Universidad de Oviedo

Programa de Doctorado en Materiales

Producción de quarks top asociada a bosones en el Modelo Estándar y sus extensiones, en estados finales leptónicos con el detector CMS en el LHC

Production of top quarks in association with bosons in the standard model and its extensions, in leptonic final states with the CMS detector at the LHC

TESIS DOCTORAL

Clara Ramón Álvarez

Enero 2024

Universidad de Oviedo

Programa de Doctorado en Materiales

Producción de quarks top asociada a bosones en el Modelo Estándar y sus extensiones, en estados finales leptónicos con el detector CMS en el LHC

Production of top quarks in association with bosons in the standard model and its extensions, in leptonic final states with the CMS detector at the LHC

TESIS DOCTORAL

Directores de tesis

Dr. D. Francisco Javier Cuevas Maestro

Dr. D. Javier Fernández Menéndez.

Universidad de Oviedo

AUTORIZACIÓN PARA LA PRESENTACIÓN DE TESIS DOCTORAL

Año Académico: 2023/2024

1.- Datos personales de la autora de la Tesis	
Apellidos: Ramón Alvarez	Nombre: Clara
DNI: ▮	Teléfono: ▮ Correo electrónico: UO245843@uniovi.es

2.- Datos académicos
Programa de Doctorado cursado: Materiales
Órgano responsable: Centro Internacional de Postgrado

Título definitivo de la Tesis	
Español: Producción de quarks top asociada a bosones en el Modelo Estándar y sus extensiones, en estados finales leptónicos con el detector CMS en el LHC	Inglés: Production of top quarks in association with bosons in the standard model and its extensions, in leptonic final states with the CMS detector at the LHC
Rama de conocimiento	☐ Arte y Humanidades X Ciencias ☐ Ciencias de la Salud ☐ Ciencias Sociales y Jurídicas ☐ Ingeniería y Arquitectura

3.- Autorización de los Directores y Tutor de la tesis	
D.: Francisco Javier Cuevas Maestro	DNI: ▮
Departamento/Instituto: Departamento de Física	
D.: Javier Fernández Menéndez	DNI: ▮
Departamento/Instituto/Institución: Departamento de Física	
Autorización del Tutor de la tesis	
D.: Francisco Javier Cuevas Maestro	DNI: ▮
Departamento/Instituto: Departamento de Física	

Autoriza la presentación de la tesis doctoral en cumplimiento de lo establecido en el Art. 32 del Reglamento de los Estudios de Doctorado, aprobado por el Consejo de Gobierno, en su sesión del día 20 de julio de 2018 (BOPA del 9 de agosto de 2018)

Oviedo, 5 de diciembre de 2023

Directores de la Tesis

CUEVAS MAESTRO   Firmado digitalmente por CUEVAS ... JAVIER - ... 23:20 +01'00'

Fdo.: Francisco J. Cuevas Maestro
(Tutor de la tesis)

FERNANDEZ ...   Firmado digitalmente por ...

Fdo.: Javier Fernández Menéndez

SR. PRESIDENTE DE LA COMISIÓN ACADÉMICA DEL PROGRAMA DE DOCTORADO EN MATERIALES

FOR-MAT-VOA-009 (Reg.2018)

RESOLUCIÓN DE PRESENTACIÓN DE TESIS DOCTORAL

Año Académico: 2023/2024

1.- Datos personales de la autora de la Tesis

Apellidos: Ramón Alvarez	Nombre: Clara	
DNI: ▓▓▓▓▓▓	Teléfono: ▓▓▓▓▓▓	Correo electrónico: UO245843@uniovi.es

2.- Datos académicos

Programa de Doctorado cursado: Materiales

Órgano responsable: Centro Internacional de Postgrado

Departamento/Instituto en el que presenta la Tesis Doctoral: Departamento de Física

Título definitivo de la Tesis

Español: Producción de quarks top asociada a bosones en el Modelo Estándar y sus extensiones, en estados finales leptónicos con el detector CMS en el LHC	Inglés: Production of top quarks in association with bosons in the standard model and its extensions, in leptonic final states with the CMS detector at the LHC

Rama de conocimiento:

☐ Arte y Humanidades   X Ciencias   ☐ Ciencias de la Salud   ☐ Ciencias Sociales y Jurídicas   ☐ Ingeniería y Arquitectura

Señale si procede:
X Mención Internacional
☐ Mención Industrial
X Idioma de presentación de la Tesis distinto al español
☐ Presentación como compendio de publicaciones

3.- Autorización del Presidente de la Comisión Académica

D: Jaime A. Viña Olay	DNI: ▓▓▓▓▓▓
Departamento/Instituto: Departamento de Ciencia de los Materiales e Ingeniería Metalúrgica	

