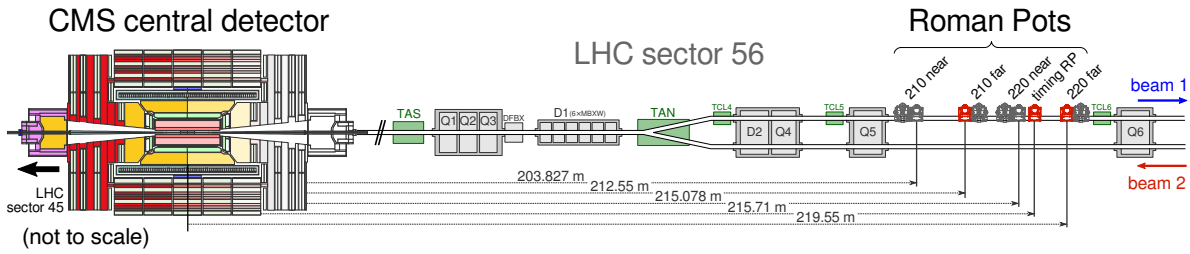

Figure 2.8: Scheme of one arm of CMS, showing the direction of the incoming and outgoing beams (red and blue arrows) and the approximate position of the RP stations. Stations labelled "near" are equipped with silicon strips, while those labelled "far" have pixel detectors (in 2017).

the RP. However, these sensors were not capable of identifying multiple proton tracks in the same event, nor sustaining the large radiation doses, and therefore were gradually replaced. In 2017, the silicon strips from one of the RP stations per arm were replaced with 3D silicon pixel sensors. The choice for 3D pixel detectors was made for their high resistance to the large irradiation flux ($\sim 3 \times 10^{15}$ neq/cm^2) that the detectors are exposed to for being so close to the beam [114], and to allow multiple proton reconstruction per event. In 2018, the strips from the other station were replaced and all tracking RPs became equipped with pixel sensors. The timing station hosts four single- and double-sided single-crystal chemical vapour deposition (scCVD) diamond sensor planes (time resolution $\sim$ 80 ps). In 2017, one of the four planes consisted of an Ultra Fast Silicon Detector (time resolution $\sim$ 30 ps). Timing information is not used for the work of this thesis, since the timing detectors were only fully operational for a small period in 2017.

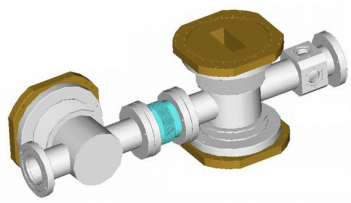

Figure 2.9: Drawing of a tracking RP station, from Ref. [115].

2.3 Data reconstruction at CMS

2.3.1 Trigger system and data recording

Events of interest are selected by CMS using a two-tiered trigger system. The first level, L1, composed of custom hardware processors, uses information from the calorimeters and muon detectors to select events at a rate of around 100 kHz [116]. The second level, known as the high-level trigger (HLT), consists of a farm of processors running a version of the full event reconstruction software optimised for fast processing, and reduces the event rate to around 1 kHz before data storage [117]. This corresponds to a data stream of 1 GB/s, which is permanently stored for later offline reconstruction. The "raw" data is stored at CERN, in the so-called Tier-0 computing centre. There, the raw data is grouped into primary datasets according to HLT information. These datasets are also stored in magnetic tapes. Then, still in Tier-0, a reconstruction of the data is performed where the detector information is converted into physics objects. The reconstructed data is distributed to the Tier-1 centres, which are large centres in CMS collaborating countries (large national labs). In these centres, the data is further calibrated and skimmed into files that can be used for physics analyses, Analysis Object Data (AOD), and its more skimmed versions, miniAOD and nanoAOD. These files are redistributed to a large set of smaller Tier-2 centres ("local" centres at universities/institutes). Files from all the different Tier-2 centres can be used in a grid-based system for analysis all over the world. The Tier-2 is also where the Monte-Carlo simulated samples that are used by the collaboration are stored [118].

2.3.2 Object and event reconstruction

For each event recorded at the CMS experiment, the energy deposits in the different detector layers are combined in an attempt to identify and reconstruct, as accurately as possible, all the particles that were produced. Deposits in the silicon tracker are used to reconstruct the curved trajectory of charged particles, such as electrons or muons. Their electric charge and momentum can be inferred from the direction and radius of curvature of the particle trajectory. These tracks are extrapolated to a crossing point, which is identified as the exact collision point, or primary vertex (PV). Most particles will then deposit their energy in the calorimeters. Energy deposits in the ECAL crystal clusters and in the HCAL towers are used to reconstruct the initial energy and direction of the particles. Muons typically do not deposit a large fraction of energy in the calorimeters and leave the detector passing through the muon chambers. Hits in these chambers, combined with information from the tracker, are used to reconstruct muons. In general, tracker, calorimeter and muon chamber information are combined to identify all particles produced in an event, using an algorithm called *Particle Flow* (PF) [119]. This algorithm starts by identifying muon candidates, as described below, and after that reconstructs electron,

photon and hadronic jet candidates, in this order. Finally, the negative $\vec{p}_{\mathrm{T}}$ sum of all PF candidates is computed and constitutes the missing transverse momentum, $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$. A brief description of the reconstruction techniques used for each type of particle is provided in the following.

Tracks and primary vertices

Hits in the tracker are first combined to form tracks. An algorithm called Combinatorial Tracker Finder is applied to the hit patterns measured in the pixel and strip trackers. Building individual particle tracks out of the hit patterns is a non-trivial exercise, given the large track density close to the IP. Figure 2.10 shows the reconstructed tracks in a typical bunch crossing, for data collected in 2016. After the track reconstruction, each track has to be associated with the corresponding PV. Tracks are extrapolated towards the z-axis and their intersection point in the vicinity of the beamline is used to identify the vertex of interaction. This is done using the deterministic annealing algorithm [120], which was developed to solve statistical mechanics problems with a large number of degrees of freedom. This algorithm results in the identification of PVs with a resolution of about 20 μm, and an efficiency close to 100% [121]. For physics analyses, only the PV with the largest energy transfer is used. It is typically chosen at a later stage of the reconstruction as the one with the highest sum of p_{T}^2 of all the associated physics objects, reconstructed with PF. All other PVs are not considered for physics analyses, and the corresponding interactions are referred to as pileup (PU).

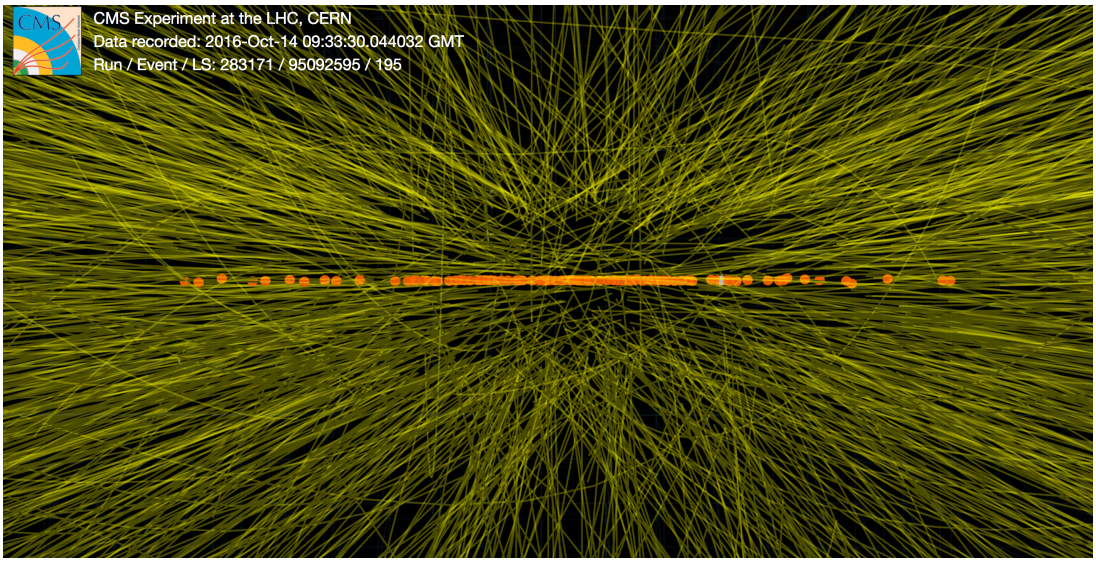

Figure 2.10: Reconstructed tracks in the CMS tracker, in a single bunch crossing during the 2016 data-taking period [122]. The interaction vertices associated with these tracks are shown in orange.

Particle reconstruction

Muons The muon reconstruction in CMS is performed using two main algorithms [112]. The first one is known as Global muon, and the second as Tracker muon. For a Global muon reconstruction, signals from the DTs and CSCs are first reconstructed to determine the muon hit positions and timing information. The hits from the different segments are combined in a fit, and the result is a so-called standalone muon. Standalone muons are matched with tracks from the tracker. When a match is found, the best trajectory is determined through a fit to the tracks and the segments in the muons detectors, and from that fit the muon p_{T} is precisely computed. The algorithm also takes into account the curvature of the muons induced by the magnetic field and the energy losses in the calorimeters. The Global muon reconstruction provides a p_{T} resolution of about 1%, for muons with $p_{\mathrm{T}} < 100\,\mathrm{GeV}$ [112]. Tracker muon reconstruction proceeds the other way around, i.e., it relies first on information from the tracker. Tracks in the inner tracker with sufficiently high p_{T} are selected. Then, they are extrapolated to the muon chambers and matched to the hits there. If a match is found, then a Tracker muon is reconstructed. Considering both algorithms, the efficiency for reconstructing a muon that is produced within the acceptance of the muon system is about 99%, for the LHC Run 2 data.

Electrons and photons During the LHC Run 2, electrons and photons in CMS are reconstructed using similar procedures [123]. Electrons typically produce hits in the silicon tracker and energy deposits in the ECAL. Photons also produce energy deposits in the ECAL, but no hits in the tracker. The same happens for neutral hadrons, which have an identical signature to photons. Before reaching the ECAL, electrons may interact with the tracker material, emitting bremsstrahlung photons, and photons may convert into electron-positron pairs. Therefore, when electrons reach the ECAL, a wide shower is produced, coming from multiple electrons and photons instead of from a single particle. For photons, on the other hand, a more collimated shower is expected. The energy reconstruction starts with the formation of clusters by grouping together consecutive ECAL crystals with energies above a predefined threshold. The seed cluster is then defined as the cluster containing the most energy in a given region. ECAL clusters within a certain area around the seed cluster define a supercluster (SC). The SC is defined in order to group together electrons and photons from bremsstrahlung and conversion, and recover the total energy of the initial object. In parallel, all tracks in the event are tested for compatibility with an electron trajectory hypothesis. Those which are compatible are matched with the SC positions based on the Gaussian sum filter (GSF) algorithm [124]. Tracks matching an SC position become electron candidates. All reconstructed SCs are photon candidates. Further identification techniques are then used to distinguish between the two types of particles, and from neutral hadrons, as will be described in section 3.3.1.

Jets In the CMS experiment, PF particles are combined to form jet candidates, using the anti-k_{T} jet clustering algorithm [125]. This is a commonly used clustering algorithm, and it considers the distances between particles, as well as their p_{T}, such that the highest-p_{T} particles are combined with nearby low-p_{T} particles into a jet. The jet momentum is computed by performing the vector sum of the momenta of the PF candidates that are clustered into the jet. To mitigate the effect of PU, charged PF candidates not originating from the PV are removed before the clustering.

Heavy flavour jets Jets originating from gluons, u, d, or s quarks are called light-flavour jets, while those originating from b or c quarks are known as heavy flavour. Identifying jets coming from b quarks (called b jets) is indispensable for top quark analyses, as the top quark decays almost exclusively into a b quark and a W boson. The b quarks hadronise immediately forming b hadrons. These hadrons have a relatively long lifetime, and after being produced travel a few mm in the tracker before decaying. This originates a secondary vertex, which is reconstructed from the tracks of the charged decay products, and used to identify b jets. Additionally, it is common for leptons, in particular muons, to be present in the b hadron decay chain. These features and others are combined into machine-learning-based classifiers, such as DeepCSV and DeepJet, mentioned in sections 4.4 and 3.3. These classifiers are designed specifically to distinguish b jets from c and light-flavour jets.

2.3.3 Intact proton reconstruction

In the PPS tracking detectors, we measure the positions (x, y) and angles with respect to the beam axis (θ_x, θ_y) of protons that left the IP intact. These positions can be translated into a proton kinematics vector $\mathbf{d} = (x, \theta_x, y, \theta_y, \xi)^T$. The proton kinematics at a given distance s from the IP relates to the proton kinematics at the IP, $\mathbf{d}^*$, via the transport equation:

$$\mathbf{d}(s) = T(s, \xi) \cdot \mathbf{d}^* \tag{2.4}$$

where ξ is the fraction of momentum lost by the protons:

$$\xi = \frac{\Delta p}{p} \quad , \tag{2.5}$$

with p being the nominal proton momentum (before the collision) and Δp being the momentum lost upon the collision. The transport matrix T encodes the LHC optics:

$$T(s,\xi) = \begin{pmatrix} v_x & L_x & m_{13} & m_{14} & D_x \\ \frac{\mathrm{d}v_x}{\mathrm{d}s} & \frac{\mathrm{d}L_x}{\mathrm{d}s} & m_{23} & m_{24} & \frac{\mathrm{d}D_x}{\mathrm{d}s} \\ m_{31} & m_{32} & v_y & L_y & D_y \\ m_{41} & m_{42} & \frac{\mathrm{d}v_y}{\mathrm{d}s} & \frac{\mathrm{d}L_y}{\mathrm{d}s} & \frac{\mathrm{d}D_y}{\mathrm{d}s} \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{2.6}$$

The terms v_x, L_x, D_x, and $m_{i,j}$ and their counterparts in the y direction are the so-called optical functions and are described in detail in Ref. [40]. They depend on several machine parameters, the most important of which is β^*, a value related to the transverse size of the particle beam at IP5. In normal physics conditions during Run 2, the LHC ran with β^* values between $0.2 - 0.4\,\mathrm{m}$. From equation 2.4, we can write the position coordinates x, y of the proton in the RPs as

$$x = x_0 + v_x \cdot x^* + L_x \cdot \theta_x^* + m_{13} \cdot y * + m_{14} \cdot \theta_y^* + D_x \cdot \xi \tag{2.7}$$

$$y = y_0 + m_{31} \cdot x * + m_{32} \cdot \theta_x^* + v_y \cdot y^* + L_y \cdot \theta_y^* + D_y \cdot \xi \tag{2.8}$$

A number of assumptions are necessary to extract the approximate proton kinematics from these equations. Namely, we assume that the terms $m_{i,j}$ are approximately 0 and can be ignored [40]. Furthermore, the calibration of the optics is done such that the beams are at $x_0 = 0, y_0 = 0$. Applying all this, equations 2.7 and 2.8 become simply:

$$x \approx v_x \cdot x^* + L_x \cdot \theta_x^* + D_x \cdot \xi \tag{2.9}$$

$$y \approx v_y \cdot y^* + L_y \cdot \theta_y^* + D_y \cdot \xi \tag{2.10}$$

D_x is the most important quantity to determine precisely. It is computed from the measured proton tracks, and calibrated using events from CEP of $\mu\mu$. Minimum bias RP data is used to determine D_y. A software called MAD-X [126] that simulates the LHC optics is used to estimate L_x and L_y as a function of ξ, $L_x(\xi)$ and $L_y(\xi)$. There are two different methods to extract the value of ξ:

- **singleRP:** The contribution from the θ_x term, as well as other subleading terms, is neglected, leading to $\xi = x/D_x$. Using the estimated value of ξ, θ_y^* is computed as $\theta_y^* = y/L_y(\xi)$. Neglecting the subleading terms leads to a degraded resolution. However, this method has the advantage that it is possible to reconstruct a proton from a single RP track and other RP(s) don't need to have a matching track. The variables x^*, y^* and θ_x^* cannot be reconstructed and are set to zero, which is a reasonable approximation in the low β^* optics.

- **multiRP:** This method aims to take advantage of the full spectrometer data. To achieve this, a fit is performed to minimise the χ^2, defined from the transport equation as

$$\chi^2 = \sum_{i:\text{RPs}} \sum_{q:x,y} \left[\frac{d_q^i - (T^i d^*)_q}{\sigma_q^i} \right]^2 \tag{2.11}$$

  With 2 RPs per arm, this is a fit to 4 measurements (two RPs times two projections). Therefore, only a maximum of 4 parameters can be extracted from the fit. To overcome this issue, x^* is set to 0, which is a good approximation in the low β^* optics, and then values of $(y^*, \theta_x^*, \theta_y^*, \xi)$ are extracted.

The difference in resolution as a function of ξ is shown in figure 2.11. As expected, the simpler method singleRP yields a lower resolution and introduces a bias in the very low and very high ξ regions. In the work of this thesis, the multiRP method is used.

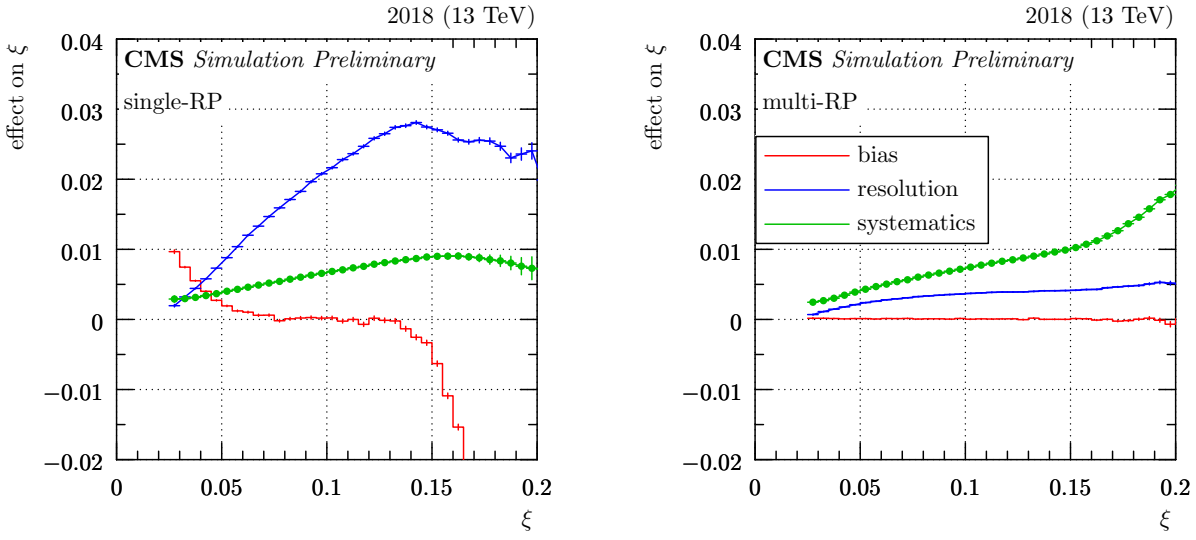

Figure 2.11: Comparison of bias, resolution and systematic uncertainties (2018 data, sector 56) between the single and multiRP methods. From [40].

The acceptance of the detectors as a function of ξ is shown for both PPS arms for the different years in figure 2.12, for the multiRP method. These plots use simulation, but the acceptance is based on calibrations done using data. In these plots, one can see the acceptance is different for both arms, from about 0.02 to 0.13 in arm 45 and 0.03 to 0.16 in arm 65, in the 2017 setup. This unintentional asymmetry is explained by the fact that the dispersion of the LHC beams is not exactly the same for both beams. In fact, the size of the asymmetry is smaller than the precision with which the LHC can control

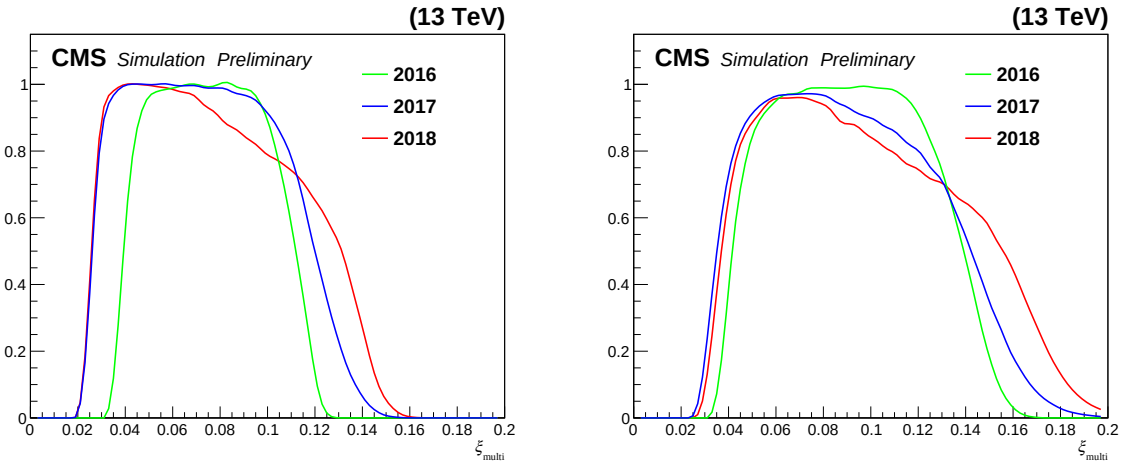

Figure 2.12: Efficiency of proton reconstruction with PPS, with the multiRP method, as a function of the fractional momentum loss (ξ), for arm 45 (left) and arm 56 (right), in Run 2 simulated data. From Ref. [40].

and measure the beams by adjusting the magnet currents. This is briefly discussed in Ref. [40].

Figure 2.13 shows the total integrated luminosity recorded by the CMS experiment when PPS was enabled, in the full Run 2. The total recorded luminosity is $111.14\,\text{fb}^{-1}$.

2.4 Luminosity measurements

Luminosity is a key property of any particle collider, as it determines the number of events that are produced. For a given process, the rate ($\frac{\mathrm{d}N_0}{\mathrm{d}t}$) at which events are produced is proportional to its cross section and the instantaneous luminosity of the collider:

$$\frac{\mathrm{d}N_0}{\mathrm{d}t} = \sigma \cdot \mathcal{L}(t) \quad . \tag{2.12}$$

The cross section σ depends on the specific physics process and can be computed from the underlying theory. In order to measure it experimentally from the observed number of events N, the integrated luminosity $L = \int \mathcal{L}\,\mathrm{d}t$ has to be known. In principle, $\mathcal{L}$ can be calculated from the knowledge of the machine parameters. For an accelerator with a revolution frequency ν, and two beams with N_1 and N_2 particles, distributed in the transverse direction to the beam according to the 2D functions $\rho_1(x, y)$ and $\rho_2(x, y)$, we can write the instantaneous luminosity of a single interacting bunch pair as:

$$\mathcal{L}_b = \nu N_1 N_2 \iint \rho_1(x, y) \cdot \rho_2(x, y)\,\mathrm{d}x\,\mathrm{d}y = \nu \cdot \frac{N_1 N_2}{A_{\text{overlap}}}, \tag{2.13}$$

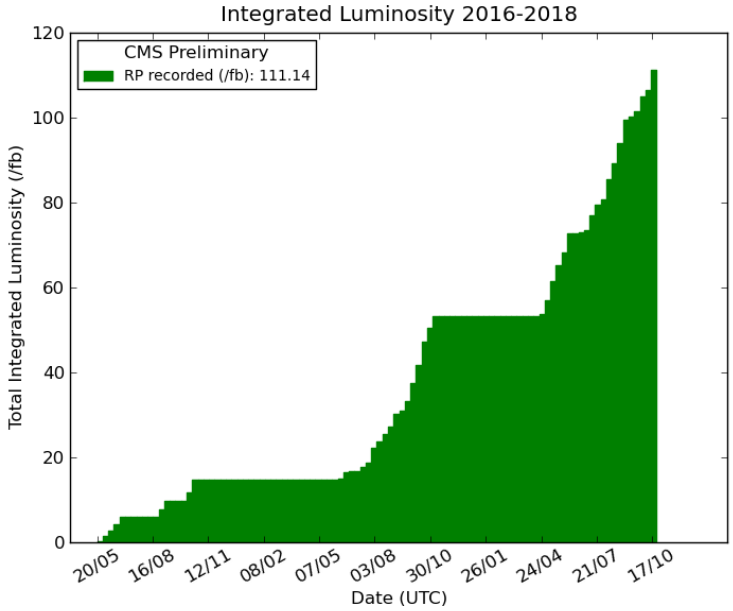

Figure 2.13: Total integrated luminosity recorded by the CMS experiment when PPS was enabled, in 2016-2018, from Ref. [127].

where A_{overlap} is the overlapping area between the two beams. Following the assumption that the beams can be described by two independent distributions in x and y, the above formula can be factorised and rewritten as:

$$\mathcal{L}_b \approx \nu N_1 N_2 \int f_1(x) \cdot f_2(x)\,\mathrm{d}x \int g_1(y) \cdot g_2(y)\,\mathrm{d}y = \nu \cdot \frac{N_1 N_2}{2\pi \Sigma_x \Sigma_y}, \tag{2.14}$$

where Σ_x and Σ_y are respectively the effective width and height of the beam overlap region. In the ideal case of two beams with exactly equal Gaussian proton distributions, this could be further simplified to $\mathcal{L}_b \approx \nu \cdot \frac{N_1 N_2}{4\pi \sigma_x \sigma_y}$, where σ_x and σ_y are the root mean squared (RMS) widths of the Gaussian beams in x and y. In real colliders such as the LHC, this is not always an accurate assumption: it is known that a correlation exists between the beam distributions in the x and y directions, and that the beam profiles are not strictly Gaussian.

Luminosity detectors, also referred to as luminometers, measure hits, tracks or clusters from the collisions. The rate $R = \frac{\mathrm{d}N_0}{\mathrm{d}t}$ measured by these detectors can be related to the instantaneous luminosity through equation 2.12. The cross section in this case is called "visible cross section", σ_{vis}, as it depends on the acceptance and efficiency of the luminometer. The σ_{vis} cannot be computed from first principles and needs to be derived from dedicated calibrations.

Luminosity detectors at the CMS experiment

At the CMS experiment, two different kinds of luminometers exist: online devices that provide luminosity in real time and bunch-by-bunch, and offline devices, where the luminosity is calculated *a posteriori* from the collected data. Online devices can also be used for offline luminosity measurement and calibrations. Figure 2.14 shows schematically the positions of the various luminometers in the CMS detector.

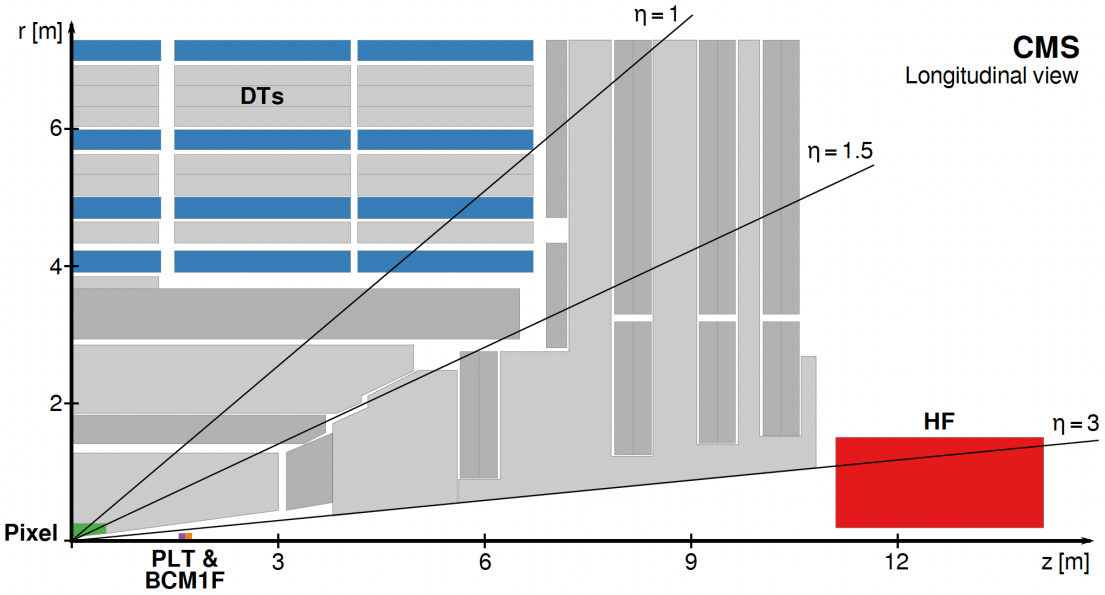

Figure 2.14: Schematic cross section through the CMS detector in the r-z plane [128]. The main luminometers in Run 2 are highlighted, showing the silicon pixel detector, PLT, BCM1F, DTs, and HF. Solid lines represent distinct η values.

The pixel luminosity telescope (PLT) [129] is a stand-alone online luminometer. It is a silicon pixel detector, located outside of the pixel endcaps, approximately 1.8 m from the IP. The detector is arranged into 16 "telescopes", eight on each side, each composed of three individual silicon sensor planes. The per-bunch instantaneous luminosity is measured by registering "triple coincidences", i.e., events where all three planes in the telescope register a hit, typically corresponding to a track from a particle originating at the IP. The triple coincidence requirement provides a strong background suppression and leads to good performance and stability.

Mounted on the same location as PLT is the fast beam conditions monitor (BCM1F) [130], a detector that measures real-time beam background and per-bunch luminosity. In Run 2, it was composed of silicon sensors as well as polycrystalline diamond sensors. For Run 3, it has been upgraded with an active cooling loop and AC-coupled read out, and

is now equipped solely with silicon sensors. The good time resolution of BCM1F helps to separate between the hits from the collision products and hits from other sources.

The HF part of the hadronic calorimeter is exploited for both online and offline luminosity measurements, using two different methods. The occupancy method (HFOC) is based on computing, for each bunch crossing and each calorimeter tower, the number of hits in each bunch crossing above a certain energy threshold. The transverse energy method (HFET), on the other hand, relies on computing the sum of the energy deposited in the calorimeter cells. The tracker can also be exploited to measure luminosity: the method known as pixel cluster counting (PCC) is used to determine luminosity from the average number of pixel clusters registering hits for one event.

The DT method for offline luminosity measurement relies on counting muon tracks in muon DT stations. Muons are low-background physics objects in the detector. In addition, the muon candidate rate has been observed to be linear with luminosity and rather stable over time. However, the rate of muons in the DTs is significantly lower than the rate for most other observables from other luminometers. Thus, there are not enough muon tracks during the vdM scans described in the next section in order to derive a calibration of σ_{vis}. As a consequence, the rate computed with the DT method is calibrated to the PCC luminosity measurement, and is only used for the monitoring of stability and linearity of the other luminometers.

The offline luminosity measurements from all the systems described above are combined to get the best possible knowledge of the integrated luminosity and its uncertainty.

2.4.1 The vdM method

To measure luminosity at the CERN Intersecting Storage Ring (ISR), the world's first proton-proton collider, in 1968, Simon van der Meer proposed a method to estimate Σ_x and Σ_y using beam separation scans [131]. In the van der Meer (vdM) method, the beams are first moved away from their nominal positions, and subsequently scanned through each other in steps, in both transverse directions. The luminometers record the event rate as a function of beam separation, in x and y, and Σ_x and Σ_y are extracted from fits to the data. This procedure is referred to as a vdM scan.

Equation 2.13, which describes the instantaneous luminosity for head-on collisions, can be generalised to:

$$\begin{aligned} \mathcal{L}_b(w,h) &= \mathcal{N} \iint \rho_1(x,y) \cdot \rho_2(x-w, y-h)\,\mathrm{d}x\,\mathrm{d}y \\ &\approx \mathcal{N} \int f_1(x) \cdot f_2(x-w)\,\mathrm{d}x \cdot \int g_1(y) \cdot g_2(y-h)\,\mathrm{d}y \quad , \end{aligned} \tag{2.15}$$

where $\mathcal{N}$ is a normalisation constant, and the weight w and height h are the separations between the beams in the x and y directions. Integrating the measured rates as a function

of the separation corresponds to integrating equation 2.15 in w and h. Technically this integral should be done between $-\infty$ and $+\infty$, but in practice, it is sufficient to ensure that the maximum separation in both directions is much larger than the beam widths - at the LHC the beams are typically displaced by a distance about 5 times larger than the beam widths. The leftmost and rightmost parts of figure 2.15 show schematically how each of the beam axes are separated during a vdM scan at the LHC.

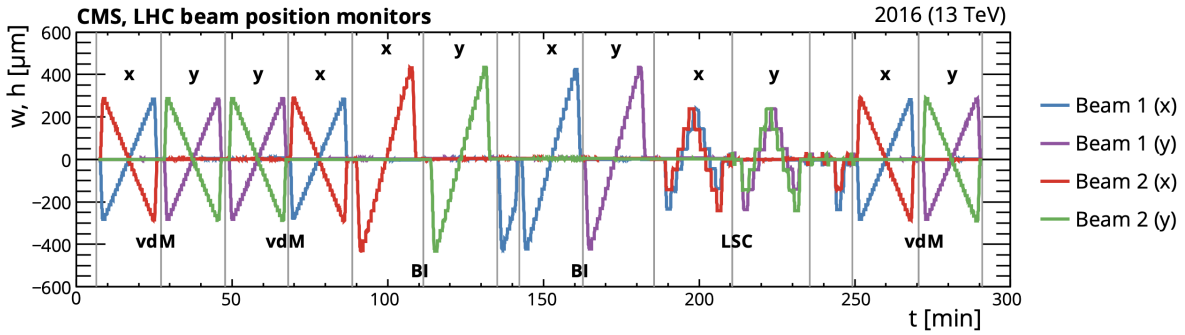

Figure 2.15: Relative change in the positions of beams 1 and 2 during the vdM scan recorded by CMS in 2016 [132], in the horizontal (x) or vertical (y) directions, as a function of the time elapsed from the beginning of the program. The first five grey vertical lines, as well as the last three, delineate the vdM scans, while the remaining delineate other types of scans, namely beam-imaging (BI) and length scale calibration (LSC) scans.

As shown in reference [131], Σ_x and Σ_y are then given by:

$$\sqrt{2\pi}\Sigma_x = \frac{\iint f_1(x) \cdot f_2(x-w)\,\mathrm{d}x\,\mathrm{d}w}{\int f_1(x) \cdot f_2(x)\,\mathrm{d}x}, \quad \sqrt{2\pi}\Sigma_y = \frac{\iint g_1(y) \cdot g_2(y-h)\,\mathrm{d}y\,\mathrm{d}h}{\int g_1(y) \cdot g_2(y)\,\mathrm{d}y} \tag{2.16}$$

or

$$\sqrt{2\pi}\Sigma_x = \frac{\int \mathcal{L}_b(w,0)\,\mathrm{d}w}{\mathcal{L}_b(0,0)}, \quad \sqrt{2\pi}\Sigma_y = \frac{\int \mathcal{L}_b(0,h)\,\mathrm{d}h}{\mathcal{L}_b(0,0)} \tag{2.17}$$

In CMS, the values of Σ_x and Σ_y in the vdM scans are extracted by fitting a function to the measured rates in the x and y scans. This function can be a single Gaussian (SG), in the simplest case, for which $\Sigma_x = \sqrt{2}\sigma_x$ and $\Sigma_y = \sqrt{2}\sigma_y$, or a more complex function, such as a sum of two or more Gaussians with specific constraints on their widths and mean values. From the values of $\Sigma_{x,y}$, the σ_{vis} is obtained simply as:

$$\sigma_{\text{vis}} = \mu_{\text{vis}} \frac{2\pi\Sigma_x\Sigma_y}{N_1 N_2} \tag{2.18}$$

In the formula above, μ_{vis} is the number of interactions per bunch and is given by the rate at $(0, 0)$. In CMS, it is computed by extracting the rate at the peak of the x and y scans from the fits and computing the average, $\frac{\text{peak}_x + \text{peak}_y}{2}$. N_1 and N_2 can be accurately measured by the LHC beam current transformers (DCBCT and FBCT).

An example vdM scan result is shown in figure 2.16, recorded with the PCC algorithm described above. This vdM scan was recorded in 2016, and the data are fitted with a double Gaussian (DG) function - the sum of two Gaussians with the same mean - including a floating constant. The parameters extracted from the fit ($\Sigma_{x,y}$ and $\text{peak}_{x,y}$) were used to calibrate the absolute luminosity in 2016.

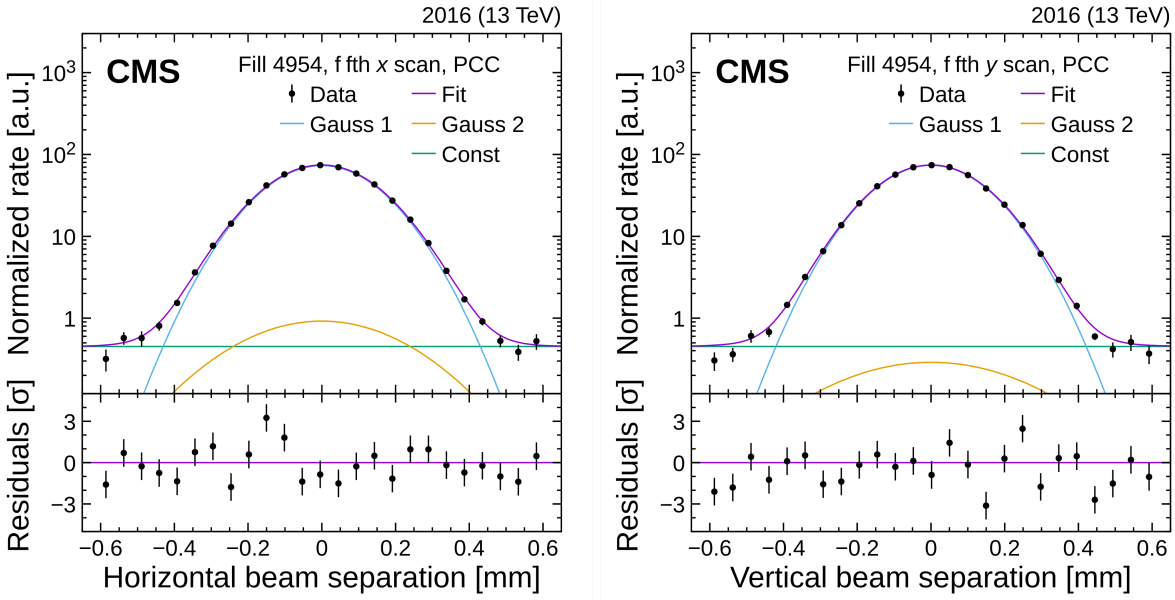

Figure 2.16: Data from vdM scan recorded by CMS in 2016 [132]. Left: scan in the horizontal direction, Right: scan in the vertical direction. The y-axis represents the normalised rate of counts recorded using the PCC algorithm.

Uncertainties on the luminosity measurements

There are several uncertainties affecting the luminosity measurements. The ones that had a larger effect in 2016 data are briefly summarised in the following. An important source of uncertainty was the limited precision with which we measure the actual **beam positions**, as they do not exactly correspond to the intended (nominal) beam positions.

Another main source of uncertainty has to do with the **non-factorisation of the beam profiles** in x and y. As mentioned earlier, the assumption that the beam shapes in the x and y planes are uncorrelated, and thus A_{overlap} from equation 2.13 can be factorised into two independent factors, is an approximation. Using a special type of

scan, called beam imaging scans, where the beams are moved one at a time, it is possible to get an estimate of the correlation between the x and y shapes of the beams, and derive a correction for σ_{vis} with an associated uncertainty. This uncertainty propagates to the measurement of the integrated luminosity. In the early days of my PhD project, I contributed to the analysis of beam imaging scans recorded in 2015 and 2016, to derive this uncertainty. That work is not described here but is published in Ref. [132].

The two beams consist of positively charged particles, and therefore they repel each other as they approach. This causes the beam shapes to be slightly altered and is referred to as **beam-beam effects**. Corrections are derived to account for this effect, and the uncertainty on these corrections is also one of the leading sources of luminosity uncertainty.

Moreover, the luminosity in CMS is measured using several luminometers, as described above. Combining the different results into one single luminosity measurement requires taking into account the **cross detector consistency**, and results in a further uncertainty. There are several other effects which have a smaller contribution to the uncertainty and are described in detail in Ref. [132].

2.4.2 Analysis of emittance scans

At the LHC, there is typically one vdM scan campaign per data-taking year. This means that, for a few days, a series of vdM scans are performed along with other types of scans such as beam imaging scans and length scale calibration scans [132].

The results from the vdM scan campaign are used to calibrate the absolute luminosity for the year. However, since the campaign happens at one single point during the year, the calibrations do not reflect possible changes in the detector conditions as a function of time and the LHC settings. Such variations include radiation damages in the detectors and non-linear behaviour as a function of the instantaneous luminosity. In order to keep track of such changes and possibly correct for such effects, on request of CMS, the LHC started to perform emittance scans in 2017.

Emittance scans are short vdM-type scans which take place at the LHC at the beginning of every fill and at the end of most of the fills. They are performed with a smaller number of steps, $\sim$10, and the maximum separation between the beams is smaller, i.e., about 3 standard deviations of the beam width, instead of 5. The smaller number of steps allows the scans to be performed in approximately five minutes, unlike a standard vdM scan that can last around 30 minutes. An example emittance scan collected in 2017 using the HFOC luminometer is shown in figure 2.17, for one bunch crossing. The data are shown after a background subtraction procedure, as well as some corrections for the so-called beam-beam effects, i.e., displacement of each beam from its nominal position due to interactions between the two beams.

The rates collected by all the luminometers in each emittance scan, in the x and y planes, for all bunch crossings, are fitted using both SG and DG functions, and the fits with the lowest χ^2 are selected. A value of σ_{vis} is extracted from the selected fits for each

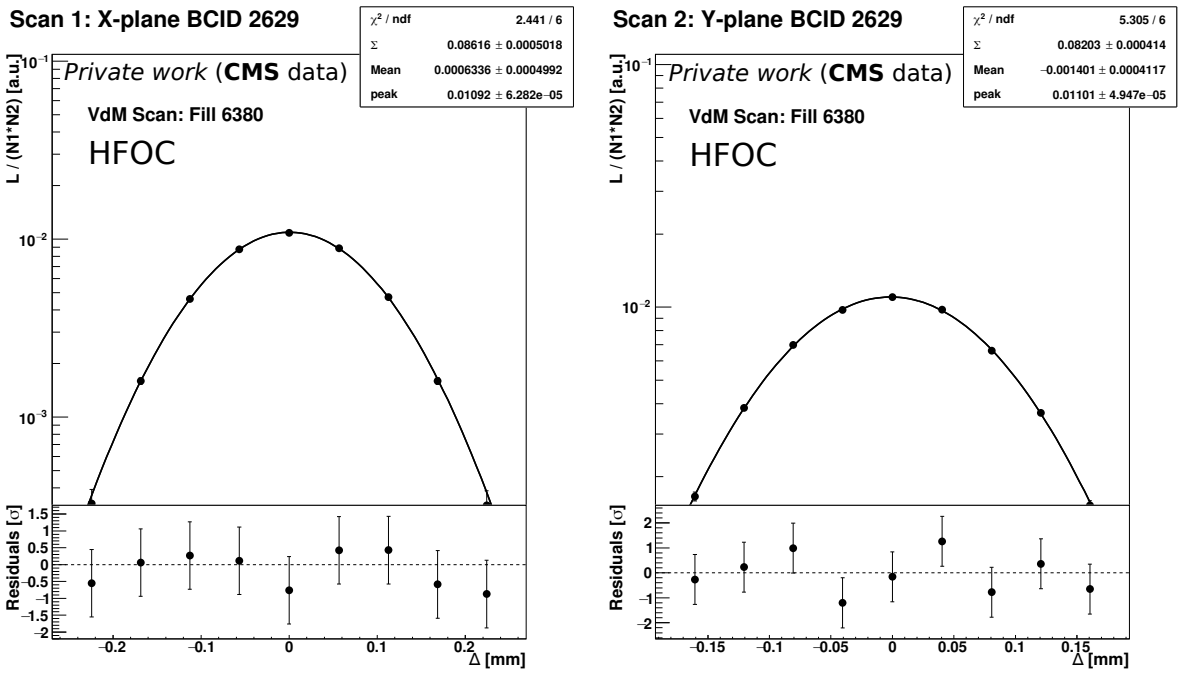

Figure 2.17: Emittance scan recorded by CMS in 2017, for one bunch crossing, in the x and y planes. The data were recorded with the HFOC luminometer, and fitted with a SG function. The y axis shows the normalised rate recorded by the luminometer, and the y axis the separation between the beams.

bunch crossing, and the average σ_{vis} for all bunch crossings is used to perform a stability analysis. We analyse how the average σ_{vis} varies with time (or equivalently, with the integrated luminosity) for each luminometer. Variations, if significant, are modelled with a linear function where the slope is used to correct luminosity measurements. As part of the work of this thesis, a stability analysis of the data-taking periods of 2017 and 2018 was performed and is presented in the following.

Stability analysis during Run 2

For the stability analysis, only emittance scans performed at the beginning of each fill are used; those performed at the end of the fills are excluded. In this way, only the time-stability effects are assessed, and not a possible dependence on the instantaneous luminosity, since at the end of the fill the instantaneous luminosity values are typically lower than at the beginning. The dependence of σ_{vis} on the instantaneous luminosity is evaluated separately, in so-called linearity analyses which are beyond the scope of this thesis.

For each fill in 2017 and 2018, the average σ_{vis} is divided by a reference σ_{vis}, which is the one computed using the scans from the vdM campaign of the corresponding year. Figure 2.18 shows the resulting values as a function of the integrated luminosity of Run 2, for 2018. The points in red (blue) correspond to scans before (after) the vdM campaign. Each group of points is fitted with a straight line, and the slope is extracted. These lines are parameterised as $y = a(x - \mathcal{L}_{\text{vdM}}) + b$, where a is the slope, b is the value of σ_{vis} at the time of the vdM scan and $\mathcal{L}_{\text{vdM}}$ is the integrated luminosity collected up to the time of the vdM scan, counting from the start of Run 2. For 2018 data, this corresponds to an integrated luminosity of $114\,\text{fb}^{-1}$. This parameterisation is chosen so that the corrections are derived relative to the vdM calibration.

If the slope is not compatible with zero within the fit uncertainties, the slope value is used to correct the data. These corrections, for 2018, are summarised in table 2.1 and and are available to be applied to the luminometer data to obtain improved offline luminosity calibrations.

The corrections are applied for each scan by changing the value of σ_{vis} to $\sigma_{\text{vis}}^{\text{corrected}} = \sigma_{\text{vis}} + a(\mathcal{L} - \mathcal{L}_{\text{vdM}})$, where $\mathcal{L}$ is the integrated luminosity collected up to the time of the scan being corrected. An improvement of up to 3% can be achieved in the values of σ_{vis} for some fills, by applying these corrections.

Figure 2.19 shows the average σ_{vis} dependence on the integrated luminosity for 2017. The BCM1F detector is not included in this figure since in 2017 it was not always operating reliably [133]. In this year, three different periods are considered, represented by the red, blue and green points. They do not correspond to the vdM scan time, which took place at approximately $50\,\text{fb}^{-1}$, but to three different data-taking conditions, as a consequence of two technical stops of the LHC where the detectors were modified.

The luminosity was found to be quite stable in 2017, i.e., all slopes were compatible

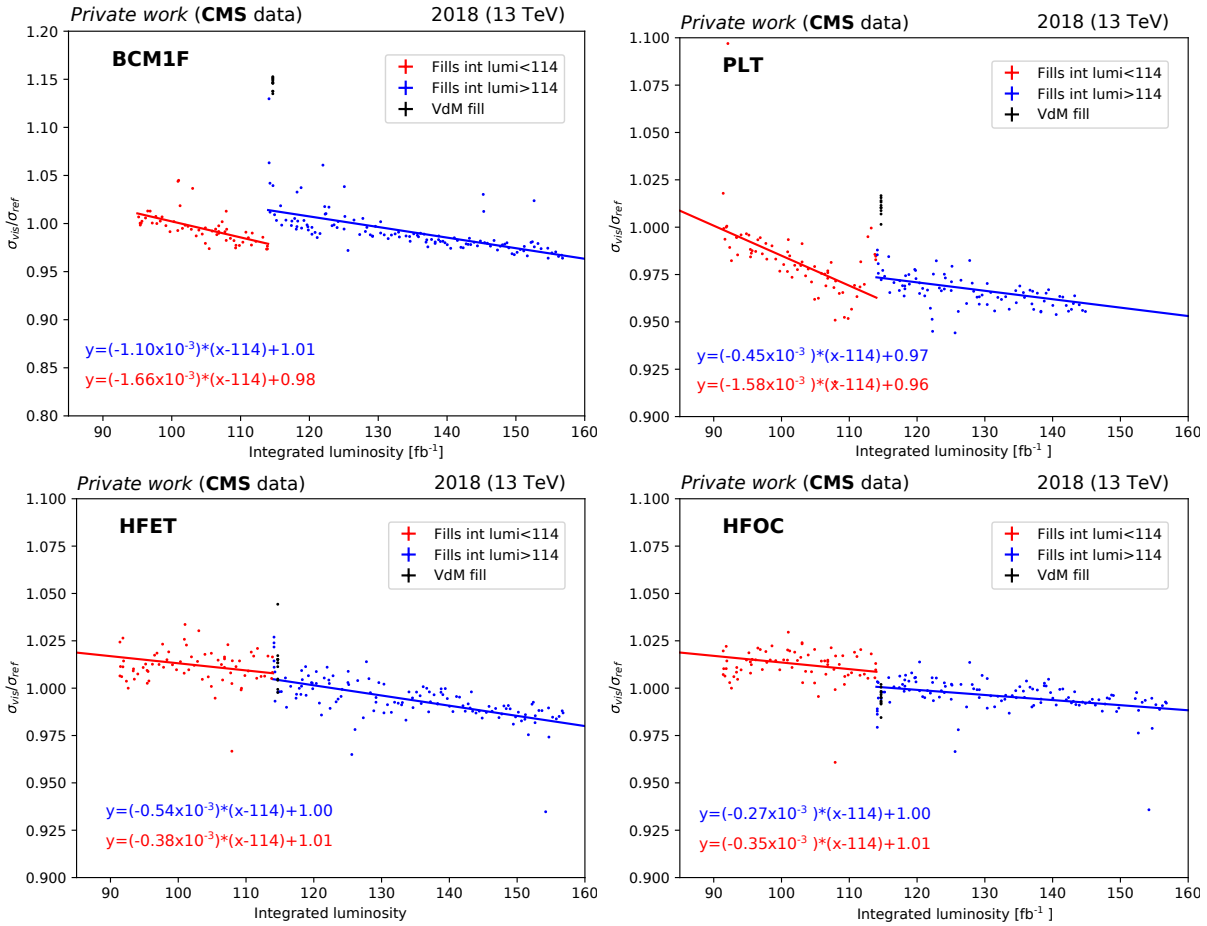

Figure 2.18: Results of the stability analysis from emittance scan data recorded with BCM1F (upper left), PLT (upper right), HFET (lower left) and HFOC (lower right), in 2018. The average σ_{vis} divided by the reference σ_{vis} from the vdM scan fill (points in black) is shown on the y axis. The x axis shows the integrated luminosity of Run 2 at which the scans were recorded. The red (blue) points represent scans recorded before (after) the vdM fill.