Resolución: La Comisión Académica del Programa de Doctorado en Materiales en su reunión de fecha 15/12/2023, acordó la presentación de la tesis doctoral a la Comisión de Doctorado, previa comprobación de que la tesis presentada y la documentación que la acompaña cumplen con la normativa vigente, según lo establecido en el Art.32.8 del Reglamento de los Estudios de Doctorado, aprobado por el Consejo de Gobierno, en su sesión del día 20 de julio de 2018 (BOPA del 9 de agosto de 2018)
Además, informa:

	Favorable	Desfavorable
• Mención Internacional	X	☐
• Mención Industrial	☐	☐
• Idioma	X	☐
• Presentación como compendio de publicaciones	☐	☐

Universidad de Oviedo

Justificación

La tesis presentada por Dña. Clara Ramón Alvarez y sus directores de tesis D. Francisco Javier Cuevas Maestro (también tutor) y D. Javier Fernández Menéndez (ambos del Departamento de Física de la Universidad de Oviedo), titulada *"Producción de quarks top asociada a bosones en el Modelo Estándar y sus extensiones, en estados finales leptónicos con el detector CMS en el LHC"*, cumple con los requisitos para ser presentada a la Comisión de Doctorado y para optar al Doctorado Internacional.

Oviedo, 15 de diciembre de 2023

Presidente de la Comisión Académica del Programa de Doctorado en Materiales

VIÑA OLAY JAIME AURELIO — Firmado digitalmente por VIÑA OLAY JAIME

Fdo.: Jaime A. Viña Olay

Contra la presente Resolución, podrá interponer recurso de alzada ante el Rectorado, en el plazo de un mes, a partir del día siguiente al de la presente notificación, de conformidad con el art. 122 de la Ley 39/2015, de 1 de octubre, de Procedimiento Administrativo Común de las Administraciones Públicas

SR. DIRECTOR DEL CENTRO INTERNACIONAL DE POSTGRADO

RESUMEN DEL CONTENIDO DE TESIS DOCTORAL

1.- Título de la Tesis	
Español: Producción de quarks top asociada a bosones en el Modelo Estándar y sus extensiones, en estados finales leptónicos con el detector CMS en el LHC	Inglés: Production of top quarks in association with bosons in the Standard Model and its extensions, in leptonic final states with the CMS detector at the LHC

2.- Autora	
Nombre: Clara Ramón Alvarez	DNI:
Programa de Doctorado: Materiales	
Órgano responsable: Centro Internacional de Postgrado	

RESUMEN (en español)

En esta tesis se han presentado medidas de producción de quarks top en asociación con un bosón de Higgs o un bosón W en estados finales con leptones. Los datos utilizados corresponden a colisiones protón-protón registradas por el experimento CMS en el LHC durante el Run-2, a una energía en centro de masas de 13 TeV. Se han presentado medidas de precisión en el contexto del Modelo Estándar (ME) y, en algunos casos, también se han realizado interpretaciones de dichas medidas en el contexto de teorías Más Allá del Modelo Estándar.

En esta tesis se ha estudiado la producción asociada de un bosón de Higgs con un par de quarks top-antitop ($t\bar{t}H$) y un solo quark top (tH), usando estados finales donde el bosón de Higgs se desintegra en WW, $\tau\tau$ y, residualmente, ZZ. Estos estudios han permitido observar la producción de $t\bar{t}H$ con una significancia observada (esperada) de 4.7 (5.2) desviaciones estándar. Además, estos procesos permiten estudiar el acoplamiento entre el top quark y el bosón de Higgs a primer orden en teoría de perturbaciones. La medida anterior ha sido interpretada en el contexto de acoplamientos anómalos, estableciendo límites en el modificador del acoplamiento entre el Higgs y el top (κ_t), que debe estar en los de los intervalos $-0.9 < \kappa_t < -0.7$ y $0.9 < \kappa_t < 1.1$ al 95% de nivel de confianza, asumiendo que el resto de los acoplamientos son los del ME.

El bosón de Higgs predicho en el ME es un pseudo-escalar invariante bajo transformaciones de Carga-Paridad (CP). En esta tesis, el proceso $t\bar{t}H$ se ha utilizado para buscar violación de la simetría CP en el acoplamiento entre el quark top y el bosón de Higgs. Se han establecido regiones de confianza bidimensionales en κ_t y $\tilde{\kappa}_t$, que son los modificadores de acoplamiento de Yukawa del quark top CP-par y CP-impar, respectivamente. Los resultados se han combinado con otras medidas para incluir sucesos donde los bosones de Higgs se desintegran en ZZ y $\gamma\gamma$, estableciendo los límites en κ_t y $\tilde{\kappa}_t$ a los intervalos (0.86, 1.26) y (-1.07, 1.07) al 95% de C.L., respectivamente. El escenario puramente CP-impar se ha excluido con 3.7 desviaciones estándar.