	2018 before vdM ($\times 10^{-3}$)	2018 after vdM ($\times 10^{-3}$)
BCM1F	-1.6 ± 0.3	-1.1 ± 0.1
PLT	-1.6 ± 0.3	-0.4 ± 0.2
HFET	-0.4 ± 0.2	-0.5 ± 0.1
HFOC	-0.4 ± 0.2	-0.3 ± 0.1

Table 2.1: Non-stability slopes for different luminometers in 2018, to be applied as corrections to the measured luminosity.

with zero within uncertainties and no correction was required. These results will be included in an upcoming publication about the integrated luminosity of 2017 and 2018.

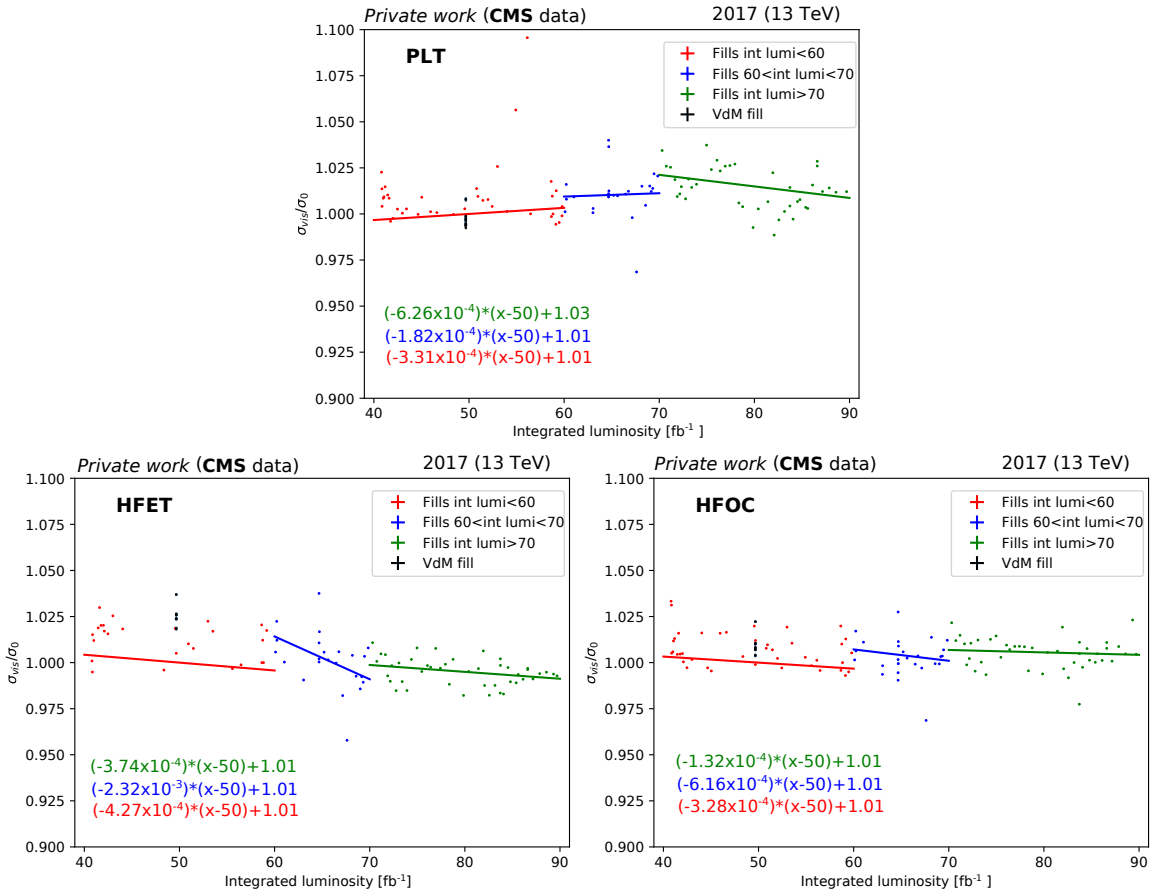

Figure 2.19: Results of the stability analysis from emittance scan data recorded with PLT (upper), HFET (lower left) and HFOC (lower right), in 2017. The average σ_{vis} divided by the reference σ_{vis} from the vdM scan fill (points in black) is shown on the y axis. The x axis shows the integrated luminosity of Run 2 at which the scans were recorded. The red, blue and green points represent scans recorded in different periods, before and after the two technical stops of the LHC in 2017.

Online calibration in the Run 3 commissioning period

Before the beginning of Run 3 in 2022, a so-called commissioning period took place, where test runs were performed to calibrate and assess the performance of all detectors after the upgrades that took place between 2018 and 2022. Moreover, the centre-of-mass energy was increased from 13 to 13.6 TeV. The online instantaneous luminosity in CMS is determined from the online luminometer rates, based on a reference σ_{vis}. Since the value of σ_{vis} depends on the centre-of-mass energy, a re-calibration of the reference σ_{vis} was needed before the start of the Run. This was first done using emittance scans in May and June 2022 and then improved at the beginning of the physics data taking in October. As part of the work of this thesis, these emittance scans were analysed and the average values of σ_{vis} were computed and then used for the online luminosity calibration.

Figure 2.20 shows an example of such a scan recorded by the BCM1F detector in one of those runs. The data are fitted with a DG function, and it is clear from observing the individual Gaussian components that a SG could not accurately describe the beam overlap area. In the y plane, even the DG function does an imperfect job at describing the data, as can be seen from the value of the χ^2 of the fit, shown in the top right corner of the figure.

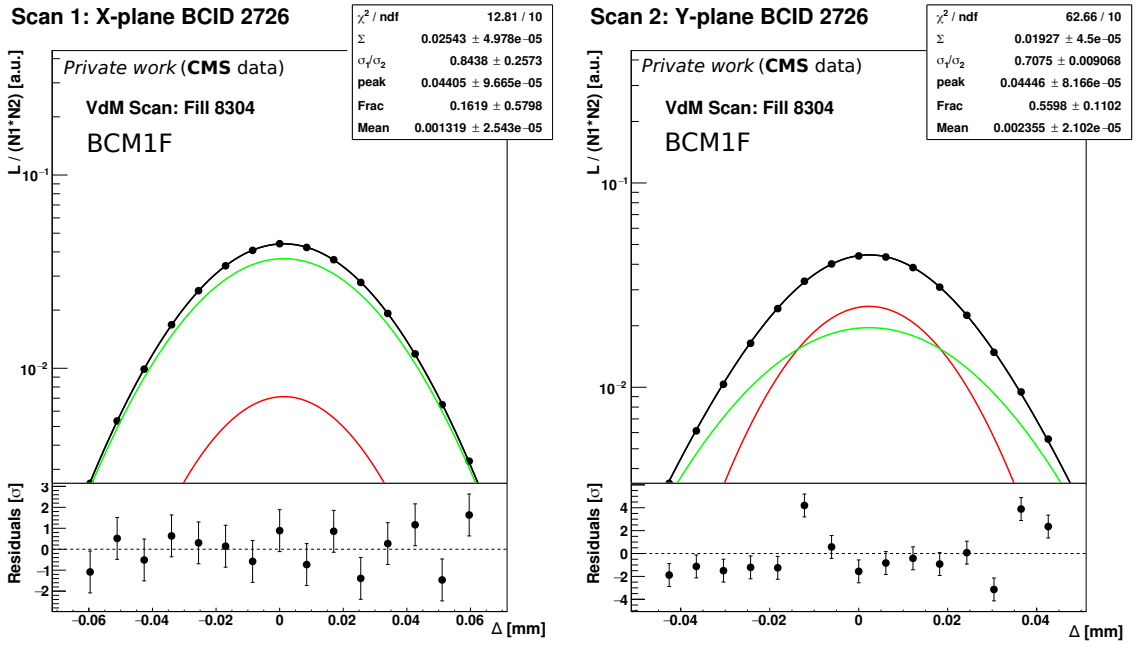

Figure 2.20: Emittance scan recorded by CMS in 2022, at the beginning of Run 3. Data were recorded with the BCM1F luminometer, for one bunch crossing, in the x and y planes. The data are fitted with a DG function. The red and green lines represent the individual Gaussian components, and the black line their sum.

We have observed that in Run 3 the beam shapes are consistently less Gaussian when compared to Run 2, and more complex functions have been necessary to accurately describe the beams. One such function is the so-called *twoMuDG*, a DG function where the two Gaussian components can have different mean values.

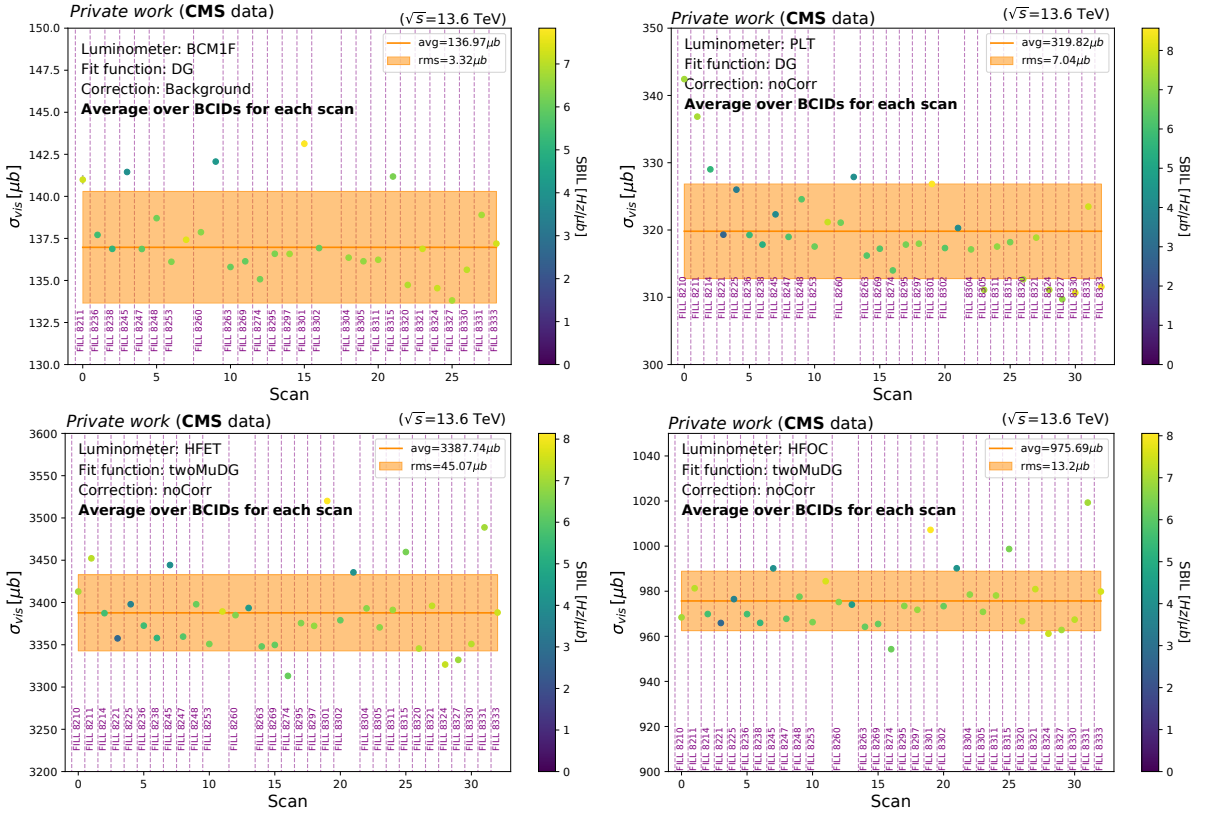

Figure 2.21: Results of the emittance scan analysis for data recorded with BCM1F (upper left), PLT (upper right), HFET (lower left) and HFOC (lower right), at the beginning of Run 3. The average σ_{vis} is indicated in the figures by the orange lines, and its uncertainty is shown as a light orange band.

Figure 2.21 shows, for BCM1F, PLT, HFET and HFOC, the average σ_{vis} per fill, for many scans performed during the calibration runs. The average of these values, indicated by the orange lines, was used as calibration for the online luminosity measurements at the start of Run 3. The light orange bands represent the uncertainty on the average, which was simply computed as the statistical uncertainty on the mean. The data from each luminometer were fitted with the function which resulted in the best χ^2 values, either a DG or a twoMuDG function. Overall, a very good stability is observed, apart from a small number of outliers, that are typically associated with poor quality fits. These results allowed for the estimation of the integrated luminosity in the beginning of Run 3,

with a relatively low uncertainty, which was crucial for the first CMS result of Run 3, the measurement of the $t\bar{t}$ cross section at 13.6 TeV [53].

CHAPTER 3

Measurement of $t\bar{t}\gamma$ and the $t\bar{t}\gamma/t\bar{t}$ ratio

In this chapter, I describe a set of differential measurements of the cross section of $t\bar{t}\gamma$, as introduced in section 1.3.3. For the first time at the LHC, the ratio between the cross sections of $t\bar{t}\gamma$ and $t\bar{t}$ is measured, inclusively and differentially.

Differential cross sections of $t\bar{t}\gamma$ are presented for several observables that have not been measured before, involving the kinematics of the top quark and the $t\bar{t}$ system. This is achieved using a $t\bar{t}$ reconstruction algorithm.

Finally, I include a study of the top quark charge asymmetry in $t\bar{t}\gamma$ events, which is the first study of this kind in the CMS experiment.

These measurements are all performed using data collected by the CMS experiment during the LHC Run 2 (2016-2018[1]), corresponding to an integrated luminosity of 138 fb^{-1}.

The chapter is organised as follows: in section 3.1, previous measurements of the $t\bar{t}\gamma$ process are described. Then, a definition of the signal process is presented, the concept of "fake" photons is discussed, and the main backgrounds are listed. Section 3.2 describes the data and simulation samples that were used to carry out the analysis. Section 3.3 focuses on the event selection, and the definition of signal and control regions. Section 3.4 shows comparisons between the data and the signal and background estimated contributions. The estimation of the background contributions, particularly that from processes with nonprompt photons, is detailed in section 3.6. Section 3.7 describes the statistical concepts used to obtain the results, including a discussion on systematic uncertainties.

The differential cross section measurements are presented in section 3.8, while the inclusive and differential ratio measurements are detailed in sections 3.9 and 3.10. The

[1]Four data-taking periods are analysed separately, corresponding to slightly different detector conditions and calibrations. Two periods are considered in 2016, referred to as 2016 preVFP and 2016 postVFP, and one period per year is considered for the other years, denoted simply as 2017 and 2018.

charge asymmetry studies are shown in section 3.11. All results are discussed in section 3.12, where an outlook is also presented.

3.1 The $t\bar{t}\gamma$ process at colliders

Experimental evidence for the production of $t\bar{t}\gamma$ was reported for the first time by the CDF collaboration at the Tevatron [134], using $p\bar{p}$ collision data. These data were collected by the CDF II detector at $\sqrt{s} = 1.96\,\text{TeV}$, corresponding to $6.0\,\text{fb}^{-1}$. In the data, 30 $t\bar{t}\gamma$ candidate events were identified, against 16 background events predicted, leading to a significance of $3\,\sigma$. The ratio between the cross sections of $t\bar{t}\gamma$ and $t\bar{t}$ is also measured to be 0.024 ± 0.009, consistent with the SM predictions.

The $5\,\sigma$ threshold to claim observation of $t\bar{t}\gamma$ was first reached at the LHC by the ATLAS collaboration, using pp collision data collected at $\sqrt{s} = 7\,\text{TeV}$ [135]. For this analysis, events where the $t\bar{t}$ pairs decay in the lepton+jets channel were used. Around 360 candidate events of $t\bar{t}\gamma$ are observed, for a background expectation of 199±65, reaching a significance of $5.3\,\sigma$and approximately 27% precision in the cross section measurement. Further inclusive cross section measurements were performed by both the ATLAS and CMS collaborations at 8 TeV [136, 137], exploring the fully hadronic and the lepton+jets decay channels and reaching a precision in the cross sections of 17% and 21%, respectively.

The data from the LHC Run 2, at $\sqrt{s} = 13\,\text{TeV}$, corresponding to $138\,\text{fb}^{-1}$, allowed both ATLAS and CMS collaborations to perform more precise inclusive measurements of $t\bar{t}\gamma$ in different final states, and for the first time to measure the cross section of $t\bar{t}\gamma$ differentially as a function of several observables [63, 138–140].

The CMS measurement reports a fiducial cross section of $798 \pm 7(\text{stat.}) \pm 48(\text{syst.})\,\text{fb}$ ("stat." refers to the statistical uncertainty and "syst." to the systematic one) in the lepton+jets channel [138], for photons reconstructed with $p_\text{T} > 20\,\text{GeV}$ and $|\eta| < 1.44$. This corresponds to a precision of 6%. Differential measurements are also presented as a function of the p_T and $|\eta|$ of the leading photon, and the ΔR between the leading photon and the charged lepton. Furthermore, SMEFT interpretations of the results are performed, where the most stringent direct limits to date are set on anomalous electromagnetic dipole moment interactions between top quark and photon. In particular, limits on the c_{tZ} and c_{tZ}^{I} coefficients, introduced in section 1.3.2, are set as shown in figure 3.1 (left).

The CMS measurement in the dilepton channel [139] reports a fiducial cross section of $175.2 \pm 2.5(\text{stat}) \pm 6.3(\text{syst})\,\text{fb}$, which corresponds to a precision of 4%. Differential cross sections are measured as a function of several kinematic observables of the photon, leptons, and jets. Three examples of such measurements are shown in figure 3.2. A SMEFT interpretation is also performed and the results are combined with the ones obtained in the lepton+jets channel to improve the limits on the Wilson coefficients c_{tZ} and c_{tZ}^{I}. Figure 3.1 (centre) shows the limits obtained in the dilepton channel, and figure 3.1 (right) shows the combination between the two channels. The observed combined intervals at

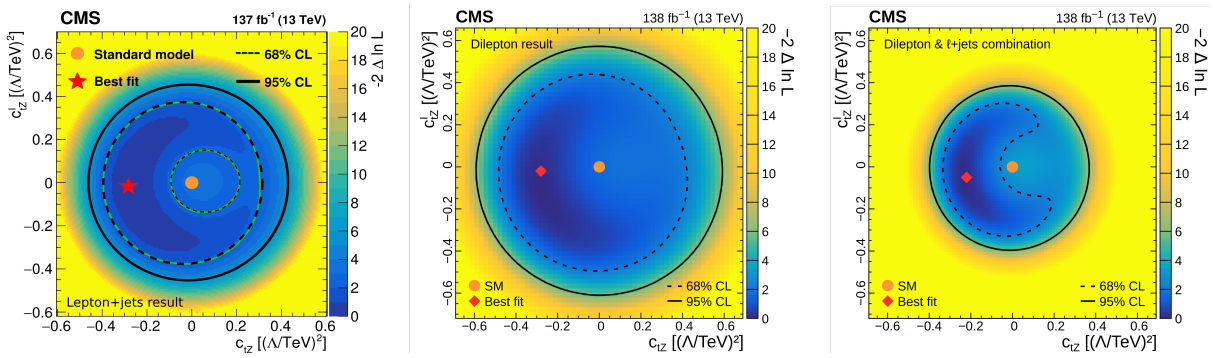

Figure 3.1: Result from the two-dimensional scan of the Wilson coefficients c_{tZ} and c_{tZ}^{I} using the photon p_T distribution from the lepton+jets analysis (left), the dilepton (centre) or the combination of the two (right). The shading quantified by the colour scale on the right reflects the negative log-likelihood difference with respect to the best fit value that is indicated by the red diamond. The 68% (dashed curve) and 95% (solid curve) CL contours are shown for the observed result. The orange circle indicates the SM prediction. Images adapted from Refs. [138, 139].

95% CL are $[-0.30, 0.00]$ for c_{tZ} (with c_{tZ}^{I} as a free parameter) and $[-0.28, 0.23]$ for c_{tZ}^{I} (with c_{tZ} as a free parameter).

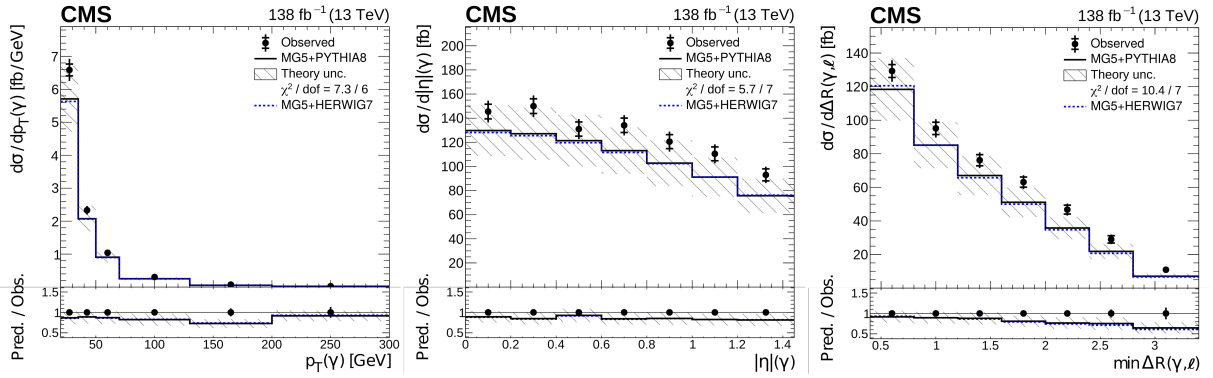

Figure 3.2: Absolute differential $t\bar{t}\gamma$ cross sections as functions of p_T (γ) (left), $|\eta|(\gamma)$ (centre) and $\min\Delta R(\gamma, \ell)$ (right) [139]. The data are represented by black points, with inner (outer) vertical bars indicating the statistical (total) uncertainties. The predictions obtained with the MadGraph5 aMC@nlo ME generator interfaced with Pythia 8 (solid lines) and Herwig 7 (dotted lines) PS simulations are shown as horizontal lines. The theoretical uncertainties in the prediction using Pythia 8 are indicated by shaded bands. The lower panels display the ratios of the predictions to the measurement.

This thesis presents inclusive and differential measurements of the $t\bar{t}\gamma/t\bar{t}$ ratio, for the first time using data from the LHC. Furthermore, differential measurements of $t\bar{t}\gamma$

including events with higher-rapidity photons that are detected in the EE of the CMS ECAL are presented for the first time. Furthermore, they are shown as a function of top quark and $\mathrm{t\bar{t}}$ observables that have not been measured before. The work uses as starting point the measurements from Ref. [139]. The event selection and the data-driven method to estimate the backgrounds are based on this previous work.

As discussed in section 1.3.5, it is also possible to use $\mathrm{t\bar{t}\gamma}$ events to measure the top quark charge asymmetry. In fact, such a study was published recently by the ATLAS collaboration [141]. The measured charge asymmetry is $A_C = -0.006 \pm 0.030$, in good agreement with the SM prediction at NLO in QCD of $A_C^{\mathrm{exp}} = -0.0139 \pm 0.001$ (where the uncertainty comes from the choice of renormalisation and factorisation scales). The large uncertainty in this measurement is mainly of statistical nature, as the asymmetry is a very small effect and more statistical power is needed in order to perform a more significant measurement. At the end of this chapter, a top quark charge asymmetry study using $\mathrm{t\bar{t}\gamma}$ events is also shown.

3.1.1 Signal and background definitions

The signal, $\mathrm{t\bar{t}\gamma}$, is defined as $\mathrm{t\bar{t}}$ in the dilepton final state associated with an energetic photon, and is generated using MadGraph, as described in the next section. The generated photon in each event must satisfy the following requirements:

- **Photon origin:** the photon must originate from the initial state quarks (*photon from ISR*), from one of the top quarks (*photon from top*) or its decay products (*photon from FSR*), but not from the hadronisation processes.
- **Transverse momentum and rapidity:** the photon must pass a p_{T} requirement at generator level of at least 10 GeV and $|\eta|$ lower than 5.
- **Isolation:** the photon must be isolated from all other final-state generated particles by $\Delta R > 0.1$, excluding neutrinos and low-momentum ($p_{\mathrm{T}} < 5\,\mathrm{GeV}$) particles.

The backgrounds can be classified as genuine photon backgrounds - processes that are expected to contain a photon in the final state, and nonprompt photon backgrounds. The "nonprompt photon" definition includes both real photons not passing the criteria above, and other objects misidentified as photons (known as "fakes"). We classify photons into five different categories, according to their origin. The categorisation is based on a matching procedure at generator level, where in each event, the leading reconstructed photon is matched to the nearest generator-level particle that has a p_{T} within 50% of that of the reconstructed photon. The matching fails in case no such generator-level is found within a ΔR cone of radius 0.3 around the photon. If the matched particle is a photon originating from either a lepton, a quark or a boson, the reconstructed photon is labelled as a genuine photon. Otherwise it is classified as nonprompt, with the following subcategories:

- Photon from hadron decay: the reconstructed photon is matched to a generator-level photon originating from a hadron decay (most commonly a π^0 meson). This type of photons are often called fragmentation or hadronic photons.

- Misidentified ("fake") electron: the reconstructed photon is matched to a generator-level electron. The contribution of this category to nonprompt photons is very small, especially after applying the pixel seed veto, described in section 3.3.1.

- Misidentified ("fake") jet: the matching procedure fails as there is no generated particle close to the reconstructed photon to carry at least 50% of its p_{T}. There are however multiple generated particles inside the ΔR cone around the reconstructed photon. These objects are not real photons, rather they correspond to hadronic jets.

- Photon from pileup: the matching procedure fails, as no particle is found within the ΔR cone. These photons are often attributed to pileup and represent a relatively large contribution to the nonprompt category. This is because photons are not reconstructed from tracks, and therefore it is not trivial to match them to the PV.

Figure 3.3 (left) shows the fractions of events with nonprompt photons in the signal region belonging to each of the categories mentioned above. Figure 3.3 (right) shows the distribution of the reconstructed photon p_{T}, for the different types of photons. It is clear that photons from pileup have the lowest p_{T} of all the categories.

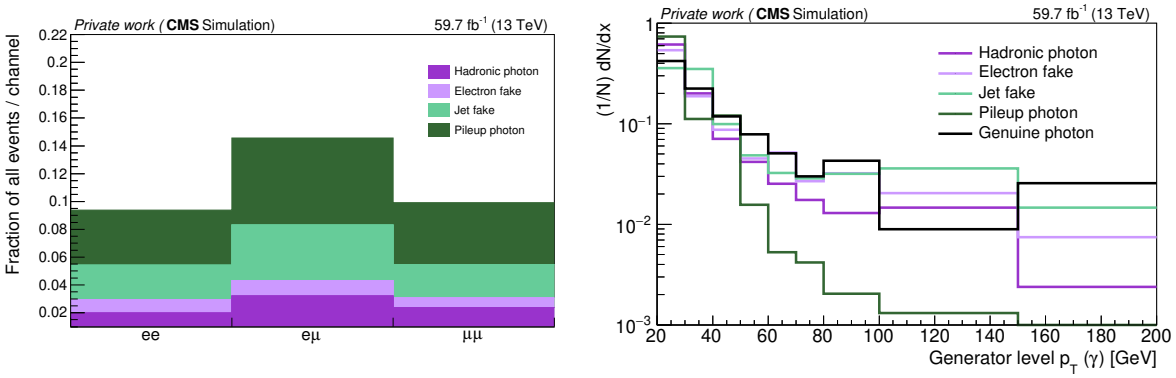

Figure 3.3: Left: Expected fractions of events in 2018 with nonprompt photons, in the signal region, for the different $\mathrm{t\bar{t}}$ decay modes. The different types of nonprompt photons are shown in different colours. The contribution from electron fakes is very small thanks to the pixel seed veto requirement. Right: Expected normalised distributions of $p_{\mathrm{T}}(\gamma)$ for the different types of nonprompt photons and for genuine photons.

3.2 Data and simulated event samples

For the measurements documented in this chapter, data from the LHC Run 2 recorded by the CMS experiment are used. The data correspond to an integrated luminosity of $\mathcal{L} = 138\,\text{fb}^{-1}$ of pp collision data at $\sqrt{s} = 13\,\text{TeV}$.

Events in data are required to pass a combination of HLT requirements that ensure the presence of charged leptons (electrons and/or muons). In order to maximise the selection efficiency, a combination of single and double lepton triggers is used to select the datasets to be included. Since the same event can be present in multiple datasets (for example, an event with two electrons can be both in the "single electron" and "double electron" datasets), a procedure is implemented to avoid double counting and make sure each event is selected only once.

To simulate the hard scattering in the signal and background processes, different MC event generators are used at different levels of accuracy. The modelling of PS and hadronisation are performed with PYTHIA 8.2 [142]. The underlying event is described using the CP5 tuning parameters [143] whereas NNPDF3.1 [30] is employed to model proton PDFs. The matching of the ME to the PS is performed using the MLM scheme [144] for samples generated at LO in perturbative QCD while for the NLO samples, the FxFx [36] merging is used.

Additional pp interactions in the same and/or neighbouring bunch crossings (pileup) are described by simulating minimum bias interactions with PYTHIA, added to the hard-scattering events. Simulated events are then reweighted to reproduce the pileup distribution in data. Details on the simulated samples are given below.

3.2.1 Signal samples

The $\mathrm{t\bar{t}}\gamma$ process is produced both at LO and NLO in perturbative QCD using MADGRAPH [145]. At NLO, $2 \to 3$ process $\mathrm{pp} \to \mathrm{t\bar{t}}\gamma$ is generated, where the top quarks are decayed to leptons using MADSPIN [146]. This sample includes diagrams where the photons originate from either top quark or from ISR. The LO sample, on the other hand, simulates the full $2 \to 7$ process $\mathrm{pp} \to \mathrm{t\bar{t}} \to \ell^+\ell^-\nu_\ell\bar{\nu}_\ell\mathrm{b\bar{b}}\gamma$, which includes also diagrams with the photon originating from FSR. Experimentally, the photon origin is indistinguishable and all possible origins must be considered in order to model the data accurately. Therefore, the $2 \to 7$ model at LO is used in this analysis, including cuts that are summarised in Table 3.1. The cuts define the fiducial phase space in which the cross section measurements are performed. Figure 3.4 (upper plot) shows the fractions of events in the $\mathrm{t\bar{t}}\gamma$ sample where the photon originates from ISR, from the top quark, and from FSR, after the event selection defined in section 3.3. Kinematic differences between events with different photon origins are also shown, for the photon p_T (lower left) and the minimum angular distance between the photon and the closest top quark, $\Delta R(\gamma, \text{closest t})$ (lower right).

The LO sample is normalised to the cross section of $\mathrm{t\bar{t}}\gamma$, evaluated at NLO, where

min. p_{T} (γ) [GeV]	max. $\lvert\eta(\gamma)\rvert$	max. $\lvert\eta(\ell)\rvert$	min. $\Delta R(\gamma, j)$	min. $\Delta R(\gamma, \ell)$
10	5	5	0.1	0.1

Table 3.1: Fiducial phase space where the $\mathrm{t\bar{t}}\gamma$ cross section is measured, based on the sample production requirements in MadGraph.

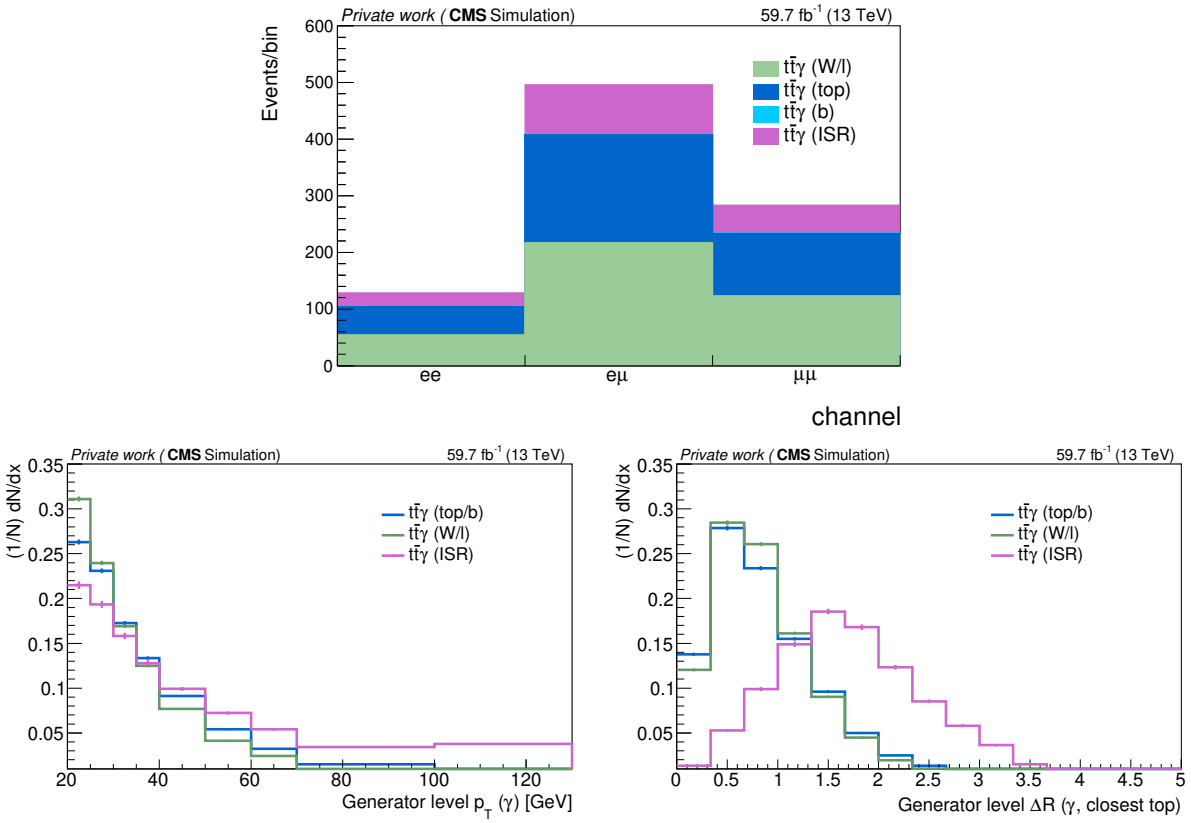

Figure 3.4: Comparison between the different possible origins of the photon in simulated $\mathrm{t\bar{t}}\gamma$ samples. The expected yields for the different final state channels, for simulation of 2018 data, for the different possible photon origins are shown in the upper plot. The kinematic distributions for the generator-level p_{T} (γ) and $\Delta R(\gamma, \text{closest top})$ are also shown in the lower left and lower right, respectively.

special attention is paid to the shape of the photon p_{T} distribution [138]. This means to derive the NLO-to-LO corrections, also known as k-factors, as a function of photon p_{T}. The NLO $\mathrm{t\bar{t}}\gamma$ sample does not contain FSR photons. This is compensated by adding events from the NLO $\mathrm{t\bar{t}}$ sample with a photon from the top quark decay. The photon p_{T} distribution is then compared to that of the LO $\mathrm{t\bar{t}}\gamma$ sample to derive the k-factor. The shape of the distribution is found to be similar between the two samples and an inclusive

k-factor of 1.49 is shown to be sufficient to bring the LO signal sample to NLO accuracy.

A sample simulating $t\bar{t}\gamma$ at LO in QCD in the lepton+jets decay channel is also included as a background source, with the same k-factor. The sample names and the respective cross sections at NLO are listed in table 3.2.

Process	Truncated sample name	Cross section [fb]
$t\bar{t}\gamma$ dilepton (signal)	TTGamma_Dilept_TuneCP5_13TeV-madgraph-pythia8	2220.37
$t\bar{t}\gamma$ ℓ + jets (background)	TTGamma_SingleLept_TuneCP5_13TeV-madgraph-pythia8	7509.17

Table 3.2: MC $t\bar{t}\gamma$ samples and production cross sections considered in the analysis.

3.2.2 Simulated background samples

The background samples and their cross sections are summarised in table 3.3. Background events from $t\bar{t}$, $t\bar{t}H$, tW, and the t-channel single top production are generated at NLO precision using POWHEG (v2.0) [147–149]. To generate the single top, $t\bar{t}$ and $t\bar{t}\gamma$ samples, a top quark mass of 172.5 GeV is used. The $t\bar{t}$ cross section is scaled to its best theoretical prediction at NNLO in QCD, $\sigma = 832 \pm 41.6$ pb [150]. All other background samples are produced with MadGraph [145] at LO or NLO accuracy in QCD. All background processes are normalised according to the measured integrated luminosity and to their most accurate theoretical cross sections. In the table, "DY+jets" refers to Drell-Yan, a process of on-shell Z boson or virtual photon production, together with additional jets.

Simulation of the tWγ background

The tWγ process, i.e., single top production in association with a W boson and a photon, is a non-negligible background contribution. This contribution is partially accounted for by using the tW samples, which contain photons from FSR. In addition to missing ISR photons, the number of tW events with a hard isolated photon is rather small, leading to a high statistical uncertainty in the estimation of this background.

A better handle on this background contribution is achieved by producing a dedicated tWγ sample at LO with MadGraph, including all possible photon origins (ISR, t, W or FSR). A $2 \to 6$ process is generated with a resonant top quark.

The cross section computed from the simulated sample is 159.2 fb. This cross section is corrected to NLO using a k-factor that is evaluated with the same method as that of $t\bar{t}\gamma$. A tWγ sample at NLO is produced using MadGraph5 aMC@nlo where top quarks are decayed via Madspin, meaning that FSR photons from decay products are not included. This is compensated using tW events with an additional photon from PS. The overlap of the NLO tWγ sample with $t\bar{t}\gamma$ is removed using the DR2 method [151]. The method removes the overlapping diagrams, while keeping the interference terms in the ME amplitude calculation. A similar strategy was employed in the recent CMS result

Process	Cross section [fb]
$t\bar{t}$ dilepton	87310
$t\bar{t}$ ℓ + jets	364360
tW top	34900
tW antitop	34900
DY+jets ($M_{\ell\ell} \in [10, 50]$ GeV)	18610000
DY+jets ($M_{\ell\ell} > 50$ GeV)	6225420
Zγ	55230
W+jets	61526700
Wγ	191000
WW	116100
ZZ	16520
WZ	47130
ZZZ	12.98
WWZ	165.1
WZZ	55.65
WWW	208.6
$t\bar{t}$Z (hadronic Z)	529.7
$t\bar{t}$Z (leptonic Z)	281.36
$t\bar{t}$W	204.3
tWZ (leptonic top and W)	1.50
tWZ (hadronic top, leptonic W)	3.004
tWZ (leptonic top, hadronic W)	3.004
tZq	94.18
$t\bar{t}$WW	11.5
$t\bar{t}$WZ	2.919
$t\bar{t}$ZZ	1.573
$t\bar{t}$H (not decaying to $b\bar{b}$)	215.1

Table 3.3: MC samples and cross sections of the backgrounds considered in the analysis.

on the production of tWZ [152]. The process card for generating the sample is shown in appendix B.

A fiducial phase space is defined in order to derive the k-factor, since the different samples have different p_{T}, $|\eta|$ and isolation requirements for the various objects, and therefore a direct comparison before any selection is not possible. The requirements of this fiducial phase space are summarised in table 3.4.

Figure 3.5 shows a comparison between the $p_{\mathrm{T}}(\gamma)$ distribution of the LO tWγ sample and that of the NLO tWγ sample, at the generator level. It is clear from the ratio at the

Cut	Generated ℓ	Generated γ
p_{T} [GeV]	15	20
$\lvert\eta\rvert$	2.5	2.5
Particle flavour	Electron, muon or tau-lepton	Photon
Origin	from W decays	not from hadronisation
Others	$m(\ell\ell) > 30$	$\Delta R(\gamma, \ell) > 0.4$, $\Delta R(\gamma, \mathrm{all}) > 0.1$

Table 3.4: Definition of objects used to define the fiducial phase space. In the last line, "all" refers to all particles with a p_{T} greater than 5 GeV, except neutrinos and other photons.

bottom of the figure that the shapes of the two distributions are compatible. An inclusive k-factor of 1.26 is extracted and subsequently applied to correct the LO cross section.

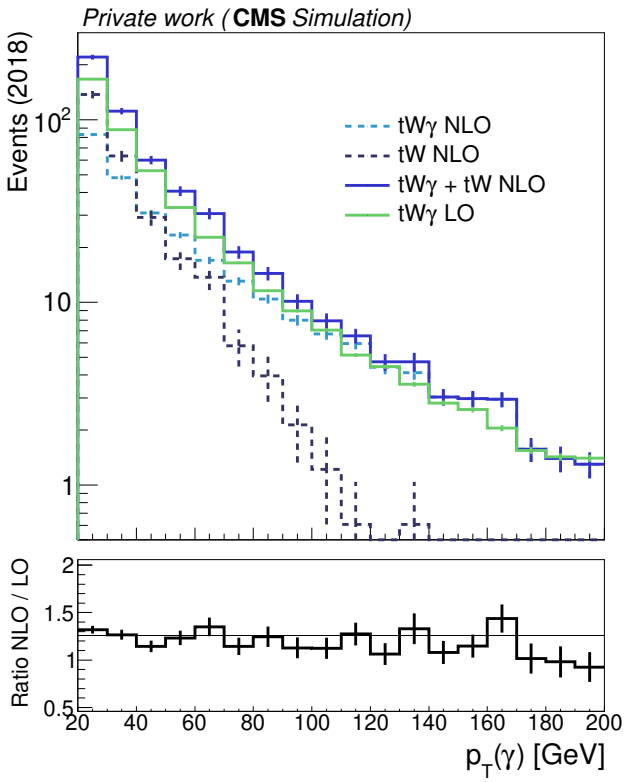

Figure 3.5: Comparison of the p_{T} (γ) distribution of the tWγ NLO sample (blue) with the tWγ LO sample used in the analysis (green). Photons from top and W decays are not present in the NLO sample and are added from the tW sample (dark blue dashed line).

3.2.3 Overlap removal

The kinematic phase space of $t\bar{t}\gamma$ overlaps with that of $t\bar{t}$, when photons appear in the decay products of top quark pair production. To remove the overlap between the samples, the signal definition from section 3.1.1 is used to perform two separate removals:

- From the $t\bar{t}$ dilepton sample, remove all events containing a photon that fulfils the signal photon definition.
- From the $t\bar{t}\gamma$ dilepton sample, remove all events with no photons fulfilling the signal photon definition. This is a small number of events, generated with photon p_T or η very close to the limits of the phase space, that after the PS have its kinematics slightly changed, such that they fall out of the phase space.

Figure 3.6 shows the fraction of simulated events passing and failing the overlap removal from the $t\bar{t}\gamma$ and $t\bar{t}$ samples, after the full event selection as defined in section 3.3.2.

The exact same procedure is employed to remove the overlap between tW and $tW\gamma$, and DY and $Z\gamma$, but modifying the photon requirements to those used to generate the $tW\gamma$ and $Z\gamma$ processes.

3.3 Event reconstruction

The $t\bar{t}\gamma$ process has an expected cross section of 2220 fb, in the final state with two leptons (including τ leptons), which translates to about 300.000 $t\bar{t}\gamma$ events, before any selection. This large number allows us to perform differential measurements, provided that a reasonable understanding of the background contributions is achieved.

The largest background source originates from events with a nonprompt photon, in particular $t\bar{t}$ events, due to the much larger cross section of $t\bar{t}$ when compared to $t\bar{t}\gamma$. For this reason, the most important aspect in the event selection is to maximise the accuracy in the reconstruction of genuine photons. Hence, several quality requirements are applied to the photons in each event, and a dedicated estimation of the nonprompt photon contribution is performed using the data.

3.3.1 Object selection

The leptons coming from the $t\bar{t}$ decay have to be identified. For that purpose, both electrons and muons are required to satisfy a number of quality criteria. A *cut-based* identification approach is employed, meaning there is a sequential application of quality requirements. For electrons, these requirements aim to identify electromagnetic showers in the ECAL arising from the primary vertex. Variables describing the shape of the shower

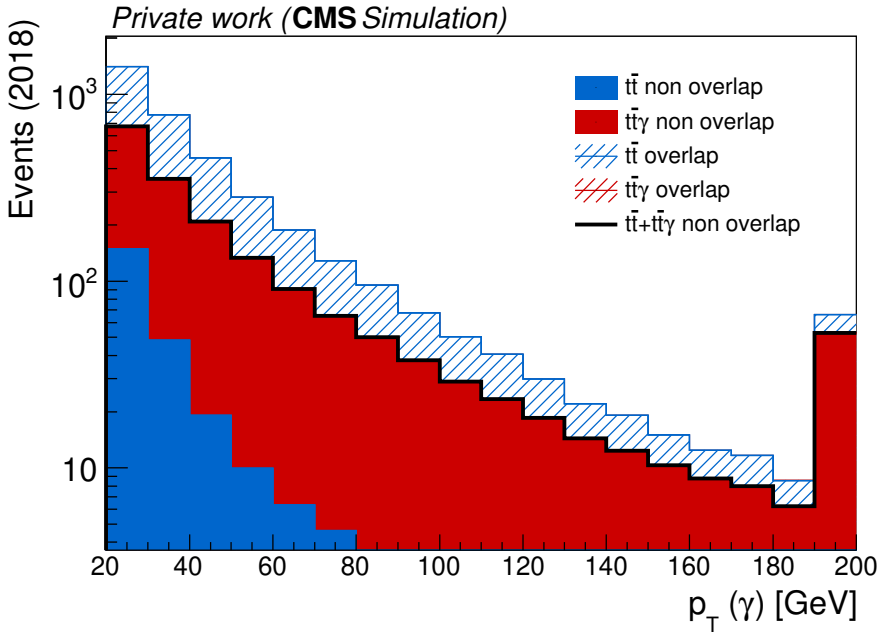

Figure 3.6: Comparison of the $p_{\mathrm{T}}(\gamma)$ distributions for the simulated $\mathrm{t\bar{t}}\gamma$ and $\mathrm{t\bar{t}}$ samples, showing the overlap between the two samples (dashed areas), which is removed before the selection. The solid black line shows the total contribution from both samples, after removing the overlap. The overlapping contribution from the $\mathrm{t\bar{t}}\gamma$ sample is very small (< 1 event) and is therefore not visible in the plot. The number of events is scaled to the luminosity collected in 2018 ($59.7\,\mathrm{fb}^{-1}$).

are combined with additional information from the HCAL and the tracker. Some of these requirements are optimised separately for electrons in the barrel and in the endcap, such that there is an overall identification efficiency of about 70% for real electrons, while keeping less than 1% of "fake" electrons (within the collaboration, this is known as the "tight" selection). The signal efficiency depends on the electron p_{T}, and can be as low as 60% for low-p_{T} electrons, but reaches a *plateau* of more than 85% for $p_{\mathrm{T}} > 60\,\mathrm{GeV}$. Muons are required to satisfy a "medium" selection, requiring them to satisfy Tracker muon requirements, as defined in section 2.3.2. Then, in case a muon is not reconstructed also as a Global muon, some additional requirements are applied to ensure that the segment from the muon stations is compatible with a hit in the tracker, such that an overall identification efficiency of 99.5% is achieved [112].

Jets are reconstructed using the PF algorithm introduced in section 2.3.2. They are further classified into light jets or jets originating from the hadronisation of b quarks. This is done by running each jet through the DeepJet algorithm, in the "medium" working point [153]. This is a b tagging algorithm which relies on modern deep learning techniques, that identifies jets from b quarks with an efficiency of about 80%, and a misidentification probability of 1% for light-flavour or gluon jets.

The photons are reconstructed as detailed in section 2.3.2. To further reduce the background, a cut-based identification algorithm is applied, based mostly on energy information from the calorimeters [123]. The variables used for the selection are the following:

- Hadronic over electromagnetic energy ratio (H/E): The H/E ratio is defined as the ratio between the energy deposited in the HCAL in a cone of radius $\Delta R = 0.15$ around the direction of the SC and the energy of the photon or electron candidate. The sources of energy in the HCAL for electromagnetic objects (electrons and photons) is noise, pileup and leakage of the objects through the calorimeter gaps. This is mainly to discriminate against neutral hadrons.

- Width of the EM shower ($\sigma_{\mathrm{i}\eta\mathrm{i}\eta}$): The $\sigma_{\mathrm{i}\eta\mathrm{i}\eta}$ is defined as the second moment of the log-weighted distribution of crystal energies in η, calculated in the 5×5 matrix around the most energetic crystal in the SC and re-scaled to units of crystal size. This distribution is expected to be narrow for electrons and single photons, and wider for double-photon signals originating from the decays of π^0 mesons.

- Charged and neutral hadron and photon isolation: The isolation variables are obtained by summing the transverse momenta of charged hadrons (I_{ch}), neutral hadrons (I_{n}) or photons (I_{γ}) inside an isolation cone of $\Delta R = 0.3$ with respect to the photon direction. The neutral hadron and photon isolation are computed as a function of the photon p_{T}.

The "medium" working point used in the analysis corresponds to an 80% selection efficiency, and the cuts which are applied are detailed in table 3.5. The cuts are different

for photons in the EB and in the EE, due to the different crystal structure of the ECAL in the two regions.

Variable	Medium WP requirement (EB)	Medium WP requirement (EE)
H/E	0.02197	0.0326
$\sigma_{i\eta i\eta}$	0.0105	0.0272
I_{ch}	1.141	1.051
I_{n}	$1.189 + 0.01512\ p_{\text{T}} + (2.259 \times 10^{-5}\ p_{\text{T}}^2$	$2.718 + 0.0117\ p_{\text{T}} + 2.3 \times 10^{-5}\ p_{\text{T}}^2$
I_{γ}	$2.08 + 0.004017\ p_{\text{T}}$	$3.867 + 0.0037\ p_{\text{T}}$

Table 3.5: Quality requirements on photons, corresponding to the medium working point (WP) for the cut-based identification algorithm described in Ref. [123]. In the formulas, p_{T} refers to the photon p_{T}.

Photons are further required to pass a pixel seed veto (PSV), which removes reconstructed photons that have a track in the silicon tracker compatible with belonging to the same particle. This cut is applied in order to reduce the number of electrons misidentified as photons.

3.3.2 Event selection

When recording the data, events are associated with "HLT paths", depending on which physics objects are identified at the HLT level, with certain p_{T}, η and isolation requirements. For the three years, events in data and simulation are required to be associated with one of the HLT paths in appendix A. These were chosen with the lowest p_{T} thresholds available, in order to maximise the selection efficiency. The paths follow a naming convention where "EleX" ("MuX") refers to the presence of an electron (muon) with p_{T} above X GeV.

After the HLT requirements, events which contain at least two oppositely charged leptons are selected. Both leptons are required to be reconstructed within $|\eta| < 2.4$. The leading lepton is required to have $p_{\text{T}} > 25\,\text{GeV}$, while the second leading lepton is required to have $p_{\text{T}} > 20\,\text{GeV}$. Furthermore, the dilepton system is required to have an invariant mass above 20 GeV ($M_{\ell\ell} > 20\,\text{GeV}$). This cut is imposed to reject any residual background from QCD multijet processes (with jets misreconstructed as leptons) as well as the low mass resonances decaying into two leptons.

For events with two reconstructed leptons of the same flavour, the invariant mass of the dilepton system is required to be outside of the Z boson mass peak, within 15 GeV ($M_{\ell\ell} \notin [76, 106]\,\text{GeV}$). Additionally, events are required to contain at least two jets and at least one b-tagged jet. They are required to have $p_{\text{T}} > 30\,\text{GeV}$ and $|\eta| < 2.4$, and to be separated from leptons by $\Delta R > 0.4$. This is mainly to avoid selecting the same object both as lepton (especially electron) and jet.