También se ha presentado la medida de la producción asociada de un bosón W con un par de quarks top-antitop ($t\bar{t}W$). La sección eficaz medida es de 868 ± 40 (estad.) ± 51 (sist.) fb. Este valor es mayor que el predicho en el modelo estándar, pero es consistente dentro de dos desviaciones estándar y representa la medida más precisa de estas secciones eficaces hasta la fecha.

La reconstrucción, identificación, aislamiento y selección de leptones son elementos importantes en todos los estudios presentados en este trabajo. En esta tesis se ha desarrollado un nuevo algoritmo basado en técnicas de análisis multivariante. Se ha mostrado que el nuevo algoritmo es más eficiente que el utilizado durante el Run-2 y tiene un comportamiento más estable en condiciones con un alto número de colisiones simultáneas.

RESUMEN (en Inglés)

This thesis presents measurements where top quarks are produced in association with a Higgs boson or a W boson in final states with leptons. The data used in these studies corresponds to proton-proton collisions recoded by the CMS experiment at the LHC during Run-2, at a centre-of-mass-energy of 13 TeV. Standard Model (SM) precision measurements are reported and, in some cases, interpretations on the context of Beyond the Standard Model theories are also carried out.

The measurement of the associated production of a Higgs boson with pair of top-antitop quarks (t$\bar{\text{t}}$H) and a single top (tH) is reported in final states targeting the Higgs decay to WW, $\tau\tau$ and, residually, ZZ. These studies allow to observe t$\bar{\text{t}}$H production with an observed (expected) significance of 4.7 (5.2) standard deviations. Additionally, these processes allow to study the top-Higgs coupling at tree-level. The measurement is interpreted in the context of anomalous couplings. The Higgs-top coupling modifier (κ_t) is found to be in either one of the intervals − 0.9 < κ_t < − 0.7 and 0.9 < κ_t < 1.1 at 95% confidence level, assuming the rest of the couplings to be those of the SM.

The SM Higgs boson is a pseudo-scalar invariant under Charge-Parity (CP) transformations. In this thesis t$\bar{\text{t}}$H process is used to search for CP violation in the top-Higgs coupling. Two-dimensional confidence regions are set on κ_t and $\tilde{\kappa}_t$, which are respectively defined as the CP-even and CP-odd top-Higgs Yukawa coupling modifiers. The results are combined with other measurements to include events where the Higgs bosons decays to ZZ and $\gamma\gamma$, allowing to set limits on the coupling modifiers, constraining κ_t and $\tilde{\kappa}_t$ to be within (0.86, 1.26) and (-1.07,1.07) at 95% C.L., respectively. The pure CP-odd scenario is excluded with 3.7 standard deviations.

The measurement of the associated production of a W boson with pair of top-antitop quarks (t$\bar{\text{t}}$W) is also reported in this thesis. The cross section is measured to be to be 868 ± 40 (stat.) ± 51 (syst.) fb. The reported value is larger than but consistent with the standard model predictions within two standard deviations and represent the most precise measurement of these cross sections to date.

Lepton reconstruction, identification, isolation and selection are important items in all the studies presented in this work. In this thesis a new algorithm, based on multivariate analysis techniques is developed and reported. The new algorithm is found to be more performant than the one used during Run-2 and shows a more stable behaviour in conditions with high number of simultaneous collisions.

SR. PRESIDENTE DE LA COMISIÓN ACADÉMICA DEL PROGRAMA DE DOCTORADO EN MATERIALES

En memoria de mi abuelo José Luis,
que siempre me escuchaba con atención
cuando hablaba de física.

Agradecimientos

Redactar estas líneas parecería la tarea más sencilla del proceso de escritura de una tesis doctoral y, sin embargo, no es así. Puede que porque, de alguna manera, es ponerle el final a un momento vital en el que he estado muy a gusto y eso me da una cierta sensación de despedida. Han sido 4 años y medio de continuo aprendizaje, de retos y de cambio, pero con lo que me quedo es con la calidad humana de las personas con las que he tenido la suerte de coincidir. La ciencia es un trabajo en equipo y debo dar las gracias a todas aquellas personas que han formado parte de este camino.