Events are finally required to have at least one reconstructed photon with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and $|\eta| < 2.5$, satisfying the criteria detailed above. The isolation requirements $\Delta R(\gamma, \ell) > 0.4$ and $\Delta R(\gamma, j) > 0.1$ are also applied to the photons. In order to further reject background coming from $Z + \gamma$ events, the invariant mass of the system composed by the two leptons and the photon is also required to be outside of the Z boson mass peak, $M_{\ell\ell\gamma} \notin [76, 106]\,\mathrm{GeV}$, for events with two same-flavour leptons. Events passing the full selection described above, and summarised in table 3.6, define the signal region (SR) for this analysis. In order to have a better control over the $\mathrm{Z} + \gamma$ background, which is the main background in the same flavour categories, a $\mathrm{Z} + \gamma$ control region (CR) is built, as described in section 3.6.1.

Object	Signal region selection	
Leptons	≥ 2 leptons	$M_{\ell\ell} > 20\,\mathrm{GeV}$
	$p_{\mathrm{T}}(\ell) > 25, 20\,\mathrm{GeV}$	$M_{\ell\ell} \notin [76, 106]\,\mathrm{GeV}$ or $\mathrm{e}\mu$
	$\lvert\eta(\ell)\rvert < 2.4\,\mathrm{GeV}$	
Jets	≥ 2 jets	≥ 1 b jets
	$p_{\mathrm{T}}(j) > 30\,\mathrm{GeV}$	$\lvert\eta(j)\rvert < 2.4\,\mathrm{GeV}$
	$\Delta R(j, \ell) > 0.4$	
Photons	≥ 1 photon	$p_{\mathrm{T}} > 20\,\mathrm{GeV}$
	$\lvert\eta\rvert < 2.5$	$M_{\ell\ell\gamma} \notin [76, 106]\,\mathrm{GeV}$ or $\mathrm{e}\mu$
	$\Delta R(\gamma, \ell) > 0.4$	$\Delta R(\gamma, j) > 0.1$

Table 3.6: Signal region requirements for the $\mathrm{t\bar{t}}\gamma$ analysis.

Corrections applied to simulation

Trigger scale factors (SFs) are applied to the simulated samples to account for the differences in efficiency of the trigger selection between data and simulation. These SFs are derived by comparing the efficiencies obtained when using single lepton and dilepton triggers, to the ones measured using a set of cross triggers which select on $p_{\mathrm{T}}^{\mathrm{miss}}$ (MET triggers) [154].

The simulated samples are also reweighted to account for the different efficiencies in selecting electron(s), muon(s) and photon(s) in data, with respect to simulation. For electrons and photons, the reweighting accounts separately for the differences in efficiency in the reconstruction and identification. For muons, the reweighting accounts for the efficiency differences in the identification and isolation criteria. For all three objects, the SFs are derived using a "tag-and-probe" method, in data and in $\mathrm{Z} \to \mathrm{e}^{+}\mathrm{e}^{-}$ and $\mathrm{Z} \to \mu^{+}\mu^{-}$ simulated events [112, 155]. Electrons and photons are very similar objects

from the reconstruction point of view, except that the former contain tracks in the tracker while the latter don't. Therefore, "photon-like" objects are created by taking electrons and "erasing" the tracks, and then these objects are used to derive the efficiencies for photons. Additionally, for photons, further SFs to account for the efficiency of the PSV [123] are applied.

The efficiencies of b tagging are computed from the MC samples, specifically for the selection of this analysis, by comparing the number of b-tagged jets to the total number of jets, per jet flavour, as a function of p_{T} and $|\eta|$ of the jet. These efficiencies are multiplied by SFs derived centrally in the CMS collaboration from QCD-enriched regions in data and dedicated MC samples [156]. The resulting factors are applied to the simulated events in this analysis.

3.4 Comparison between data and simulation

After the full event reconstruction, the final expected signal and background contributions are compared to the data for several distributions, to ensure that a good description is achieved. In the following comparisons, the "NP (data)" contribution corresponds to the background from nonprompt photons, estimated with the method that will be described in section 3.6.2. Figures 3.7 and 3.8 show the distribution of the leading lepton p_{T} for the different data-taking years separately, for the SR of the analysis. The $\mathrm{Z}\gamma$ contribution is separated in $\mathrm{Z}\gamma$ FSR and ISR, as explained in section 3.6.1. Only the statistical uncertainty is shown, both from the data (black error bars) and the simulated samples (black dashed area).

Figure 3.9 shows the distributions of several kinematic variables, this time for data collected in the full Run 2. In general, a good modelling is achieved with the signal and background templates, and no significant shape mismodelling is observed. However, for all years except 2018, there is an excess of data with respect to the predictions. This is partially covered by the systematic uncertainties, but not fully, and will be reflected on the results shown at the end of this chapter.

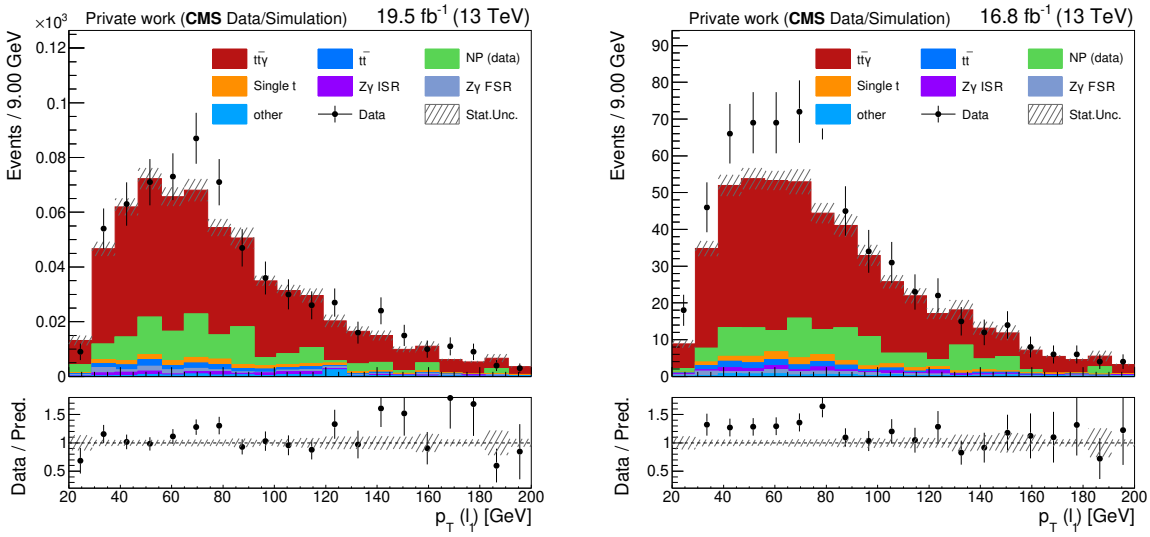

Figure 3.7: Comparison between data and simulated events for the leading lepton p_{T}, for each of the data-taking eras separately: 2016 preVFP (left), and 2016 postVFP (right). Only the statistical uncertainty is shown.

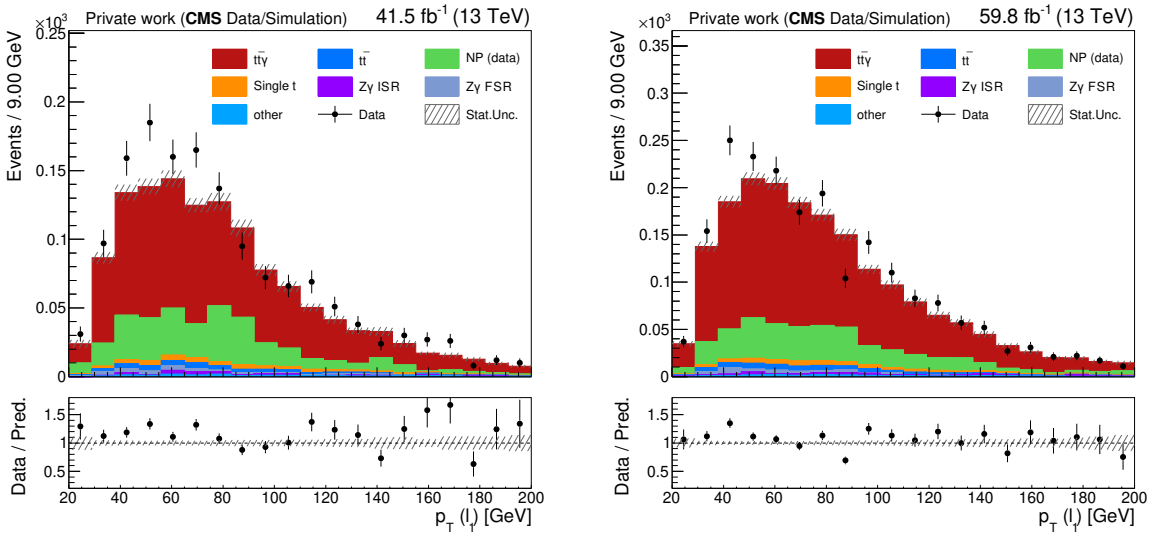

Figure 3.8: Comparison between data and simulated events for the leading lepton p_{T}, for each of the data-taking eras separately: 2017 (left), and 2018 (right). Only the statistical uncertainty is shown.

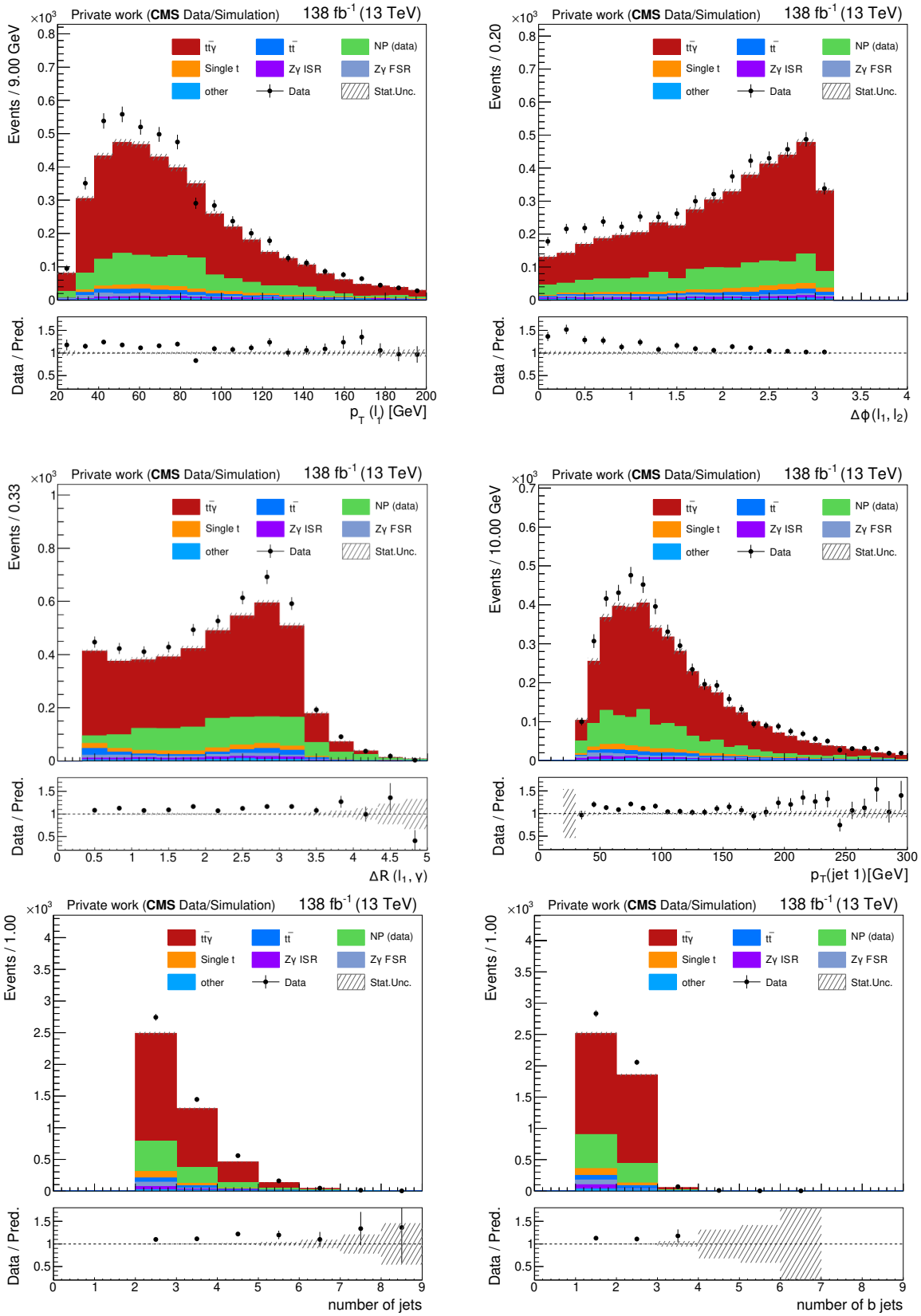

Figure 3.9: Comparison between data and simulated events for the leading lepton p_{T} (upper left), $\Delta\phi(\ell,\ell')$ (upper right), ΔR(lead. ℓ,γ) (centre left), p_{T} (leading jet) (centre right), number of jets (lower left), and number of b jets (lower right), for the full Run 2 data. Only the statistical uncertainty is shown.

3.5 Reconstruction of the $t\bar{t}$ system

Assuming that the only source of $E_{\mathrm{T}}^{\mathrm{miss}}$ are the two neutrinos, it is possible to reconstruct the $t\bar{t}$ system. In the literature, two types of analytical algorithms to reconstruct $t\bar{t}$ kinematics in the dilepton channel are described: the solution developed by Sonnenschein [157] and the one developed by Betchart and collaborators [158]. In the CMS collaboration, a method known as *kinReco* was developed, based on the prescription by Sonnenschein, which is used in a number of $t\bar{t}$ measurements (see, e.g. Refs. [159, 160]). Six kinematic constraints are applied, in order to determine the 4-momenta of the two neutrinos. Equations 3.1 to 3.6 are solved analytically, resulting in a maximum of four solutions per event. The solution with lowest invariant mass of the $t\bar{t}$ system is chosen.

$$E_x = p_{x,\nu} + p_{x,\bar{\nu}} \tag{3.1}$$

$$E_y = p_{y,\nu} + p_{y,\bar{\nu}} \tag{3.2}$$

$$m^2_{W^+} = (E_{l^+} + E_\nu)^2 - (p_{x,l^+} + p_{x,\nu})^2 - (p_{y,l^+} + p_{y,\nu})^2 - (p_{z,l^+} + p_{z,\nu})^2 \tag{3.3}$$

$$m^2_{W^-} = (E_{l^-} + E_{\bar{\nu}})^2 - (p_{x,l^-} + p_{x,\bar{\nu}})^2 - (p_{y,l^-} + p_{y,\bar{\nu}})^2 - (p_{z,l^-} + p_{z,\bar{\nu}})^2 \tag{3.4}$$

$$m^2_t = (E_{l^+}+E_\nu+E_b)^2-(p_{x,l^+}+p_{x,\nu}+p_{x,b})^2-(p_{y,l^+}+p_{y,\nu}+p_{y,b})^2-(p_{z,l^+}+p_{z,\nu}+p_{z,b})^2 \tag{3.5}$$

$$m^2_{\bar{t}} = (E_{l^-}+E_{\bar{\nu}}+E_{\bar{b}})^2-(p_{x,l^-}+p_{x,\bar{\nu}}+p_{x,\bar{b}})^2-(p_{y,l^-}+p_{y,\bar{\nu}}+p_{y,\bar{b}})^2-(p_{z,l^-}+p_{z,\bar{\nu}}+p_{z,\bar{b}})^2 \tag{3.6}$$

To improve reconstruction efficiency, the energies and directions of jets and leptons are smeared according to the detector resolution, and a weighted average is taken. The reconstruction efficiency of this algorithm is about 90%, i.e., the algorithm finds a physical solution for $t\bar{t}$ kinematics in about 90% of the events that pass the $\geq$2 leptons and $\geq$2 b-tagged jets selection. For about 65% of $t\bar{t}$ simulated events, the solution which is found is the "correct" one, in the sense that the leptons and b jets are correctly associated.

The photon in the top quark pair reconstruction

The reconstruction described thus far ignores the presence of additional particles, namely photons, when computing the $t\bar{t}$ kinematics. Hence, it is important to check that an appropriate reconstruction can still be achieved when there is an additional photon in the event, i.e., in $t\bar{t}\gamma$ events. While the photon has no influence in the result when it is

emitted from the initial state quarks, this is not the case if the photon originates from one of the t quarks or their decay products. The latter constitutes, as described in section 1.3.3, a sizeable portion of events.

In order to assess whether applying the standard $t\bar{t}$ reconstruction without including the photon provides suitable results, the performance of the algorithm was compared for the $t\bar{t}$ and the $t\bar{t}\gamma$ samples. Figure 3.10 (left) shows the efficiency of the reconstruction (fraction of events for which a solution is found) as a function of the leading lepton p_T, for $t\bar{t}$ and $t\bar{t}\gamma$ events after the event selection without requirement on the presence of reconstructed photons. The efficiency of the reconstruction is also shown for events in data. It is clear that the efficiency for $t\bar{t}\gamma$ events is the same as for $t\bar{t}$, and therefore the presence of the photon does not reduce the efficiency.

Figure 3.10 (right) also shows the efficiency of the reconstruction for both samples and the data, but for events satisfying the full event selection. The efficiency is shown as a function of the photon p_T, showing that there is also no reduction of efficiency when highly energetic photons are present.

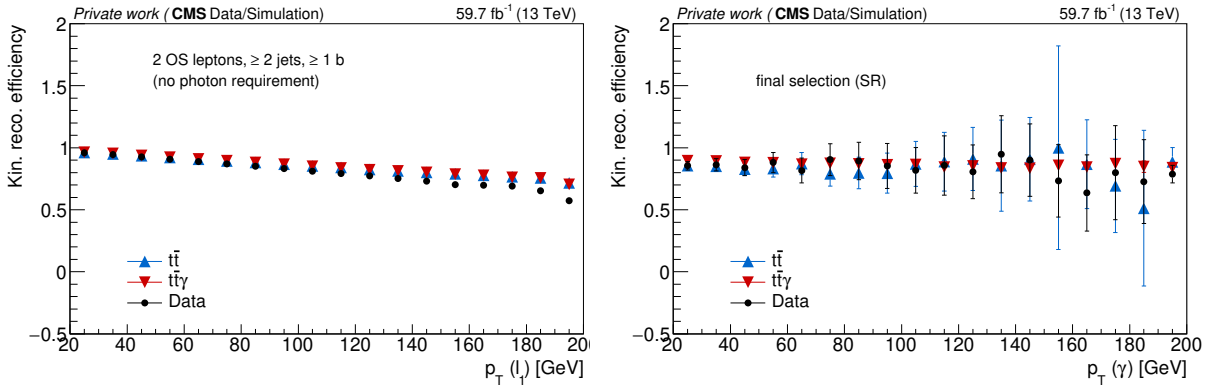

Figure 3.10: Efficiency of the $t\bar{t}$ kinematic reconstruction as a function of the leading lepton p_T (left) and the photon p_T (right), for $t\bar{t}\gamma$, $t\bar{t}$, and data events from 2018.

Apart from the efficiency of the reconstruction, it is also important to prove that the resolution of the reconstruction is unaffected by the presence of the photon. This can be quantified by computing the resolution of the reconstructed quantities with respect to the generated ones. For the leading top p_T, we can define the resolution as in equation 3.7.

$$\text{Resolution} = \frac{p_T(t)^{\text{reco}} - p_T(t)^{\text{gen}}}{p_T(t)^{\text{gen}}} \tag{3.7}$$

Figure 3.11 shows this resolution for $t\bar{t}$ and $t\bar{t}\gamma$ simulation, for events passing the lepton and b jet selection (left) and the full event selection (right). A 31% resolution is obtained for both processes for the basic selection, which worsens to 34% when requiring the full event selection. No difference in the resolution is found between the $t\bar{t}\gamma$ and $t\bar{t}$ samples. However, it is visible from the ratio in the lower part of the plots that in

the $t\bar{t}\gamma$ sample the reconstruction tends towards slightly lower values of reconstructed top p_T. This is expected, because in case the photon is emitted from the top quark decay products, it carries some of the p_T of the system and the reconstruction has this contribution missing.

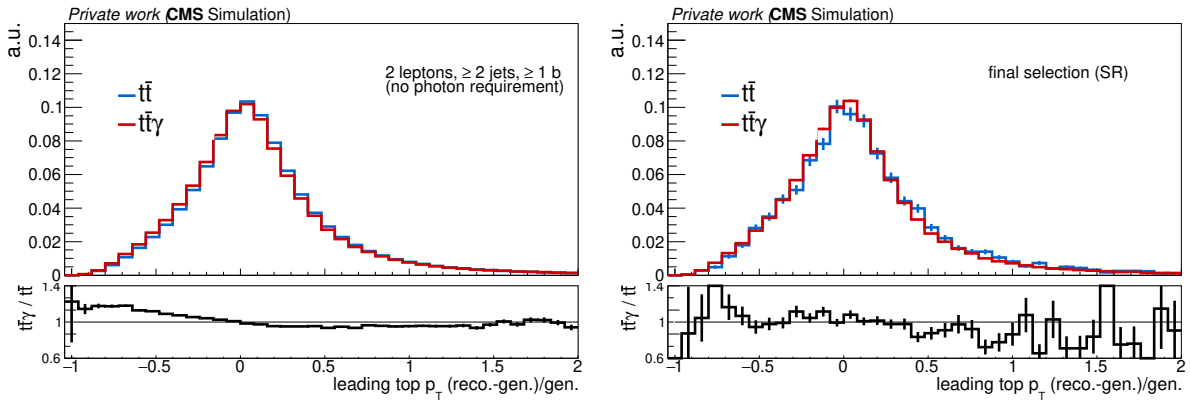

Figure 3.11: Resolution of the reconstructed leading top p_T for $t\bar{t}\gamma$ and $t\bar{t}$ simulated events. The figure on the left (right) shows events passing the lepton and b jet (final event) selection.

One remaining question is whether this resolution varies with the p_T of the photon. It could be the case that for high p_T photons the reconstruction would be increasingly biased and the resolution would be higher (worse). In order to test this, the graph in figure 3.11 (right) was repeated for different bins of photon p_T, and Gaussian functions were fitted to all the graphs. The widths of the resulting Gaussian functions for each photon p_T bin are shown in figure 3.12. It is clear that the resolution in fact improves with photon p_T, being lower (better) for higher photon p_T.

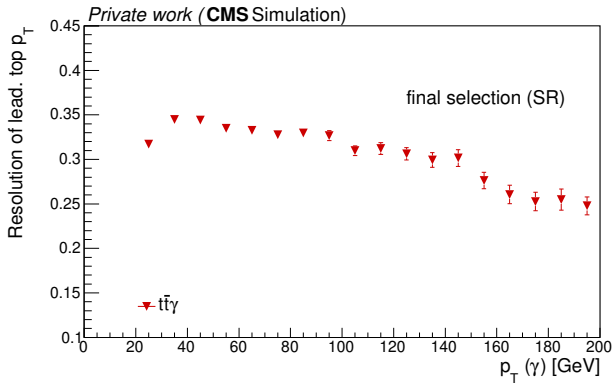

Figure 3.12: Resolution of the reconstructed leading top p_T for $t\bar{t}\gamma$ simulated events, as a function of the photon p_T.

After these studies, the conclusion is that it is reasonable to apply the standard $t\bar{t}$ reconstruction algorithm to $t\bar{t}\gamma$ events. Figure 3.13 shows the comparison between the data and the signal and background templates for several kinematic distributions involving the reconstructed top quarks and $t\bar{t}$ system, for the full Run 2 data.

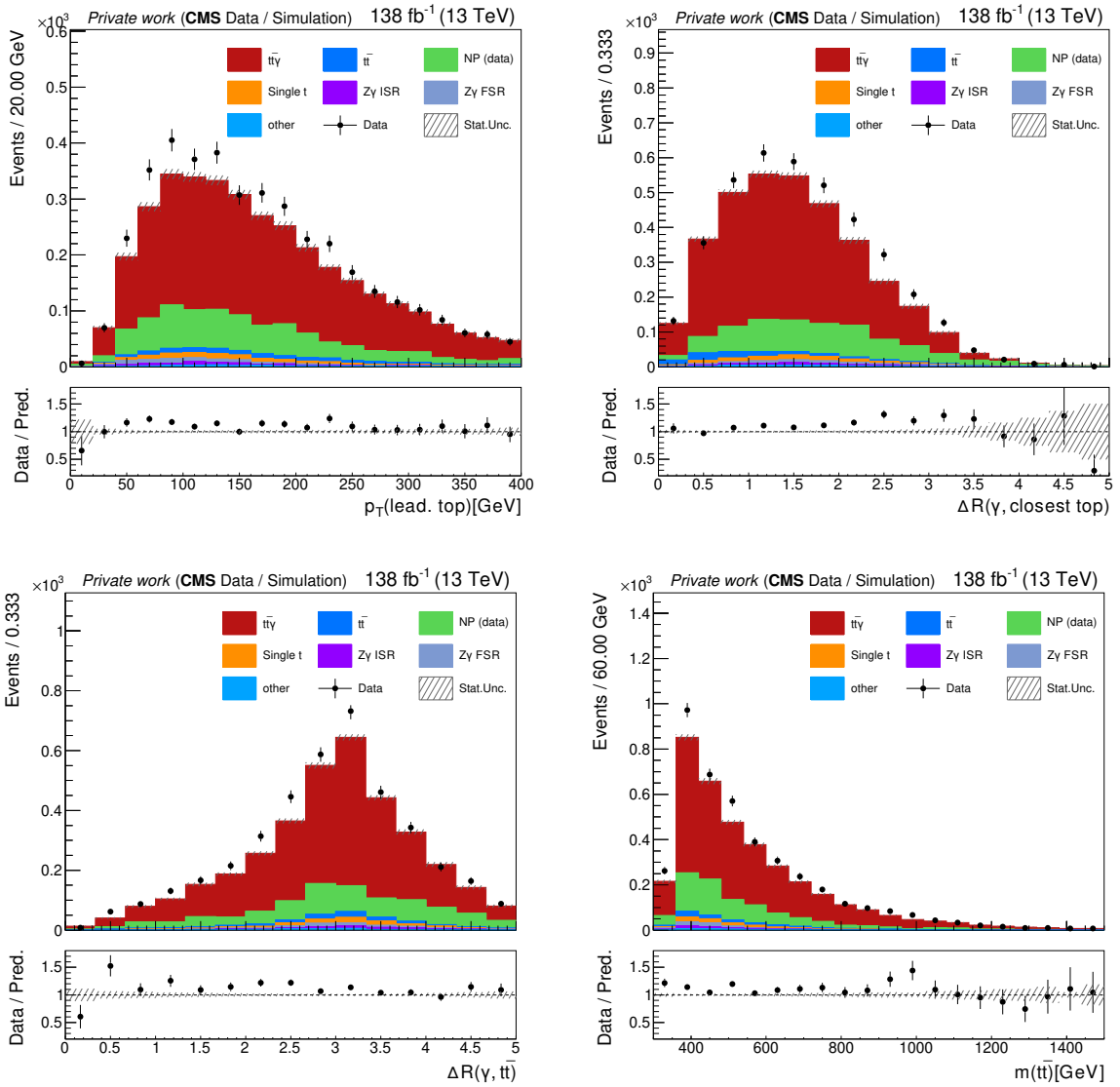

Figure 3.13: Comparison between data and simulated events for the leading top p_{T} (upper left), $\Delta R(\gamma, \text{closest top})$ (upper right), $\Delta R(\gamma, t\bar{t})$ (lower left), and $m(t\bar{t})$ (lower right), for the full Run 2 data. Only the statistical uncertainty is shown.

In Ref. [161], the authors propose a likelihood-fit based method to perform the reconstruction of the $t\bar{t}$ kinematics while considering the photon explicitly. The most crucial step of this reconstruction is to correctly assign the photon to the particle which origi-

nates it. In this thesis, such a method is not applied, but it could be considered in future studies.

3.6 Background estimation

3.6.1 The Zγ background

A CR is defined for the largest non-top-quark background contribution in the categories with same flavour leptons, which comes from the Zγ process. This CR is defined by inverting the $m_{\ell\ell\gamma}$ requirement, i.e., by requiring $76 < m_{\ell\ell\gamma} < 106\,\text{GeV}$ and two same-flavour leptons. In the Zγ process, much like in the signal, the photon can have different origins: it can originate from ISR or from FSR (not from the Z boson, which is electrically neutral). This CR especially targets photons from FSR, since those can typically be reconstructed with $m_{\ell\ell\gamma} \in [76, 106]\,\text{GeV}$. Therefore, it is effectively a Zγ FSR CR, and to account for that, the simulated Zγ process is split in its ISR and FSR components, using generator-level information. Figure 3.14 shows the distributions of several kinematic variables in the CR.

3.6.2 Background from nonprompt photons

After the selection of section 3.3, the largest remaining background consists of $t\bar{t}$ events where a photon has been reconstructed, but it is a nonprompt photon, as defined in section 3.1.1. Additionally, there is a sizeable contribution from single top and DY production, also in association with a nonprompt photon.

These nonprompt contributions are included in the simulated background samples, but given the good background suppression, a very large number of events would have to be generated to accurately simulate them. Moreover, it is not guaranteed that the simulation contains all possible sources of nonprompt photons from the data.

Therefore, in order to reduce the statistical uncertainty and to ensure an appropriate modelling of this background, its contribution is extracted directly from the data. The method used here is normally referred to as an ABCD method and is used in a number of high energy physics analyses. The method presented here is the same method used in the previously published $t\bar{t}\gamma$ analysis by CMS [162], with some modifications, which are described in detail throughout this section.

The method consists in defining four orthogonal regions of events (denoted as A, B, C and D), two of them forming the measurement region (C and D) and the other two the application region (A and B, where A is the SR of the analysis). The strategy is to measure a nonprompt photon rate using regions C and D, which are built to be enriched in nonprompt photons, and then apply it to regions A and B. In order to define the

four regions, two photon observables which provide good discrimination between genuine photons and nonprompt ones are chosen: I_{ch} and $\sigma_{\text{i}\eta\text{i}\eta}$.

Photons satisfying the cut-based identification criteria described in section 3.3 are selected, but without the $\sigma_{\text{i}\eta\text{i}\eta}$ and I_{ch} requirements. The application and measurement regions are defined as follows:

Measurement region:

- $1.141 \leq I_{\text{ch}} < 15$ for EB photons
- $1.051 \leq I_{\text{ch}} < 15$ for EE photons
- $N_j \geq 1$ or $e\mu$ final state

Application region:

- $I_{\text{ch}} < 1.141$ for EB photons
- $I_{\text{ch}} < 1.051$ for EE photons
- $N_j \geq 2$ and $N_b \geq 1$

Regions A and C are then defined as events passing the $\sigma_{\text{i}\eta\text{i}\eta} < 0.01015$ (0.027) requirement for EB (EE) photons, while regions B and D consist of events with $\sigma_{\text{i}\eta\text{i}\eta} > 0.012$ (0.034). These requirements are chosen based on the medium working point of the cut-based photon identification algorithm, such that region A corresponds exactly to the SR. Between 0.01015 (0.027) and 0.012 (0.034) there is a gap to reduce the contribution of genuine photons in regions B and D. The regions are represented schematically in figure 3.15.

Figures 3.16 and 3.17 show the distribution of the $\sigma_{\text{i}\eta\text{i}\eta}$ in the four regions, including the simulated genuine and simulated nonprompt contributions, for events corresponding to data collected in 2018, with an identical selection to the SR, but with relaxed requirements on $\sigma_{\text{i}\eta\text{i}\eta}$ and I_{ch} for the photon(s).

The nonprompt rate is defined as the ratio between the number of events with nonprompt photons passing the $\sigma_{\text{i}\eta\text{i}\eta}$ cut and those falling in the sideband. In the measurement region, this is the ratio between the number of nonprompt-photon events in region C and that in region D. This quantity will be denoted as fake rate ($\text{FR} = \text{Data}_C/\text{Data}_D$). The FR is then applied as a weight to the data in region B, in order to estimate the nonprompt photon contribution in region A, the SR. In this way, the data from the SR is not used in the estimation of the FR, but only data from the other regions - this is to avoid biasing the measurement.

The measurement region is built such that the contribution from genuine photons is minor. However, as seen in figures 3.16 and 3.17, it is not zero. Thus, before estimating the FR, one needs to remove any genuine photon contribution from the data. This can be done by subtracting the genuine photon contribution (from simulation) from the data in these regions, and performing the ratio computation using the remaining events.

The nonprompt photon contribution depends strongly on the p_{T} and η of the photon. This is expected, since highly energetic and central photons are less likely to be nonprompt, but also because the ECAL resolution is not uniform, and in the EE ($\eta > 1.5$),

a higher nonprompt photon rate is expected. Hence, instead of extracting one global FR, it is more precise to do so in bins of p_{T} and η of the photon. The binning in η is chosen to be the same as in Ref. [162] in the EB. This division corresponds to the edges of the different ECAL modules, so that the non-uniform detector response is taken into account. In the EE, the ECAL crystal edges are defined in the x and y axis, and do not correspond to unique η values, therefore the choice of binning is not so trivial, and the binning from the EE laser calibration plots in Ref. [163] is used. In photon p_{T}, the binning was chosen considering a trade-off between differentiating as much as possible while keeping enough statistical precision in each bin. The bin edges were then defined to be:

- η = [0,0.435,0.783,1.13,1.5,1.8,2.1,2.5]
- $p_{\mathrm{T}} = [20, 25, 30, 45, 70, \infty]$ GeV

The FR per bin (i,j) of photon p_{T} and η can then be extracted according to equation 3.8:

$$\mathrm{FR}^{ij} = \frac{\mathrm{Data}_C^{ij} - \mathrm{Genuine\ MC}_C^{ij}}{\mathrm{Data}_D^{ij} - \mathrm{Genuine\ MC}_D^{ij}} \quad . \tag{3.8}$$

The traditional ABCD method works on the assumption that the relation

$$\frac{n_A}{n_B} = \frac{n_C}{n_D} \tag{3.9}$$

holds, where n_X is the number of nonprompt photon events in region X. This relation holds only if the two variables are uncorrelated between each other. This is not the case in this analysis. The expected distributions of $\sigma_{\mathrm{i}\eta\mathrm{i}\eta}$ in the measurement and application regions, using nonprompt photons from the $\mathrm{t\bar{t}}$ simulated samples, is shown in figure 3.18. It is clear that the distributions don't match exactly, especially for EE photons, showing that there is a dependence between $\sigma_{\mathrm{i}\eta\mathrm{i}\eta}$ and I_{ch}.

The physical reason for this mismatching is that "fake" photons tend to have both high $\sigma_{\mathrm{i}\eta\mathrm{i}\eta}$ values and be poorly isolated from charged particles around it, for example if they are misreconstructed jets. In order to compensate for this, an additional shape correction factor is computed from $\mathrm{t\bar{t}}$ MC (k_{MC}), using the same binning as for the FR, following equation 3.10:

$$k_{MC}^{ij} = \left(\frac{\text{nonprompt MC}_A^{ij}}{\text{nonprompt MC}_B^{ij}} \right) / \left(\frac{\text{nonprompt MC}_C^{ij}}{\text{nonprompt MC}_D^{ij}} \right) \tag{3.10}$$

This method was also used in Ref. [138]. The final nonprompt photon rate in the SR is then estimated as:

$$\text{nonprompt } \gamma \text{ contribution} = \sum_{i,j}(\text{Data}_B^{ij} \times \text{FR}^{ij} \times k_{MC}^{ij}) - \sum_{i,j}(\text{genuine MC}_B^{ij} \times \text{FR}^{ij} \times k_{MC}^{ij}) \tag{3.11}$$

The FRs computed from data are shown in figure 3.19 (left) and the k_{MC} factors, computed from simulated $t\bar{t}$ events are shown in figure 3.19 (right). Simulated events corresponding to the 2016, 2017 and 2018 eras are considered all together for one single derivation, in order to reduce the statistical uncertainty, after verifying that using events from each year individually provides compatible results.

For each observable to be measured, the data in region B are scaled by FR*k_{MC}, depending on the photon p_{T} and η of each individual event. The same procedure is applied to simulated events that contain a genuine photon, also in region B. Subsequently, the two histograms are subtracted from each other. The resulting histogram is the nonprompt photon background estimate for the SR.

Closure tests in simulation

In order to ensure that the method presented above provides consistent results, a closure test is performed using simulated $t\bar{t}$ events. The FR is derived using $t\bar{t}$ simulated events instead of data, such that:

$$\text{FR}^{ij}(\text{closure test}) = \frac{\text{nonprompt } t\bar{t} \text{ MC}_C^{ij}}{\text{nonprompt } t\bar{t} \text{ MC}_D^{ij}} \tag{3.12}$$

The obtained FRs are shown in figure 3.20.

These FRs are applied to nonprompt $t\bar{t}$ simulated events in region B and the resulting distributions are compared to the direct nonprompt $t\bar{t}$ simulation in region A. The results of these closure tests are shown in figures 3.21 to 3.22, for several variables of interest, for events corresponding to the 2018 period. Good closure is obtained for all bins, i.e., the distributions generally agree within the statistical uncertainty. Flat uncertainties are assigned to cover for the fluctuations observed in some of the bins: an overall 10% uncertainty is assigned to all events, and an additional 30% uncertainty is applied for high p_{T} (>80 GeV) photons.

3.6.3 Other backgrounds

Several other processes contribute to the background through genuine photons. The single top tW production in association with a prompt photon is estimated using a dedicated MC sample, as discussed in the previous section. Other minor backgrounds like diboson and triboson events with a photon - Wγ, $t\bar{t}$Zγ, tZqγ, and others - are estimated from MC and included in the "Others" category. Contributions from these processes through nonprompt photons are included in the data-driven estimation described in section 3.6.2.

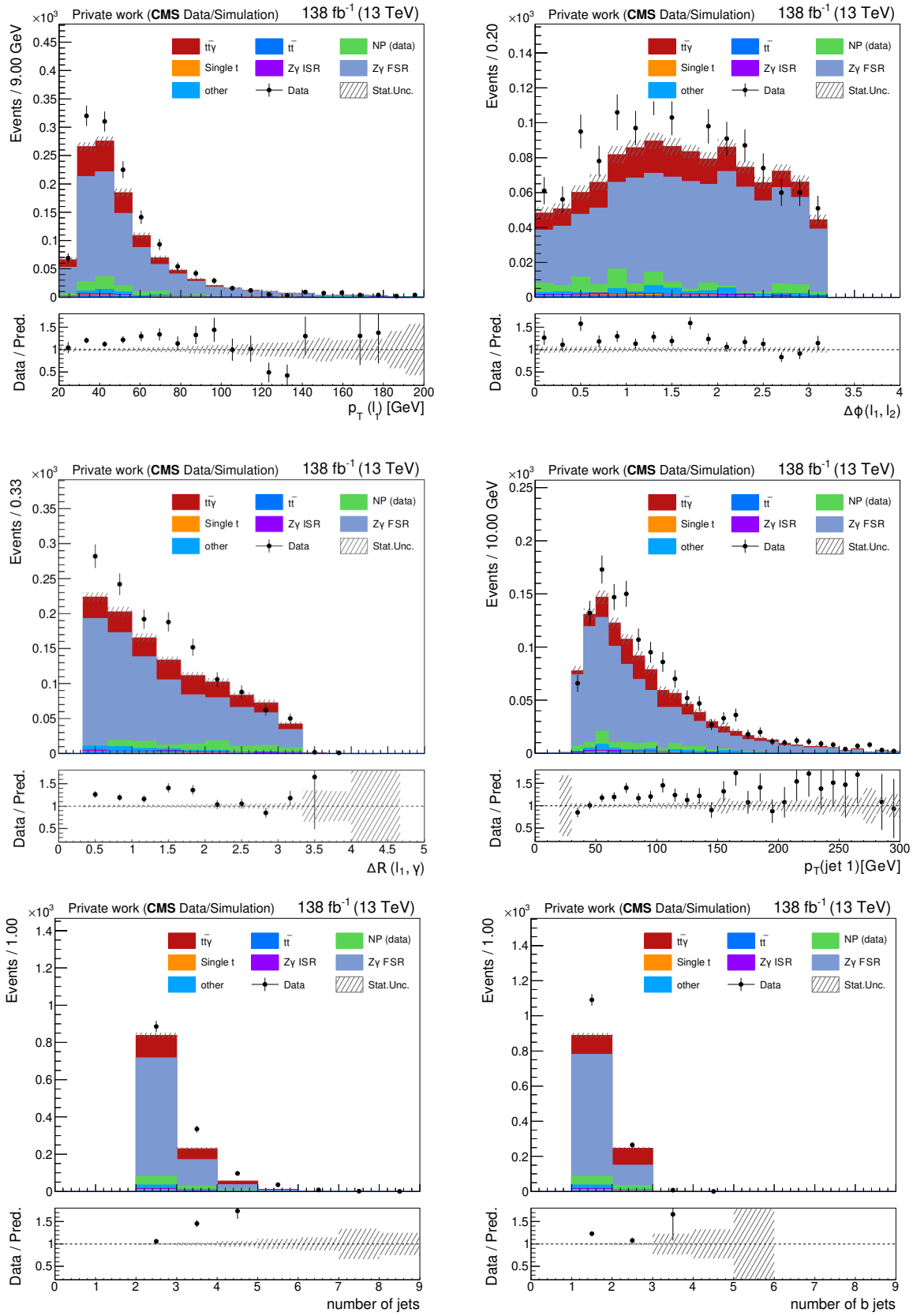

Figure 3.14: Comparison between data and simulated events in the Zγ CR. The leading lepton p_{T} (upper left), $\Delta\phi(\ell,\ell')$ (upper right), ΔR(lead. ℓ,γ) (centre left), p_{T} (leading jet) (centre right), number of jets (lower left), and number of b jets (lower right) are shown, for the full Run 2 data. Only the statistical uncertainty is shown.

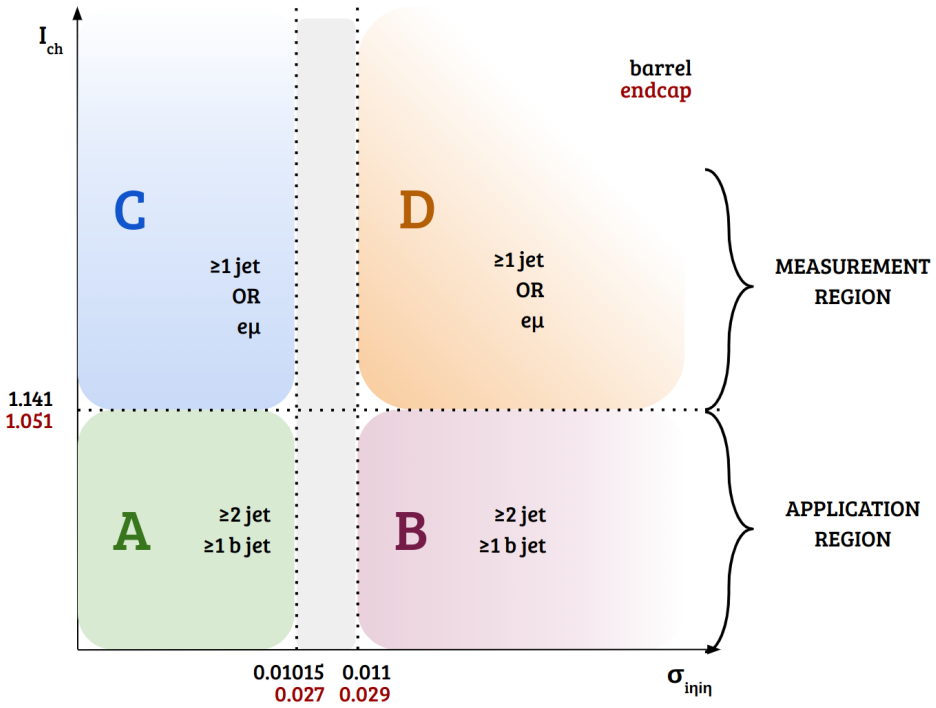

Figure 3.15: Scheme showing the ABCD regions used for the nonprompt estimations. Cuts which are not mentioned are the same as in the SR. In black the requirements for EB photons are shown, and in dark red those for EE photons.

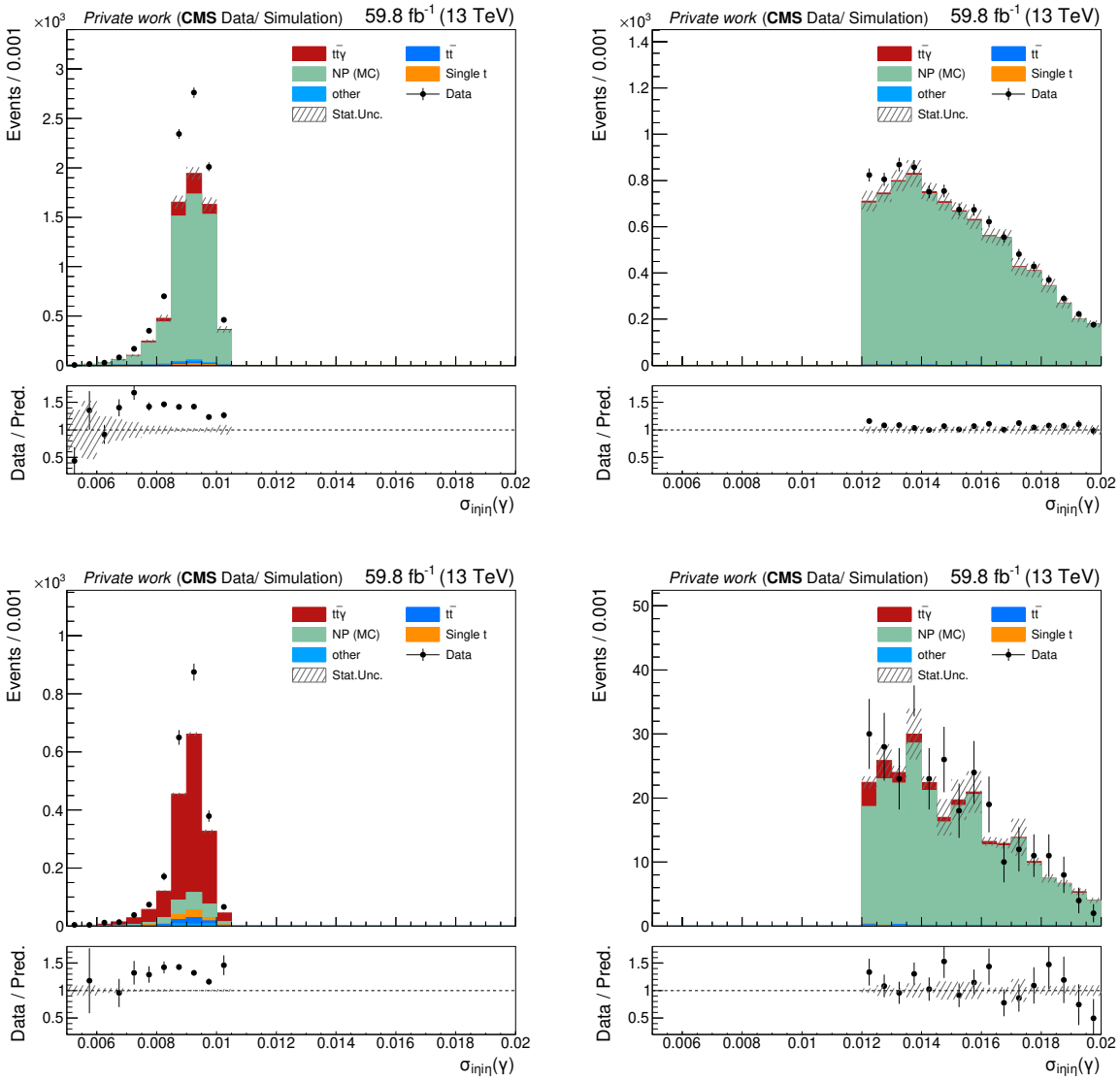

Figure 3.16: Distribution of the $\sigma_{i\eta i\eta}$ in the A (lower left), B (lower right), C (upper left), and D (upper right) regions used for the nonprompt photon estimation, in 2018, for photons in the EB. The positions of each figure reflect the scheme of figure 3.15. "NP (MC)" refers to the nonprompt photon contribution taken from simulation. Regions B and D, the sidebands of the $\sigma_{i\eta i\eta}$ cut, show good agreement between data and simulation, while regions A and C which are enriched in genuine photon events show an excess of data, same as observed in section 3.4.

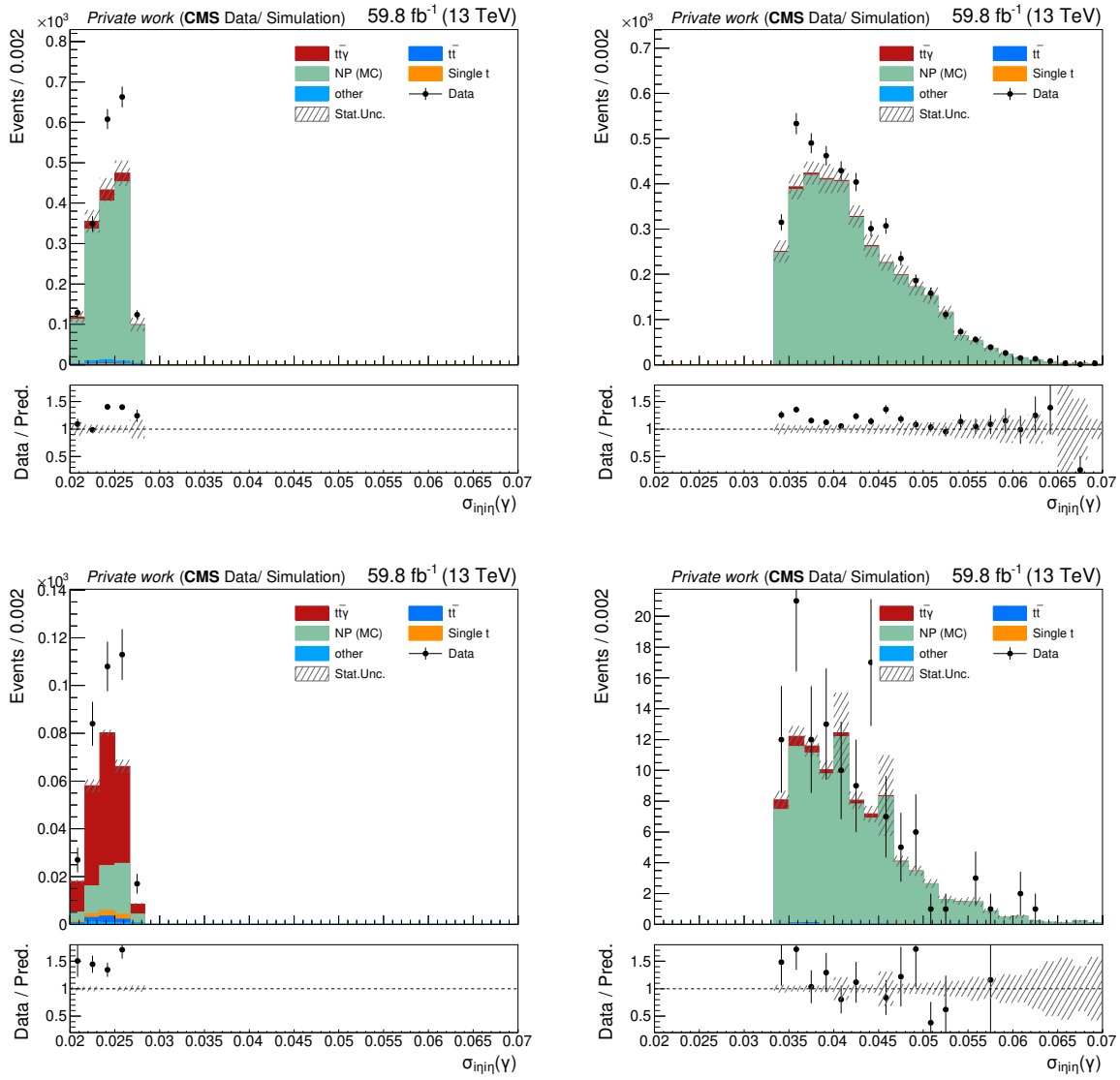

Figure 3.17: Distribution of the $\sigma_{i\eta i\eta}$ in the A (lower left), B (lower right), C (upper left), and D (upper right) regions used for the nonprompt photon estimation, in 2018, for photons in the EE. The positions of each figure reflect the scheme of figure 3.15. "NP (MC)" refers to the nonprompt photon contribution taken from simulation. Regions B and D, the sidebands of the $\sigma_{i\eta i\eta}$ cut, show good agreement between data and simulation, while regions A and C which are enriched in genuine photon events show an excess of data, same as observed in section 3.4.

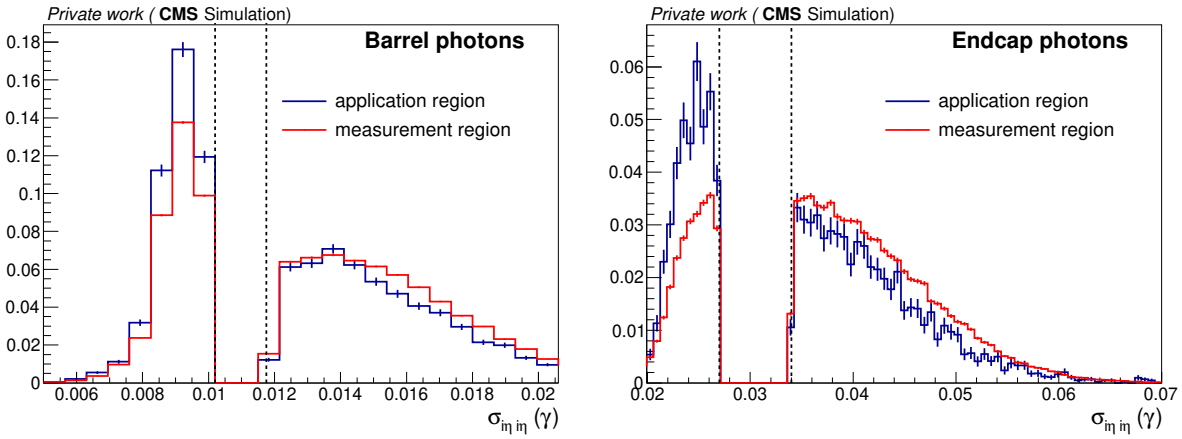

Figure 3.18: Comparison of the $\sigma_{i\eta i\eta}$ distribution shape of nonprompt photons between the measurement and application regions in 2018, using nonprompt photons from the $t\bar{t}$ simulated samples. The dotted lines indicate the subregions passing the $\sigma_{i\eta i\eta}$ requirement, those falling in the sideband, and the gap between them.

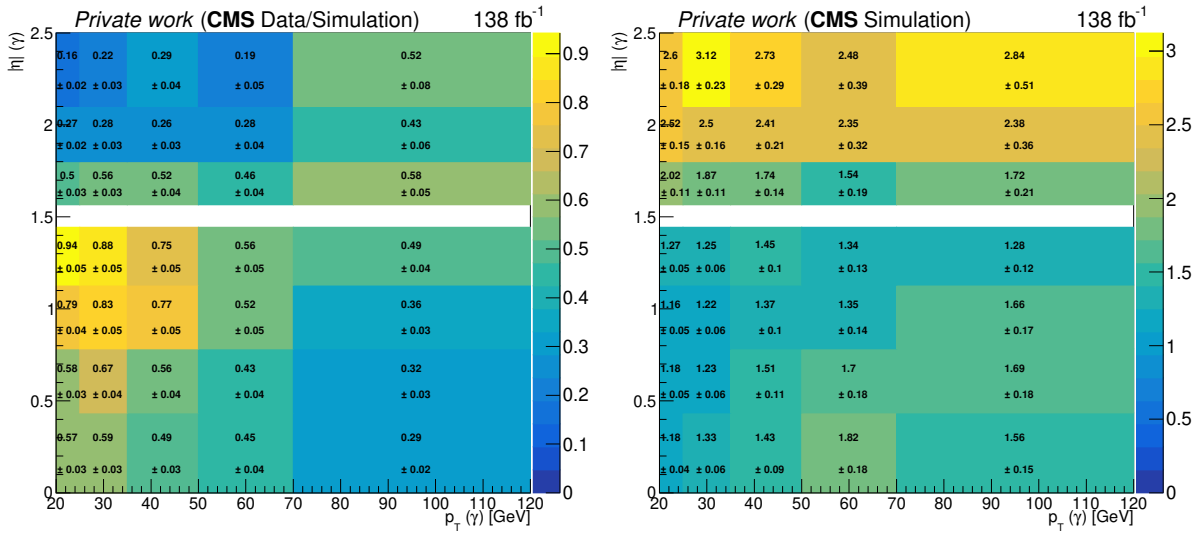

Figure 3.19: Left: FR as a function of photon p_T and η, extracted from the full Run 2 data, after subtracting the prompt contribution from MC. Right: MC correction factor k_{MC} extracted from the full Run 2 $t\bar{t}$ MC. The white gaps in the plots correspond to the transition region between the ECAL EB and EE. Those events are removed because the photon reconstruction is not well understood in that region.

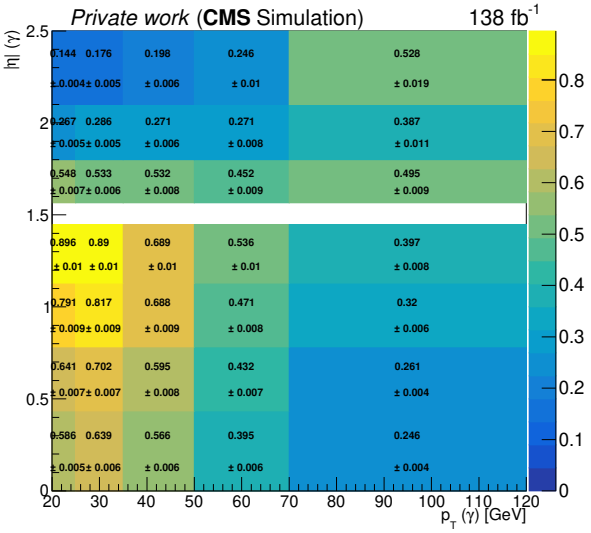

Figure 3.20: FR from MC (for comparison) as a function of photon p_{T} and η, extracted from the full Run 2 nonprompt $\mathrm{t\bar{t}}$ simulation.

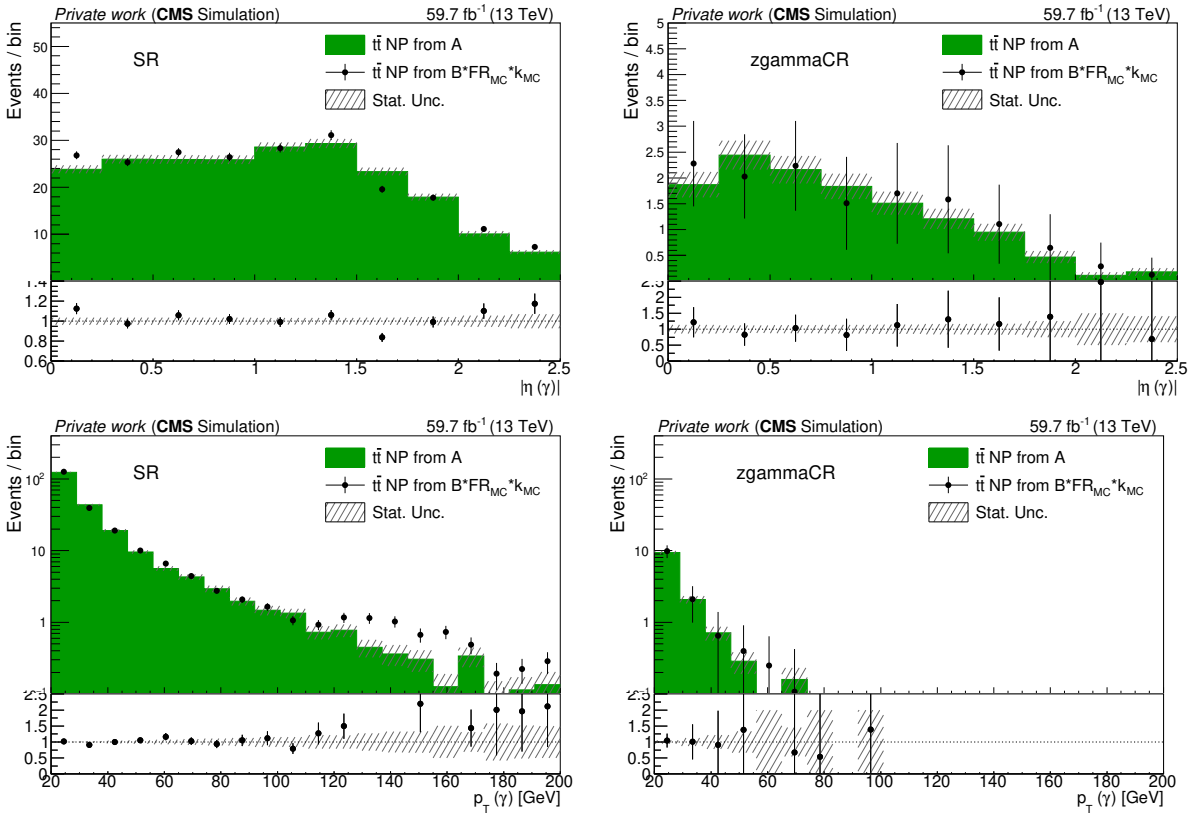

Figure 3.21: Closure test using 2018 $t\bar{t}$ MC, for the photon $|\eta|$, in the SR (upper left) and Zγ CR (upper right), and for the photon p_{T}, in the SR (lower left) and Zγ CR (lower right).

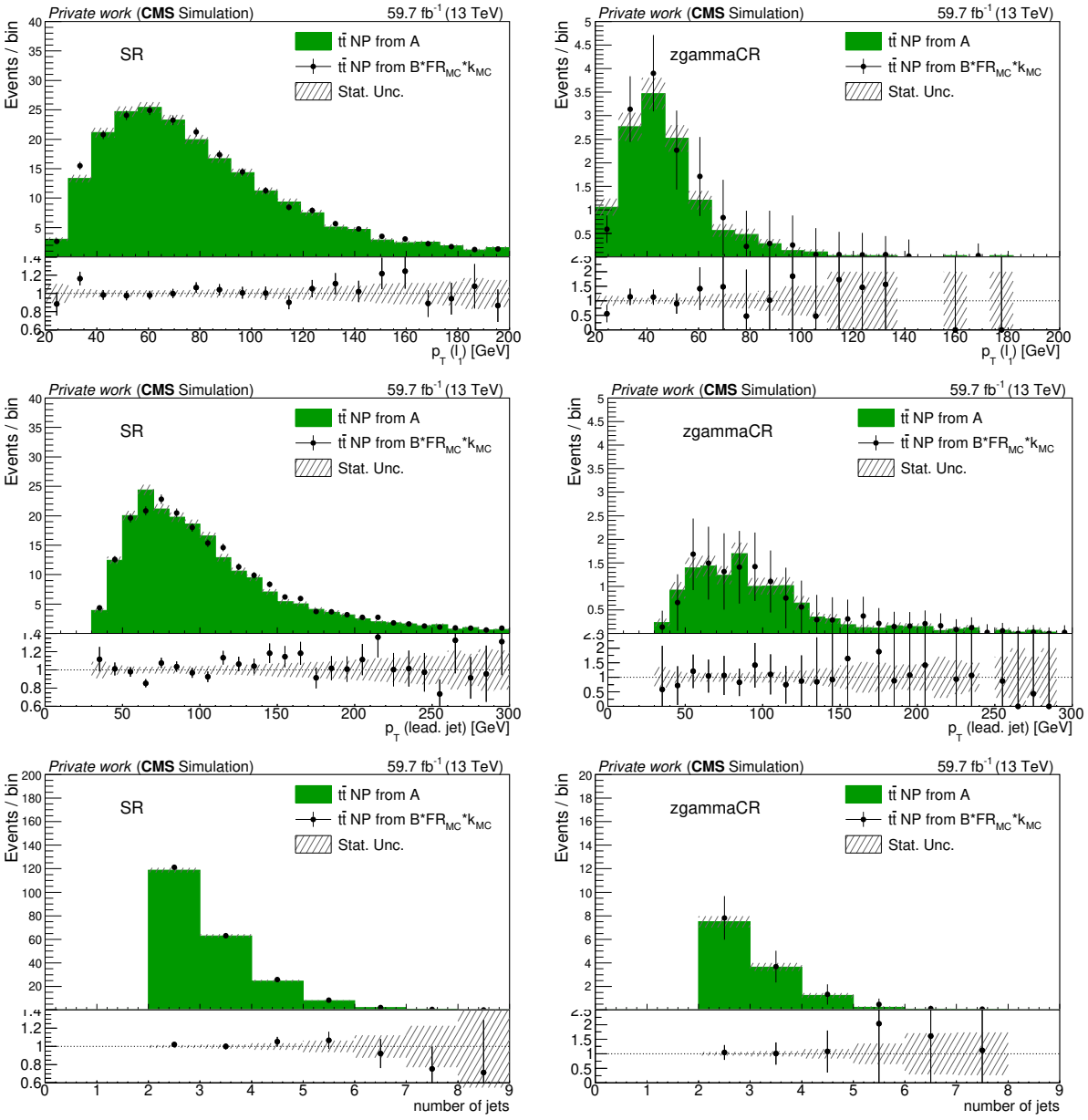

Figure 3.22: Closure test using 2018 $t\bar{t}$ MC, for the leading lepton p_{T}, in the SR (upper left) and Zγ CR (upper right), for the leading jet p_{T}, in the SR (centre left) and Zγ CR (centre right), and for the number of jets, in the SR (lower left) and Zγ CR (lower right).

3.7 Statistical analysis

Events passing all the selection steps described above are subject to a statistical analysis, to extract the differential cross section of $t\bar{t}\gamma$ as a function of several observables, as well as the ratio between the cross sections of $t\bar{t}\gamma$ and $t\bar{t}$.

In order to describe how to extract differential cross sections, it is useful to first understand how to measure an inclusive cross section. In this section, the basic statistical concepts needed for such a measurement are introduced. Subsection 3.7.2 lists the systematic uncertainties which are considered, from theoretical and experimental sources.

3.7.1 Signal extraction

In order to measure the inclusive cross section of $t\bar{t}\gamma$ from a fit to the data, the signal strength μ is defined as the ratio between the observed cross section of the process and the one predicted by the SM:

$$\mu = \frac{\sigma(t\bar{t}\gamma)}{\sigma(t\bar{t}\gamma)^{\mathrm{SM}}} \tag{3.13}$$

This is called a *parameter of interest (POI)* of the fit. The POI parameter is extracted through a maximum likelihood fit to the data, binned as a function of some observable, in N bins. In each bin i of the distributions, the probability to observe n_i events in the data, when $\lambda_i(\mu)$ are expected, in the absence of systematic uncertainties, is given by the Poisson distribution:

$$P_i(n_i|\mu) = \frac{\lambda_i(\mu)\mathrm{e}^{-\lambda_i(\mu)}}{n_i!} \tag{3.14}$$

where the number of expected events $\lambda_i(\mu)$ is given by:

$$\lambda_i(\mu) = \mu \cdot s_i + \sum_j^{N_{\mathrm{bkg}}} b_{i,j} \tag{3.15}$$

In this formula, s_i is the number of expected signal events and $b_{i,j}$ the number of expected background events for the j^{th} source of background. The *likelihood* is a function quantifying the probability of the observed data for a given value of μ, and is obtained by taking the product of equation 3.14 for all N bins of the distributions:

$$\mathcal{L}(\mathbf{n}|\mu) = \prod_{i=1}^{N} \frac{\lambda_i(\mu)\mathrm{e}^{-\lambda_i(\mu)}}{n_i!} \tag{3.16}$$

The vector $\mathbf{n} = n_1, ..., n_N$ represents the observed number of entries per bin. In the presence of systematic uncertainties, these can be included as a set of "nuisance"

parameters, $\boldsymbol{\Theta}$, whose values are not taken as known *a priori* but rather must be fitted from the data, as described in Ref. [164]. The number of expected events, in this case, becomes a function of μ and the nuisance parameters, $\lambda_i(\mu, \boldsymbol{\Theta})$, and the likelihood function is extended to include our prior knowledge about the nuisance parameters:

$$\mathcal{L}(\mathbf{n}|\mu, \boldsymbol{\Theta}) = \prod_{i=1}^{N} \frac{\lambda_i(\mu, \boldsymbol{\Theta}) \mathrm{e}^{-\lambda_i(\mu, \boldsymbol{\Theta})}}{n_i!} \cdot \prod_{m=1}^{M} f(\Theta_m) \tag{3.17}$$

The second product runs over M nuisance parameters, and $f(\Theta_m)$ is the probability distribution function (p.d.f.) constraining the m^{th} nuisance parameter. For most nuisances, these p.d.f.s are taken as Gaussian constraints, such that

$$f(\Theta_m) = \mathrm{e}^{-\frac{1}{2}(\Theta_m - \Theta_{0m})^2/(\Delta\Theta_m)^2} \tag{3.18}$$

where Θ_m and Θ_{0m} are the prior uncertainty on the nuisance parameter and the initial estimate for its value, respectively. For parameters with a large uncertainty or those which are positively-defined, the Gaussian distribution would be truncated at 0. In such cases, the parameters are instead constrained with a log-normal function, that is, a function whose logarithm is normally distributed. The implementation of Eqs. 3.13-3.18 is made in RooStats, using the Higgs Combination Tool [165].

The *maximum likelihood estimation (MLE)* method is used to estimate the parameters. It consists in finding the values of the parameters for which the likelihood is maximum, i.e., the parameters for which the observed data is the most "likely" to occur. Maximising the likelihood is equivalent to maximising its logarithm, and in fact for computational simplicity what is normally done is to minimise $-2\ln(\mathcal{L})$, the *negative log-likelihood*. This minimisation is implemented with the Minuit tool [166].

In order to interpret the data in terms of its compatibility with a given physical hypothesis $\mathcal{H}$, a *test statistics* must be defined. The one used for this analysis is the ratio of profiled likelihoods (PLR):

$$q_\mu = \frac{\mathcal{L}(\mathbf{n}|\mu, \hat{\boldsymbol{\Theta}}_\mu)}{\mathcal{L}(\mathbf{n}|\hat{\mu}, \hat{\boldsymbol{\Theta}})} \tag{3.19}$$

"Profiled" refers to maximising the likelihood with respect to all parameters except the signal strength, which is fixed. In the numerator, μ is fixed, and the nuisance parameters are optimised such that the likelihood is maximum for that value of μ: $\hat{\boldsymbol{\Theta}}_\mu$ is the conditional estimator for a particular value of μ. In the denominator, $\hat{\mu}$ and $\hat{\boldsymbol{\Theta}}$ are the global maximum likelihood estimators, and so by construction the denominator is larger than the numerator and the PLR is bound between 0 and 1. Values close to 1 indicate that the data favours the assumed value of μ. The prescription by Feldman and Cousins [167] modifies the test statistics in Eq. 3.19 to ensure consistent behaviour in the case where

the estimator $\hat{\mu}$ is smaller than 0, by replacing the denominator with $\mathcal{L}(\mathbf{n}|0, \hat{\hat{\boldsymbol{\Theta}}})$ in that case.

When quoting a value of $\hat{\mu}$, it is common to quote a confidence interval (CI), which is an interval likely to include the POI. How likely the interval is to contain the POI is determined by the confidence level (CL). For example, a CL of 95% means that the quoted interval contains the true value of the POI with 95% probability.

p-value and significance

The *p-value* $p(q_{\text{obs}}|\mathcal{H})$ is the probability to get a result as discrepant as the observed one or even more so, given a hypothesis $\mathcal{H}$ [168]. Assuming a background-only hypothesis $\mathcal{H}_0$ is true, the probability to measure a result of equal or greater incompatibility with the predictions of $\mathcal{H}_0$ is

$$p_0 = \int_{q_{\text{obs}}}^{\infty} f(q_0|\mathcal{H}_0) dq_0 \tag{3.20}$$

where $f(q_0|\mathcal{H}_0)$ is the p.d.f. of the test statistic assuming $\mathcal{H}_0$. The hypothesis is rejected if this probability is found to be below a certain threshold, that is defined before carrying out the test. In HEP, conventional thresholds to reject the background-only hypothesis and claim an evidence or a discovery are set at 1.4×10^{-3} and 2.9×10^{-7}. These numbers are normally stated in terms of significance in number of Gaussian standard deviations σ. The significance is obtained directly from the p-value as $Z = \Phi^{-1}(1 - p_0)$, where Φ^{-1} is the inverse of the cumulative distribution of the Gaussian function. The thresholds above correspond to $3\,\sigma$and $5\,\sigma$, respectively. For rejecting a signal model, the threshold for p-value is typically set at $p_\mu < 0.05$, corresponding to $1.64\,\sigma$.

Asymptotic approximation

In practice, determining the full distribution of the PLRs can be difficult, and often requires the simulation of a large number of MC "toy" experiments, which is computationally very expensive. Conveniently, Wilks' theorem [169] and the approximate formulas by Wald [170] tell us that for large sample sizes, the distribution of $-2\ln(\mathcal{L})$ approaches that of a χ^2 function, with the same number of degrees of freedom as the number of POIs. This allows us to simply compute the value of the PRL for the observed data and compare the result to the χ^2 value corresponding to the desired statistical significance.

An important consequence of this, as shown in Ref. [165], is that the median expected significance can be approximated without the need to simulate MC "toys", by using an "Asimov" dataset. This is a representative dataset with the property that when evaluating the likelihood for this dataset, the estimates of the parameters correspond to their true values.

3.7.2 Systematic uncertainties

Uncertainties coming from the signal and background modelling, as well as from the experimental limitations, are considered. Some uncertainties concern only the normalisation of the MC templates, and are evaluated by scaling the entire contributions up and down by a given percentage. Those will be referred to as *rate* uncertainties. Others, however, affect the shape of the distribution, as well as its normalisation, and are called *shape* uncertainties in the following. In the latter case, alternative MC templates are built with the varied shapes. The different contributions to the uncertainty and the way they are obtained are described in the following. Then, table 3.7 details the process(es) to which each uncertainty is applied and whether they are considered as correlated or uncorrelated between the processes.

This analysis is based on data collected by CMS during four different data-taking periods, 2016 preVFP, 2016 postVFP, 2017, and 2018. During these eras, the detector conditions have changed and many corrections are derived separately per era, so some uncertainties have to be considered separately, i.e., as uncorrelated between the years of data-taking. On the other hand, other uncertainties such as theoretical ones, are independent of the era and should be correlated across the whole dataset. Table 3.7 also includes the information of whether each uncertainty is considered as correlated or uncorrelated between data-taking periods (referred to as "years"). The following uncertainty sources are included:

Renormalisation and factorisation scales: The uncertainties on the choice of μ_F and μ_R for the ME calculation are evaluated by varying these scales by a factor of 2 and 0.5. They are considered as shape uncertainties, for the $\mathrm{t\bar{t}\gamma}$, $\mathrm{t\bar{t}}$ and $\mathrm{Z\gamma}$ processes. These uncertainties are considered to be correlated between the $\mathrm{t\bar{t}\gamma}$ and $\mathrm{t\bar{t}}$ processes, as at ME level the two process are identical, both based on the strong interaction. They are however treated as uncorrelated with the $\mathrm{Z\gamma}$ process, as its production mechanism is different, based on the electroweak interaction.

PDF and α_S: Replicas of the nominal PDFs used to generate the MC samples, containing 100 different variations, are evaluated, to assess the impact of the PDF choice on the signal acceptance. Additionally, the uncertainty on the choice of α_S is included by scaling α_S up and down by its uncertainty.

Parton shower modelling: Uncertainties in the modelling of ISR and FSR are included. To assess them, the value of α_S in the simulation is varied up and down by a factor 2 for ISR and $\sqrt{2}$ for FSR. The uncertainties on FSR are considered to be correlated among processes that involve top quarks ($\mathrm{t\bar{t}}$, $\mathrm{t\bar{t}\gamma}$, single t), and uncorrelated among the other processes; the ones on ISR are correlated between the $\mathrm{t\bar{t}}$ and $\mathrm{t\bar{t}\gamma}$ processes, but uncorrelated with single top and other processes.

Background normalisation: The uncertainties on the cross section of each of the background contributions that are estimated from MC simulation are included as an overall rate uncertainty per process, affecting only the corresponding process. The assigned values for these uncertainties are presented in table 3.7.

Luminosity: The uncertainties on the luminosity have several sources, as described in section 2.4. Some of these sources are correlated across the years and some are not, and therefore a partial correlation scheme is adopted as detailed in Ref. [171].

L1 prefiring: In 2016 and 2017 data, a problem in early L1 electron/photon triggers lead to the recording of many uninteresting events and the loss of interesting ones. This effect is only present in the data, not in the simulation. To account for this, the MC events are reweighted to match the data, and corresponding uncertainties for this reweighting are included.

Lepton selection: The SFs applied due to the selection efficiency of electrons and muons have an associated uncertainty, both statistical, due to the limited size of the samples used to estimate them, and systematic, due to the choice of background modelling in the "tag-and-probe" methods. The values of these efficiencies are varied by the corresponding $\pm 1\,\sigma$ distributions, and the resulting varied spectra are included in the fit, as shape uncertainties, separately for electrons and muons [112, 155].

b tagging: Uncertainties on b tagging SFs have several sources, such as the statistical uncertainties on the b, c and light-flavour components of the samples used to compute them, and the estimated contamination from light-flavour (heavy-flavour) jets in heavy-flavour (light-flavour) measurement regions [156].

Jet energy corrections: Uncertainties on the jet energy scale (JES) are taken into account by rescaling the jet momenta as a function of p_{T} and η, for the simulated samples. The JES uncertainties are split according to their sources, and the dominant components come from pileup, non-uniformity of the detector response, and residual data-simulation JES differences. The resulting uncertainties are below 3% for the phase space considered in this analysis [172]. They are estimated using events with dijet topologies, such as photon+jets and Z+jets. These uncertainties are considered as being partially correlated across the years. The uncertainties on the jet energy resolution (JER) are estimated in the same way, but varying the width of the jet momenta distribution, instead of its mean. The values of the JER uncertainties vary between 2-4% in the tracker covered regions, and increase up to 6% in the endcaps. They are fully correlated across the years.

Unclustered $\boldsymbol{p_{\mathrm{T}}^{\mathrm{miss}}}$: The uncertainties on all the individual components contributing to the $p_{\mathrm{T}}^{\mathrm{miss}}$ (momenta of jets, muons, electrons, as well as unclustered energy) are propagated onto the value of $p_{\mathrm{T}}^{\mathrm{miss}}$, and the respective $\pm 1\,\sigma$ variations are considered as a shape uncertainty.

Trigger SFs: The trigger SFs mentioned in section 3.3 have an associated uncertainty, which a combination of a statistical uncertainty (from the limited size of the samples used to derive the SFs) and a number of systematic uncertainties, arising from the imperfect modelling of the pileup and the number of jets distributions, as well as a dependence of the SFs on the data-taking era.

MC statistical uncertainty: The statistical uncertainty associated with the finite size of the simulated samples is taken into account using the Beeston-Barlow method [173]. This method assigns a single nuisance parameter to each bin of the fitted distribution, to account for the overall statistical uncertainty on the predicted number of events, for all processes. This nuisance parameter is built assuming a Poisson or Gaussian prior, depending on the amount of events.

Photon selection: The efficiency in selecting photons is derived in the same way as the one for electrons [123], from DY events in the e^+e^- final state. This procedure has an associated uncertainty, both statistical and systematic, arising from the "tag-and-probe" method. Additionally, there is an uncertainty associated to the efficiency of the PSV. The values of these efficiencies are varied by $\pm 1\,\sigma$, and the resulting varied distributions are included in the fits, as shape uncertainties.

Nonprompt photon normalization: For the nonprompt photon contribution, estimated from data, an overall normalization uncertainty of 10% is assumed for all nonprompt events, and an additional uncertainty of 30% for the normalisation on the high photon p_{T} ($p_{\mathrm{T}} > 80\,\mathrm{GeV}$) region. These account for the residual differences observed in the closure tests, as detailed in section 3.6.2.

Pileup: The uncertainty on the distribution of the number of vertices (pileup) is taken into account by using two sets of alternative weights derived with a variation of 4.6% on the total inelastic cross section [174].

3.8 Differential measurements

The goal of the analysis is to measure the cross section of $t\bar{t}\gamma$ differentially, as a function of several observables of interest x. In practice, this observable is split in N bins, at generator level, so instead of measuring one single cross section value σ, we measure the differential cross sections σ_k, with $k = 1, ..., N$, defined as:

$$\sigma_k = \int_{x_k^{\mathrm{low}}}^{x_k^{\mathrm{high}}} \frac{\mathrm{d}\sigma}{\mathrm{d}x}, \tag{3.21}$$

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

When reconstructing events as described in section 3.3, the values of x which are reconstructed do not correspond to the true values of x, as they suffer from hadronisation effects and imprecision in the measurement. Indeed, the observed x spectrum is a smeared version of the true spectrum. Furthermore, the number of measured events is reduced by the limited detector acceptance as well as reconstruction and selection efficiency. In order to recover the true distribution of x from the reconstruction level distribution, we use a technique known as unfolding. The unfolding procedure consists in trying to correct for the smearing effects and recovering the true values of σ_k.

3.8.1 Likelihood-based unfolding

The unfolding is performed using a maximum-likelihood method, as introduced in section 3.7.1, implemented in *RooStats*, using the Higgs Combination Tool [165]. For a measurement with N generator level bins, one defines the signal strength parameters $\vec{\mu}$ as:

$$\vec{\mu} = \begin{pmatrix} \sigma_0/\sigma_0^{\mathrm{SM}} \\ \sigma_1/\sigma_1^{\mathrm{SM}} \\ \dots \\ \sigma_N/\sigma_N^{\mathrm{SM}} \end{pmatrix}, \tag{3.22}$$

where the σ_k^{SM} denotes the expected SM cross section in each generator level bin k, at parton level, as defined later in this section.

Then, the negative log-likelihood is minimised in order to obtain the best estimate for the values of all μ_k simultaneously. In this way, the bin-to-bin migrations - the fact that the smearing or reconstruction inefficiency can cause events which are generated in bin k to be reconstructed in bin j, $j \neq k$ - is taken into account.

In order to treat the systematic uncertainties appropriately, with the correlations detailed in section 3.7.2, the histograms with data are built separately for the different data-taking periods, 2016 preVFP, 2016 postVFP, 2017 and 2018. For each histogram, templates are built for the signal and background process, considering the nominal event selection and the systematic variations. For the signal, N separate templates are built, one per generator-level bin of the observable to be measured.

3.8.2 Choice of observables

Several observables are selected to measure the differential cross section of $\mathrm{t\bar{t}\gamma}$. Three photon and lepton observables (leading photon p_{T}, leading lepton p_{T} and $\Delta\phi$ between the two leptons) that were also measured in the work of Ref. [162] (see figure 3.2) are selected, in order to validate the fit strategy against the existing measurement. The results are not directly comparable, since the work of this thesis includes photons reconstructed in the

EE and not only the EB, and the binning choice is not exactly the same, but a qualitative assessment of the consistency between the two analyses is still possible.

The observables which are selected are:

1. Leading photon p_{T};
2. Leading lepton p_{T};
3. $\Delta\phi$ between the two leptons ($\Delta\phi(\ell, \ell')$).
4. Leading top quark p_{T};
5. ΔR between the photon and the $\mathrm{t\bar{t}}$ system ($\Delta R(\gamma, \mathrm{t\bar{t}})$). This observable is sensitive to the origin of the photon, and therefore to modelling aspects of $\mathrm{t\bar{t}}\boldsymbol{\gamma}$.

Variables 4 and 5 listed above are measured here for the first time, while the remaining variables were already measured in Ref. [162], but only for photons in the EB. Measuring the cross section of $\mathrm{t\bar{t}}\boldsymbol{\gamma}$ as a function of different p_{T} spectra is meaningful when looking for BSM effects, since as motivated in chapter 1, such effects are expected to affect the high-p_{T} tails of the distributions. Angular distributions of charged leptons are of importance since they reflect spin correlations of the top quark pair and can be used to probe the charge-parity (CP) structure of the SM [78]. Variables such as $\Delta R(\gamma, \mathrm{t\bar{t}})$ are very sensitive to the origin of the photon, and are therefore useful in order to study regions which are richer in events where the photon originates from the top quark, where there is enhanced sensitivity to the $\mathrm{t}\gamma$ coupling.

3.8.3 Parton level definitions

The results of a measurement should be extracted in a way that is as independent as possible from the experiment, such that they can be compared to other experiments or to theory predictions. In order to do this, different definitions of the physical objects to be measured are possible, normally denoted as "parton-level" and "particle-level" definitions.

At the **parton level**, the measurement is performed in the full phase space of $\mathrm{t\bar{t}}\boldsymbol{\gamma}$ events with two leptons in the final state. Objects are defined based on outgoing particles generated in the hard interaction after ISR and FSR but before hadronisation. This includes short-lived particles such as the top quark. This definition does not take into account the energy loss of the particles due to gluon and photon radiation, which can be a significant effect, especially for low mass particles, but is typically not so relevant for very massive particles such as the top quark.

At the **particle level**, only "stable" particles, with a lifetime longer than 30 ps, are considered. Additionally, the energy loss due to gluon and photon radiation is taken into account, and for example leptons are clustered together with stable photons produced

around them, creating what is called "dressed leptons". The objects are redefined according to a number of acceptance requirements, that allow to reduce the extrapolation to the detector level. Hence, a fiducial phase space is defined by signal events that pass those requirements and the measurement is performed in this fiducial phase space. This also reduces the MC generator dependence, but comes with drawbacks: in particular, it is not possible to recover the top quark properties, since the top quark is unstable and thus not well-defined at particle level.

In the work of this thesis, as one of the main objectives is to measure the cross section as a function of the top quark kinematics, the parton-level definition is used.

3.8.4 Binning definition

For each observable, the number of generator-level bins to fit is optimised based on the amounts of data available and the behaviour of each individual observable (i.e. how the reconstruction-level spectrum compares to the generator-level one). In order to assess this behaviour, response matrices of the simulated $t\bar{t}\gamma$ events in the SR are built, comparing the reconstructed- and generator-level spectra of the different observables. The choice of binning is made based on a trade-off between keeping a sufficient number of $t\bar{t}\gamma$ events in each detector-level bin, a reasonably low condition number, and high enough stability and purity of the response matrices.

The condition number is a function that measures how much the output value of a given function can change for a small change in the input argument. In this case, how much the unfolded observables are sensitive to small change of the generator-level observables. If this number is high, typically above 10^4, the problem is said to be ill-conditioned and a technique known as regularisation needs to be applied [175]. This technique consists on introducing a bias term in the likelihood that encourages solutions which are closer to the truth-level distribution. In this thesis, the binning is always chosen such that the condition number is kept small (~ 10) and regularisation is not needed.

The stability s_i is defined as the fraction of all reconstructed events generated in the generator-level bin y_i that are also observed in the reconstruction-level bin x_i:

$$s_i = \frac{N(\text{events in } x_i \text{ and } y_i)}{N(\text{events in } y_i)} \tag{3.23}$$

In a response matrix, this corresponds to, for each generator-level bin, dividing the diagonal term x_i, y_i by the full column $\sum x_j, y_i$, with $j = 1, ..., N$.

The purity p_i is defined as the fraction of all reconstructed events which are observed in the reconstruction-level bin x_i that were also generated in the generator-level bin y_i:

$$p_i = \frac{N(\text{events in } x_i \text{ and } y_i)}{N(\text{events in } x_i)} \tag{3.24}$$

In the response matrix, this corresponds to, for each reconstruction-level bin, dividing the diagonal term x_i, y_i by the full line $\sum x_i, y_j$, with $j = 1, ..., N$.

The strategy to choose the binning is defined such that both purity and stability are kept above 0.4, while maintaining a comparable number of expected signal events in each reconstruction-level bin.

By doing so, the response matrices in figures 3.23 to 3.25 were built. The colours indicate the transition probability, i.e., the probability that an event generated in bin i also gets reconstructed in the same reconstruction-level bin i. The lower panels show the stability and purity for each bin, as defined in equations 3.23 and 3.24. The binning and condition numbers obtained after optimisation are summarised in table 3.8. In this work, the response matrices are only used to define the appropriate binning, the unfolding is done using a maximum likelihood fit to the data, as introduced earlier.

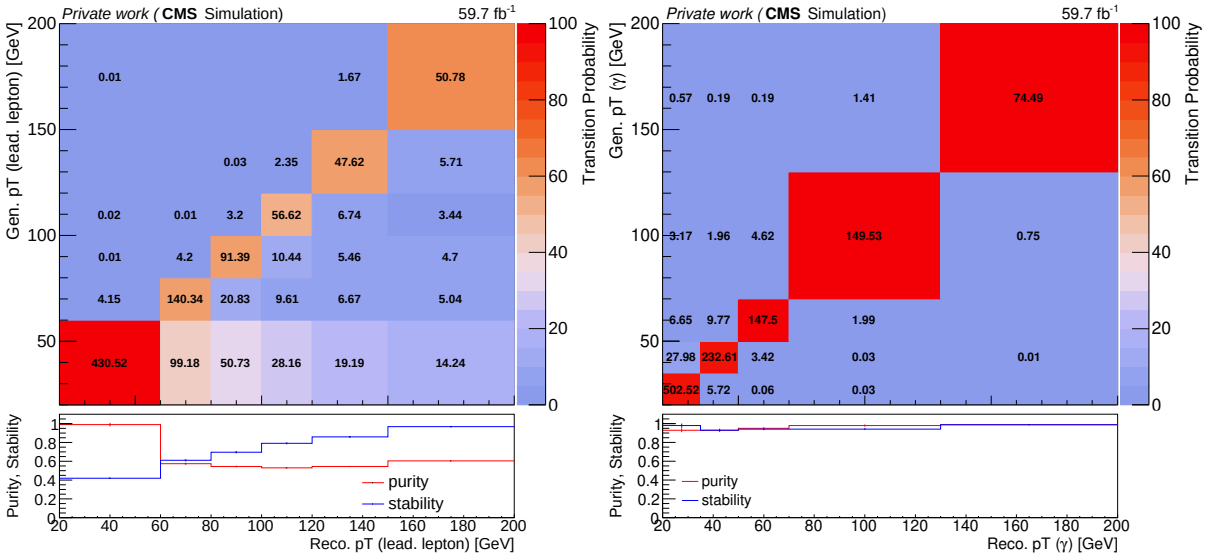

Figure 3.23: Response matrices for the leading lepton p_{T} (left) and the photon p_{T} (right). The numbers on the two-dimensional plots indicate the number of expected signal events with the luminosity collected in 2018.

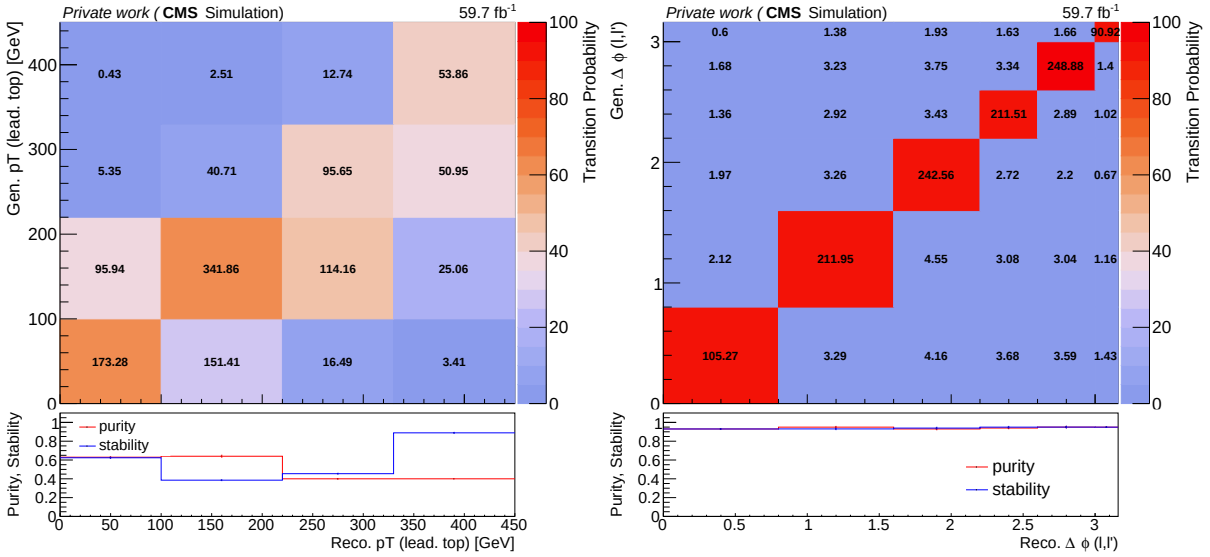

Figure 3.24: Response matrices for the leading top quark p_{T} (left) and the $\Delta\phi(\ell, \ell')$ (right). The numbers on the two-dimensional plots indicate the number of expected signal events with the luminosity collected in 2018.

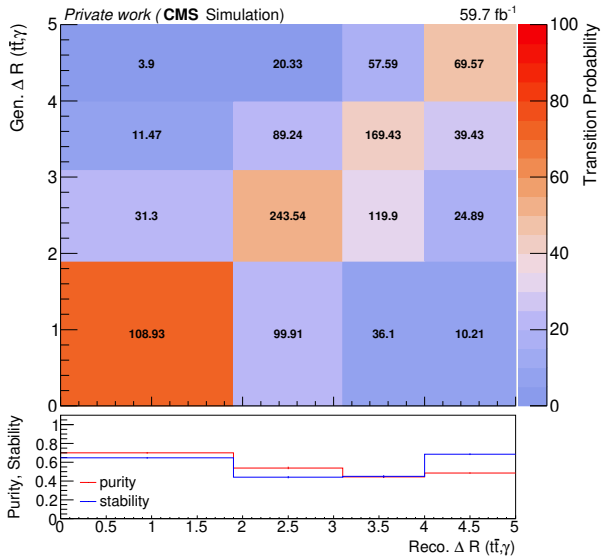

Figure 3.25: Response matrix for the $\Delta R(\mathrm{t\bar{t}}, \gamma)$. The numbers on the two-dimensional plot indicate the number of expected signal events with the luminosity collected in 2018.

Uncertainty	Type	Value	Process	Year	Proc. corr.	Year corr.
Luminosity	rate	-	all	all	yes	partial
$\sigma_{t\bar{t}}$	rate	5%	$t\bar{t}$	all	-	yes
σ_{others}	rate	30%	others	all	-	yes
$\sigma_{Z\gamma}$	rate	20%	$Z\gamma$	all	-	yes
$\sigma_{Z\gamma\ \text{FSR}}$	rate	20%	$Z\gamma$ FSR	all	-	yes
μ_R	shape+rate	-	$t\bar{t}$, signal, $Z\gamma$	all	partial	yes
μ_F	shape+rate	-	$t\bar{t}$, signal, $Z\gamma$	all	partial	yes
ISR	shape+rate	-	all	all	yes	yes
FSR	shape+rate	-	all	all	yes	yes
PDF	shape+rate	-	all	all	yes	yes
α_s	shape+rate	-	all	all	yes	yes
Electron Reco	shape+rate	-	all	all	yes	yes
Electron ID	shape+rate	-	all	all	yes	yes
Muon ID	shape+rate	-	all	all	yes	yes
Photon ID	shape+rate	-	all	all	yes	yes
Photon PSV	shape+rate	-	all	all	yes	no
L1prefire	shape+rate	-	all	2016,2017	yes	yes
Pileup	shape+rate	-	all	all	yes	yes
b tagging	shape+rate	-	all	all	yes	partial
JER	shape+rate	-	all	all	yes	no
JEC	shape+rate	-	all	all	yes	partial
Unclustered MET	shape+rate	-	all	all	yes	no
Trigger SFs	shape+rate	-	all	all	yes	no
Nonprompt photons	shape+rate	-	nonprompt	all	-	yes

Table 3.7: Sources of systematic uncertainties considered in the analysis. The type of uncertainty is described as being "rate" - constant normalisation factor - or "shape+rate" - variations that modify both the normalisation and the shape of the distribution. The processes affected by each uncertainty are also listed. "all" processes refers to all except the nonprompt contribution, which is derived from data.

	Bins	Condition number
Leading lepton p_{T}	$[20, 60, 80, 100, 120, 150, \infty[$ GeV	10.1
Photon p_{T}	$[20, 35, 50, 70, 130, \infty[$ GeV	6.8
Leading top quark p_{T}	$[0, 110, 220, 330, \infty[$ GeV	12.1
$\Delta\phi(\ell, \ell')$	$[0, 0.8, 1.6, 2.2, 2.6, 3.0, 3.16]$	2.8
$\Delta R(t\bar{t}, \gamma)$	$[0, 1.9, 3.1, 4, \infty[$	7.4

Table 3.8: Condition numbers for all observables to be measured, with the chosen binning.

3.8.5 Results

For each observable, a fit is performed simultaneously to two distributions: in the SR, the distribution of the observable being measured, and in the Zγ CR, the number of jets. For each generator-level bin of the signal process, one separate template is given as input to the fit. The signal strength parameters μ_i, for each bin i, are extracted directly from the fit and can be converted into a differential cross section by multiplying by the expected cross section, in that bin, obtained from the simulated signal.

The expected cross sections per bin are computed at generator-level, by evaluating the fractions of events in each generator-level bin, compared to the total number of simulated events, and multiplying these fractions by the total expected signal cross section. The uncertainties on the theoretical predictions include the uncertainty on the μ_F and μ_R scales (ME uncertainties), on the choice of PDFs, and on the choice of α_S. The ME uncertainties are estimated by taking the envelope of all variations, i.e., varying μ_R and μ_F up and down, individually and simultaneously, and for each bin, taking the largest variation. The PDF uncertainties are computed by taking the RMS of all possible 100 variations. The uncertainty on α_S is simply computed by varying the value of α_S up and down. The total uncertainty is then computed by summing the three components in quadrature, as they are considered to be largely uncorrelated sources of uncertainty. Tables 3.9 to 3.13 summarise the expected cross sections per bin, with the respective theoretical uncertainties, as well as the observed values of μ_i and the observed absolute differential cross sections.

p_{T} (γ) [GeV]	Expected cross section [fb]	Observed μ_i	Observed cross section [fb]
$[20, 35[$	534^{+155}_{-113}	$1.173^{+0.095}_{-0.094}$	626^{+50}_{-51}
$[35, 50[$	184^{+54}_{-39}	$1.177^{+0.090}_{-0.088}$	216^{+16}_{-17}
$[50, 70[$	98^{+29}_{-21}	$1.160^{+0.095}_{-0.092}$	113^{+11}_{-11}
$[70, 130[$	71^{+21}_{-15}	$1.095^{+0.100}_{-0.098}$	78^{+7}_{-7}
$[130, \infty[$	19^{+6}_{-4}	$1.157^{+0.122}_{-0.116}$	21^{+2}_{-2}

Table 3.9: Expected and observed absolute differential cross sections of $\mathrm{t\bar{t}}\gamma$ as a function of the p_{T} (γ), for the full Run 2 data.

p_{T} (ℓ_1) [GeV]	Expected cross section [fb]	Observed μ_i	Observed cross section [fb]
$[20, 60[$	1194^{+340}_{-249}	$1.196^{+0.084}_{-0.082}$	1428^{+100}_{-97}
$[60, 80[$	265^{+76}_{-56}	$1.182^{+0.097}_{-0.094}$	265^{+26}_{-24}
$[80, 100[$	147^{+42}_{-31}	$0.841^{+0.095}_{-0.094}$	124^{+14}_{-13}
$[100, 120[$	81^{+24}_{-17}	$1.101^{+0.113}_{-0.109}$	89^{+9}_{-9}
$[120, 150[$	60^{+18}_{-13}	$1.075^{+0.122}_{-0.118}$	64^{+7}_{-7}
$[150, \infty[$	52^{+16}_{-11}	$1.063^{+0.118}_{-0.116}$	55^{+6}_{-6}

Table 3.10: Expected and observed absolute differential cross sections of $\mathrm{t\bar{t}\gamma}$ as a function of the p_{T} (ℓ_1), for the full Run 2 data.

$\Delta\phi(\ell, \ell')$	Expected cross section [fb]	Observed μ_i	Observed cross section [fb]
$[0, 0.8[$	235^{+67}_{-49}	$1.656^{+0.152}_{-0.151}$	389^{+36}_{-35}
$[0.8, 1.6[$	268^{+77}_{-56}	$1.279^{+0.116}_{-0.116}$	342^{+31}_{-31}
$[1.6, 2.2[$	251^{+72}_{-53}	$1.192^{+0.109}_{-0.109}$	299^{+27}_{-27}
$[2.2, 2.6[$	206^{+59}_{-43}	$1.141^{+0.097}_{-0.110}$	235^{+20}_{-23}
$[2.6, 3.0[$	237^{+69}_{-50}	$1.073^{+0.098}_{-0.098}$	254^{+23}_{-23}
$[3.0, 3.16[$	89^{+26}_{-19}	$1.025^{+0.126}_{-0.123}$	91^{+11}_{-11}

Table 3.11: Expected and observed absolute differential cross sections of $\mathrm{t\bar{t}\gamma}$ as a function of the $\Delta\phi(\ell, \ell')$, for the full Run 2 data.

p_{T} (t_1) [GeV]	Expected cross section [fb]	Observed μ_i	Observed cross section [fb]
$[0, 110[$	907^{+251}_{-185}	$1.240^{+0.139}_{-0.134}$	1124^{+126}_{-121}
$[110, 220[$	1009^{+293}_{-213}	$1.158^{+0.115}_{-0.111}$	1168^{+116}_{-112}
$[220, 330[$	229^{+70}_{-50}	$1.054^{+0.144}_{-0.137}$	241^{+32}_{-31}
$[300, \infty[$	74^{+24}_{-16}	$1.086^{+0.170}_{-0.163}$	80^{+13}_{-12}

Table 3.12: Expected and observed absolute differential cross sections of $\mathrm{t\bar{t}\gamma}$ as a function of the p_{T} (t_1), for the full Run 2 data.

$\Delta R(\gamma, \mathrm{t\bar{t}})$	Expected cross section [fb]	Observed μ_i	Observed cross section [fb]
$[0, 1.9[$	494^{+143}_{-105}	$1.219^{+0.160}_{-0.150}$	602^{+79}_{-74}
$[1.9, 3.1[$	770^{+222}_{-162}	$1.427^{+0.158}_{-0.151}$	1098^{+122}_{-116}
$[3.1, 4[$	556^{+158}_{-116}	$0.882^{+0.198}_{-0.195}$	490^{+110}_{-108}
$[4, \infty[$	401^{+113}_{-83}	$1.196^{+0.248}_{-0.241}$	480^{+99}_{-97}

Table 3.13: Expected and observed absolute differential cross sections of $\mathrm{t\bar{t}\gamma}$ as a function of the $\Delta R(\gamma, \mathrm{t\bar{t}})$, for the full Run 2 data.

For the photon p_{T} and the leading lepton p_{T}, the pre- and post-fit plots are shown in figures 3.26 and 3.27. The generator-level bins of $\mathrm{t\bar{t}}\boldsymbol{\gamma}$ are shown in different shades of orange, from dark to light. They do not correspond exactly to the reconstruction-level bins due to bin-to-bin migration. In the pre-fit distributions, a slight excess of data is observed in the SR. This is consistent with the control plots of section 3.4, and as discussed there, is mainly caused by the data collected during 2016 and 2017. In the 2018 data, very good agreement between data and simulation is observed. The post-fit distribution shows good agreement. In the $\mathrm{Z}\boldsymbol{\gamma}$ CR, the description of the data by the simulation is not good. The $\mathrm{Z}\boldsymbol{\gamma}$ process is shown to be severely mismodelled. Including this region in the fit with a floating normalisation parameter allows for the fit to partially correct for this mismodelling.

Figures 3.28 to 3.30 show the observed unfolded distributions for all the observables that are measured. All measured distributions agree with the predictions, within the theoretical uncertainties, except for the first bin of the $\Delta\phi(\ell, \ell')$. For this distribution, an excess is observed in the first bin. This is probably due to an imperfect modelling of the photon emission in the $\mathrm{t\bar{t}}\boldsymbol{\gamma}$ samples.

Figure 3.31 shows, for each of the 15 leading systematic uncertainties, the expected and observed pull and constraint with respect to the prior value assigned to the nuisance parameter. The right hand side of the figure shows the impact of each of these uncertainties on the signal strength parameters, when performing a fit to the Asimov dataset (blue- and red-shaded bars) and to data (blue and red lines).

The leading systematic uncertainties are those on the electron and photon identification SFs, the uncertainty on the nonprompt photon estimation, and on the normalisation of some of the backgrounds, in particular the $\mathrm{tW}\boldsymbol{\gamma}$ background and other $\mathrm{t\bar{t}}$ processes. The fact that the $\mathrm{tW}\boldsymbol{\gamma}$ normalisation plays such an important role in the results motivates a dedicated study of the $\mathrm{tW}\boldsymbol{\gamma}$ process in the future. For the other variables, the impacts of the main uncertainties are not shown since they are very similar.

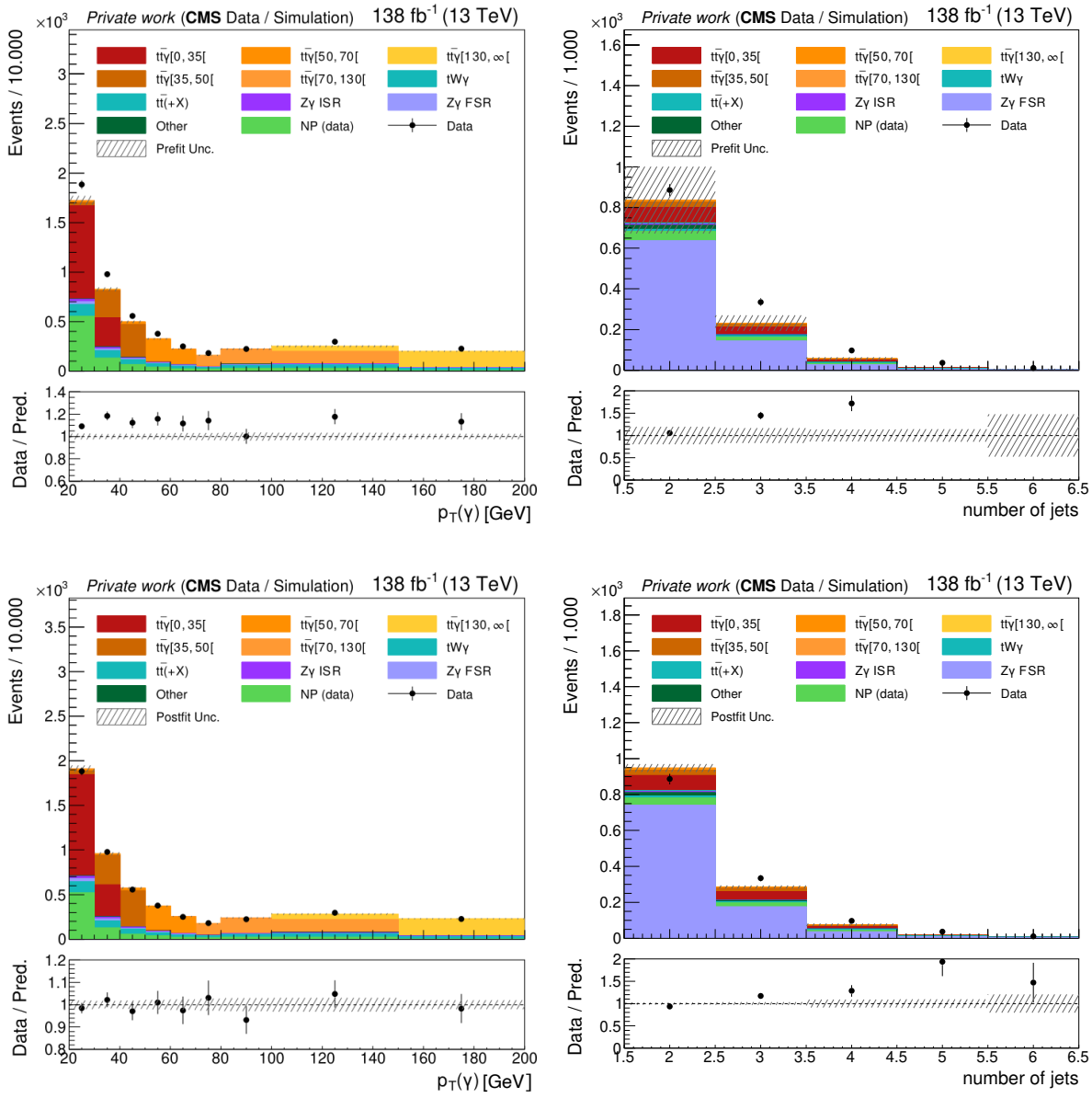

Figure 3.26: Pre-fit (upper) and post-fit (lower) distributions of the photon p_{T} in the SR (left) and Zγ CR (right), for full Run 2 data. The different generator-level bins of $p_{\mathrm{T}}(\gamma)$ for t$\bar{\mathrm{t}}\gamma$ are shown in different shades of orange, from dark to light.

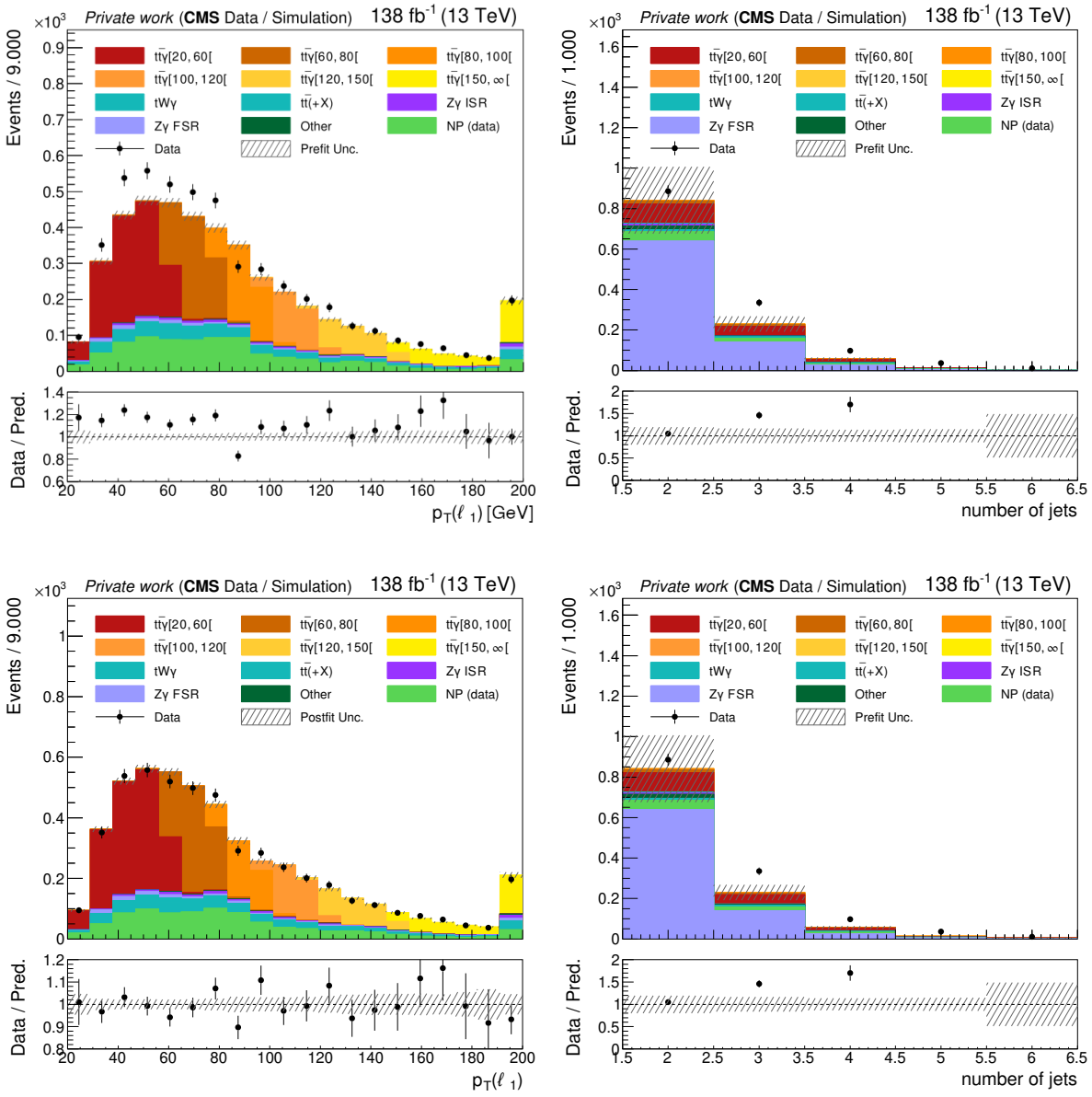

Figure 3.27: Pre-fit (upper) and post-fit (lower) distributions of the leading lepton p_{T} in the SR (left) and Zγ CR (right), for full Run 2 data. The different generator-level bins of $p_{\mathrm{T}}(\ell_1)$ for t$\bar{\mathrm{t}}\gamma$ are shown in different shades of orange, from dark to light.

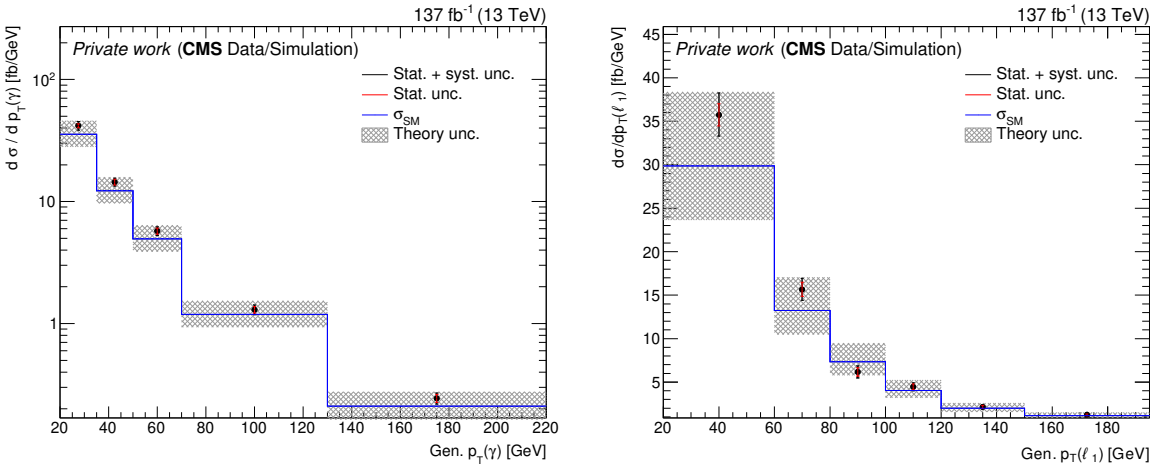

Figure 3.28: Absolute differential distributions of photon p_{T} (left) and leading lepton p_{T} (right). The blue lines show the theoretical predictions, and the grey shaded areas represent the theoretical uncertainties in the predictions. The black points represent the measured values, with the total uncertainty, while the red line shows the results considering only the statistical uncertainty.

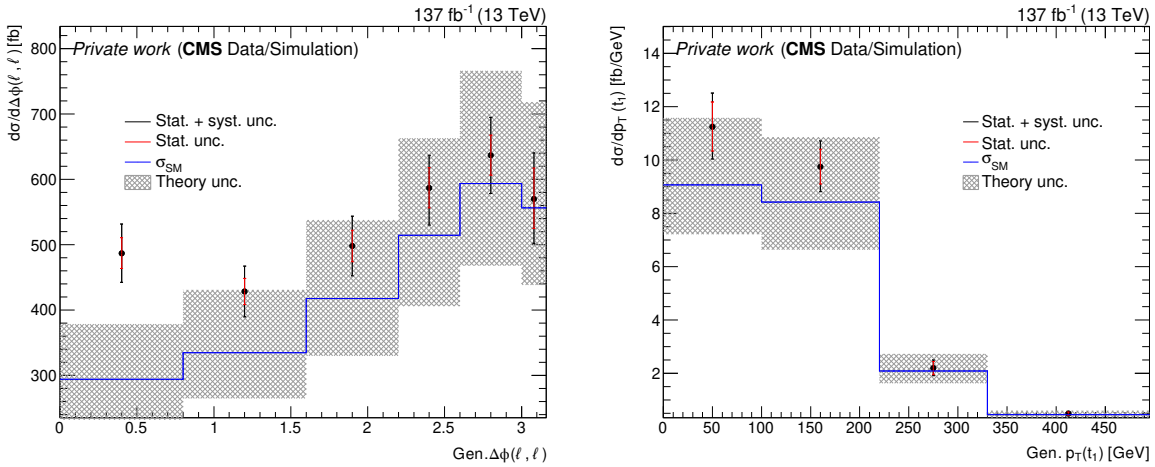

Figure 3.29: Absolute differential distributions of the $\Delta\phi$ between the leptons (left), and leading top quark p_{T} (right).

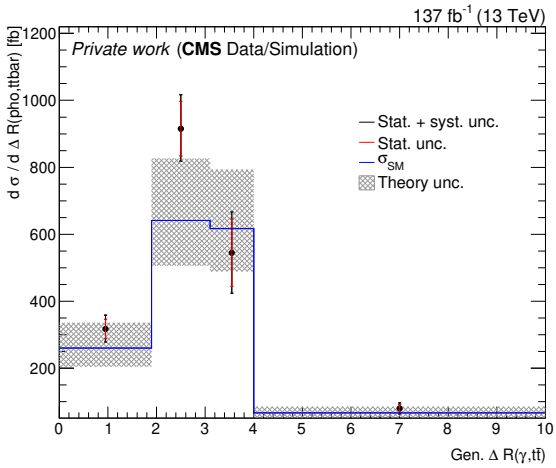

Figure 3.30: Absolute differential distributions of the $\Delta R(\gamma, \mathrm{t\bar{t}})$.

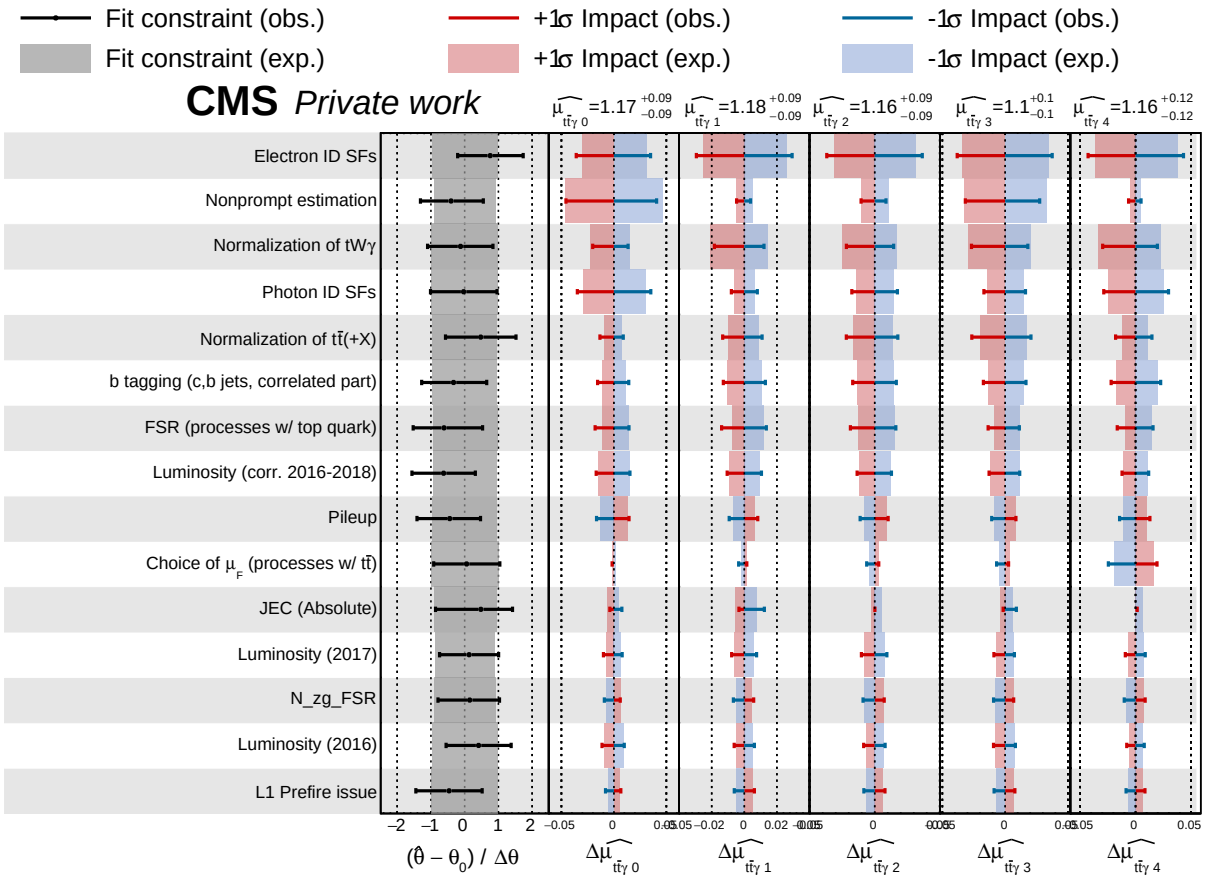

Figure 3.31: Expected and observed impacts of the leading systematic uncertainties on the differential measurement as a function of the photon p_{T}. Each column shows the impacts of the uncertainties on each of the signal strength parameters μ_i. The black line shows, for each parameter, the relative constraint and pull after the fit. The blue and red shaded areas (lines) show the expected (observed) impacts of shifting each parameter to its -1 and $+1\,\sigma$ values, respectively.

3.9 Measurement of the inclusive ratio between the $t\bar{t}\gamma$ and $t\bar{t}$ cross sections

Additionally, I present a measurement of the ratio R_γ between the cross sections of $t\bar{t}\gamma$ and $t\bar{t}$, as motivated in section 1.3.4. This is the first analysis aiming to do a direct measurement of this ratio at the LHC, both inclusively and differentially.

3.9.1 Definition of the ratio

The signal ($t\bar{t}\gamma$) process is defined as in section 3.1.1, while the $t\bar{t}$ process is defined as $t\bar{t}$ production with zero associated photons satisfying the signal definition. The ratio R_γ is the ratio between $t\bar{t}\gamma$ and the sum of $t\bar{t}\gamma$ and the $t\bar{t}$. With this definition, we can write R_γ as:

$$R_\gamma = \frac{\sigma_{t\bar{t},\geq 1\gamma}}{\sigma_{t\bar{t},0\gamma} + \sigma_{t\bar{t},\geq 1\gamma}} \tag{3.25}$$

Similarly to the signal strength defined in section 3.7.1, one can define a "ratio strength parameter" μ_R as

$$\mu_R = \frac{R}{R^{\text{exp}}} \quad , \tag{3.26}$$

where R^{exp} is the expected ratio between the inclusive cross section of $t\bar{t}\gamma$ and the total $t\bar{t}$ (including $t\bar{t}\gamma$) production.

For the measurement, we define three different regions: the $t\bar{t}$ 0-photon region, identical to the SR from the differential fits, except for inverting the photon requirement (i.e. requiring exactly 0 reconstructed photons), the $t\bar{t}$ $\geq$1-photon region, or $t\bar{t}\gamma$ region, which is identical to the SR in the differential fits, and the $Z\gamma$ CR, also identical to the CR in the differential fits.

The $\Delta\phi(\ell, \ell')$ distribution for all three regions, using the full Run 2 data, is shown in figure 3.32. This distribution was chosen for the fit, to extract the inclusive R_γ, by comparing the expected uncertainty obtained when fitting different distributions (with the Asimov dataset), and choosing a distribution that yields the lowest uncertainty while being reasonably well modelled in all regions. Statistical and systematic uncertainties before the fit are shown as hashed bands in the figures.

The signal strength parameters for the $t\bar{t}$ (μ_0) and $t\bar{t}\gamma$ (μ_1) processes are defined as follows:

$$\mu_0 = \frac{\sigma_{t\bar{t},0\gamma}}{\sigma^{\text{exp}}_{t\bar{t},0\gamma}} \quad \mu_1 = \frac{\sigma_{t\bar{t},\geq 1\gamma}}{\sigma^{\text{exp}}_{t\bar{t},\geq 1\gamma}} \tag{3.27}$$

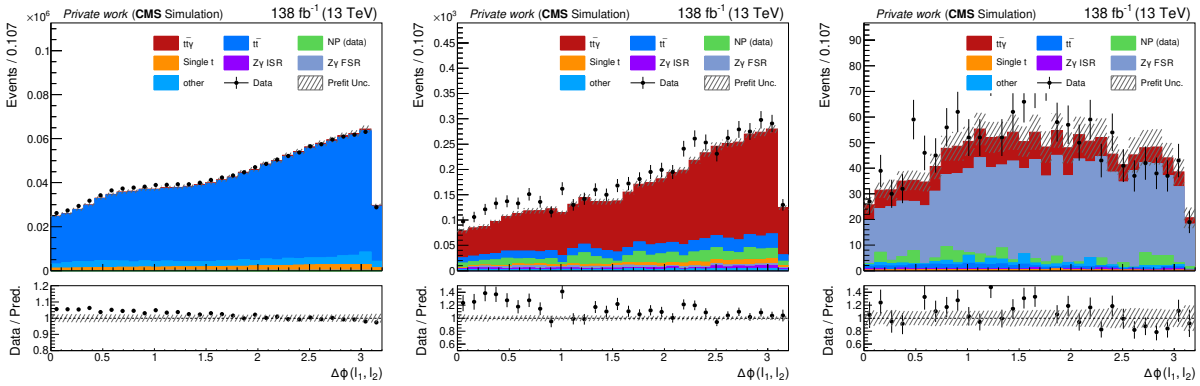

Figure 3.32: Distribution of the $\Delta\phi(\ell, \ell')$ for the $t\bar{t}$ region (left), the SR (centre) and the $Z\gamma$ CR (right), for full Run 2 data, before performing any fit.

where σ are the measured cross sections and σ^{exp} are the SM expected cross sections, computed from simulation. With some simple algebra, μ_0 can be rewritten as a function of μ_1 and μ_R:

$$\mu_0 = \frac{\mu_1}{\mu_R} \times \left(1 + \frac{\sigma^{exp}_{t\bar{t},\geq 1\gamma}}{\sigma^{exp}_{t\bar{t},0\gamma}}(1 - \mu_R)\right) \tag{3.28}$$

This reparametrization allows the direct extraction of μ_R from the fit.

The nonprompt photon background estimation from section 3.6.2 cannot be applied to this measurement directly. In this measurement, there is a 0-photon region, that is composed in more than 90% of $t\bar{t}$ events. These events are accurately modelled using a $t\bar{t}$ simulated sample. The nonprompt photon contribution in the 1-photon region is composed mostly of $t\bar{t}$ events. By using the data-driven estimation in the 1-photon region and the simulation in the 0-photon region, these processes would be treated as fully uncorrelated, when they are in fact almost fully correlated. To circumvent this issue, the $t\bar{t}$ simulation is used in both regions, and the remaining (non-$t\bar{t}$) nonprompt contribution is taken from the data-driven estimation, after subtracting the $t\bar{t}$ contribution from simulation.

3.9.2 Results

From the fit, we extract the following parameters:

$$\mu_R : 1.086^{+0.051}_{-0.050}\ (68\%\ \text{CL}) \text{ or } 1.086^{+0.020}_{-0.020}\ (\text{stat.})\ ^{+0.047}_{-0.046}\ (\text{syst.}) \tag{3.29}$$

$$\mu_1 : 1.093^{+0.045}_{-0.043}\ (68\%\ \text{CL}) \text{ or } 1.093^{+0.021}_{-0.020}\ (\text{stat.})\ ^{+0.040}_{-0.038}\ (\text{syst.}) \tag{3.30}$$

where the statistical uncertainty is obtained by fitting each POI separately and freezing the other one, as well the nuisance parameters, to their nominal value. We obtain the following value for R:

$$R = 0.02779^{+0.0013}_{-0.0013}\ (68\%\ \text{CL}) \quad , \tag{3.31}$$

compatible with the SM prediction of $0.02543^{+0.0092}_{-0.0113}$. This prediction is computed by taking the ratio between the cross section predictions for $\mathrm{t\bar{t}\gamma}$ (at NLO in QCD) and $\mathrm{t\bar{t}}$ (at NNLO in QCD). The theoretical uncertainties on the prediction are computed in the same way as in section 3.8, but propagating the variations through the ratio and only then taking the envelope/RMS of the different contributions and summing them in quadrature. The cross section of $\mathrm{t\bar{t}\gamma}$ in the dilepton channel, in the fiducial region defined in table 3.4, is measured to be:

$$\sigma_{t\bar{t},\geq 1\gamma} = (2426.86^{+100}_{-95})\ \text{fb}\ (68\%\ \text{CL}) \quad , \tag{3.32}$$

compatible with the prediction of (2220.37^{+638}_{-466}) fb. The likelihood scans for both parameters that are measured are shown in figure 3.33.

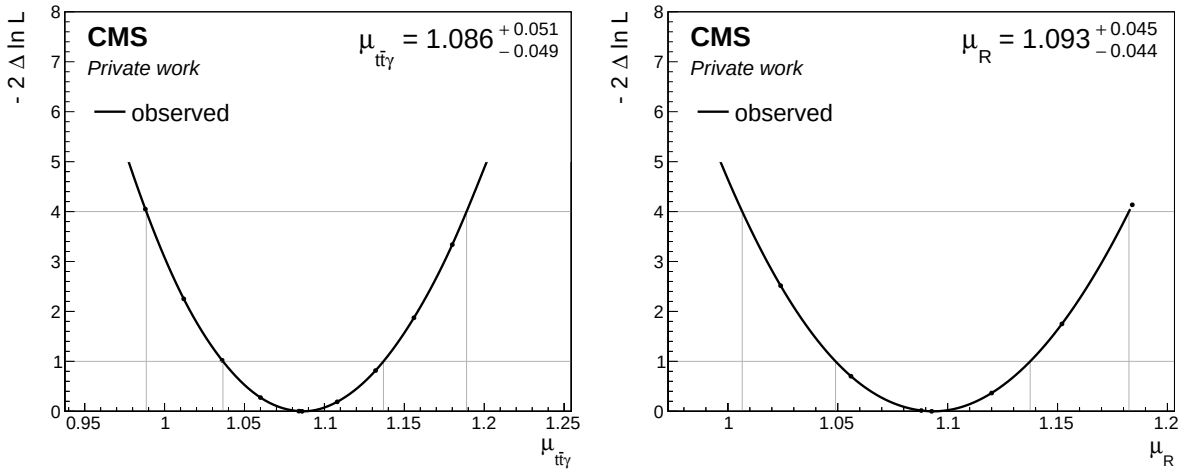

Figure 3.33: Likelihood scans for the two observed fits to the two POIs, $\mu_{\mathrm{t\bar{t}\gamma}}$ (left) and μ_R (right). The vertical lines mark the 68 and 95% confidence bands, corresponding to 1 and $2\,\sigma$, respectively.

Figure 3.34 shows the post-fit distributions of the $\Delta\phi(\ell, \ell')$ in the three regions. Figure 3.35 shows the expected and observed impacts of the leading systematic uncertainties. The most relevant uncertainties are related to the photon identification and the nonprompt photon estimation, which is expected since a change in the photon selection can lead to a change in the number of photons, and consequently, migration from one region to the

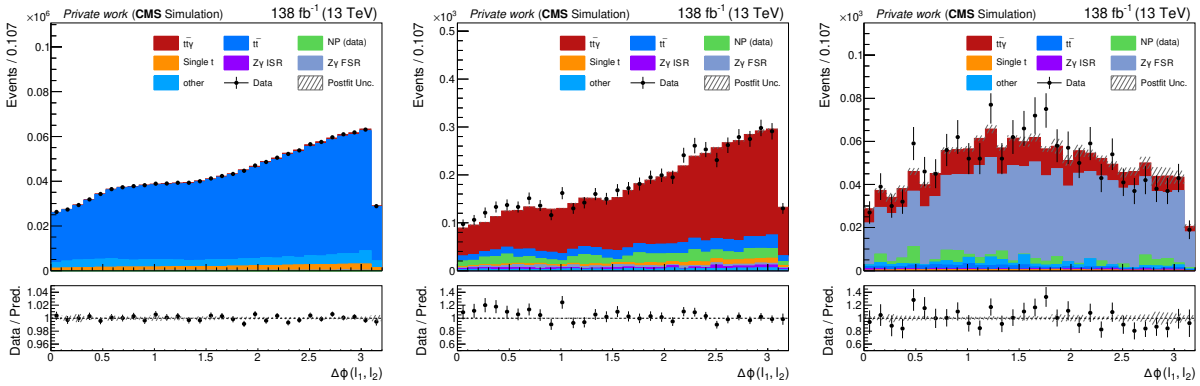

Figure 3.34: Distribution of the $\Delta\phi(\ell, \ell')$ for the $t\bar{t}$ region (left), the SR (centre) and the $Z\gamma$ CR (right), for full Run 2 data, post-fit.

other. On the other hand, ISR and FSR play an important role because they modify the photon emission, and also cause migrations between the 0-photon and 1-photon regions.

3.10 Measurement of differential ratios between the $t\bar{t}\gamma$ and $t\bar{t}$ cross sections

We perform the differential ratio measurement as a function of the leading lepton p_{T} and the p_{T} of the leading t quark. For that, we apply the same reparametrization as for the inclusive ratio measurement, but instead of doing it inclusively, we do so per generator-level bin of the variable to be measured. As a result, one parameter μ_{1i} and one parameter μ_{Ri} will be extracted for each gen-level bin i. Here, no $Z\gamma$ CR is included, as the fit has many POIs and became too computationally expensive. It was shown that removing the CR did not significantly affect the measurements.

For the leading lepton p_T, 6 generator-level bins are defined for the fit, while for the leading top p_T, 4 generator-level bins are defined. Tables 3.14 and 3.15 show the expected ratio between $t\bar{t}\gamma$ and $t\bar{t}$, computed from simulation at generator-level, per bin, for the two observables.

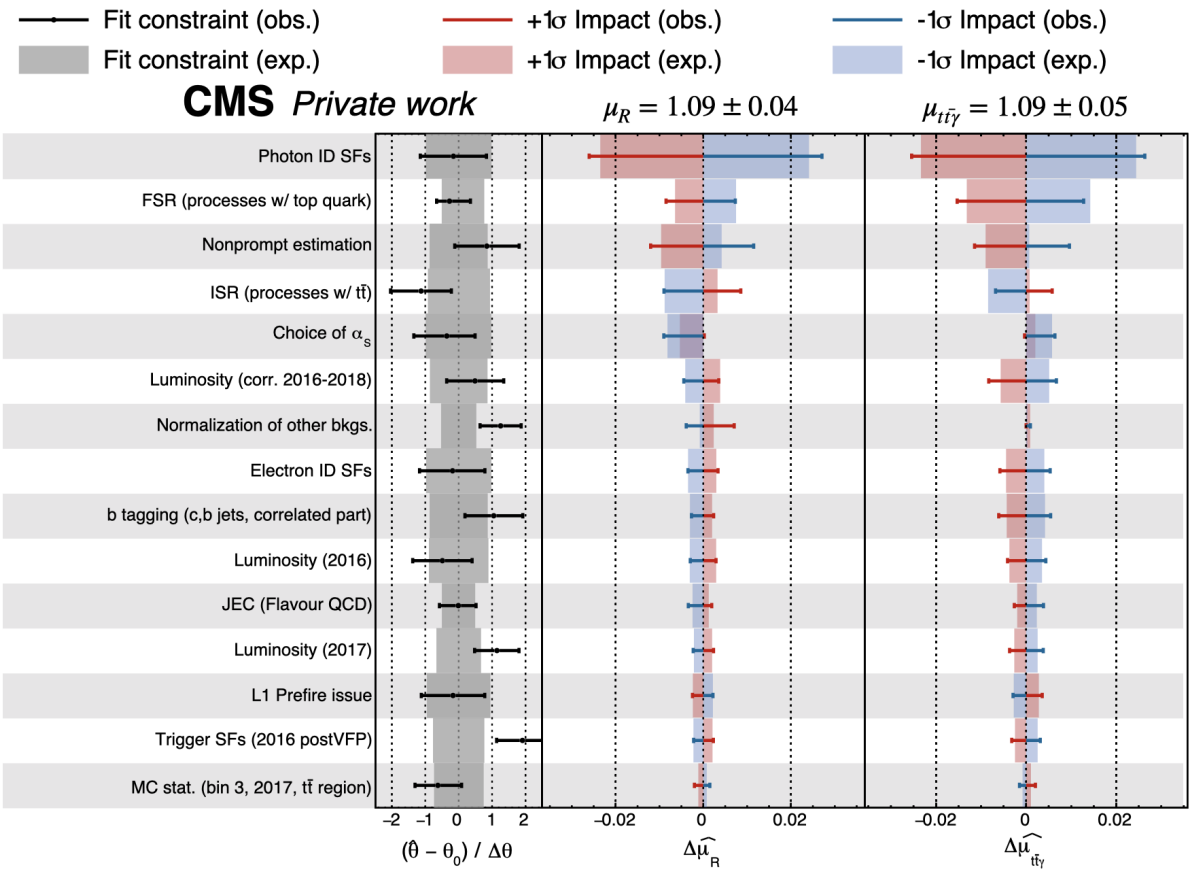

Figure 3.35: Expected and observed impacts on each of the leading systematic uncertainty on the signal strength parameter, $\mu_{t\bar{t}\gamma}$ and the ratio parameter μ_R, where the expected value $\mu = 1$ corresponds to the SM expected cross sections.

Generated $p_T(\ell_1)$ [GeV]	Expected diff. ratio
20-60	0.0231
60-80	0.0221
80-100	0.0225
100-120	0.0233
120-150	0.0243
150-∞	0.0268

Table 3.14: Expected differential ratio $t\bar{t}\gamma/t\bar{t}$ per generator-level bin of the leading lepton p_T, computed from the simulated signal samples.

Generated $p_T(\mathrm{t}_1)$ [GeV]	Expected diff. ratio
0-100	0.02288
100-220	0.02675
220-330	0.03199
330-∞	0.03653

Table 3.15: Expected differential ratio $\mathrm{t\bar{t}\gamma}/\mathrm{t\bar{t}}$ per generator-level bin of the leading top p_T, computed from the simulated signal samples.

3.10.1 Results

Leading lepton p_{T}

From the fit, we extract the following parameters:

$$\mu_{10} : 1.175^{+0.071}_{-0.067} \quad \mu_{R0} : 1.138^{+0.061}_{-0.058} \tag{3.33}$$
$$\mu_{11} : 1.251^{+0.083}_{-0.085} \quad \mu_{R1} : 1.236^{+0.074}_{-0.078} \tag{3.34}$$
$$\mu_{12} : 0.917^{+0.087}_{-0.086} \quad \mu_{R2} : 0.912^{+0.082}_{-0.081} \tag{3.35}$$
$$\mu_{13} : 1.172^{+0.104}_{-0.104} \quad \mu_{R3} : 1.189^{+0.099}_{-0.100} \tag{3.36}$$
$$\mu_{14} : 1.130^{+0.117}_{-0.105} \quad \mu_{R4} : 1.147^{+0.113}_{-0.101} \tag{3.37}$$
$$\mu_{15} : 1.102^{+0.105}_{-0.095} \quad \mu_{R5} : 1.145^{+0.103}_{-0.093} \tag{3.38}$$

which result in the observed unfolded distribution of figure 3.36 (right). The figure (left) shows the expected distribution, obtained using an Asimov dataset. Figures 3.37 and 3.38 show the pre- and post-fit distributions of $p_{\mathrm{T}}(\ell_1)$ in both regions. The expected and observed impacts of the leading systematic uncertainties are shown in figure 3.39. As expected, the leading systematic uncertainties are those related to photon identification and the nonprompt photon estimation, similarly to the inclusive ratio case.

Leading top quark p_{T}

From the fit, we extract the following parameters:

$$\mu_{10} : 1.194^{+0.149}_{-0.128} \quad \mu_{R0} : 1.205^{+0.135}_{-0.121} \tag{3.39}$$
$$\mu_{11} : 1.062^{+0.090}_{-0.116} \quad \mu_{R1} : 1.077^{+0.079}_{-0.110} \tag{3.40}$$
$$\mu_{12} : 1.196^{+0.159}_{-0.113} \quad \mu_{R2} : 1.200^{+0.152}_{-0.107} \tag{3.41}$$
$$\mu_{13} : 0.873^{+0.128}_{-0.153} \quad \mu_{R3} : 0.941^{+0.131}_{-0.160} \tag{3.42}$$
$$\tag{3.43}$$

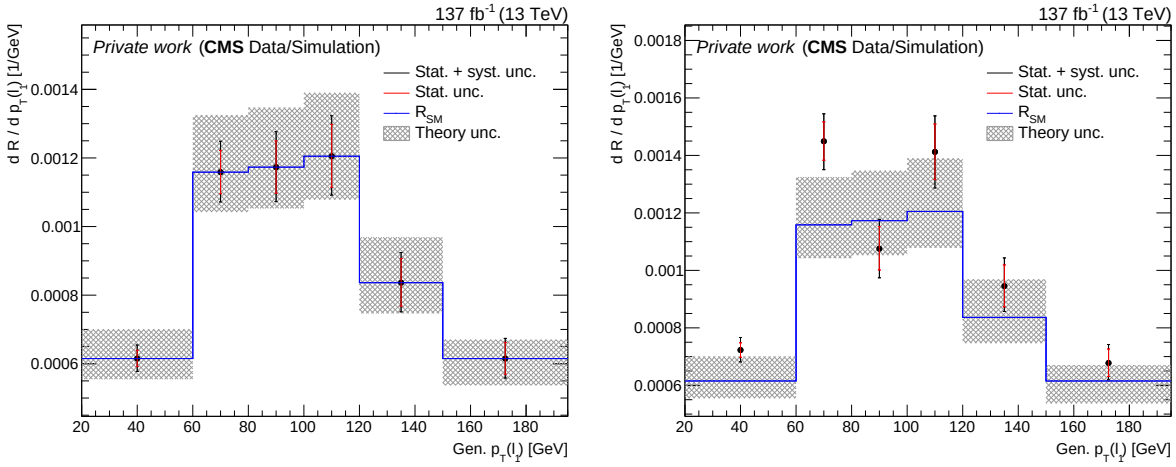

Figure 3.36: Expected (left) and observed (right) unfolded distributions of the ratio between $t\bar{t}\gamma$ and $t\bar{t}$ as a function of generator-level $p_T(\ell_1)$.

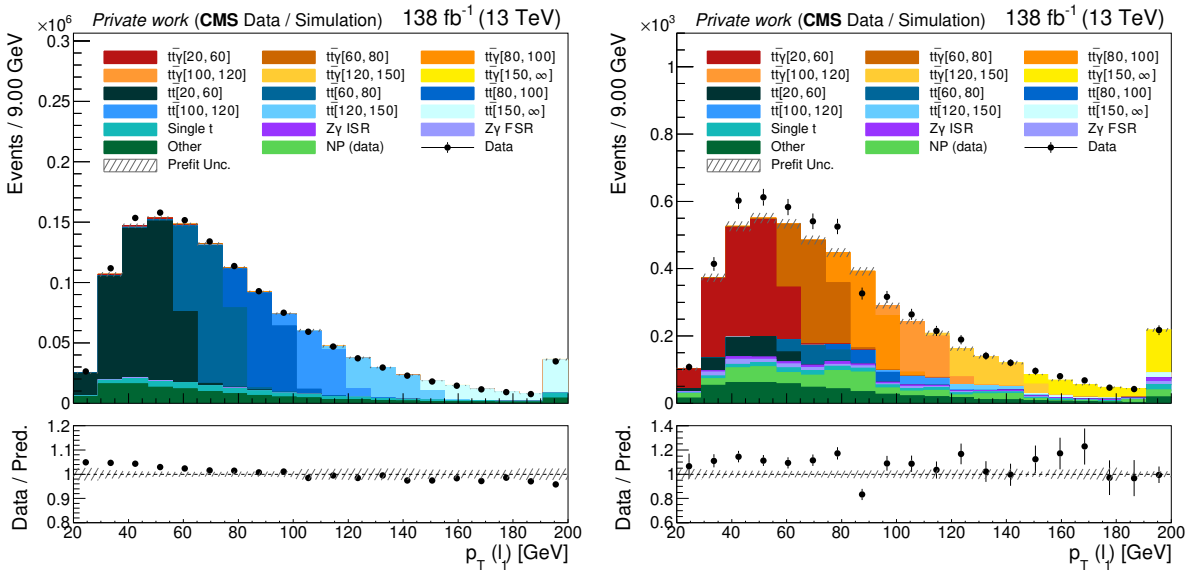

Figure 3.37: Distribution of the $p_T(\ell_1)$ for the $t\bar{t}$ region (left) and the SR (right), for full Run 2 data, pre-fit. The different generator-level bins of $p_T(\ell_1)$ for $t\bar{t}$ ($t\bar{t}\gamma$) are shown in different shades of blue (orange), from dark to light.

which result in the observed unfolded distribution of figure 3.40 (right). The figure (left) shows the expected distribution, obtained using an Asimov dataset. Figures 3.41 and 3.42 show the pre- and post-fit distributions of $p_T(t_1)$ in both regions. The impacts of the leading systematic uncertainties are not shown here, as they are identical to the

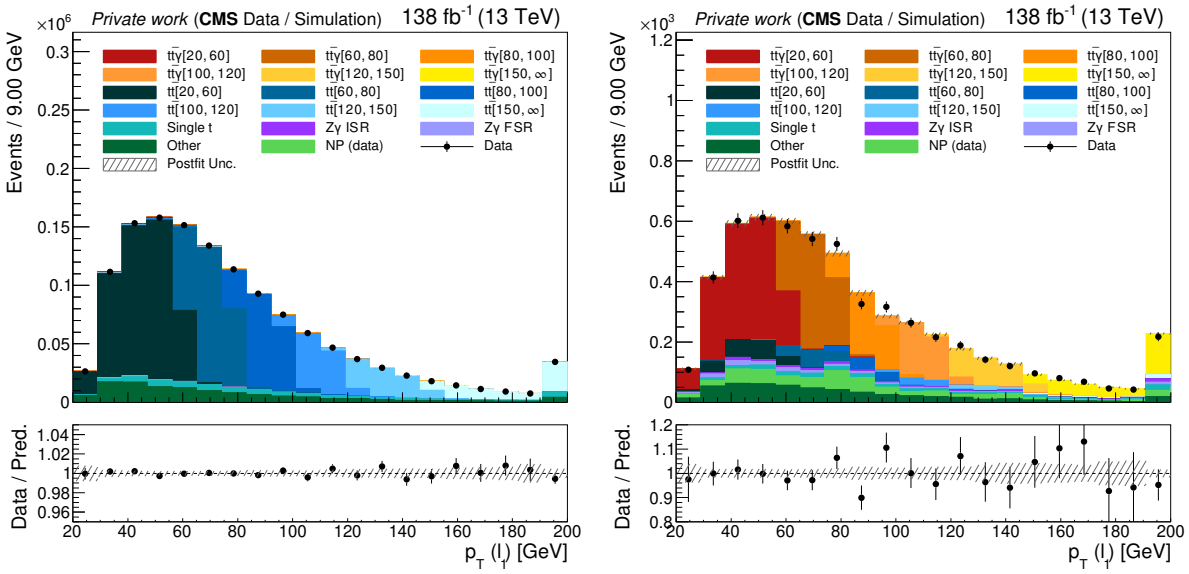

Figure 3.38: Distribution of the $p_{\mathrm{T}}(\ell_1)$ for the $\mathrm{t\bar{t}}$ region (left) and the SR (right), for full Run 2 data, post-fit. The different generator-level bins of $p_{\mathrm{T}}(\ell_1)$ for $\mathrm{t\bar{t}}$ ($\mathrm{t\bar{t}\gamma}$) are shown in different shades of blue (orange), from dark to light.

lepton p_{T} case shown above.

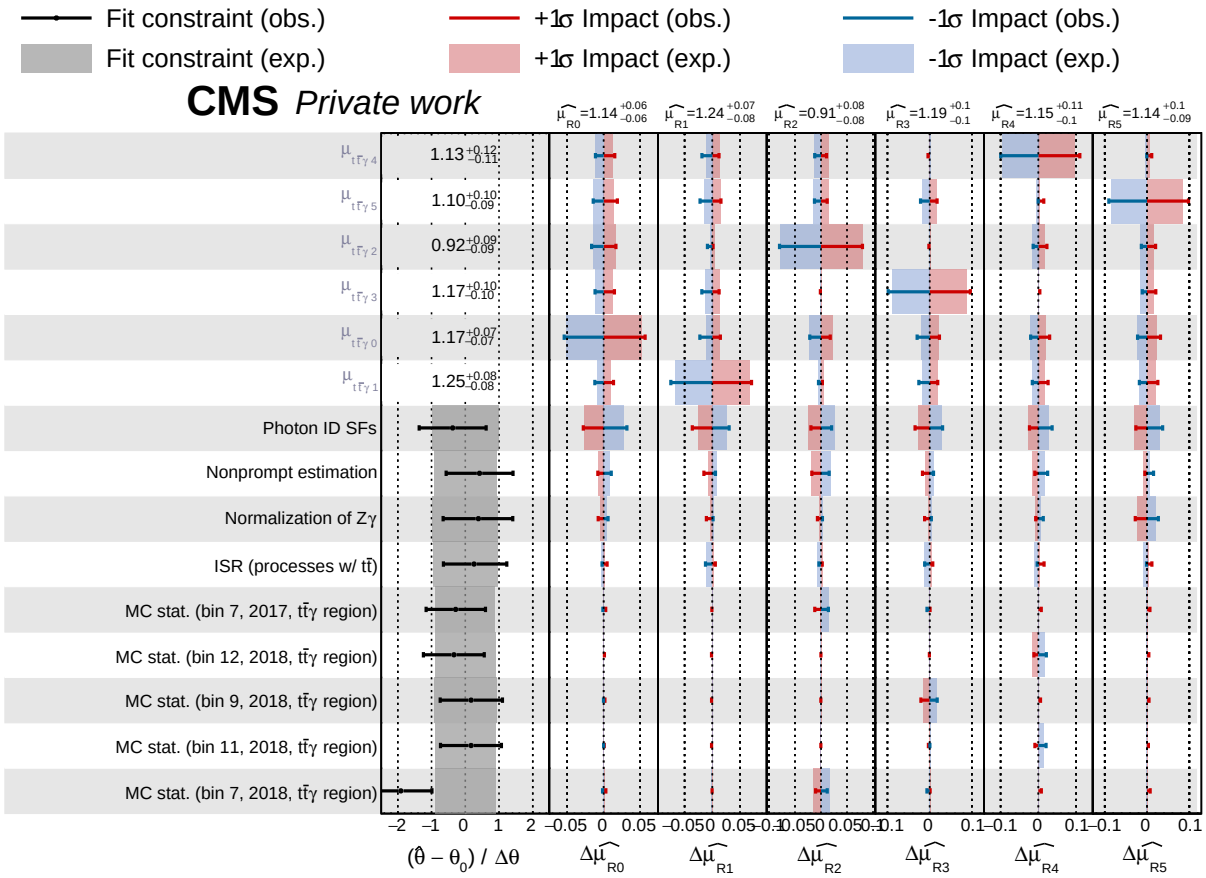

Figure 3.39: Expected and observed impacts of the leading systematic uncertainties on the different ratio parameters.

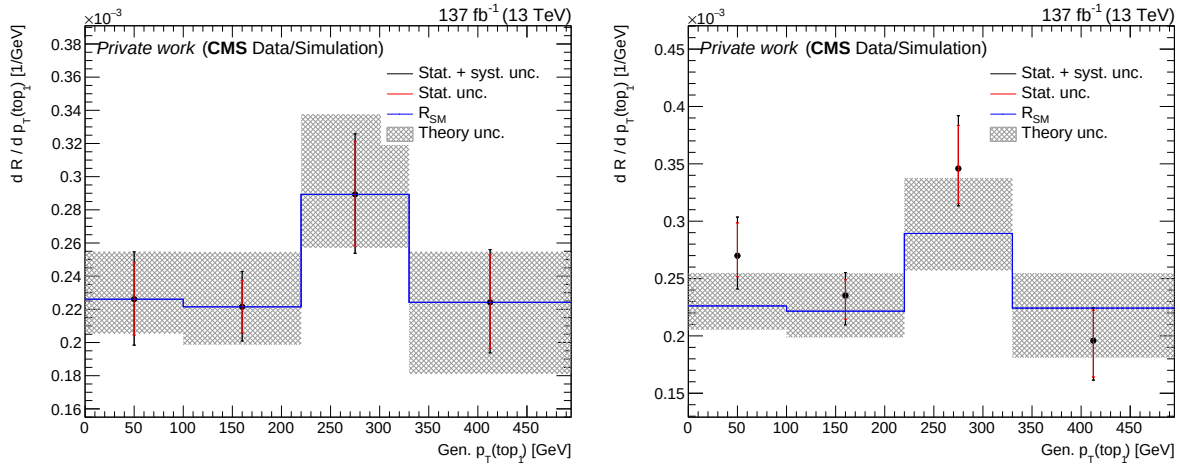

Figure 3.40: Expected (left) and observed (right) unfolded distributions of the ratio between $t\bar{t}\gamma$ and $t\bar{t}$ as a function of generator-level $p_{\mathrm{T}}(\mathrm{t}_1)$.

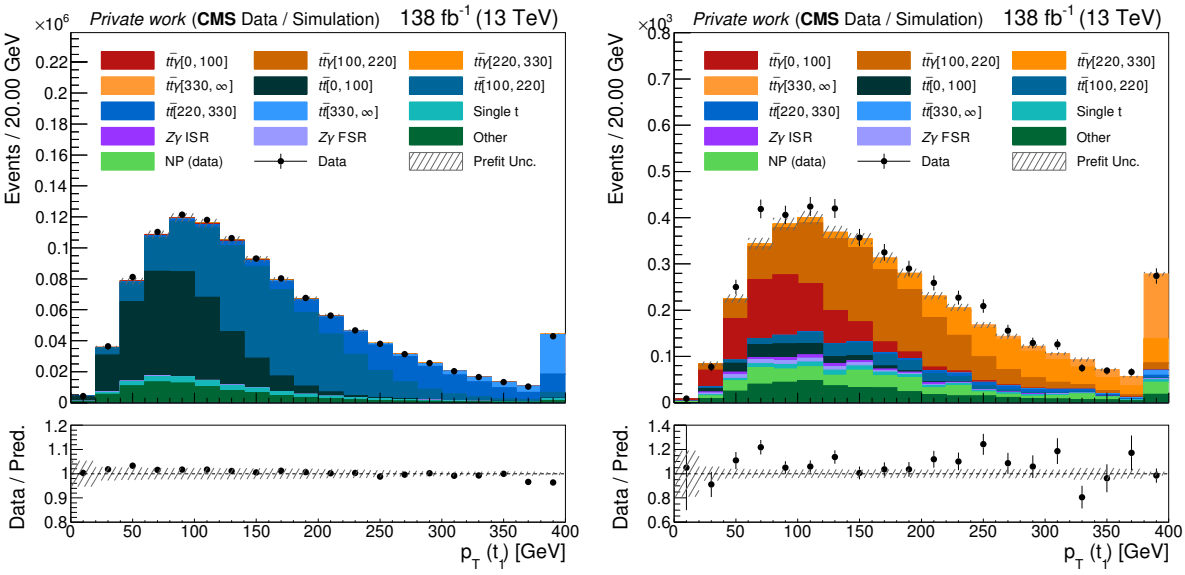

Figure 3.41: Distribution of the $p_{\mathrm{T}}(\mathrm{t}_1)$ for the $t\bar{t}$ region (left) and the SR (right), for full Run 2 data, pre-fit. The different generator-level bins of $p_{\mathrm{T}}(\mathrm{t}_1)$ for $t\bar{t}$ ($t\bar{t}\gamma$) are shown in different shades of blue (orange), from dark to light.

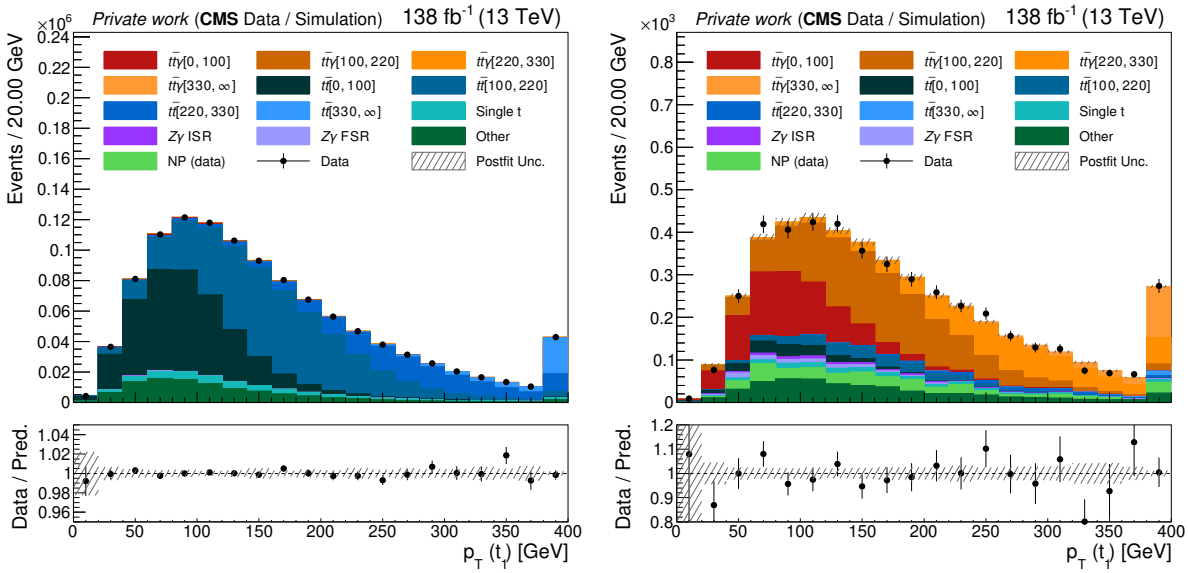

Figure 3.42: Distribution of the $p_{\mathrm{T}}(\mathrm{t}_1)$ for the $\mathrm{t\bar{t}}$ region (left) and the SR (right), for full Run 2 data, post-fit. The different generator-level bins of $p_{\mathrm{T}}(\mathrm{t}_1)$ for $\mathrm{t\bar{t}}$ ($\mathrm{t\bar{t}\gamma}$) are shown in different shades of blue (orange), from dark to light.

3.11 The top quark charge asymmetry in $t\bar{t}\gamma$ events

The top quark charge asymmetry, as defined in section 1.3.5, can be extracted from the data in a similar way to the signal strengths presented in the previous sections. In a way, measuring the charge asymmetry is equivalent to doing a differential measurement of the variable $\Delta|y| = |y_t| - |y_{\bar{t}}|$. The data can be divided in two bins of this variable, such that the first bin ("bin−") has $\Delta|y| < 0$ and the second one ("bin+") has $\Delta|y| > 0$. For the two bins, we assign the signal strength parameters μ_- and μ_+, respectively.

In order to extract the charge asymmetry directly from the signal strengths, μ_+ can be reparametrised as

$$\mu_+ = \alpha\mu_- \frac{100 + A_C(\%)}{100 - A_C(\%)} \tag{3.44}$$

which is obtained directly from the definition in section 1.3.5 with some algebraic manipulation and using

$$\alpha = \frac{\sigma_-^{\text{exp}}}{\sigma_+^{\text{exp}}} \tag{3.45}$$

which is a constant, taken from simulation at the generator level. Then, the same maximum-likelihood estimation can be used as for the differential measurements, and the A_C in percentage can be extracted directly. The advantage of performing this reparametrization is that the correlations between the different systematic uncertainties can be taken into account directly by the minimisation procedure, and no *a posteriori* error propagation is needed.

3.11.1 Results

The expected (SM) value for the charge asymmetry, computed from the signal simulation, is -0.514%. The fit to the full Run 2 distributions of $\Delta y(\mathrm{t}, \bar{\mathrm{t}})$ results in an expected uncertainty of 4.5% for the charge asymmetry. The observed values of the POIs (A_C and μ_-) are:

$$\mu_- : 1.158^{+0.096}_{-0.089} \quad A_C : (4.686^{+4.385}_{-4.398})\% \tag{3.46}$$

Figure 3.43 shows the likelihood scans for the fits to the two POIs. The pre- and post-fit plots for the $\Delta y(\mathrm{t}, \bar{\mathrm{t}})$ with the full Run 2 data are shown in figure 3.44.

The impacts of the leading systematic uncertainties are shown in figure 3.45. The statistical uncertainty due to the limited size of the MC samples (labelled as "MC stat." in the figure) is the leading contribution in the extraction of the charge asymmetry, together with the statistical uncertainty in the data. In practice, this means that in order

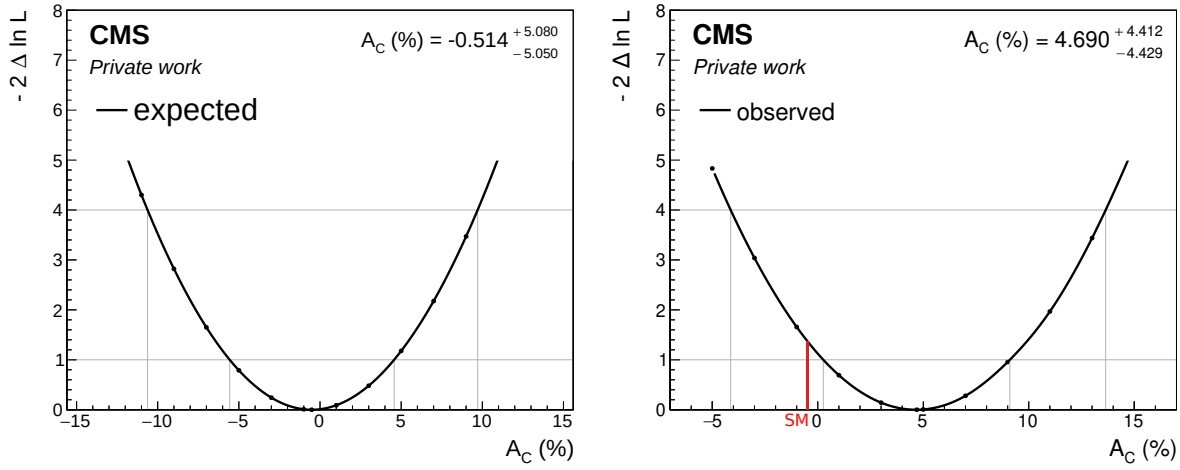

Figure 3.43: Likelihood scans for the expected (left) and observed (right) fits to A_C.

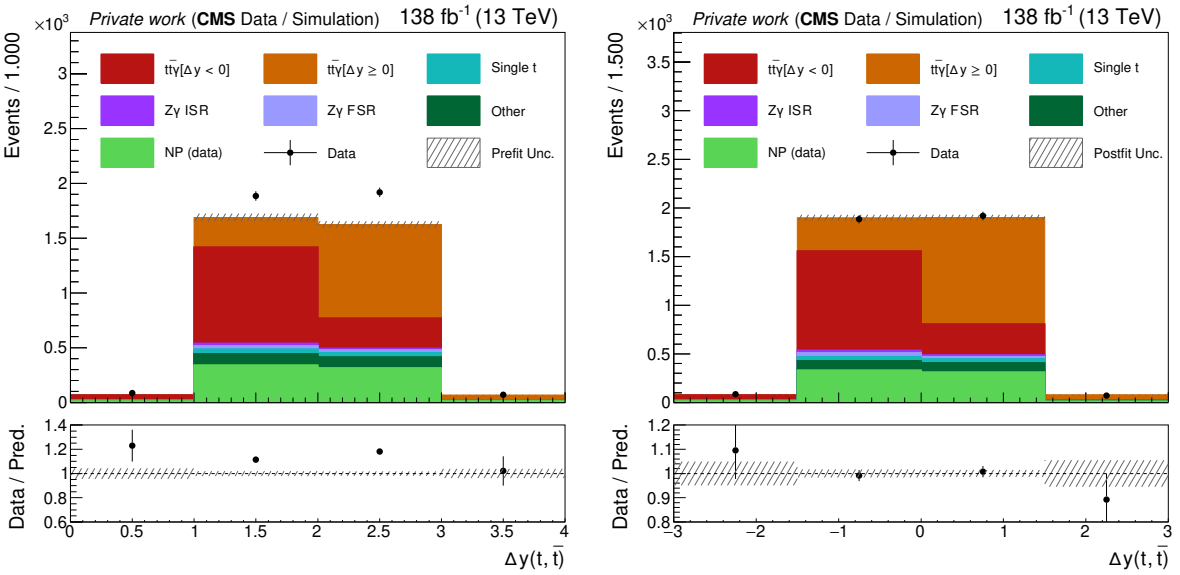

Figure 3.44: Distribution of the $\Delta y(\mathrm{t},\bar{\mathrm{t}})$ for the SR, for full Run 2 data, pre- (left) and post-fit (right). The two generator-level bins of $\Delta y(\mathrm{t},\bar{\mathrm{t}})$ $\mathrm{t\bar{t}\gamma}$ are shown in red and orange.

to disentangle the very small relative uncertainty, a very precise simulation of $\mathrm{t\bar{t}\gamma}$ events, and more data, are needed. The current data allows us to measure $A_C = (4.686^{+4.385}_{-4.398})\%$, which differs from 0 with a significance of about $1\,\sigma$, and is compatible with the SM expectation also within about $1\,\sigma$, as can be seen in figure 3.43 (right).

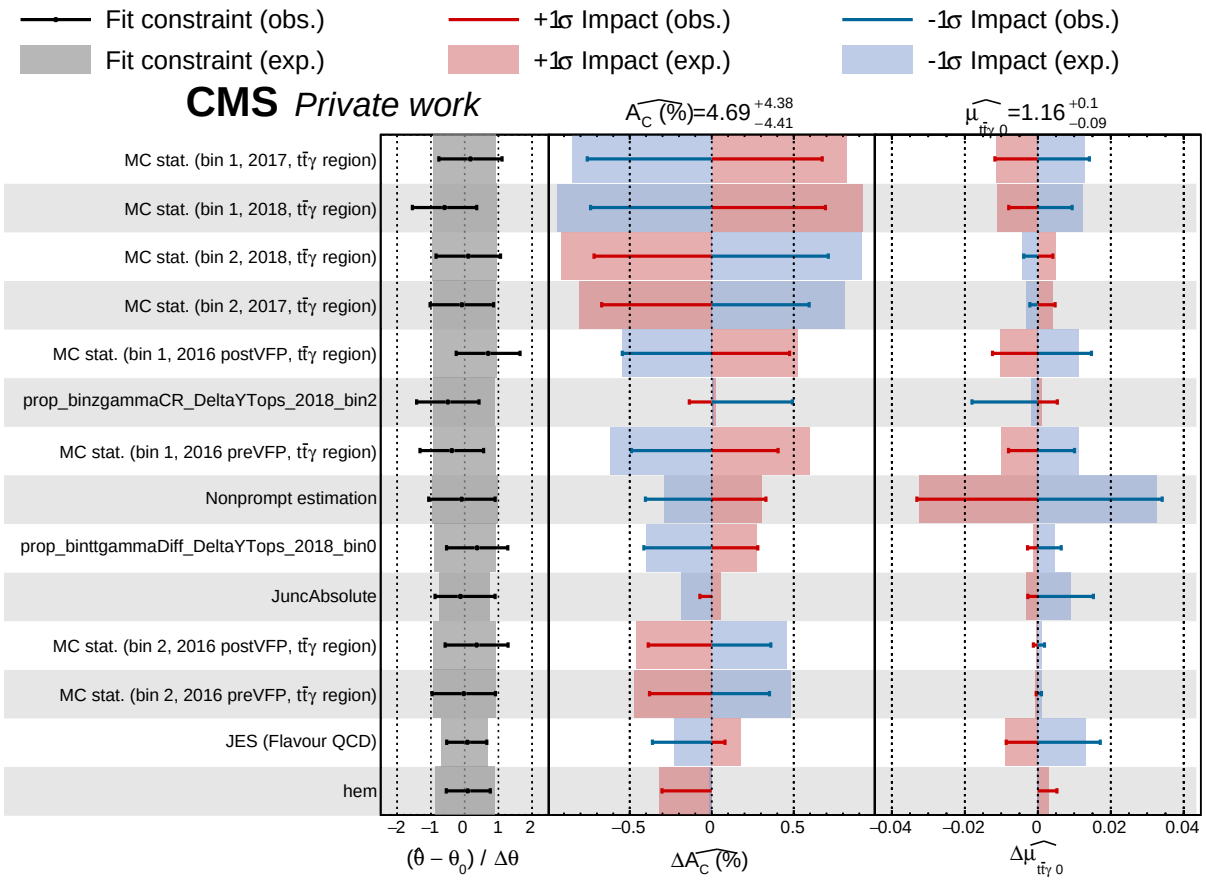

Figure 3.45: Expected and observed impacts of the leading systematic uncertainties on the measured A_C and the parameter μ_-.

3.12 Discussion and outlook

In this chapter, data collected by the CMS experiment between 2016 and 2018 at a centre-of-mass energy of 13 TeV are analysed to measure differential cross sections of $t\bar{t}\gamma$, using parton level definitions, and the inclusive and differential ratio between the cross sections of $t\bar{t}\gamma$ and $t\bar{t}$. The work has been documented within the CMS collaboration in an internal analysis note, and a publication reporting the results is in preparation. Section 1.3.2 motivates the study of $t\bar{t}\gamma$ and the $t\bar{t}\gamma/t\bar{t}$ ratio in terms of their potential for interpretations in the context of SMEFT. While this is outside the scope of this PhD thesis, it would be a logical next step to this work. The publication that is in preparation will contain all necessary information to allow such an interpretation to be performed *a posteriori* from the results.

Measurements of $t\bar{t}\gamma$ can also be combined with other processes which also offer sensi-

tivity to the $t\gamma$ vertex, and consequently to the same EFT parameters. A prime example is $t\bar{t}Z$, which offers complementary sensitivity to $t\bar{t}\gamma$, as motivated in section 1.3.4. Therefore, a future simultaneous measurement of $t\bar{t}\gamma$ and $t\bar{t}Z$, or even better the ratio between the $t\bar{t}\gamma$ and $t\bar{t}Z$ cross sections, could provide even more powerful inputs to theory.

The differential cross section measurements in this thesis are affected by statistical and systematic uncertainties almost in the same measure, with the systematic uncertainties being slightly dominating in some bins. This means that in order to reach better precision levels, it is not sufficient to simply collect more data, but it is also crucial to reduce, where possible, the systematic uncertainties, as well as to improve the theory predictions.

In particular, the uncertainty on the measurements related to top quark observables can be reduced by improving the reconstruction of the $t\bar{t}$ system. One way to improve it is to include the photon in the reconstruction, either with the method proposed in Ref. [161], or using machine learning techniques. Deep neural networks to perform $t\bar{t}$ reconstruction have been developed and shown to bring improvements with respect to the traditional analytical algorithms, for example in Ref. [176]. Those networks can be modified such that the photon kinematics is given as an input feature, without having to make any *a priori* assumption about the photon origin.

One of the main systematic uncertainties affecting all the differential cross section measurements of $t\bar{t}\gamma$ is the uncertainty on the cross section of $tW\gamma$, a non-negligible background to $t\bar{t}\gamma$, which has not yet been observed. An extensive dedicated study of $tW\gamma$, taking into account the interference between this process and $t\bar{t}\gamma$, would be a very interesting addition to the work of this thesis, and will potentially bring improvements to the measurements presented.

The charge asymmetry measurement, instead, is heavily dominated by its statistical uncertainty. This includes both the uncertainty due to insufficient amounts of data, and the one due to the limited size of the simulated samples. This means that in order to improve this measurement, first of all, more data is needed. This can be achieved, for example, by combining the data from Run 2 and Run 3. However, that is not enough, as a much higher number of simulated events is also needed. Generating large enough samples to reduce the uncertainty would come at a very high computational cost, and is most likely not feasible in the near future. A recently developed technique uses machine learning to overcome limitations due to limited size of simulated samples: by reweighting the existing simulations to different model parameters with a neural network, a large number of events can be simulated, without the need for simulating the detector response multiple times [177, 178]. This model could in principle be adapted to generate a large number of $t\bar{t}\gamma$ events in a phase space where the charge asymmetry is enhanced. Samples with different values of charge asymmetry could be generated, such that they can be compared to the data and potentially disentangle any BSM effect that could be present.

The HL-LHC is expected to collect about 5 times the luminosity of all the LHC runs so far, at a centre-of-mass energy of 14 TeV. For $t\bar{t}\gamma$, this brings an increase in cross section, and a drastic reduction in the statistical uncertainty. This is particularly relevant for the

ratio measurements, where the statistical uncertainty dominated. The $t\bar{t}\gamma$ cross section increase is about 18%, while for $t\bar{t}$ it is only about 6.7% [150, 179, 180], which means the ratio between $t\bar{t}\gamma$ and $t\bar{t}$ is enhanced, which allows for a even more precise measurement.

A future higher-energy hadron collider, such as the FCC-hh [181], if approved, will represent a large step both in the energy and in the precision frontiers. The FCC-hh is designed to collect 5 to 25 times the luminosity of the LHC, at a centre-of-mass energy of 100 TeV. At 100 TeV, the $t\bar{t}\gamma$ cross section is expected to be about 55 times larger than at 13 TeV, as shown in figure 3.46 (left). The $t\bar{t}$ cross section is expected to increase about 30 times, as shown in figure 3.46 (right), so the ratio will almost double, in case the SM predictions still hold at those energies.

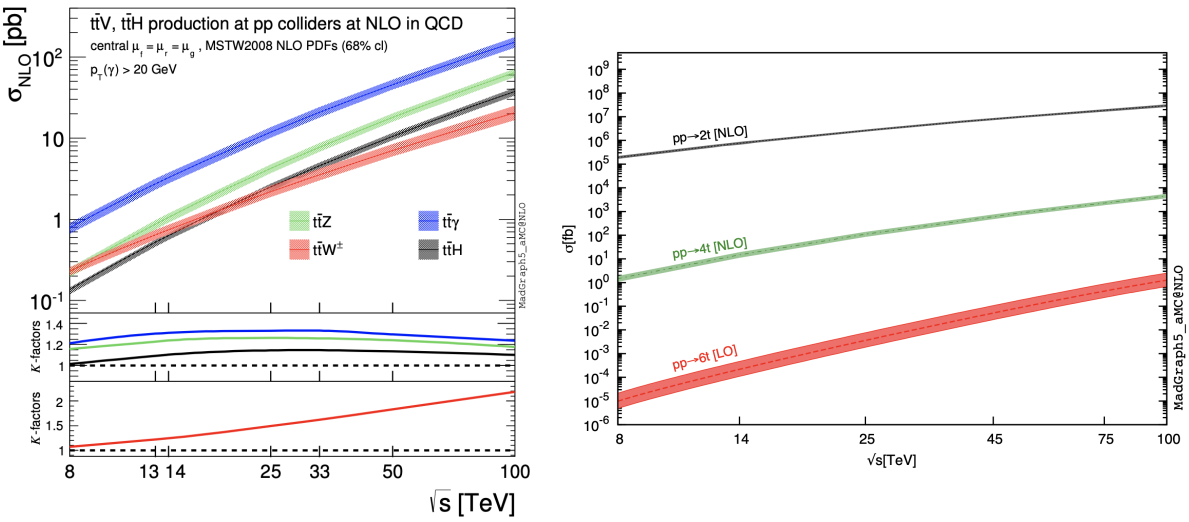

Figure 3.46: Left: NLO total cross sections for $t\bar{t}V$ processes from 8 to 100 TeV [179]; Right: Multitop production cross sections from 8 to 100 TeV [180]. The simulations were carried out using MadGraph 5 MadGraph5 aMC@nlo and the error bands reflect the scale and PDF uncertainties.

Figure 3.47, shows two different scenarios for the improvement of an inclusive $t\bar{t}\gamma/t\bar{t}$ ratio measurement at the HL-LHC and at the FCC-hh: a pessimistic one where the systematic uncertainties are the same as the ones measured in Run 2, and a more realistic one where they are reduced by 30%. The first scenario is unlikely, as significant detector upgrades will be carried out, which will improve the precision of the event reconstruction, and also because our understanding of theoretical aspects such as the PDFs improves as more data is collected.

A number of assumptions has to be made, in order to make this calculation. First of all, since no projections are available for all the other processes that are not $t\bar{t}$ or $t\bar{t}\gamma$, an assumption was made that their cross section scales in the same way as the $t\bar{t}$ cross section.

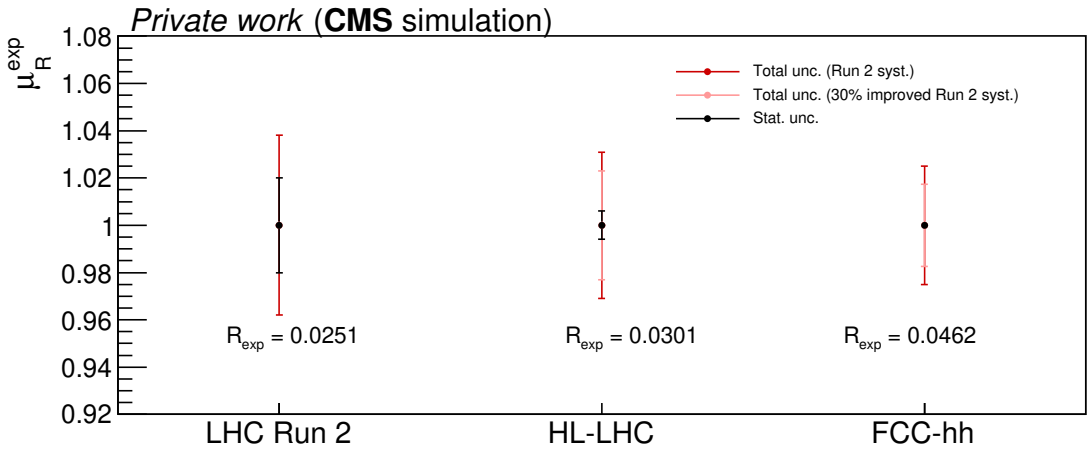

Figure 3.47: Projections for the inclusive $t\bar{t}\gamma/t\bar{t}$ ratio measurement, for the HL-LHC and the FCC-hh. The y-axis represents the expected ratio parameter μ_R, as defined in section 3.9. The black lines represent the statistical uncertainty, the red line represents the systematic uncertainties with their current magnitude, and the pink lines represent the systematic uncertainties in case they get reduced by 30% in future experiments.

This is a gross oversimplification, but since these processes have a minor contribution, it is not expected to bias the results significantly. Additionally, the predictions for the HL-LHC and the FCC-hh are made based on simulated data from CMS, simply by increasing the luminosity and the cross section. This is not accurate, since those will be very different machines and detectors, so dedicated simulations will have to be made in order to get accurate predictions. It is, however, enough to have a rough idea of the order of magnitude of the improvements that can be expected.

The top quark charge asymmetry measurement, which is currently heavily limited by the statistical uncertainty, will also benefit from the higher amounts of data collected in the HL-LHC and, potentially, the FCC-hh [181].

CHAPTER 4

Search for exclusive production of top quark pairs

In this chapter, I describe a search for CEP $t\bar{t}$, the process which was introduced in section 1.3.6. While an observation is only expected to become possible at the HL-LHC [182], the LHC Run-2 data provides the opportunity for the first investigations. I have conducted a search using pp collision data collected by the CMS and TOTEM experiments at the LHC in 2017, including intact proton data from PPS. These data correspond to an integrated luminosity of $29.4\,\mathrm{fb}^{-1}$. The work is published in Ref. [183].

This chapter is organised as follows: section 4.1 contains the definition of the CEP $t\bar{t}$ signal, as well as the characterisation of the most relevant backgrounds; section 4.2 details the data and simulation samples that were used in the analysis, whereas section 4.4 focuses on the analysis strategy, including the event selection and the background estimation; the reconstruction of the signal event hypothesis is described in section 4.5; the multivariate analysis technique used for background discrimination is explored in section 4.6; the statistical analysis and the results are presented in sections 4.7 and 4.8, respectively.

4.1 Signal and background definitions

The signal process, CEP $t\bar{t}$, is defined in detail in section 1.3.6. For the purpose of this analysis, only the QED-induced diagram, figure 1.18 (left), is simulated, since the contribution of the QCD diagram is expected to be negligible in comparison. Furthermore, the analysis focuses on events where the $t\bar{t}$ system decays dileptonically or to one lepton and jets (ℓ+jets channel). The word "leptons", in the following, always refers to electrons and muons, e and μ, respectively. Final states where the electron or muon appear in the decay of tau leptons from W bosons are also considered as part of the signal.

Figure 4.1 shows an event display of an event recorded by CMS in 2017, containing a typical CEP $t\bar{t}$ event candidate, from the ℓ + jets channel.

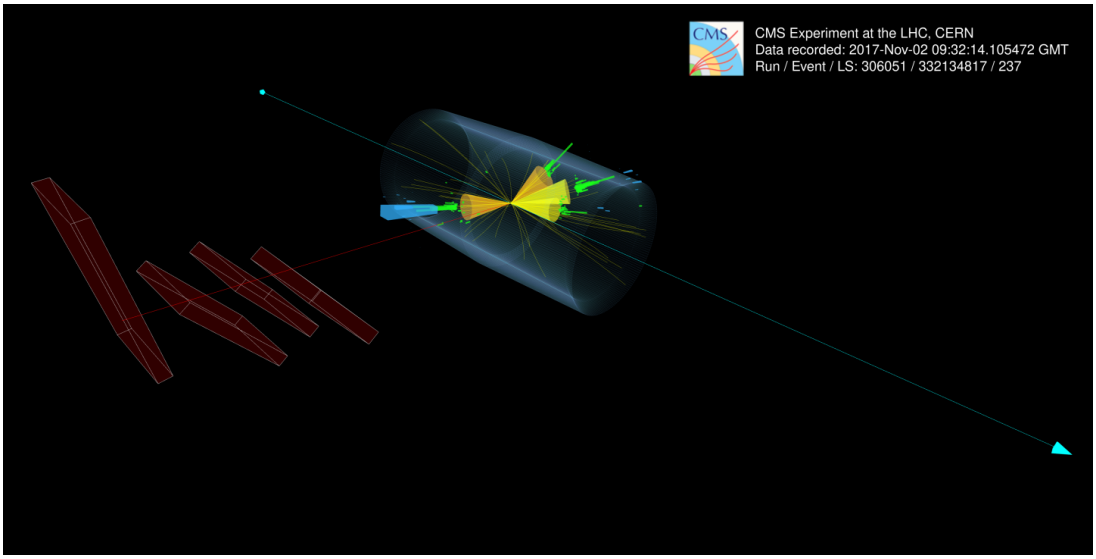

Figure 4.1: Event display of a possible CEP $t\bar{t}$ candidate, recorded by the CMS experiment in 2017 [184]. The outgoing cyan lines represent the outgoing protons, detected in PPS. The orange cones represent two b-tagged jets. The yellow cones represent two reconstructed light flavour jets from the decay of a W boson, while the red lines represents a reconstructed muon. The green and blue areas represent energy deposits in the calorimeters. The yellow lines show particles identified as coming from pileup events.

This thesis focuses mainly on the dilepton channel, and on the combination with the ℓ + jets channel. The full analysis setup for the ℓ + jets final state will however not be described, as that analysis was carried out by a different team. The dilepton channel, though having the lowest branching ratio, provides very high purity in $t\bar{t}$ and a clean signature for the CEP signal. In particular, the CEP signal in the dilepton channel is characterised by very low hadronic activities: the only jets expected in the event are those originating in b quarks from the top decays.

In the dilepton channel, the signal is characterised by a dileptonic $t\bar{t}$ decay signature in the central CMS detector, and two intact protons to be detected in PPS. The kinematics of the two intact protons is strongly correlated with that of the $t\bar{t}$ system.

The largest contribution to the background, by far, is non-exclusive production of $t\bar{t}$, which we will denote QCD $t\bar{t}$. This process, as detailed in chapter 1, occurs abundantly at the LHC, with a cross section about a million times larger than that of the signal process, of 88.29 pb in the dilepton final state. The QCD $t\bar{t}$ signature is identical to that of the signal process except for the fact that intact protons in PPS should in principle not be present for QCD $t\bar{t}$ events. However, PPS has several sources of background protons,

such as protons coming from pileup and from the beam halo, and there is a probability that two proton tracks are reconstructed accidentally, coincidentally with a QCD $t\bar{t}$ event in CMS. In addition, processes that are a background to QCD $t\bar{t}$ such as production of Z bosons (Z+jets) and single top quarks, contribute to the backgrounds, when combined with random proton tracks.

4.2 Data and simulated samples

The analysis is based on data collected by the CMS and PPS detectors in 2017, during the LHC Run 2. Over the course of this year, the strip and pixel detectors in PPS were not fully operational at all times. Only runs where all PPS detectors were operational are considered in the analysis: this results in a sample corresponding to an integrated luminosity of $29.4\,\text{fb}^{-1}$ [185]. The beam crossing angle at the IP, α_X, defined as the angle between the LHC axis and one of the beams, was set at different values throughout the year, with most data being recorded at $\alpha_\text{X} = 120$, 130, 140, or 150 μrad. The remaining data are not included in this analysis.

Similarly to the $t\bar{t}\gamma$ analysis, a combination of single and double lepton triggers is used. Events must pass one of several trigger paths requiring the presence of either a single electron (isolated muon) with p_T greater than 35 (24) GeV, or two electrons (muons) with p_T greater than 23 and 12 (17 and 8) GeV, or an electron and a muon with p_T greater than 12 and 23 GeV, or 23 and 8 GeV, respectively. Appendix C contains the full list of trigger paths. In addition, depending on which HLT requirements are fulfilled for each event, a procedure is implemented to make sure each event is selected only once and avoid double counting.

To simulate the signal and background processes, different MC event generators are used. For all processes, the response of the central CMS detector is simulated using the GEANT4 package [37] as described in section 1.2.2.

A signal $pp \to p\gamma\gamma p \to pt\bar{t}p$ sample is produced at LO in QED, using FPMC [186] as the ME generator, with the equivalent photon approximation for the photon flux [187] and a proton survival probability of 0.9. Events are generated with fractional momentum loss $0.02 < \xi < 0.20$. Top quark decays are simulated with MADSPIN [146]. The PPS response is simulated using a fast forward-proton simulation which includes beam divergence and vertex smearing at the IP, along with their dependence on α_X [40]. The outgoing protons are propagated through the beamline from the IP to the RPs. The aperture limitations for a given crossing angle are included in the simulation, and hits in the detectors are generated taking into account sensor acceptance and resolution. The simulated hits are then used to reconstruct proton tracks using the PPS reconstruction algorithms, as described in section 2.3.2.

The QCD $t\bar{t}$ sample is simulated at NLO precision using the POWHEG (v2.0) [147–149] event generator. The inclusive $t\bar{t}$ production cross section is scaled before the fit to the

best available theoretical prediction at NNLO in QCD, amounting to 832 pb [150]. For the QCD $t\bar{t}$ processes, the p_T spectra of top quarks in simulated samples are reweighted according to predictions at NNLO QCD accuracy [77]. For both signal and background event generation, a top quark mass of 172.5 GeV is used. Other processes with small contributions are generated at NLO or LO, depending on the sample.

For all processes, events emerging from the ME simulation are interfaced with PYTHIA 8.2 [142] in order to include particles from parton shower and hadronisation, as discussed in section 1.2.2. All simulated background samples are listed in table 4.1, with the respective assumed cross sections. In the table, "DY+jets" refers to Drell-Yan.

Process	Truncated sample name	Cross section [pb]
W+jets	W[2,3,4]JetsToLNu_TuneCP5_13TeV	61526.7
$t\bar{t}$ 1ℓ	TTToSemiLeptonic_TuneCP5_13TeV	365.34
$t\bar{t}$ 2ℓ	TTTo2L2Nu_TuneCP5_13TeV	88.29
tW	ST_tW_top_5f_NoFullyHadronicDecays_TuneCP5_13TeV	35.85
$\bar{t}W$	ST_tW_antitop_5f_NoFullyHadronicDecays_TuneCP5_13TeV	35.85
DY+jets	DY[2,3,4]JetsToLL_M-50_TuneCP5_13TeV	460 (total)
WW	WW_TuneCP5_13TeV-pythia8	75.88
WZ	WZ_TuneCP5_13TeV-pythia8	27.57

Table 4.1: Samples and production cross sections of the considered backgrounds.

4.3 Selection of physics objects

After the HLT requirements, a set of additional offline criteria are applied to the reconstructed physics objects. Many of these requirements are the same as applied in section 3.3, so only the differences with respect to the $t\bar{t}\gamma$ analysis are highlighted here.

The charged leptons coming from the $t\bar{t}$ decay are identified with a cut-based identification approach. The only difference is that muons must pass the "tight" identification criteria. This means that in addition to the aforementioned requirements in section 3.3.1, the muon Global track must be matched to segments in at least two muon stations, and must be compatible with coming from the PV. The muon reconstruction efficiency with this algorithm is about 96% in both data and simulation, with a background efficiency lower than 4%. After the identification requirements, leptons are required to be reconstructed with $p_T > 20$ GeV and $|\eta| < 2.4$.

Jets are reconstructed using the PF algorithm introduced in section 2.3.2. They are required to have $p_T > 30$ GeV and $|\eta| < 2.4$, and to be separated from leptons by $\Delta R > 0.4$. The b tagging is done using the DEEPCSV algorithm [188], unlike in the $t\bar{t}\gamma$ analysis, which used DEEPJET. The DEEPCSV algorithm was developed before DEEPJET, and has

a slightly worse performance, as shown in figure 4.2. Jets from b quarks are selected with an efficiency of about 70% for correctly identified b quark jets, and a misidentification probability of 12 (1)% for c quark (light-flavour or gluon) jets.

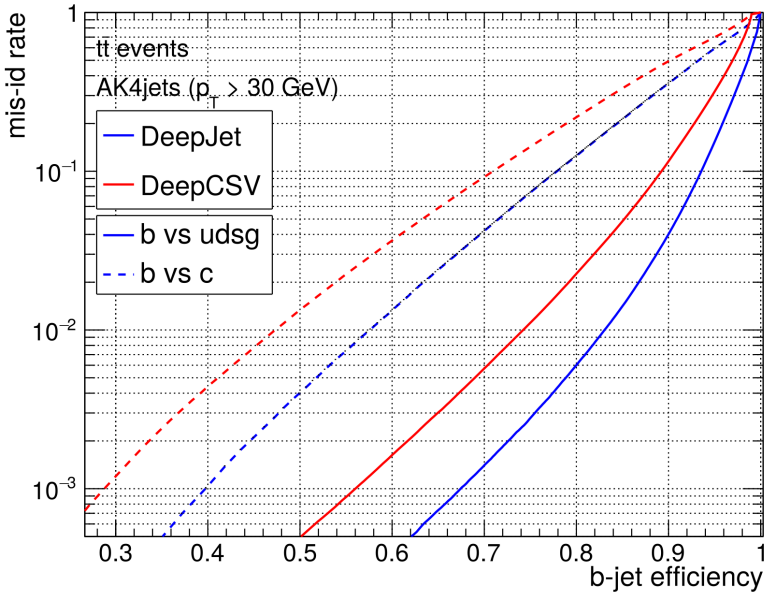

Figure 4.2: Performance of the DEEPJET and DEEPCSV b-tagging algorithms on $t\bar{t}$ simulated events with both top quarks decaying hadronically. The jets are required to have $p_T > 30\,\text{GeV}$. The performance is shown for both b vs. c classification (dashed lines), and b vs. light (solid lines). Taken from Ref. [153].

Protons are selected using the multiRP algorithm described in section 2.3.2.

4.4 Event selection

The final selection requires the presence of at least two charged leptons. The lepton with the highest p_T must have $p_T > 30\,\text{GeV}$ and $|\eta| < 2.1$; the lepton with the second highest value of p_T must have $p_T > 20\,\text{GeV}$ and $|\eta| < 2.4$. The two leptons are required to have opposite charge, and the dilepton system they form is required to have an invariant mass $m_{\ell\ell} > 20\,\text{GeV}$. For the events with two reconstructed leptons of the same flavour, the invariant mass of the dilepton system is further required to be outside a 30 GeV window around the Z boson mass peak: $(m_{\ell\ell} < 76\,\text{GeV}) \cup (m_{\ell\ell} > 106\,\text{GeV})$, in order to reject events from DY production. Events are categorised, according to the flavour of the leading leptons, as electron-electron (ee), electron-muon (eμ), or muon-muon ($\mu\mu$).

Only events with at least two b-tagged jets are retained. Energy corrections are applied to the jets as described in section 2.3.2. The simulated samples are reweighted to match the selection efficiency in data, and different weights are applied to account for the different efficiencies in selecting electrons, muons, and b jets, as described in the previous chapter.

The events are further required to have exactly one proton object reconstructed in each arm of PPS. The requirement of exactly one proton per arm introduces a reduction in selection efficiency of about 90% for the signal process. This is explained by two reasons: the first is that one or both signal protons may fall out of the acceptance of PPS or not be correctly reconstructed by the multiRP algorithm; the other is that additional proton tracks may be present due to pileup. Nevertheless, this cut is necessary because in 2017, the PPS detector was equipped with both pixel and strip detectors, and the strip detectors only allow for one proton track per event. Having multiple tracks would mean using only the pixels, with the singleRP method, and in that case the resolution of the reconstruction would be significantly worse, as shown in section 2.3.2.

4.4.1 Pileup reweighting

In order to match the pileup conditions for simulated events to those in collision data, a further reweighting procedure is applied to simulated events, based on the number of reconstructed interaction vertices (n_{vtx}). The distribution of n_{vtx} varies significantly with the data-taking conditions, therefore it is determined for the data separately for each of the 20 (era, α_{X}) regions introduced in the next section, $P^{\text{data}}(n_{\text{vtx}} \mid \text{era}, \alpha_{\text{X}})$. For MC, one single distribution is determined, $P^{\text{MC}}(n_{\text{vtx}})$. A weight

$$w_{\text{PU}} = P^{\text{data}}(n_{\text{vtx}} \mid \text{era}, \alpha_{\text{X}}) / P^{\text{MC}}(n_{\text{vtx}})$$

is assigned depending on the sampled region, and applied to simulation.

4.4.2 Estimation of background from pileup protons

As mentioned earlier, there is the possibility that some protons which originate from pileup reach PPS. These are not related to the PV where the $\text{t}\bar{\text{t}}$ system is reconstructed. The probability to have at least one proton in the acceptance of one of the arms of PPS, for any bunch crossing, ranges between about 40 and 70%, depending on the LHC optics settings and instantaneous luminosity [40]. However, while the pileup activity in the central detector can be modelled with adequate accuracy, no simulation has been validated so far to include additional protons from uncorrelated diffractive events. Consequently, in the MC samples, background events contain no forward protons, while signal events contain exactly two forward protons on opposite sides (though not necessarily within the acceptance).

The presence of pileup protons, uncorrelated with the event reconstructed in the central detector, has two effects:

- a background event may be selected as signal candidate because exactly one random proton per arm has been reconstructed in PPS;
- a signal event may be rejected because of the multi-proton reconstruction inefficiency, or it may be wrongly reconstructed because a pileup proton is selected instead of the signal one that went undetected as a result of detector inefficiency or limited acceptance.

The combinatorial background is modelled by appending pileup protons obtained from a pool of data events to the MC sample. The pool consists of events selected similarly to the signal region events except for the b jet multiplicity requirement, which is omitted. Figure 4.3 displays a flowchart of the this procedure.

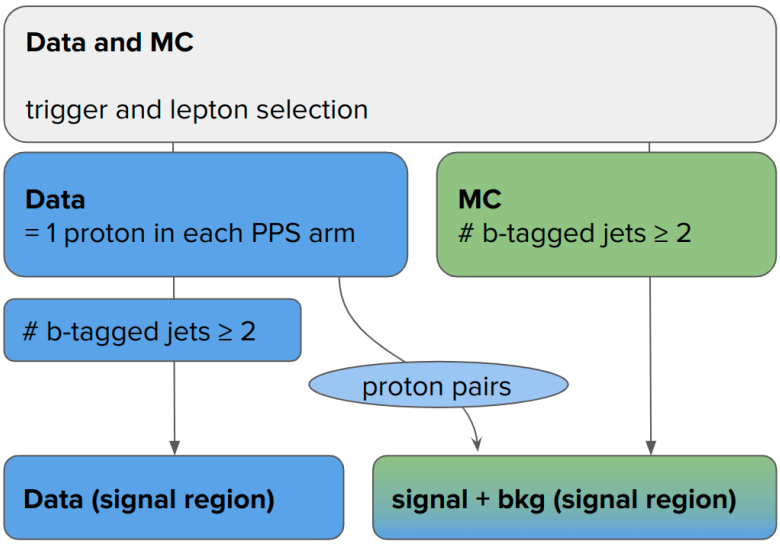

Figure 4.3: Flowchart depicting the background mixing procedure. The signal and background templates are obtained by merging the MC samples with pileup protons from data.

The acceptance of PPS changed between different data-taking periods and beam conditions. Consequently, the proton ξ distributions are different for each of these regions in data. Therefore, the pool is divided into 20 orthogonal regions: five categories corresponding to different data-taking periods (eras B, C, D, E, or F) and four categories corresponding to different α_X values: 120 μrad, 130 μrad, 140 μrad, and 150 μrad. Then,

for every MC event that passes the final event selection on the central CMS detector, a pair of protons and a corresponding region in era and α_X are randomly sampled from the proton pool and are assigned to that event.

The algorithm is executed in two steps:

Step 1: using the data events, the following quantities are computed:

- $P(\text{era}, \alpha_X)$: the probability for an event to be found in a certain era and with a certain crossing angle (α_X):
- $P(11 \mid \text{era}, \alpha_X)$: the probability to find an event with two reconstructed protons (one in each PPS arm) for a given era and crossing angle.
- $P(10 \mid \text{era}, \alpha_X)$ and $P(01 \mid \text{era}, \alpha_X)$: the probability to find an event with 1 proton in a given arm and zero protons in the second arm (and *vice versa*) for the selected (era, α_X) region.
- $P(00 \mid \text{era}, \alpha_X)$: the probability to find an event with 0 protons in both PPS arms for a given (era, α_X) region.

Step 2: for each simulated event, a pair of protons from the proton pool is randomly sampled (using the cumulative distribution function obtained from $P(\text{era}, \alpha_X)$). Besides the protons, an era and α_X are also assigned to the MC event. The background and the signal samples are treated differently, in the following way:

- Background: a pair of protons is assigned to every MC event, and the MC event weight is multiplied by the proton tagging probability: $\omega \to \omega \times P(11 \mid \text{era}, \alpha_X)$.
- Signal: events are separated into 4 categories based on the event topology. Events with 2 reconstructed protons, only 1 reconstructed proton in the positive arm, 1 reconstructed proton in the negative arm, and events with no reconstructed protons. Then, each proton is randomly kept or discarded, to account for the limited reconstruction efficiency of PPS. This is done by extracting a set of SFs from data, and then drawing a random number r between 0 and 1: if $r <$SF the proton is kept, otherwise it is discarded. This will result in a new categorisation of signal events. Then depending on the final event topology, the following weights are applied:
  - Signal events with two reconstructed protons are multiplied by the zero pileup proton probability: $\omega \to \omega \times P(00)$.
  - Signal events with one reconstructed proton: a pileup proton is added to the opposite arm. Then, the event weight is multiplied by the probability to measure zero pileup protons in the reconstructed arm and exactly one proton in the other arm.

– Signal events with no reconstructed protons are treated as background events.

In figure 4.4 the per-event probability of having no pileup protons, as well as the probability of having one proton in each arm are shown as a function of era and beam crossing angle, for data events collected in 2017.

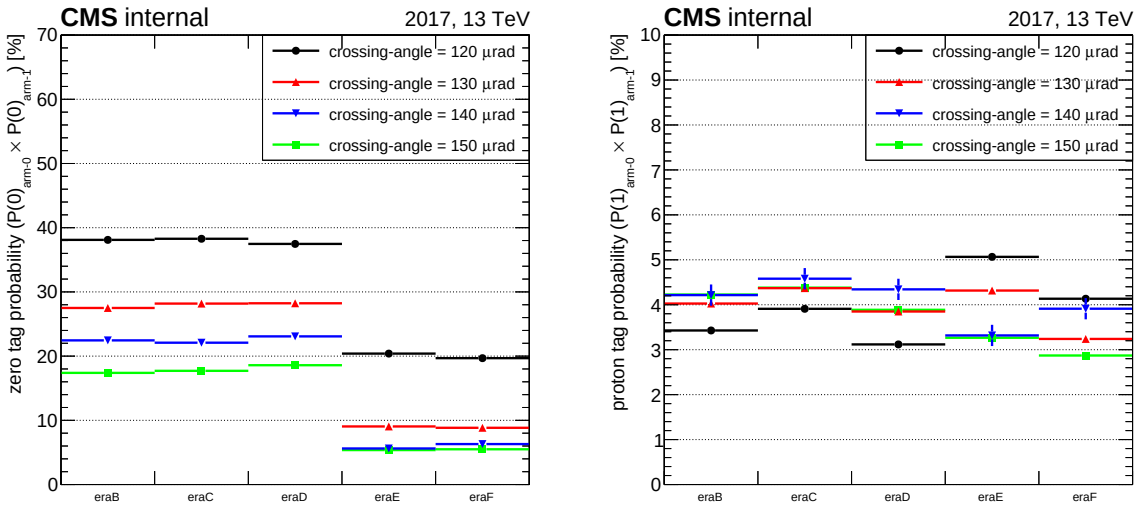

Figure 4.4: Combined probability of finding zero ($P\,(00 \mid \text{era}, \alpha_X)$, left) or 1 ($P\,(11 \mid \text{era}, \alpha_X)$, right) reconstructed protons in each PPS arm, computed from data. The difference between eras B-D and E-F is due to the replacement of some PPS detector constituents during a technical stop of the LHC.

4.4.3 Comparison between data and simulation

To assess the validity of the background model obtained in section 4.4.2, the data and simulation are compared for several distributions. Figure 4.5 shows the number of data and simulation events per (era, α_X) region. Figure 4.6 shows the distributions of ξ in arm 0 and 1, denoted as ξ_0 and ξ_1, after applying the background model. The signal peaks towards lower values of ξ_0 and ξ_1, since those are the values corresponding to the bulk of the $m_{t\bar{t}}$ distribution. For the backgrounds, ξ_0 and ξ_1 have almost uniform distributions, only modelled by the PPS detector acceptance, since they correspond to pileup protons.

Additionally, figure 4.7 shows the p_T distribution of the leading b-tagged jet (left) and the number of light-flavour jets (right). The number of light jets is typically zero or very low for the signal, since there is only the $t\bar{t}$ system being produced and no proton remnants, while for QCD $t\bar{t}$ and other backgrounds, a higher number of jets is expected. Distributions of estimated backgrounds follows well those of data.

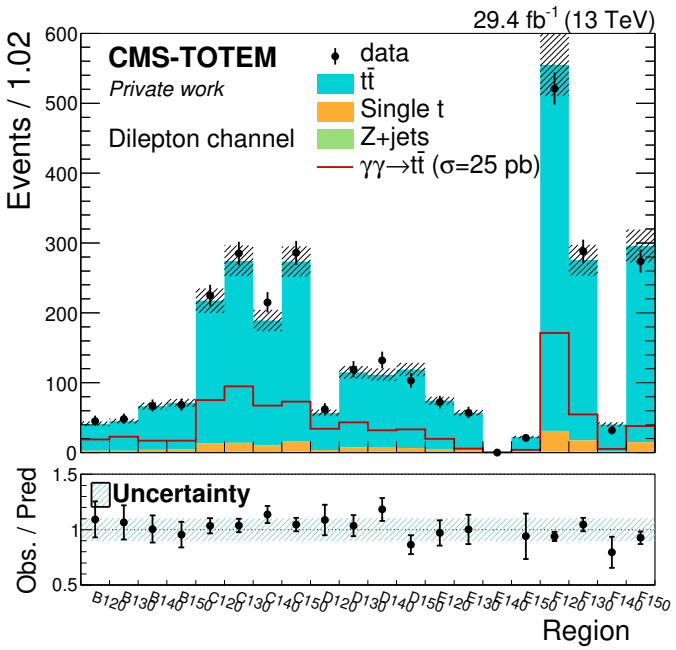

Figure 4.5: Distribution of the number of observed and simulated events per (era, α_X) region. Region 140E has no events after selection, because the PPS was only fully functional for a small fraction of this data-taking period. The signal is scaled to a cross-section of 25 pb in order to be visible.

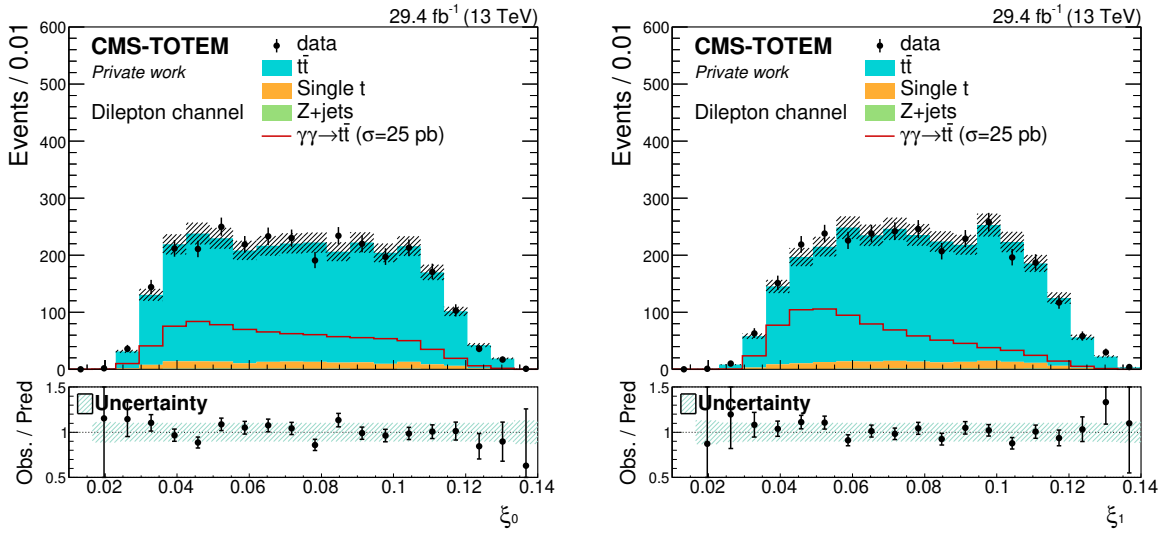

Figure 4.6: Distribution of the ξ_0 and ξ_1 of the reconstructed protons. The uncertainty band includes both statistical and systematic uncertainties (pre-fit). The signal is scaled to a cross-section of 25 pb in order to be visible.

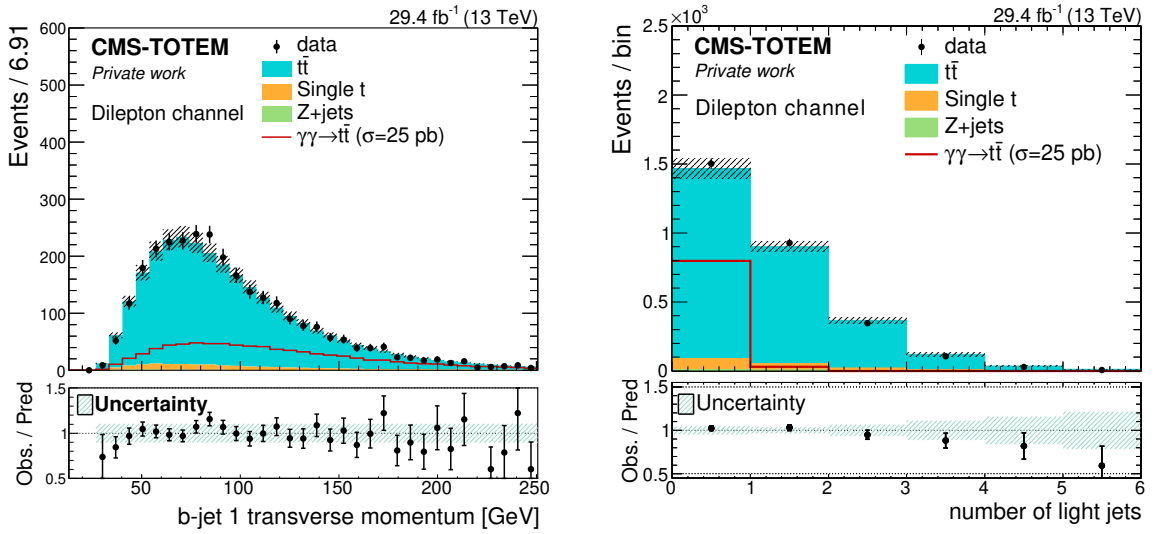

Figure 4.7: Distribution of the p_{T} of the leading b-tagged jet (left) and the number of light-flavour jets (right). The uncertainty band includes both statistical and systematic uncertainties (pre-fit). The signal is scaled to a cross-section of 25 pb in order to be visible.

4.5 Reconstruction of the $t\bar{t}$ system

In PPS, when looking at CEP events, if at least one proton is reconstructed on each arm of CMS, one can extract ξ_0 and ξ_1 using the method described in section 2.3.2 and use them to reconstruct the invariant mass and rapidity of the system X (m_X and y_X). The m_X and y_X are related to ξ_0 and ξ_1 by the approximate formulas:

$$m_X = \sqrt{s\xi_0\xi_1} \quad , \tag{4.1}$$

$$y_X = \frac{1}{2}\ln\left(\frac{\xi_0}{\xi_1}\right) \quad , \tag{4.2}$$

where the subscripts 0 and 1 refer to the arm index. A comparison between data and simulation for these quantities is shown in figure 4.8. In case of CEP $t\bar{t}$ production, $m_{t\bar{t}}$ and $y_{t\bar{t}}$ should match m_X and y_X, while for QCD $t\bar{t}$ background there should be no correlation. This is a powerful way to disentangle the two contributions, which have an otherwise very similar signature.

Reconstructing the full kinematics of the $t\bar{t}$ system plays a central role for discriminating between signal CEP $t\bar{t}$ and QCD $t\bar{t}$ background. The reconstruction algorithm described in section 3.5 is also used here. The resolution of the $t\bar{t}$ mass reconstructed with this algorithm is defined as in section 3.5 and is shown in figure 4.9 for simulated signal events.

Figure 4.10 show the distributions of the resulting reconstructed $m_{t\bar{t}}$ and $y_{t\bar{t}}$. Good agreement between data and simulation is observed for both variables.

4.6 Multivariate analysis with boosted decision trees

In order to achieve the best possible separation between signal and background, several observables which offer discriminating power are combined by means of a multivariate analysis technique. A boosted decision tree (BDT) algorithm [189] is used, implemented with the TMVA toolkit [190]. The samples used for the training of the model consist of signal events with both protons reconstructed in PPS, and QCD $t\bar{t}$ production events with two pileup protons added from the collision data, as described in section 4.4.2. Other backgrounds are not included in the training, as they are minor in comparison to QCD $t\bar{t}$.

The signal and QCD $t\bar{t}$ background samples used for developing the BDT are a random subset of the MC samples used in the analysis after the full selection, with 50.000 events for training and 50.000 events for testing the classifier, for either the signal and the QCD $t\bar{t}$. Only signal events where both protons from the CEP are reconstructed by PPS are considered for the training, i.e., signal events with zero or one signal protons in total, where the remaining proton(s) come from PU are excluded. This decision was made

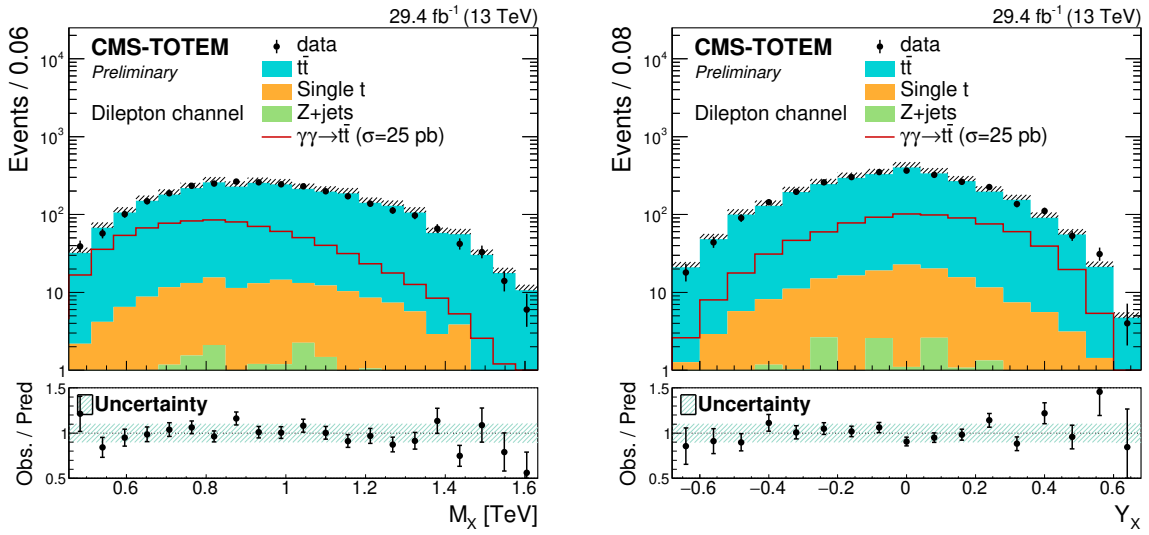

Figure 4.8: Distribution of the mass and rapidity of the central system (m_X and y_X), computed from the two proton ξ values. The uncertainty band includes both statistical and systematic uncertainties (pre-fit). The signal is scaled to a cross-section of 25 pb in order to be visible.

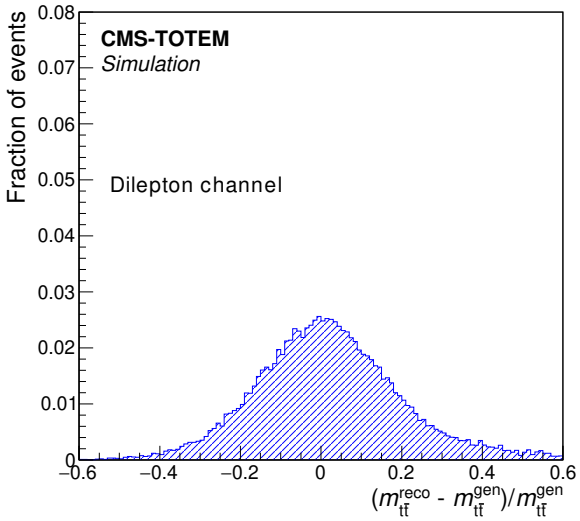

Figure 4.9: Normalised distribution of the relative invariant mass resolution of the reconstructed $t\bar{t}$ system in simulated signal events.

because events with two signal protons are better modelled and are independent of the proton mixing strategy.

Several options were tested for the training parameters, and those giving the best discrimination power without showing "overtraining" were used in the final classification. The chosen parameters are summarised in table 4.2. A complete description of what the parameters represent can be found in Ref. [190].

	Option
NTrees	100
MaxDepth	4
BoostType	AdaBoost
SeparationType	GiniIndex
nCuts	20

Table 4.2: Training parameters of the BDT.

The main purpose of the BDT classifier is to exploit the fact that for signal events a strong correlation between the proton kinematics (m_X and y_X) and the $t\bar{t}$ system ($m_{t\bar{t}}$ and $y_{t\bar{t}}$) is expected, while for the background no correlation should be present. Furthermore, signal events are expected to have less extra activity due to the absence of proton remnants, and therefore observables that quantify this extra activity, such as the number of light jets, are also useful for the classification. Finally, and for the same

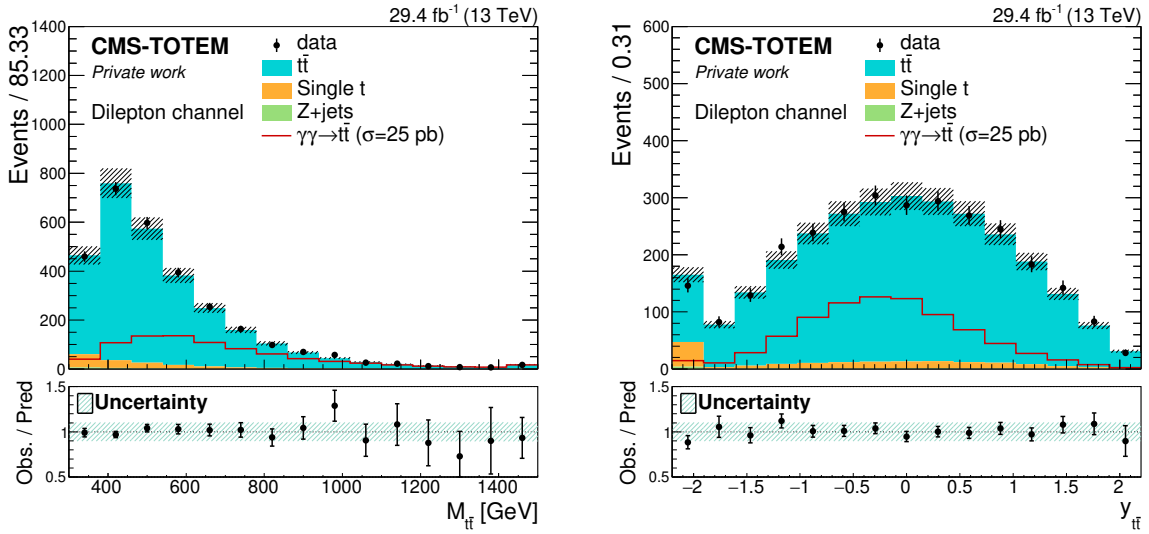

Figure 4.10: Distribution of the mass and rapidity of the reconstructed $t\bar{t}$ system ($m_{t\bar{t}}$ and $y_{t\bar{t}}$). The uncertainty band includes both statistical and systematic uncertainties (pre-fit). The signal is scaled to a cross-section of 25 pb in order to be visible. The first bin on the right hand side distribution shows mostly the events that did not pass the $t\bar{t}$ reconstruction, as those are given a dummy negative value of $y_{t\bar{t}}$.

reason, in the signal events, the top quark and antiquark are expected to be produced more "back-to-back", since there are no other particles for the $t\bar{t}$ system to recoil against. With this logic in mind, the 15 kinematic variables listed in table 4.3 were chosen to be included in the classifier. This list includes some high-level variables that combine information from the full $t\bar{t}$ system and extra event activity.

Variable	Definition (where needed)	Disc. power
y_{vis}	Rapidity of the vector-sum of the four-momenta of the visible particles from $t\bar{t}$ decay: $y(b_1 + b_2 + l_1 + l_2)$	0.149
m_X	-	0.123
y_X	-	0.122
number of light-flavour jets	-	0.114
$m_{t\bar{t}}$	Output of the $t\bar{t}$ reconstruction	0.099
E^2	$(E(\text{b}_1) + E(\text{b}_2) + E(\ell_1) + E(\ell_2) + p_{\text{T}}^{\text{miss}})^2$	0.077
$y_{\text{sum}}^{\text{extra}}$	Sum of the absolute values of rapidity of each visible particle that is not from $t\bar{t}$ decay (extra jets)	0.072
$\Delta R(\ell, \ell)$	-	0.066
y_{sum}	Sum of the absolute values of rapidity of each visible particle from $t\bar{t}$ decay: $\lvert y(\text{b}_1)\rvert + \lvert y(\text{b}_2)\rvert + \lvert y(\ell_1)\rvert + \lvert y(\ell_2)\rvert$	0.037
Δy_{min}	Minimum $(\lvert y(\text{b}_1 + \ell_1)\rvert - \lvert y(\text{b}_2 + \ell_2)\rvert, \lvert y(\text{b}_1 + \ell_2)\rvert - \lvert y(\text{b}_2 + \ell_1)\rvert)$	0.030
$y_{t\bar{t}}$	Output of the $t\bar{t}$ reconstruction	0.029
$\lvert\Delta\phi_{\text{b},\bar{\text{b}}}\rvert$	-	0.029
$p_{\text{T}}^{\text{miss}}$	-	0.020
$m_{\ell\ell}$	-	0.020
y of extra system	Rapidity of the system of additional particles, i.e. those not associated to the $t\bar{t}$ system	0.149

Table 4.3: Training parameters of the BDT, with the respective definitions, where needed. The variables are shown in descending order of discriminating power. The last column shows the discriminating power of each variable, as given by TMVA.

The discriminating power presented in the last column of table 4.3 is provided by the

TMVA package and quantifies the relative importance of each variable in the training. The number is computed for each variable by counting how often the variable is used to split decision tree nodes, and then weighting each split occurrence by the gain in signal purity it has achieved and the number of events in the node. The distribution of all input variables in the BDT training samples, for signal and $t\bar{t}$ background, are shown on figures 4.11 and 4.12, for the training samples. It is clear that the variables which TMVA ranks higher are also the ones where a different spectrum between signal and background is more apparent.

The correlation matrices in figure 4.13 show the linear correlation coefficients (in percentage) between all variables, for signal and for $t\bar{t}$ background.

Looking at the matrices, we can see that, as expected, for the signal (left) there is a high correlation between the forward variables (m_X and y_X, designated *mpp* and *ypp* in Fig. 4.13) and the reconstructed $t\bar{t}$ kinematics, i.e. reconstructed $m_{t\bar{t}}$ and $y_{t\bar{t}}$ (*kinreco_mtt* and *kinreco_ytt* in the figure); while for the background (right), no significant correlations are observed, since the information in PPS is random and has origin in pileup. Some variables, such as $y_{\mathrm{sum}^{\mathrm{extra}}}$, are heavily correlated with other variables that rank higher in terms of discriminating power. The performance of the classifier without those variables was tested and it was shown to be slightly worse. For this reason, the variables were kept.

As a measure of the BDT classifier performance, the ROC curve is presented in figure 4.14. This curve shows the fraction of rejected background events (background rejection, on the y-axis) as a function of the fraction of accepted signal events (signal efficiency, on the x-axis). The larger the area under this curve, the better the performance of the classifier.

The output of the classifier, comparing training and test samples, is shown in figure 4.15. Events which are more "signal-like" are attributed a larger BDT output (closer to 1), while "background-like" events are classified with a low BDT output (closer to -1). The training and test samples show compatible distributions, and there is no clear sign of overtraining. TMVA performs a series of Kolmogorov-Smirnov (K-S) tests comparing the training and test samples and outputs the average result for signal and for background, as shown in the figure. The K-S test is a statistical test which compares the cumulative distributions (in this case, from the training and test distributions), evaluating how well they agree with each other. For random samples (not drawn from the same distribution), the test outputs a uniform distribution between 0 and 1, and therefore the average should be 0.5. In practice, it is difficult to obtain precisely 0.5 with limited number of events, so we consider that there is no overtraining if the test result is of the order of 0.5 and it is visually apparent that the two distributions agree.

The output weights from the BDT model are stored and applied to data and all simulated events, in the ee, $e\mu$ and $\mu\mu$ final states, and the obtained distribution is shown in figure 4.16. The separation between signal and background contributions is visible in the distribution, and this will be used for the statistical analysis described in the next section. It is worth emphasising that the background overwhelmingly large, around 1

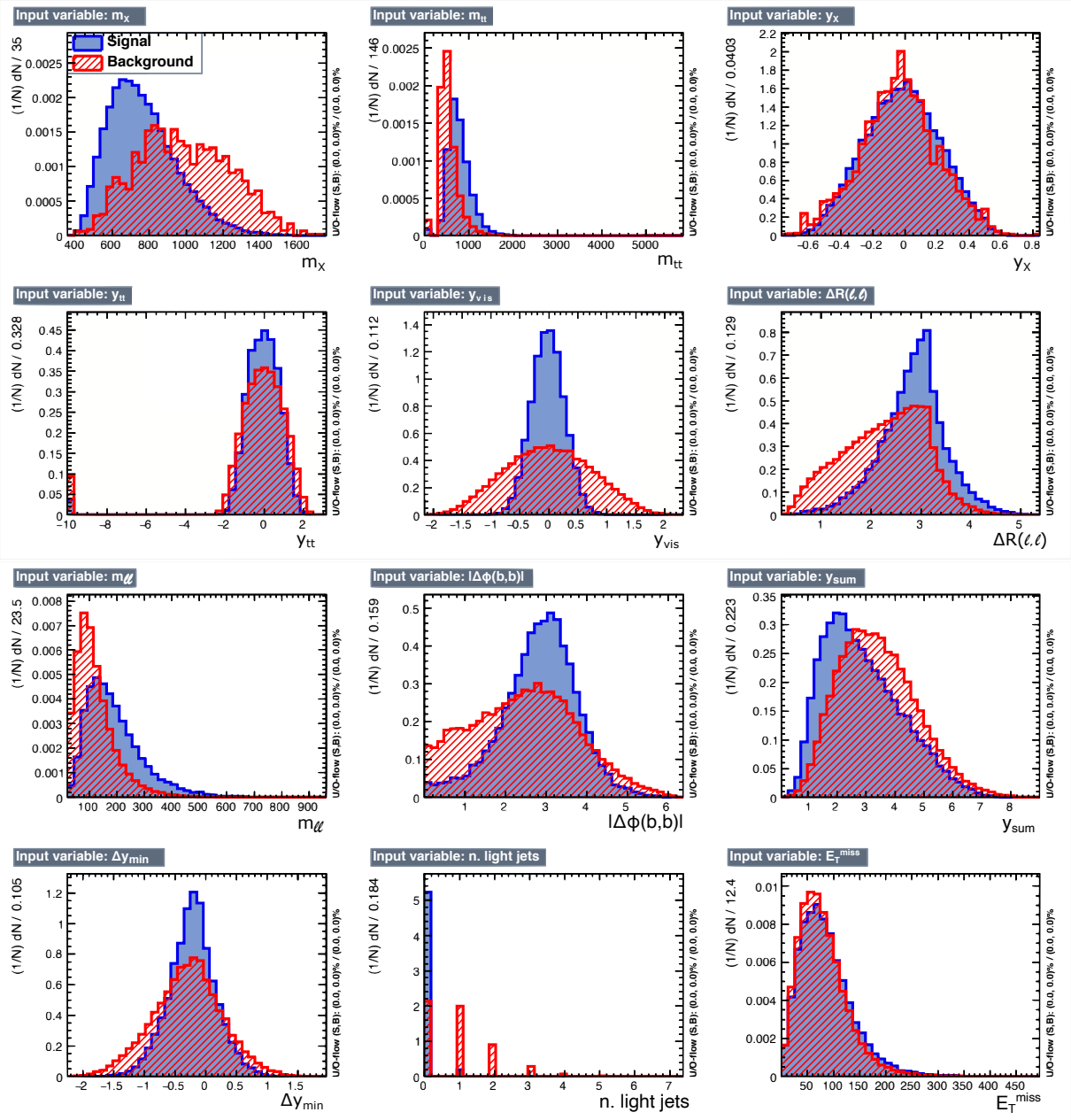

Figure 4.11: Distribution of the first 12 out of 15 input variables in the training sample, for signal (blue) and $t\bar{t}$ background (red).

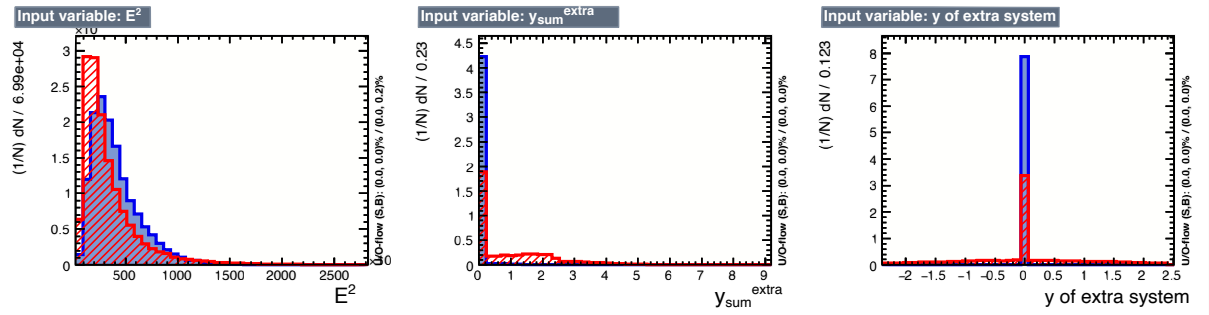

Figure 4.12: Distribution of the last 3 out of 15 input variables in the training sample, for signal (blue) and $t\bar{t}$ background (red).

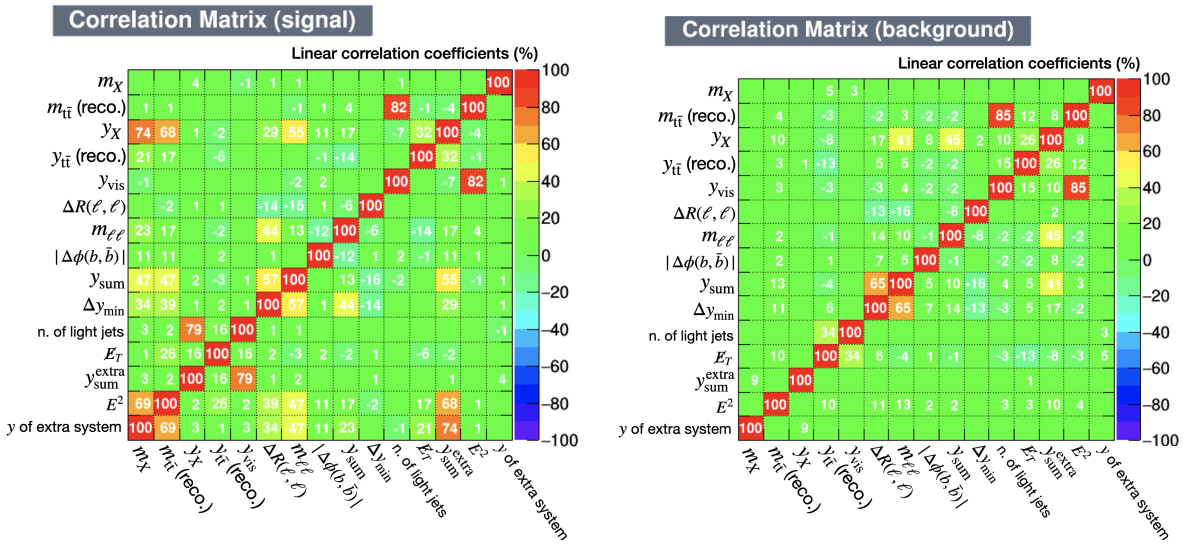

Figure 4.13: Correlation matrices, showing the linear correlation coefficients (in percentage) between all pairs of variables, for signal and background (see text).

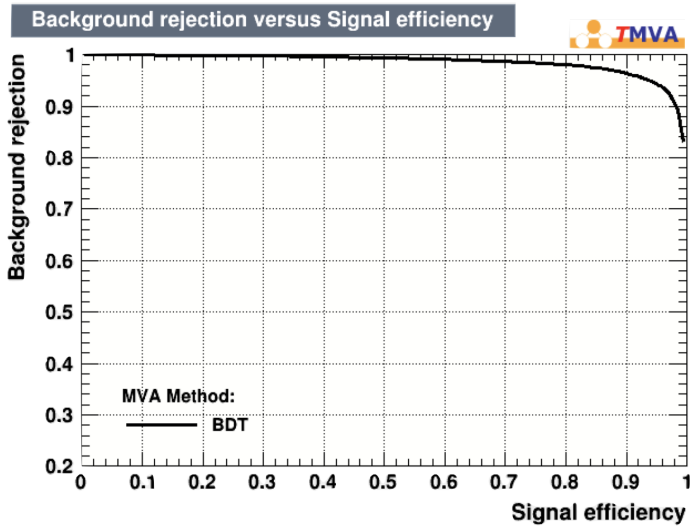

Figure 4.14: ROC curve produced by ROOT TMVA, quantifying the performance of the BDT classifier. The larger the area of this curve, the better the discriminating power.

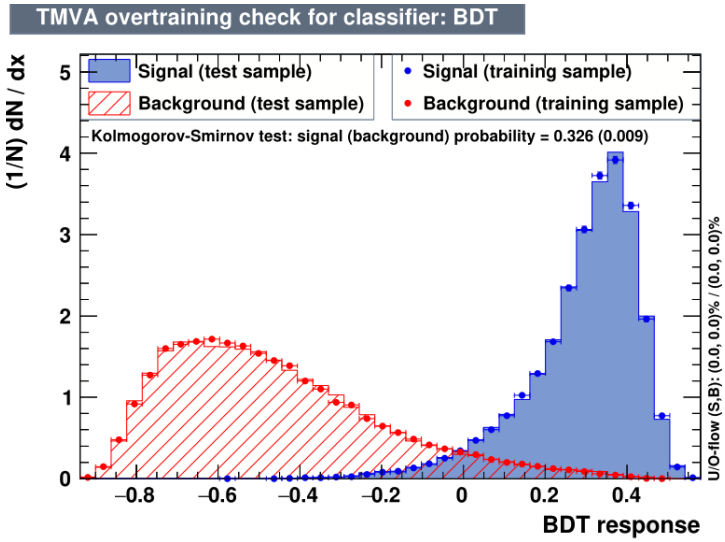

Figure 4.15: Comparison between the response of the BDT classifier for training and test samples. No overtraining is observed.

million times larger than the expected signal, and even in the "signal-like" region, the background events are the dominant contribution. In the figure, the signal contribution is scaled to a cross-section of 25 pb ($\sim 10^4\ \sigma_{\mathrm{SM}}$) in order to be visible.

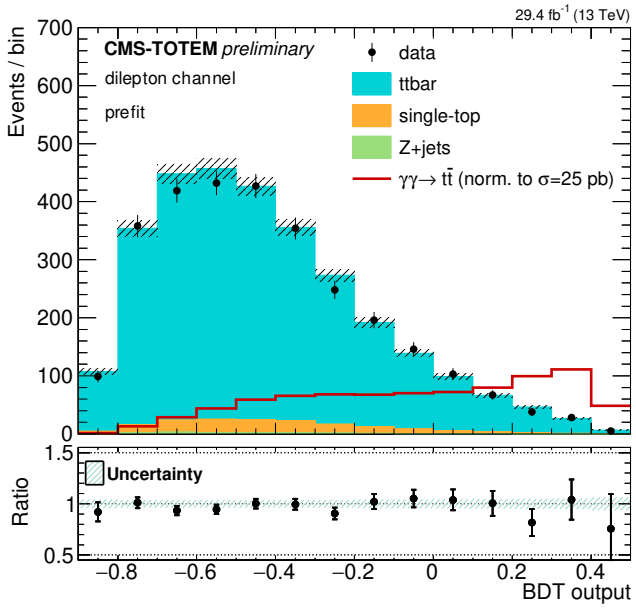

Figure 4.16: Output of the BDT classifier, when applied to the full data and simulation samples. The uncertainty band includes both statistical and systematic uncertainties (pre-fit). The signal is scaled to a cross-section of 25 pb in order to be visible. Signal events cluster in larger BDT output values and background events in lower BDT output values.

4.7 Statistical analysis

In order to extract an upper limit on the cross section of the $\mathrm{pp} \to \mathrm{p}\gamma\gamma\mathrm{p} \to \mathrm{pt}\bar{\mathrm{t}}\mathrm{p}$ process, a simultaneous fit is performed to the BDT outputs from dilepton and ℓ+jets final states. The BDT of the ℓ + jets channel is trained independently and details can be found in Ref. [183].

The same statistical concepts described in section 3.7 are applied here. The POI in this case is the signal strength parameter for the cross section of $\mathrm{pp} \to \mathrm{pt}\bar{\mathrm{t}}\mathrm{p}$, and is defined as:

$$\mu = \frac{\sigma(\mathrm{pp} \to \mathrm{pt}\bar{\mathrm{t}}\mathrm{p})}{\sigma(\mathrm{pp} \to \mathrm{pt}\bar{\mathrm{t}}\mathrm{p})^{\mathrm{SM}}} \quad . \tag{4.3}$$

The likelihood is the one of equation 3.17, except that in this analysis, the normalization of the QCD $t\bar{t}$ backgrounds is assigned a uniform prior with a quasi-infinite range, so that these parameters are directly inferred and constrained from the data. This corresponds to modifying equation 3.15 such that:

$$\lambda_i(\mu, \boldsymbol{\Theta}) = \mu \cdot s_i(\boldsymbol{\Theta}) + \sum_{j=1}^{N_1} b_{i,j}(\boldsymbol{\Theta}) + \sum_{k=1}^{2} \alpha_k \cdot c_{i,k}(\boldsymbol{\Theta}) \tag{4.4}$$

where $j = 1, ..., N_1$ runs over the other (non $t\bar{t}$) background sources, and $k = 1, 2$ runs over the QCD $t\bar{t}$ backgrounds. The normalization of the ℓ + jets and dilepton decay channels are left to float independently. The rate parameter $\alpha_{1(2)}$ scales the normalization of QCD $t\bar{t}$ in the ℓ + jets (dilepton) channel [1]. The MLE method is then used to estimate the POI, μ.

4.7.1 Systematic uncertainties

Most of the systematic uncertainties detailed in chapter 3 are relevant in this analysis, too. Obvious exceptions are the uncertainties that are directly related to the selection of the photon, together with those associated with the estimation of backgrounds. The CEP analysis is in addition affected by systematic uncertainties associated with the identification of intact protons. One has to also consider that the analysis presented in this chapter is based on data collected by CMS in 2017 only. Therefore, considering the correlations between the different data-taking years, as detailed in the previous chapter, is not necessary. This section provides details for uncertainties affecting the CEP analysis only. A summary of all uncertainties, including the process(es) they influence, their effect being on the shape of the BDT output or only the rate of the signal, and whether they are treated as correlated between the two channels, is presented in table 4.4.

Parton shower modelling: Uncertainties in the modelling of ISR and FSR are included. Unlike in the $t\bar{t}\gamma$ analysis, separate uncertainties are considered for the different components of the PS, in particular, different splitting types are considered independently - g $\rightarrow$ gg, g $\rightarrow$ q$\overline{\text{q}}$, q $\rightarrow$ qg, x $\rightarrow$ xg, where q refers to light flavour quarks and x to b or t quarks. The motivation behind splitting this uncertainty in its individual components instead of simply including the overall contribution, was the fact that these uncertainties were strongly constrained by the fit, before splitting. Such a constraint was not expected and is probably due to incorrectly assumed correlations between the sources; when including all sources separately, the constraint is no longer observed. A possible

[1] In the ℓ + jets channel, there is another background contribution, from QCD multijet production, which is not included directly in the fit. Instead, it was shown that the BDT classifier output distribution is identical in QCD multijet and in QCD $t\bar{t}$ events, so the normalization parameter α_1 also accounts for this background.

explanation for the fact that this analysis is very sensitive to ISR and FSR effects, unlike the $t\bar{t}\gamma$ analysis, is that these effects heavily impact several inputs to the BDT classifiers, such as the number of additional jets, and the rapidity of several particles.

Background normalisation: The uncertainties on the cross section of each of the background contributions are included and summarised in table 4.4. The normalisation of the QCD $t\bar{t}$ process in the dilepton and ℓ + jets channels is included in the fit as rate parameters, which are allowed to float freely, as described in section 4.7.

Luminosity: The uncertainty on the integrated luminosity collected in 2017 is considered as a rate uncertainty on all processes [171].

Proton reconstruction: The uncertainty on the reconstructed proton ξ is applied as a shape uncertainty affecting only the signal. This uncertainty is derived by varying the fractional momentum loss ξ of the protons within its uncertainty, as estimated in Ref. [40], and reevaluating the BDT distributions with the modified ξ values. The variation in the reconstructed momentum loss of the protons is evaluated for each of the two arms of PPS independently, resulting in two uncorrelated systematic uncertainties.

PPS efficiency: The uncertainty on the efficiency of the silicon strips and of the multiRP reconstruction method is considered as a shape uncertainty, affecting only the signal.

Pileup proton rates: The background arising from the existence of pileup protons is estimated by enriching the MC samples with pileup protons from the data, as described in section 4.4.2. The simulated events are then normalised according to the pileup proton rate measured in each region, before the requirement of at least 2 b-tagged jets. Any bias in the proton tag rate arising from loosening the b-jet requirement is taken into account by measuring the proton tag rate again after requiring at least 1 b-tagged jet, and taking the difference between the two as a systematic uncertainty. This is a rate uncertainty, affecting the normalization of all processes.

This analysis is heavily limited by its statistical uncertainty, both in data, due to the low number of expected signal events, and in simulation. This is very different from, for example, the precision measurement reported in the previous chapter, where the systematic uncertainties assume a more relevant role. For this reason, the varied shape templates are subject to an additional treatment, in order to stabilise the fit and remove possible unphysical spikes in the distributions caused by the limited number of simulated events. The treatment is summarised in three steps:

- **Shape significance test:** For each uncertainty, a single K-S test is performed between the alternative shapes and the nominal one to check if the alternative

template is compatible with being drawn from the same distribution as the nominal. The shape systematic is retained if the K-S test yields a value lower than 95%, for either of the upwards or downwards fluctuated shape. Otherwise, it means the shape variation is not significant and it is converted to a rate uncertainty, which is then considered in the fit.

- **Smoothing:** For some variations, the presence of some very large or very small weights for particular simulated events was observed, which leads to unphysical shapes with spikes. In order to mitigate this effect, a smoothing procedure was applied to the shape of the systematic variations. The ratio of the varied to the nominal shape (separately for upwards and downwards variations) is smoothed, using the TH1::Smooth(1) method from the ROOT package [191]. The smoothed shapes of variations are obtained by multiplying the nominal template by the smoothed ratio.

- **Pruning:** All systematic uncertainties which are shown to have an impact lower than 0.1% in the final limit are considered to be negligible and are omitted.

The uncertainties after this treatment are used in the fit, and a summary is presented in table 4.5. There, "lnN" refers to log-normal or rate uncertainties, while "shape" refers to shape+rate uncertainties.

4.8 Results

From the selection procedures in the two $t\bar{t}$ decay channels, 23 BDT shape distributions are constructed, each containing a signal-enriched and a background-enriched region. From the dilepton analysis, 3 distributions are obtained: for the e e, e μ, and μ μ final states. From the ℓ + jets analysis, 20 distributions are included, one for each (era,α_X) region. The final results are obtained from a simultaneous fit to all distributions and, given that no statistically significant sign of the signal is observed, they are interpreted with upper limits.

Setting an upper limit is equivalent to quoting a CI containing $\mu = 0$, CI=$[0,\mu_{up}]$, using the notation from section 3.7. Using the test statistics q_μ, it is possible to compute the p-value for a range of values of μ, and reject those with $p_\mu < 0.05$. This method, however, can lead to the exclusion of a signal model in cases where the sensitivity is low and there is a downward fluctuation of the observed data. The CLs prescription [192] is a method to avoid this issue by relying not directly on p_μ, but on the ratio $p_\mu/(1-p_0)$.

The simultaneous fit to the 23 categories, considering all systematic uncertainties as nuisance parameters and using the CLs prescription, yields a combined upper limit on the production cross section of CEP $t\bar{t}$ of

$$\sigma < 0.59\,\text{pb} \quad (1.10^{+1.22}_{-0.55}\ \text{expected @ 95\% confidence level}) \tag{4.5}$$

Figure 4.17 demonstrates the BDT output distribution in data and simulation. In the dilepton channel, the e e, e μ, and μ μ categories are shown together, and in the ℓ + jets channel all (era,α_X) categories are shown together.

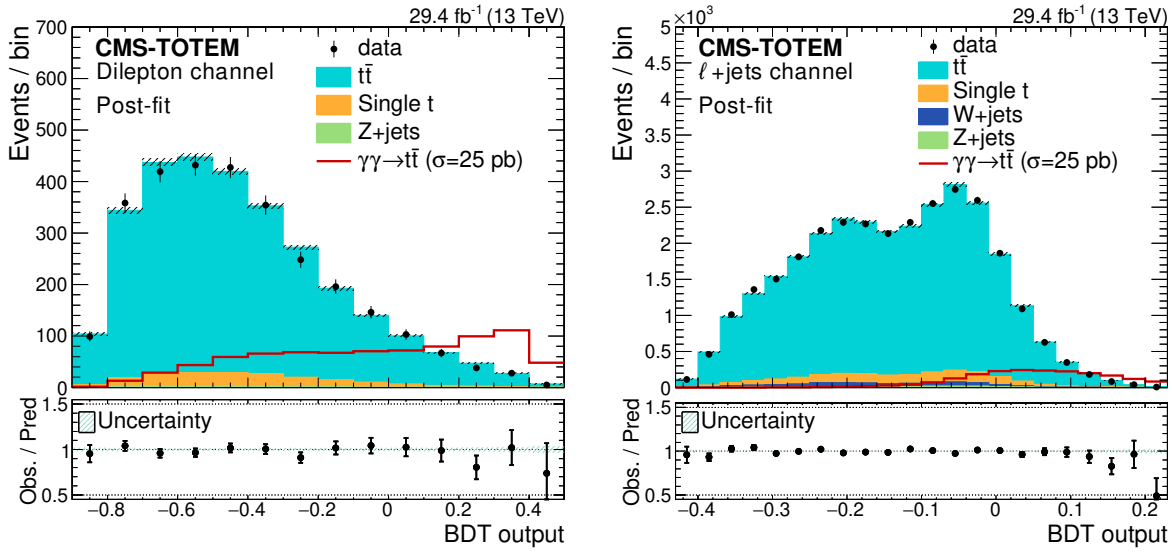

Figure 4.17: Distribution of the BDT score in the signal region for simulated events after the fit, and for data. Left: dilepton channel; right: ℓ + jets channel. The different ranges of the two BDT output distributions are a consequence of the different architectures of the algorithms. The solid histograms show the expected background contributions, while the red open histograms show the expected signal shapes, normalised to a cross section of 25 pb, approximately 10^5 larger than the SM cross section prediction from Ref. [93]; points with statistical error bars represent collision data. For both decay channels, all signal regions are combined. The lower panels show the data-to-prediction ratios; the hatched bands represent the relative uncertainty in the predictions.

Figure 4.18 shows the expected limits obtained by fitting each channel separately, as well as the combined result. The median expected value is shown as a dashed black line, the $\pm1\sigma$ and $\pm2\sigma$ bands are shown in yellow and green, respectively, while the observed value is shown as a solid black line.

The ℓ + jets and the dilepton analyses achieve comparable levels of sensitivity, with slightly better sensitivity for the first. While the ℓ + jets channel benefits from the higher branching ratio, and therefore higher number of events, as well as the less ambiguous reconstruction of the $t\bar{t}$ system (since there is only one neutrino present), the dilepton

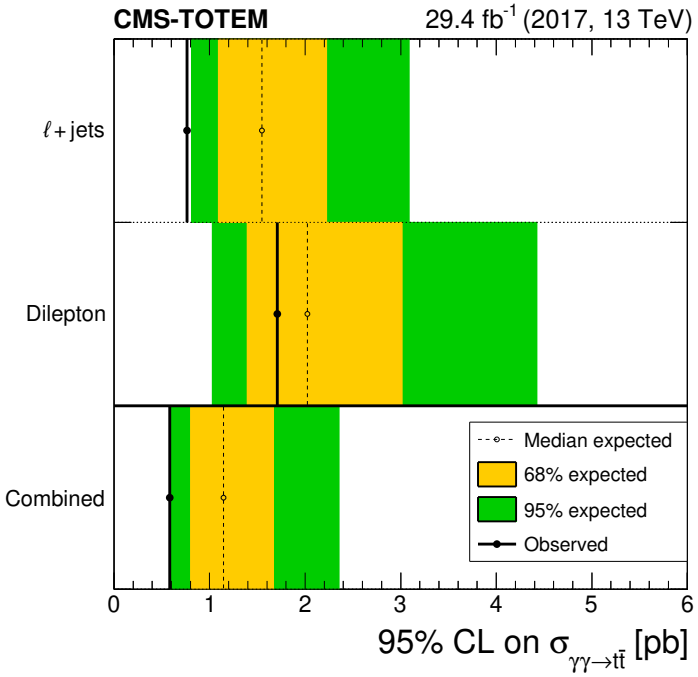

Figure 4.18: Expected upper limits and respective $\pm 1\sigma$ and $\pm 2\sigma$ bands, obtained when fitting the two channels separately, and when combining the two measurements. Observed limit superimposed as a solid black line.

channel benefits from the higher purity in $t\bar{t}$, as well as the low number of jets, that results in lower uncertainties and a better separation performance of the BDT.

Figure 4.19 shows the expected impact of the leading systematic uncertainties, as well as the expected pulls and constraints. These pulls and constraints are computed based on a fit where the expected cross-section value is fixed to 1 pb. The expected impact of the leading uncertainties on the final limit is also showed in percentage on table 4.6.

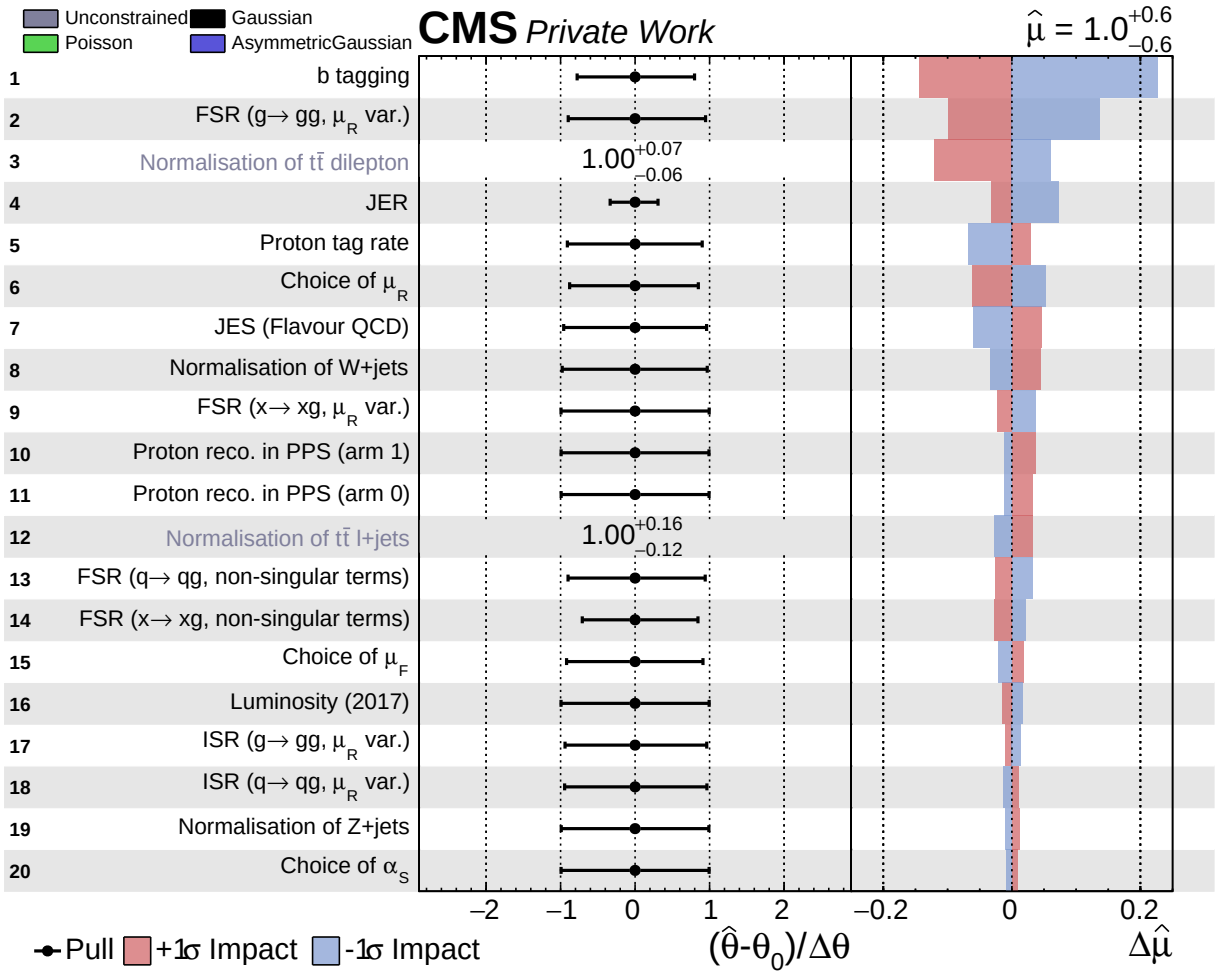

Figure 4.19: Expected impacts on each of the leading systematic uncertainty on the signal strength parameter, r, where the expected value $r = 1$ corresponds to a cross-section scenario of 1 pb.

Figure 4.20 shows the observed pulls and constraints on the signal strength when performing a fit to data. In order to perform this fit, the signal strength is allowed to take negative values. The impacts on the upper limit are shown in 4.6.

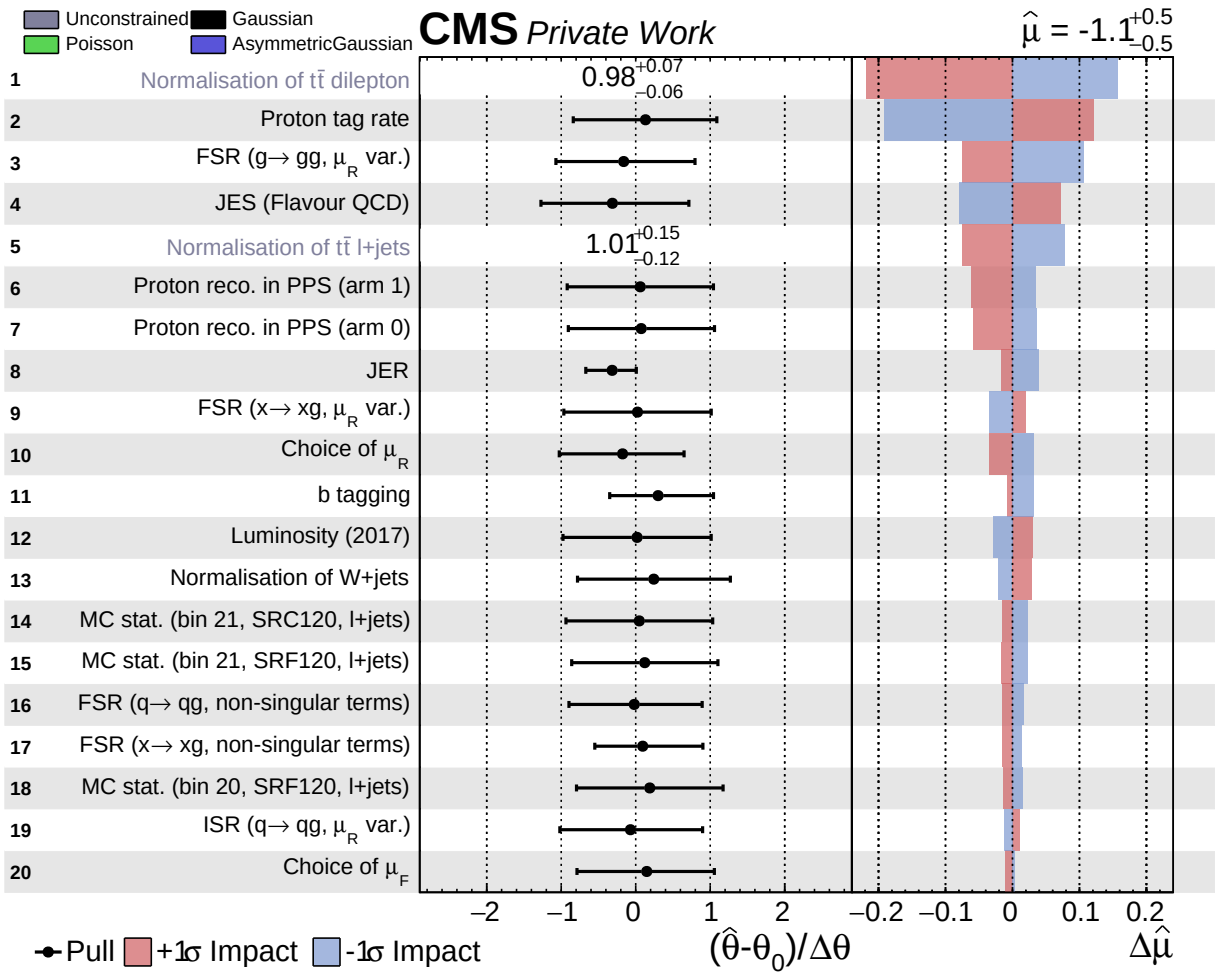

Figure 4.20: Observed impacts on each of the leading systematic uncertainty on the signal strength parameter, r, where the expected value $r = 1$ corresponds to a cross-section scenario of 1 pb.

4.9 Discussion and outlook

In this chapter, data collected by the CMS experiment in 2017 are analysed to measured the first upper limit on the cross section of $pp \to pt\bar{t}p$. This result was made public as an analysis summary by the CMS collaboration [183]. A publication is in preparation to be submitted to JHEP.

The main factor limiting the sensitivity of this measurement is the statistical uncertainty. The data collected by the LHC so far are not sufficient to observe the CEP process, in case the cross section matches the one predicted by the SM, which is $< 1\,\text{fb}$. Nevertheless, the systematic uncertainties also play a significant role, having a total impact of approximately 11% on the upper limit, as shown in table 4.6. In particular, experimental uncertainties such as the uncertainties associated with the b tagging and JER have a significant impact on the signal acceptance and, consequently, on the upper limit. This is expected because the analysis relies on the selection of at least two b-tagged jets. Also, JER has a large impact because it affects the b jet kinematics and the number of selected light jets, which are important inputs to the BDT classifiers in both channels. The leading modelling uncertainty is associated with FSR, that heavily impacts the number of light jets. This, in turn, affects the performance of the BDT classifiers.

This analysis is the first result targeting the top quark sector using intact protons. Outside the top quark sector, the CMS and TOTEM collaborations recently published a number of results [193–197] exploiting partially or fully the more than $100\,\text{fb}^{-1}$ data collected by CMS and PPS during the Run 2 of the LHC, including the observation of the CEP of dilepton pairs (e^+e^- and $\mu^+\mu^-$) [196].

A logical future improvement for this analysis would be to exploit the data collected during the full Run 2. This requires taking into account the different configurations of PPS in 2016, 2017, and 2018, and was not included in this thesis because the calibrations of the 2018 data were not finalised at the time of this work. In 2018, PPS was equipped with pixel detectors both in the near and in the far RPs, which makes it possible to include events in the selection with multiple proton tracks per arm. Including these events drastically increases the signal efficiency since, as mentioned in section 4.4, the single proton track requirement introduces an efficiency loss of $\sim 90\%$.

The data being collected during the Run 3 will bring even more statistical power. Additionally, PPS in Run 3 is equipped with functioning timing detectors. Measuring the proton's time of arrival will provide a powerful way to reject the pileup proton background, that is by far the dominant background in the current analysis.

The HL-LHC, with its upgraded detectors and unprecedented amounts of data, will make it possible to access very rare processes, with cross sections of the same order as the expected cross section of $pp \to pt\bar{t}p$. In case the PPS detector upgrade plan [198] is approved and carried out successfully, it will be possible to significantly improve the

measurement described in this chapter. In this case, enough precision will be reached such that an independent test of the top-photon coupling will be possible, which would be complementary to what is currently reachable through the measurement of $\mathrm{t\bar{t}}\boldsymbol{\gamma}$, described in chapter 3.

Uncertainty	Channel	Type	Value	Process
Luminosity	both	rate - log-normal	2.3%	all
σ_{DY}	both	rate - log-normal	30%	DY
σ_{Wjets}	ℓ + jets	rate - log-normal	30%	others
σ_{tW}	both	rate - log-normal	5%	others
μ_R	both	shape	-	QCD $\mathrm{t\bar{t}}$
μ_F	both	shape	-	QCD $\mathrm{t\bar{t}}$
ISR	both	shape	-	QCD $\mathrm{t\bar{t}}$
FSR	both	shape	-	all processes
PDF	both	shape	-	all backgrounds
α_S	both	shape	-	all backgrounds
Electron ID	both	shape	-	all
Muon ID	both	shape	-	all
Trigger SFs	both	shape	-	all
L1 prefire	both	shape	-	all
b tagging	both	shape	-	all
JER	both	shape	-	all
JEC	both	shape	-	all
Unclustered MET	both	shape	-	all
Proton reconstruction	both	shape	-	signal
PPS efficiency	both	shape	-	signal
Proton tag rate	both	rate	-	all
$\mathrm{t\bar{t}}$ normalization dilepton	dilepton	rate parameter	-	QCD $\mathrm{t\bar{t}}$
$\mathrm{t\bar{t}}$ normalization ℓ + jets	ℓ + jets	rate parameter	-	QCD $\mathrm{t\bar{t}}$

Table 4.4: Sources of systematic uncertainties considered in the combination, before the pruning. The column "channel" shows whether the uncertainty is affecting the dilepton channel, the ℓ + jets final state or both. The "type" of uncertainty is described as "rate" (constant normalization factor) or "shape" (variations that modify the shape and the normalization of the distribution). The processes affected by each uncertainty are also listed.

Uncertainty	Initial type	Type after treatment	
		dilepton	ℓ + jets
Luminosity	lnN	lnN	lnN
σ_{DY}	lnN	lnN	lnN
σ_{Wjets}	lnN	-	lnN
σ_{tW}	lnN	lnN	lnN
μ_R	shape	-	lnN
μ_F	shape	-	lnN
ISR g $\to$ gg, μ_R	shape	-	lnN/shape
ISR other sources	shape	-	lnN
FSR $x \to x$g, (non-singular terms)	shape	shape	shape
FSR q $\to x$q, (non-singular terms)	shape	lnN	lnN/shape
FSR $x \to x$g, μ_R	shape	lnN	lnN/shape
FSR other sources	shape	lnN	lnN
Choice of PDF	shape	-	lnN
Choice of α_s	shape	-	-
Electron ID	shape	lnN	lnN
Muon ID	shape	lnN	lnN
Trigger SFs	shape	lnN	-
L1 prefire	shape	-	-
b tagging (u,d,s quarks)	shape	lnN	lnN
b tagging (c,b quarks)	shape	shape	lnN
JER	shape	lnN/shape	shape
JES (Absolute)	shape	lnN	-
JES (Relative)	shape	lnN	-
JES (Flavour QCD)	shape	lnN	-
JES (Time)	shape	lnN	-
JES (High p_{T})	shape	-	-
JES (Pileup)	shape	-	-
Unclustered $p_{\mathrm{T}}^{\mathrm{miss}}$	shape	lnN	lnN
Proton reco. (arm 0)	shape	shape	shape
Proton reco. (arm 1)	shape	shape	shape
PPS efficiency	shape	lnN	lnN
Proton tag rate	shape	lnN	lnN

Table 4.5: Input sources of systematic uncertainties considered in the final analysis setup, after the additional treatment, for each channel. The type of uncertainty is described as being "lnN" (constant normalization factor) or "shape" (variations that modify the shape and normalization of the distribution).

Table 4.6: Impact of the leading systematic uncertainties on the expected and observed 95% CL upper limit on the cross section of $pp \to p t\bar{t} p$, for the combined result as well as for the two individual decay channels.

Systematic variation	obs. [exp.] limit change (%)					
	combined		dilepton		ℓ + jets	
Experimental						
Proton reconstruction efficiency	0.1	[0.0]	1.6	[1.9]	0.0	[0.0]
Jet energy resolution	8.2	[0.4]	2.6	[5.6]	16.2	[3.8]
Efficiency of b tagging	2.3	[7.2]	3.9	[7.1]	0.0	[0.3]
Proton tag rate	0.4	[0.4]	0.1	[0.8]	0.0	[0.3]
Modelling						
QCD scales	0.5	[0.0]	0.1	[0.4]	1.4	[1.6]
Final state radiation	7.9	[3.1]	0.3	[0.4]	11.4	[6.8]
All systematic uncertainties	10.9	[14.5]	11.2	[22.3]	21.2	[13.9]
MC statistical uncertainty	0.9	[0.0]	0.1	[0.4]	0.9	[0.6]

Conclusion

A set of differential measurements of the associated production of top quark pairs with a photon ($t\bar{t}\gamma$), with the CMS experiment, is presented in this thesis. This process has been measured before, inclusively and differentially as a function of lepton and photon observables. This thesis contains a number of new results that constitute substantial experimental progress in the understanding of the process. In particular, differential measurements as a function of top quark and top quark pair ($t\bar{t}$) observables are presented for the first time, by using a $t\bar{t}$ reconstruction algorithm. Moreover, for the first time at the LHC, the ratio between the cross sections of $t\bar{t}\gamma$ and $t\bar{t}$ is measured, inclusively and differentially. These ratios are an important input to theory, as they can be used to understand modelling aspects of $t\bar{t}\gamma$, and to probe the top-photon coupling, for example through interpretations in the context of the Standard Model Effective Field Theory. To achieve the results presented, proton-proton collision data recorded at a centre-of-mass energy of 13 TeV are analysed. These data were collected during the second Run of the CERN LHC, during the years 2016, 2017, and 2018, corresponding to an integrated luminosity of $138\,\text{fb}^{-1}$. A publication of these results is in preparation and currently undergoing internal review within the CMS collaboration.

Events with two oppositely charged leptons are selected. One photon, at least 2 jets, and at least 1 jet originating from a b hadron decay are required. The main background to $t\bar{t}\gamma$, after these selection requirements, is $t\bar{t}$ production in association with a nonprompt photon. This can be either a misidentified object or a real photon originating from hadronisation processes or from pileup. The contribution from nonprompt photons is estimated using a data-driven method. Another important background is $Z\gamma$ production. This background is suppressed by imposing the requirement that the invariant mass of the system formed by the two leptons and the photon ($m_{\ell\ell\gamma}$) be at least 15 GeV away from the Z mass. The $Z\gamma$ background is further controlled by introducing a control region, defined by applying the same cuts as the signal region but inverting the $m_{\ell\ell\gamma}$ requirement.

By performing a simultaneous profile likelihood fit to the signal and control regions, the reconstructed quantities can be unfolded such that the generator-level quantities are

recovered. This is done for 5 different variables, the leading photon p_{T}, the leading lepton p_{T}, the $\Delta\phi$ between the two leptons ($\Delta\phi(\ell,\ell')$), the leading top quark p_{T}, and the ΔR between the photon and the $\mathrm{t\bar{t}}$ system ($\Delta R(\gamma,\mathrm{t\bar{t}})$). From the unfolded quantities, differential cross sections of $\mathrm{t\bar{t}\gamma}$ are extracted. A precision of about 8-15% is achieved for lepton and photon observables, and about 11-25% for top quark and $\mathrm{t\bar{t}}$ observables.

Additionally, the ratio between the cross sections of $\mathrm{t\bar{t}\gamma}$ and $\mathrm{t\bar{t}}$ is measured to be

$$R_{\gamma} = 0.02779^{+0.0013}_{-0.0013}(68\%) \quad , \tag{4.6}$$

corresponding to a 5% precision. This ratio is measured differentially as a function of two different parton-level observables, with a 6-15% precision, depending on the bin and distribution.

A study of the top quark charge asymmetry in $\mathrm{t\bar{t}\gamma}$ events is also presented, where a value of $(4.686^{+4.385}_{-4.398})\%$ is measured, in agreement with the SM predictions at NLO.

As a second project, a search for the central exclusive production of $\mathrm{t\bar{t}}$ with the CMS experiment is presented. A subdetector of CMS, the Precision Proton Spectrometer (PPS), is used to detect intact protons leaving the CMS interaction point. The analysis uses proton-proton collision data collected at a centre-of-mass energy of 13 TeV, during the year 2017, corresponding to an integrated luminosity of $24.9\,\mathrm{fb}^{-1}$. This search was initiated as a master thesis project [199]. Here, in this PhD thesis, I present the completed analysis and the first upper limit on the cross section of exclusive $\mathrm{t\bar{t}}$. The results were published by the CMS Collaboration in Ref. [183]. Events with a typical $\mathrm{t\bar{t}}$ signature, in the dilepton channel, are selected. The same $\mathrm{t\bar{t}}$ algorithm used in the $\mathrm{t\bar{t}\gamma}$ analysis is employed to reconstruct the $\mathrm{t\bar{t}}$ system. The reconstructed system is matched with protons tagged in PPS. A data-driven method is designed and implemented to estimate the background from pileup protons, which are not present in the simulated data. A classification algorithm based on Boosted Decision Trees is developed to separate the signal from the overwhelming $\mathrm{t\bar{t}}$ background, taking advantage of the correlations between the central system and the intact protons, for the signal. Finally, a combination between the dilepton and lepton+jets final states is performed, taking all nuisance parameters into account. The measured upper limit on the cross section of central exclusive $\mathrm{t\bar{t}}$ is:

$$\sigma < 0.59\,\mathrm{pb} \quad (1.10^{+1.22}_{-0.55}\ \text{expected @ 95\% confidence level}) \tag{4.7}$$

Both of these processes offer sensitivity to the coupling between the top quark and the photon. At the same time, they reflect very different ways of doing physics at the LHC: while the $\mathrm{t\bar{t}\gamma}$ process is a precision measurement, where the large number of events allows for stringent tests of the SM predictions, the CEP $\mathrm{t\bar{t}}$ is an extremely rare process that is explored for the first time here and for which an observation will only be possible with much larger amounts of data.

The planned upgrade of the LHC, the HL-LHC, is expected to collect about five times more data than those collected so far, leading to increased precision in almost all the

existing measurements, and opening the door to the observation of more rare processes that have so far been out of reach.

For the farther future, future colliders such as the Future Circular Collider have been proposed and are being considered, reaching an order of magnitude higher centre-of-mass energy. In case new physics effects manifest at those energies, it may be possible to discover them and improve our understanding of the fundamental particles and their interactions.

Bibliography

[1] M. E. Peskin and D. V. Schroeder, *An introduction to quantum field theory*. Westview Press (1995).

[2] J. Schwichtenberg, *Physics from symmetry*. Springer (2018).

[3] E. Noether, "Invariante Variationsprobleme." *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* **1918** (1918) 235–257.

[4] P. A. M. Dirac, "The quantum theory of the electron." *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **117** (1928), no. 778 610–624.

[5] R. P. Feynman, "Space-time approach to quantum electrodynamics." *Phys. Rev.* **76** (1949) 769–789.

[6] M. Gell-Mann, "Symmetries of baryons and mesons." *Phys. Rev.* **125** (1962) 1067–1084.

[7] A. Deur, S. J. Brodsky, and G. F. de Teramond, "The QCD Running Coupling." *Nucl. Phys.* **90** (2016) 1, [arXiv:1604.08082].

[8] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, "Experimental test of parity conservation in beta decay." *Phys. Rev.* **105** (1957) 1413–1415.

[9] S. L. Glashow, "The renormalizability of vector meson interactions." *Nucl. Phys.* **10** (1959) 107–117.

[10] A. Salam and J. C. Ward, "Weak and electromagnetic interactions." *Nuovo Cim.* **11** (1959) 568–577.

[11] S. Weinberg, "A Model of Leptons." *Phys. Rev. Lett.* **19** (1967) 1264–1266.

[12] K. Nishijima, "Charge Independence Theory of V Particles." *Progress of Theoretical Physics* **13** (1955), no. 3 285–304, [https://academic.oup.com/ptp/article-pdf/13/3/285/5425869/13-3-285.pdf].

[13] Super-Kamiokande Collaboration, "Evidence for oscillation of atmospheric neutrinos." *Phys. Rev. Lett.* **81** (1998) 1562–1567, [hep-ex/9807003].

[14] Z. Maki, M. Nakagawa, and S. Sakata, "Remarks on the unified model of elementary particles." *Prog. Theor. Phys.* **28** (1962) 870–880.

[15] B. Pontecorvo, "Inverse beta processes and nonconservation of lepton charge." *Zh. Eksp. Teor. Fiz.* **34** (1957) 247.

[16] KATRIN Collaboration, "Direct neutrino-mass measurement with sub-electronvolt sensitivity." *Nature Phys.* **18** (2022), no. 2 160–166, [arXiv:2105.08533].

[17] UA1 Collaboration, "Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s}$=540 GeV." *Physics Letters B* **122** (1983), no. 1 103–116.

[18] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons." *Phys. Rev. Lett.* **13** (1964) 508–509.

[19] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons." *Phys. Rev. Lett.* **13** (1964) 321–323.

[20] J. Ellis, "Higgs Physics." in *2013 European School of High-Energy Physics*, pp. 117–168, 2015. arXiv:1312.5672.

[21] Particle Data Group Collaboration, "Review of Particle Physics." *PTEP* **2022** (2022) 083C01.

[22] N. Cabibbo, "Unitary Symmetry and Leptonic Decays." *Phys. Rev. Lett.* **10** (1963) 531–533.

[23] M. Kobayashi and T. Maskawa, "CP Violation in the Renormalizable Theory of Weak Interaction." *Prog. Theor. Phys.* **49** (1973) 652–657.

[24] M. L. Perl et al., "Evidence for anomalous lepton production in e^+e^- annihilation." *Phys. Rev. Lett.* **35** (1975) 1489–1492.

[25] H. Burkhardt and J. Steinberger, "Tests of the electroweak theory at the Z resonance." *Ann. Rev. Nucl. Part. Sci.* **41** (1991) 55–96.

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

[27] CMS Collaboration, "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC." *Phys. Lett. B* **716** (2012) 30–61, [arXiv:1207.7235].

[28] J. C. Collins, D. E. Soper, and G. F. Sterman, "Factorization of Hard Processes in QCD." *Adv. Ser. Direct. High Energy Phys.* **5** (1989) 1–91, [hep-ph/0409313].

[29] J. D. Bjorken, "Asymptotic Sum Rules at Infinite Momentum." *Phys. Rev.* **179** (1969) 1547–1553.

[30] NNPDF Collaboration, R. D. Ball et al., "Parton distributions from high-precision collider data." *Eur. Phys. J. C* **77** (2017) 663, [arXiv:1706.00428].

[31] G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language." *Nucl. Phys. B* **126** (1977) 298–318.

[32] Y. L. Dokshitzer, "Calculation of the structure functions for deep inelastic scattering and e^+e^- annihilation by perturbation theory in Quantum Chromodynamics.." *Sov. Phys. JETP* **46** (1977) 641–653.

[33] V. N. Gribov and L. N. Lipatov, "Deep inelastic ep scattering in perturbation theory." *Sov. J. Nucl. Phys.* **15** (1972) 438–450.

[34] C. Bierlich et al., "A comprehensive guide to the physics and usage of PYTHIA 8.3." arXiv:2203.11601.

[35] Hoeche et al., "Matching parton showers and matrix elements." in *HERA and the LHC: A Workshop on the Implications of HERA for LHC Physics: CERN - DESY Workshop 2004/2005*, pp. 288–289, 2005. hep-ph/0602031.

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

[37] GEANT4 Collaboration, S. Agostinelli et al., "GEANT4: A simulation toolkit." *Nucl. Instrum. Meth.* **A506** (2003) 250.

[38] P. D. B. Collins, *An Introduction to Regge Theory and High-Energy Physics*. Cambridge Monographs on Mathematical Physics. Cambridge Univ. Press, Cambridge, UK (2009).

[39] V. Barone and E. Predazzi, *High-Energy Particle Diffraction*, vol. v.565 of *Texts and Monographs in Physics*. Springer-Verlag, Berlin Heidelberg (2002).

[40] TOTEM, CMS Collaboration, "Proton reconstruction with the CMS-TOTEM Precision Proton Spectrometer." arXiv:2210.05854.

[41] H. Harari, "A new quark model for hadrons." *Physics Letters B* **57** (1975), no. 3 265–269.

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

[43] CMS Collaboration, "Measurement of the inclusive $\mathrm{t\bar{t}}$ cross section in pp collisions at $\sqrt{s} = 5.02$ TeV using final states with at least one charged lepton." *JHEP* **03** (2018) 115, [arXiv:1711.03143].

[44] ATLAS Collaboration, "Measurement of the top quark pair production cross-section with ATLAS in the single lepton channel." *Phys. Lett. B* **711** (2012) 244, [arXiv:1201.1889].

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

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

[47] CMS Collaboration, "Measurements of the $\mathrm{t\bar{t}}$ production cross section in lepton+jets final states in pp collisions at 8 TeV and ratio of 8 to 7 TeV cross sections." *Eur. Phys. J. C* **77** (2017) 15, [arXiv:1602.09024].

[48] ATLAS Collaboration, "Measurement of the $\mathrm{t\bar{t}}$ production cross-section using $\mathrm{e\mu}$ events with b-tagged jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector." *Phys. Lett. B* **761** (2016) 136, [arXiv:1606.02699]. [Erratum: doi:10.1016/j.physletb.2017.09.027].

[49] CMS Collaboration, "Measurement of the top quark pair production cross section in proton-proton collisions at $\sqrt{s} = 13$ TeV." *Phys. Rev. Lett.* **116** (2016) 052002, [arXiv:1510.05302].

[50] CMS Collaboration, "Measurement of the $\mathrm{t\bar{t}}$ production cross section using events in the $\mathrm{e\mu}$ final state in pp collisions at $\sqrt{s} = 13$ TeV." *Eur. Phys. J. C* **77** (2017) 172, [arXiv:1611.04040].

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

[52] LHCb Collaboration, "Measurement of forward top pair production in the dilepton channel in pp collisions at $\sqrt{s} = 13$ TeV." *JHEP* **08** (2018) 174, [arXiv:1803.05188].

[53] CMS Collaboration, "First measurement of the top quark pair production cross section in proton-proton collisions at $\sqrt{s} = 13.6$ TeV." arXiv:2303.10680.

[54] ATLAS, CMS Collaboration, "Combinations of single-top-quark production cross-section measurements and $|f_{LV}\mathrm{V}_{tb}|$ determinations at $\sqrt{s} = 7$ and $8\,\text{TeV}$ with the ATLAS and CMS experiments." *JHEP* **05** (2019) 088, [arXiv:1902.07158].

[55] ATLAS Collaboration, "Measurement of the inclusive cross-sections of single top-quark and top-antiquark t-channel production in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector." *JHEP* **04** (2017) 086, [arXiv:1609.03920].

[56] CMS Collaboration, "Measurement of differential cross sections and charge ratios for t-channel single top quark production in proton–proton collisions at $\sqrt{s} = 13$ TeV." *Eur. Phys. J. C* **80** (2020) 370, [arXiv:1907.08330].

[57] ATLAS Collaboration, "Measurement of the cross-section for producing a W boson in association with a single top quark in pp collisions at $\sqrt{s} = 13$ TeV with ATLAS." *JHEP* **01** (2018) 063, [arXiv:1612.07231].

[58] LHCTopWG, "LHCTopWG Summary Plots." (2023).

[59] M. Fael and T. Gehrmann, "Probing top quark electromagnetic dipole moments in single-top-plus-photon production." *Phys. Rev. D* **88** (2013) 033003, [arXiv:1307.1349].

[60] CMS Collaboration, "Evidence for the associated production of a single top quark and a photon in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$." *Phys. Rev. Lett.* **121** (2018), no. 22 221802, [arXiv:1808.02913].

[61] ATLAS Collaboration, "Observation of single-top-quark production in association with a photon using the ATLAS detector." arXiv:2302.01283.

[62] 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, [arXiv:2111.02860].

[63] ATLAS Collaboration, "Measurements of inclusive and differential cross-sections of combined $t\bar{t}\gamma$ and $tW\gamma$ production in the $e\mu$ channel at 13 TeV with the ATLAS detector." *JHEP* **09** (2020) 049, [arXiv:2007.06946].

[64] ATLAS Collaboration, "Observation of the associated production of a top quark and a Z boson in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector." *JHEP* **07** (2020) 124, [arXiv:2002.07546].

[65] CMS Collaboration, "Evidence for tWZ production in proton-proton collisions at $\sqrt{s} = 13$ TeV in multilepton final states." tech. rep., CERN, Geneva, 2023.

[66] ATLAS, CMS Collaboration, "Combination of the W boson polarization measurements in top quark decays using ATLAS and CMS data at $\sqrt{s} = 8$ TeV." *JHEP* **08** (2020), no. 08 051, [arXiv:2005.03799].

[67] CMS Collaboration, "Measurement of the top quark polarization and $t\bar{t}$ spin correlations using dilepton final states in proton-proton collisions at $\sqrt{s} = 13$ TeV." *Phys. Rev. D* **100** (2019), no. 7 072002, [arXiv:1907.03729].

[68] ATLAS Collaboration, "Measurements of top-quark pair spin correlations in the $e\mu$ channel at $\sqrt{s} = 13$ TeV using pp collisions in the ATLAS detector." *Eur. Phys. J. C* **80** (2020), no. 8 754, [arXiv:1903.07570].

[69] J. A. Aguilar-Saavedra et al., "Interpreting top-quark LHC measurements in the standard-model effective field theory." arXiv:1802.07237.

[70] G. Bevilacqua, H. B. Hartanto, M. Kraus, T. Weber, and M. Worek, "Hard Photons in Hadroproduction of Top Quarks with Realistic Final States." *JHEP* **10** (2018) 158, [arXiv:1803.09916].

[71] G. Bevilacqua, H. B. Hartanto, M. Kraus, T. Weber, and M. Worek, "Off-shell vs on-shell modelling of top quarks in photon associated production." *JHEP* **03** (2020) 154, [arXiv:1912.09999].

[72] U. Baur, M. Buice, and L. H. Orr, "Direct measurement of the top quark charge at hadron colliders." *Phys. Rev. D* **64** (2001) 094019, [hep-ph/0106341].

[73] CDF Collaboration, "Exclusion of exotic top-like quarks with -4/3 electric charge using jet-charge tagging in single-lepton ttbar events at CDF." *Phys. Rev. D* **88** (2013), no. 3 032003, [arXiv:1304.4141].

[74] ATLAS Collaboration, "Measurement of the top quark charge in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector." *JHEP* **11** (2013) 031, [arXiv:1307.4568].

[75] ATLAS Collaboration, "Search for anomalous couplings in the Wtb vertex from the measurement of double differential angular decay rates of single top quarks produced in the t-channel with the ATLAS detector." *JHEP* **04** (2016) 023, [arXiv:1510.03764].

[76] CMS Collaboration, "Constraints on the Top-Quark Charge from Top-Pair Events." tech. rep., CERN, Geneva, 2012.

[77] M. Czakon, D. Heymes, A. Mitov, D. Pagani, I. Tsinikos, and M. Zaro, "Top-pair production at the LHC through NNLO QCD and NLO EW." *JHEP* **10** (2017) 186, [arXiv:1705.04105].

[78] G. Bevilacqua, H. B. Hartanto, M. Kraus, T. Weber, and M. Worek, "Precise predictions for $\mathrm{t\bar{t}\gamma}/\mathrm{t\bar{t}}$ cross section ratios at the LHC." *JHEP* **01** (2019) 188, [arXiv:1809.08562].

[79] M. Schulze and Y. Soreq, "Pinning down electroweak dipole operators of the top quark." *Eur. Phys. J. C* **76** (2016), no. 8 466, [arXiv:1603.08911].

[80] J. H. Kuhn and G. Rodrigo, "Charge asymmetry in hadroproduction of heavy quarks." *Phys. Rev. Lett.* **81** (1998) 49–52, [hep-ph/9802268].

[81] K. Cheung, W.-Y. Keung, and T.-C. Yuan, "Top Quark Forward-Backward Asymmetry." *Phys. Lett. B* **682** (2009) 287–290, [arXiv:0908.2589].

[82] P. Ferrario and G. Rodrigo, "Heavy colored resonances in $\mathrm{t\bar{t}}$ + jet at the LHC." *JHEP* **02** (2010) 051, [arXiv:0912.0687].

[83] G. Rodrigo and P. Ferrario, "Charge asymmetry: A Theory appraisal." *Nuovo Cim. C* **033** (2010), no. 4 221–228, [arXiv:1007.4328].

[84] A. Djouadi, G. Moreau, F. Richard, and R. K. Singh, "The Forward-backward asymmetry of top quark production at the Tevatron in warped extra dimensional models." *Phys. Rev. D* **82** (2010) 071702, [arXiv:0906.0604].

[85] S. Jung, H. Murayama, A. Pierce, and J. D. Wells, "Top quark forward-backward asymmetry from new t-channel physics." *Phys. Rev. D* **81** (2010) 015004, [arXiv:0907.4112].

[86] CDF Collaboration, "Forward-Backward Asymmetry in Top Quark Production in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96\,\mathrm{TeV}$." *Phys. Rev. Lett.* **101** (2008) 202001, [arXiv:0806.2472].

[87] D0 Collaboration, V. M. Abazov et al., "First measurement of the forward-backward charge asymmetry in top quark pair production." *Phys. Rev. Lett.* **100** (2008) 142002, [arXiv:0712.0851].

[88] J. H. Kuhn and G. Rodrigo, "Charge asymmetries of top quarks at hadron colliders revisited." *JHEP* **01** (2012) 063, [arXiv:1109.6830].

[89] J. Bergner and M. Schulze, "The top quark charge asymmetry in $t\bar{t}\gamma$ production at the LHC." *Eur. Phys. J. C* **79** (2019), no. 3 189, [arXiv:1812.10535].

[90] J. de Favereau de Jeneret et al., "High energy photon interactions at the LHC." 2009.

[91] D. d'Enterria and J. Lansberg, "Study of Higgs boson production and its $b\bar{b}$ decay in $\gamma\gamma$ processes in proton-nucleus collisions at the LHC." *Phys. Rev. D* **81** (2010) 014004, [arXiv:0909.3047].

[92] S. Fayazbakhsh, S. T. Monfared, and M. Mohammadi Najafabadi, "Top quark anomalous electromagnetic couplings in photon-photon scattering at the LHC." *Phys. Rev. D* **92** (2015) 014006, [arXiv:1504.06695].

[93] M. Łuszczak, L. Forthomme, W. Schäfer, and A. Szczurek, "Production of $t\bar{t}$ pairs via $\gamma\gamma$ fusion with photon transverse momenta and proton dissociation." *JHEP* **02** (2019) 100, [arXiv:1810.12432].

[94] V. P. Gonçalves, D. E. Martins, M. S. Rangel, and M. Tasevsky, "Top quark pair production in the exclusive processes at the LHC." *Phys. Rev. D* **102** (2020) 074014, [arXiv:2007.04565].

[95] J. Howarth, "Elastic potential: a proposal to discover elastic production of top quarks at the Large Hadron Collider." 2020.

[96] H.-S. Shao and D. d'Enterria, "gamma-UPC: automated generation of exclusive photon-photon processes in ultraperipheral proton and nuclear collisions with varying form factors." *JHEP* **09** (2022) 248, [arXiv:2207.03012].

[97] C. Baldenegro, A. Bellora, S. Fichet, G. von Gersdorff, M. Pitt, and C. Royon, "Searching for anomalous top quark interactions with proton tagging and timing detectors at the LHC." *JHEP* **08** (2022) 021, [arXiv:2205.01173].

[98] S. C. Inan and A. A. Billur, "Polarized top pair production in extra dimension models via photon-photon fusion at the CERN LHC." *Phys. Rev. D* **84** (2011) 095002.

[99] *LEP design report*. CERN, Geneva (1984). Copies shelved as reports in LEP, PS and SPS libraries.

[100] "High-Luminosity LHC Project Schedule." , accessed on September 9th, 2022.

[101] J. Vollaire et al., *Linac4 design report*, vol. 6/2020 of *CERN Yellow Reports: Monographs*. CERN, Geneva (2020).

[102] "Facts and figures about the LHC." , accessed on November 3rd, 2022.

[103] T. Petterson, T. Sven and P. Lefèvre (LHC Study Group), "The Large Hadron Collider: Conceptual design.".

[104] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider." *JINST* **3** (2008) S08003.

[105] LHCb Collaboration, "LHCb technical proposal.".

[106] ALICE Collaboration, "The ALICE experiment at the CERN LHC." *JINST* **3** (2008) S08002.

[107] "About CMS." https://cms.cern/index.php/detector, accessed on March 7th, 2023.

[108] "CMS slice image view." http://cds.cern.ch/record/2204863, accessed on August 29th, 2022.

[109] CMS Collaboration, "The CMS Experiment at the CERN LHC." *JINST* **3** (2008) S08004.

[110] CMS Tracker Group Collaboration, W. Adam et al., "The CMS Phase-1 Pixel Detector Upgrade." *JINST* **16** (2021), no. 02 P02027, [arXiv:2012.14304].

[111] CMS Collaboration, "The CMS electromagnetic calorimeter project: Technical Design Report.".

[112] CMS Collaboration, "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$." *JINST* **13** (2018), no. 06 P06015, [arXiv:1804.04528].

[113] CMS, TOTEM Collaboration, M. Albrow et al., "CMS-TOTEM Precision Proton Spectrometer." Tech. Rep. CERN-LHCC-2014-021, TOTEM-TDR-003, CMS-TDR-13, 2014.

[114] F. Ravera, *3D silicon pixel detectors for the CT-PPS tracking system*, Ph.D. thesis (2017). PhD Thesis, Presented 16 Jun 2017.

[115] M. Deile, F. Caspers, T. Kroyer, M. Oriunno, E. Radermacher, A. Soter, and F. Roncarolo, "Beam Coupling Impedance Measurement and Mitigation for a TOTEM Roman Pot." *Conf. Proc.* **C0806233** (2008) TUPP029, [arXiv:0806.4974]. arXiv:0806.4974.

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

[117] CMS Collaboration, "The CMS trigger system." *JINST* **12** (2017) P01020, [arXiv:1609.02366].

[118] CMS Collaboration, A. Petrilli and A. Hervé, "CMS Computing Model: The CMS Computing Model RTAG.".

[119] CMS Collaboration, "Particle-flow reconstruction and global event description with the CMS detector." *JINST* **12** (2017), no. 10 P10003, [arXiv:1706.04965].

[120] K. Rose, "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems." *Proceedings of the IEEE* **86** (1998), no. 11 2210–2239.

[121] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker." *JINST* **9** (2014), no. 10 P10009, [arXiv:1405.6569].

[122] CMS Collaboration, T. Mc Cauley, "Collisions recorded by the CMS detector on 14 Oct 2016 during the high pile-up fill." CMS Collection., 2016.

[123] CMS Collaboration, "Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC." *JINST* **16** (2021), no. 05 P05014, [arXiv:2012.06888].

[124] W. Adam, R. Fruhwirth, A. Strandlie, and T. Todorov, "Reconstruction of electrons with the Gaussian sum filter in the CMS tracker at LHC." *eConf* **C0303241** (2003) TULT009, [physics/0306087].

[125] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm." *JHEP* **04** (2008) 063, [arXiv:0802.1189].

[126] H. Grote and F. Schmidt, "Mad-x - an upgrade from mad8." in *Proceedings of the 2003 Particle Accelerator Conference*, vol. 5, pp. 3497–3499, 2003.

[127] CMS, TOTEM Collaboration, "CTPPS plots: 2017 performance update and 2018 start of data taking." Tech. Rep. CMS DP-2018/021, 2018.

[128] CMS Collaboration, "Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ in 2015 and 2016 at CMS." *Eur. Phys. J. C* **81** (2021), no. 9 800, [arXiv:2104.01927].

[129] CMS BRIL Collaboration, P. Lujan, "The Pixel Luminosity Telescope: A detector for luminosity measurement at CMS using silicon pixel sensors." arXiv:2206.08870.

[130] CMS Collaboration, "BCM1F and Luminosity calibration.".

[131] S. van der Meer, "Calibration of the effective beam height in the ISR." tech. rep., CERN, Geneva, 1968.

[132] CMS Collaboration, "Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ in 2015 and 2016 at CMS." *Eur. Phys. J. C* **81** (2021), no. 9 800, [arXiv:2104.01927].

[133] V. Scheurer, *First measurement of the $t\bar{t}Z$ production cross section in association with a Z boson decaying into a quark pair at the CMS Experiment*, Ph.D. thesis, Hamburg U., Dept. Math. (2022).

[134] CDF Collaboration, "Evidence for $t\bar{t}\gamma$ Production and Measurement of $\sigma_{t\bar{t}\gamma}/\sigma_{t\bar{t}}$." *Phys. Rev. D* **84** (2011) 031104, [arXiv:1106.3970].

[135] ATLAS Collaboration, "Observation of top-quark pair production in association with a photon and measurement of the $t\bar{t}\gamma$ production cross section in pp collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector." *Phys. Rev. D* **91** (2015), no. 7 072007, [arXiv:1502.00586].

[136] ATLAS Collaboration, "Measurement of the $t\bar{t}\gamma$ production cross section in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector." *JHEP* **11** (2017) 086, [arXiv:1706.03046].

[137] CMS Collaboration, "Measurement of the semileptonic $\mathrm{t\bar{t}} + \gamma$ production cross section in pp collisions at $\sqrt{s} = 8$ TeV." *JHEP* **10** (2017) 006, [arXiv:1706.08128].

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

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

[140] ATLAS Collaboration, "Measurements of inclusive and differential fiducial cross-sections of $t\bar{t}\gamma$ production in leptonic final states at $\sqrt{s} = 13$ TeV in ATLAS." *Eur. Phys. J. C* **79** (2019), no. 5 382, [arXiv:1812.01697].

[141] ATLAS Collaboration, "Measurement of the charge asymmetry in top-quark pair production in association with a photon with the ATLAS experiment." arXiv:2212.10552.

[142] T. Sjöstrand et al., "An introduction to PYTHIA 8.2." *Comput. Phys. Commun.* **191** (2015) 159, [arXiv:1410.3012].

[143] CMS Collaboration, "Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements." *Eur. Phys. J. C* **80** (2020) 4, [arXiv:1903.12179].

[144] Alwall, J. et al., "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions." *Eur. Phys. J. C* **53** (2008) 473, [arXiv:0706.2569].

[145] 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, [arXiv:1405.0301].

[146] P. Artoisenet, R. Frederix, O. Mattelaer, and R. Rietkerk, "Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations." *JHEP* **03** (2013) 015, [arXiv:1212.3460].

[147] P. Nason, "A new method for combining NLO QCD with shower Monte Carlo algorithms." *JHEP* **11** (2004) 040, [hep-ph/0409146].

[148] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method." *JHEP* **11** (2007) 070, [arXiv:0709.2092].

[149] 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, [arXiv:1002.2581].

[150] M. Czakon, P. Fiedler, and A. Mitov, "Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_S^4)$." *Phys. Rev. Lett.* **110** (2013) 252004, [arXiv:1303.6254].

[151] S. Frixione, B. Fuks, V. Hirschi, K. Mawatari, H.-S. Shao, P. A. Sunder, and M. Zaro, "Automated simulations beyond the Standard Model: supersymmetry." *JHEP* **12** (2019) 008, [arXiv:1907.04898].

[152] CMS Collaboration, "Evidence for tWZ production in proton-proton collisions at $\sqrt{s}$ = 13 TeV in multilepton final states." Tech. Rep. CMS-PAS-TOP-22-008, CERN, Geneva, 2023.

[153] E. Bols, J. Kieseler, M. Verzetti, M. Stoye, and A. Stakia, "Jet Flavour Classification Using DeepJet." *JINST* **15** (2020), no. 12 P12012, [arXiv:2008.10519].

[154] CMS Collaboration, N. Trevisani et al., "TOP Trigger (Run2)." twiki, 2022.

[155] CMS Collaboration, "Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s}$ = 8 TeV." *JINST* **10** (2015), no. 06 P06005, [arXiv:1502.02701].

[156] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV." *JINST* **13** (2018), no. 05 P05011, [arXiv:1712.07158].

[157] L. Sonnenschein, "Analytical solution of ttbar dilepton equations." *Phys. Rev. D* **73** (2006) 054015, [hep-ph/0603011]. [Erratum: Phys.Rev.D 78, 079902 (2008)].

[158] B. Betchart, R. Demina, and A. Harel, "Analytic solutions for neutrino momenta in decay of top quarks." 1305.1878.

[159] M. Savitskyi, *Measurements of differential cross sections for $t\bar{t}$ production in proton-proton collisions at $\sqrt{s} = 13$ TeV using events containing two leptons with the CMS experiment*, Ph.D. thesis (2018).

[160] CMS Collaboration, "Measurement of differential cross sections for the production of top quark pairs and of additional jets in pp collisions at $\sqrt{s}$ = 13 TeV." tech. rep., CERN, Geneva, 2022.

[161] D. Dobur, J. Knolle, G. Mestdach, and K. Skovpen, "Photon radiation effects in kinematic reconstruction of top quarks." *JINST* **16** (2021), no. 12 P12001, [arXiv:2107.07586].

[162] CMS Collaboration, "Measurement of the inclusive and differential $\mathrm{t\bar{t}}\gamma$ cross sections in the dilepton channel and effective field theory interpretation in proton-proton collisions at $\sqrt{s}$ = 13 TeV." *Journal of High Energy Physics* **2022** (2022), no. 5 [arXiv:2201.07301].

[163] CMS Collaboration, "CMS ECAL Response to Laser Light." tech. rep., 2019.

[164] J. S. Conway, "Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra." in *PHYSTAT 2011*, pp. 115–120, 2011. arXiv:1103.0354.

[165] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics." *The European Physical Journal C* **71** (2011), no. 2.

[166] F. James and M. Roos, "Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations." *Comput. Phys. Commun.* **10** (1975) 343–367.

[167] G. J. Feldman and R. D. Cousins, "A Unified approach to the classical statistical analysis of small signals." *Phys. Rev. D* **57** (1998) 3873–3889, [physics/9711021].

[168] G. Cowan, *Statistical data analysis* (1998).

[169] S. S. Wilks, "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses." *Annals Math. Statist.* **9** (1938), no. 1 60–62.

[170] A. Wald, "Tests of statistical hypotheses concerning several parameters when the number of observations is large." *Transactions of the American Mathematical Society* **54** (1943), no. 3 426–482.

[171] CMS Collaboration, "CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV." CMS Physics Analysis Summary CMS-PAS-LUM-17-004, 2018.

[172] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV." *JINST* **12** (2017), no. 02 P02014, [arXiv:1607.03663].

[173] R. J. Barlow and C. Beeston, "Fitting using finite Monte Carlo samples." *Comput. Phys. Commun.* **77** (1993) 219.

[174] ATLAS Collaboration, "Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV with the ATLAS detector at the LHC." *Phys. Rev. Lett.* **117** (2016) 182002, [arXiv:1606.02625].

[175] A. Hocker and V. Kartvelishvili, "SVD approach to data unfolding." *Nucl. Instrum. Meth. A* **372** (1996) 469–481, [hep-ph/9509307].

[176] J. Rubenach, *Search for heavy Higgs bosons in conjunction with neural-network-driven reconstruction and upgrade of the Fast Beam Condition Monitor at the CMS experiment*, Ph.D. thesis, Hamburg U., Dept. Math. (2023).

[177] CMS Collaboration, "Machine learning approaches for parameter reweighting in MC samples of top quark production in CMS.".

[178] A. Andreassen and B. Nachman, "Neural Networks for Full Phase-space Reweighting and Parameter Tuning." *Phys. Rev. D* **101** (2020), no. 9 091901, [arXiv:1907.08209].

[179] F. Maltoni, D. Pagani, and I. Tsinikos, "Associated production of a top-quark pair with vector bosons at NLO in QCD: impact on $t\bar{t}H$ searches at the LHC." *JHEP* **02** (2016) 113, [arXiv:1507.05640].

[180] A. Deandrea and N. Deutschmann, "Multi-tops at the LHC." *JHEP* **08** (2014) 134, [arXiv:1405.6119].

[181] FCC Collaboration, A. Abada et al., "FCC Physics Opportunities: Future Circular Collider Conceptual Design Report Volume 1." *Eur. Phys. J. C* **79** (2019), no. 6 474.

[182] CMS Collaboration, "The CMS Precision Proton Spectrometer at the HL-LHC – Expression of Interest." 2021.

[183] CMS, TOTEM Collaboration, "Search for central exclusive production of top quark pairs in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ with tagged protons." tech. rep., CERN, Geneva, 2022.

[184] CMS Collaboration, "Display of an event from a search for exclusive production of top and antitop quark pairs in proton-proton collisions at 13 TeV and with tagged protons." CMS Collection., 2022.

[185] CMS Collaboration, "CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV." CMS Physics Analysis Summary CMS-PAS-LUM-17-004, 2018.

[186] M. Boonekamp, A. Dechambre, V. Juranek, O. Kepka, M. Rangel, C. Royon, and R. Staszewski, "FPMC: a generator for forward physics." 2011.

[187] V. M. Budnev, I. F. Ginzburg, G. V. Meledin, and V. G. Serbo, "The two photon particle production mechanism. Physical problems. Applications. Equivalent photon approximation." *Phys. Rept.* **15** (1975) 181.

[188] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV." *JINST* **13** (2018), no. 05 P05011, [arXiv:1712.07158].

[189] L. Breiman, J. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*. Chapman and Hall/CRC (1984).

[190] A. Hocker et al., "TMVA – Toolkit for Multivariate Data Analysis." 2007.

[191] R. Brun et al., "root-project/root: v6.18/02." tech. rep., 2019.

[192] A. L. Read, "Presentation of search results: The CL(s) technique." *J. Phys. G* **28** (2002) 2693–2704.

[193] CMS, TOTEM Collaboration, "First Search for Exclusive Diphoton Production at High Mass with Tagged Protons in Proton-Proton Collisions at $\sqrt{s} = 13\,\text{TeV}$." *Phys. Rev. Lett.* **129** (2022), no. 1 011801, [arXiv:2110.05916].

[194] CMS, TOTEM Collaboration, "A search for new physics in central exclusive production using the missing mass technique with the CMS detector and the CMS-TOTEM precision proton spectrometer." arXiv:2303.04596.

[195] CMS, TOTEM Collaboration, "Search for exclusive diphoton production with intact protons in PPS." tech. rep., CERN, Geneva, 2022.

[196] CMS, TOTEM Collaboration, "Observation of proton-tagged, central (semi)exclusive production of high-mass lepton pairs in pp collisions at 13 TeV with the CMS-TOTEM precision proton spectrometer." *JHEP* **07** (2018) 153, [arXiv:1803.04496].

[197] CMS, TOTEM Collaboration, "Search for high-mass exclusive $\gamma\gamma \to$ WW and $\gamma\gamma \to$ ZZ production in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$." arXiv:2211.16320.

[198] CMS Collaboration, "The CMS Precision Proton Spectrometer at the HL-LHC – Expression of Interest." tech. rep., CERN, Geneva, 2020. 88 pages.

[199] B. Ribeiro Lopes, "Search for exclusively produced top quark pairs at the LHC." (2019). MSc thesis.

Acknowledgements

I start by expressing my heartfelt gratitude to my supervisor, Nadjieh Jafari. Her guidance illuminated my path throughout this journey. With each conversation, she not only shared her knowledge but also inspired me with new ideas and solutions to the challenges at hand. Her expertise, but also her very human approach with her students, are an inspiration to me.

Elisabetta Gallo stood by me with unwavering support, especially during the trying times of the pandemic. Her meticulous feedback on my thesis significantly enhanced its quality. Her role as both a supporter and a critical reader has been invaluable.

Working alongside Andreas Meyer on luminosity measurements was a pleasure and an honour. His profound understanding of physics and his ability to navigate challenges helped the work move forward. The two months at CERN, made possible by his support, were a very enriching experience.

A big thanks to Maria, for the constant presence and willingness to assist, for the cheerful support and insightful feedback that consistently propelled my work forward.

The collaboration with colleagues like Michele Gallinaro and Pedro Silva has been a bedrock of support and expertise since 2018. Their guidance led to the exclusive $t\bar{t}$ analysis, culminating in a publication. Enrico Robutti and Michael Pitt's contributions within this context are also acknowledged with gratitude.

To my office mates over these four years – Joscha Knolle, Valerie Scheurer, Michele Mormile, Federica Colombina, and Konstantin Sharko – I extend my heartfelt appreciation. You turned office 01.115 into the best place in the world to work. Thanks for all the laughs and conversations, about work but about life as well, because a PhD is also a life adventure, not just a scientific one. Michele, thanks for more than I can put into words - from the ever-insightful physics discussions to the daily emotional support.

The "topV" group, particularly Alessia Saggio and David Walter, provided pivotal guidance. Alessia helped me set up the analysis, provided help and ideas, and was someone I knew I could always count on whenever I felt stuck (which of course happened many times). David Walter was a brilliant source of inspiration for my work, provided many scripts and knowledge, and had a lot of patience to explain things to me. Thanks also to Jonas Rübenach, for being a good friend and a programming genius, who developed and maintained the *Pepper* framework, that made our work much easier and more efficient.

A particular thanks to the people from the second best office, Alexis, Benno and Gabriele, for making me laugh and forcing me to tell them all the DESY gossip (while providing good coffee).

Some friends deserve a special mention: Ana Ventura Barroso, for being always there, for all the fun moments and the not so fun ones, and for teaching me that there's two kinds of people, normal ones and... cool ones like you. And to Charlie, for bringing beauty into my life, for being my non-physicist best friend, and reminding me that there's more to life than the particles of the SM. He also shared the secret to a successful PhD: "just do it like you're supposed to do it".

I have to dedicate some lines to Nicolas. He taught me a lot, spent hours of his life helping me with my work, always with a smile. His enthusiasm for high energy physics, for statistics, for life in general, helped me find my own.

I acknowledge the collective contribution of all DESY colleagues, secretaries, and staff who supported me in my PhD pursuit.

Even if it's unconventional, I want to thank the city of Hamburg, the place I've had the pleasure to call home for these four years, and from where I keep so many memories.

Finally, I would like to thank my family for the unconditional love, for feeding my curiosity since my childhood and supporting all my choices, even the choice to leave them and go to a rainy place to study strange things.

APPENDIX A

List of triggers used in the $t\bar{t}\gamma$ analysis

Primary dataset	Trigger name (2016)
Single electron	HLT_Ele27_WPTight_Gsf
Single muon	HLT_IsoMu24
	HLT_IsoTkMu24
Double muon	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL (2016B-G and MC)
	HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL (2016B-G and MC)
	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ (only 2016H)
	HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ (only 2016H)
Double electron	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ
	HLT_DoubleEle33_CaloIdL_MW
	HLT_DoubleEle33_CaloIdL_GsfTrkIdVL
Electron muon	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL (2016B-G and MC)
	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL (2016B-G and MC)
	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ (only 2016H)
	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ (only 2016H)

Table A.1: Trigger paths considered in this analysis in 2016.

Primary dataset	Trigger name (2017)
Single electron	HLT_Ele35_WPTight_Gsf
Single muon	HLT_IsoMu27
Double muon	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ (only 2017B) HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass8 HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8
Double electron	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL HLT_DoubleEle33_CaloIdL_MW

Table A.2: Trigger paths considered in this analysis in 2017.

Primary dataset	Trigger name (2018)
Single muon	HLT_IsoMu24
Electron muon	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL
Double muon	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8
"EGamma" (Single and double electron)	HLT_Ele32_WPTight_Gsf HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ

Table A.3: Trigger paths considered in this analysis in 2018.

APPENDIX B

Production card for the tWγ sample in MadGraph

```
import model loop_sm-no_b_mass
generate p p > t w- a [QCD]
add process p p > t~ w+ a [QCD]
```

APPENDIX C

List of triggers used in the exclusive $t\bar{t}$ analysis

Primary dataset	Trigger name
Single electron	HLT_Ele35_WPTight_Gsf_v
Single muon	HLT_IsoMu24_v
	HLT_IsoMu24_2p1_v
	HLT_IsoMu27_v
Double muon	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ
	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass8_v
	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8_v
Double electron	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_v
	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v
Electron muon	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_v
	HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL_DZ_v
	HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_v
	HLT_Mu12_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v
	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_v
	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ_v

Table C.1: Triggers used in the event selection, in the exclusive $t\bar{t}$ analysis in the dilepton channel.