Quiero dar las gracias a mis directores de tesis Javier Cuevas y Javier Fernández. En primer lugar, por depositar su confianza en mí, ver mi potencial y darme la oportunidad de comenzar el doctorado. Sin su continua orientación, liderazgo y paciencia durante estos años, nunca habría sido capaz de desarrollar en su plenitud esta tesis. Además, he tenido la suerte de contar con el mejor grupo que podía haber imaginado, y no solo por las galletas. Doy las gracias a Bárbara, Kike, Santi y Chiqui de los que he aprendido –por poner sólo algunos ejemplos– a trabajar en equipo, mejorar mi sentido crítico o ser más ordenada en mis tareas. Sus aportaciones me han hecho crecer como científica y docente. Todos sois para mí un ejemplo a seguir. También quiero agradecer a los jóvenes-ya-no-tan-jóvenes que han pasado (y algunos regresado) al grupo de HEP-Uniovi: Xuan, Sergio, Carlos, Víctor, Nicolò y Pietro. Gracias por enseñarme a dar mis primeros pasos en este mundillo y hacer que sobreviviera, al menos durante el primer año. Gracias también por las interminables charlas sobre física y muchas otras cosas (trenes y castañas sobre todo), por escucharme siempre y darme vuestros consejos. Seguís siendo una fuente de inspiración y aprendo de vosotros cada día. No puedo olvidarme de los jóvenes del grupo. En primer lugar de la "inminente doctora" Andrea Trapote; desde la carrera sabía que trabajar con ella era un placer, pero no podía haber imaginado una mejor compañera para esta aventura: conferencias, estancias, momentos buenos y no tan buenos. Al final hemos salido del "estanque"; sin ti no sé si hubiera podido. Luego tenemos a Carlos y Alejandro, para completar el cuarteto cómico. Gracias por tantos buenos momentos, por siempre echar una mano en lo profesional y personal. Alejandro, a ver a quién le robas los pañuelos de papel ahora. Carlos, confío en tí para cuidar del cactus cuando Andrea y yo ya no estemos. No puedo dejar fuera a Pelayo que, aunque me chinche, es capaz de sacarme siempre una sonrisa. Puede que sin vosotros sí hubiera salido adelante (tenía que poner una pullita), pero esto no habría sido lo mismo. Ahora tenéis la tarea de cuidar de las nuevas generaciones: Javier Prado, Javier del Riego, Miguel y Jorge. Espero que entre todos no os peleéis mucho por los puestos del despacho.

Let me switch to English to thank the people from around the world with whom I've had the privilege of collaborating. Firstly, I want to thank our colleges in Ghent and Louvain, that worked with us on the $t\bar{t}W$ analysis: Andrea, Didar, Pieter, Tu Thong, Kirill, Joscha and, more recently Luka and David. Your dedication, and insightful discussions were key on our success. Secondly, I want also to acknowledge all the brilliant people in the $t\bar{t}H$ multilepton team; in particular Marco, Xanda, Aurelijus, Sascha, Angela and Anshul. Thank you for your effort, team work all the interesting discussions. I can't forget the people from the muon POG: I want to thank Georgios, Jan and Federica Primavera for their leadership. As well as Federica Simone and Federica Colombina, it was a pleasure to work side by side with them.

Además, esta tesis me ha permitido conocer a gente con la que, si bien no he trabajado (o no de forma tan cercana), ha sido un auténtico placer coincidir. Celia, Pablo, Clara, Sergio, Rubén, Pedro, Dalia, Won, Marina (sorry for the Spanish, but at this point I think both of you should be proficient[1]), Gorka, Emanuela, Adrián y David. Gracias por los cafés, cervezas, cenas, risas, charlas, consejos y demás apoyo.

A mis amigas Ceci, Inés y Carla, que siempre han estado ahí celebrando mis éxitos y apoyándome en mis momentos malos; aunque a veces estemos lejos os siento muy cerca. Esta tesis también lleva un pedacito vuestro.

Finalmente, a mis padres Teo y María Jesús y a mi hermano Tomás. Gracias por ser mi continuo soporte desde que tengo memoria, y ya incluso antes. Por dejarme total libertad para tomar mis decisiones pero siempre orientándome y apoyándome con cariño. Gracias por aguantar mis momentos de agobio, que en estos años no han sido pocos, por alegraros incluso más que yo de cada pequeña victoria, por recogerme o llevarme aeropuerto en cada viaje, por tener siempre lista la tortilla de los viernes... podría seguir enumerando. Gracias a toda mi familia: mis abuelos y abuelas, tías y tíos, y primas y primo, por darme siempre vuestro cariño y por interesaros alguna vez por estas cosas tan raras que hago.

---

[1] I'll translate: thank you for all the coffee breaks, shared beers and dinners, laughs, small talks and advice you gave me

Contents

Introduction

Physics has always aimed to explain the Universe with the highest grade of detail and precision possible and at all scales. This is an enormous task, which is usually divided in smaller, interconnected fields of study. Specifically, the realm of particle physics aims to explain which are the fundamental constituents of the Universe and how they interact. In this context, the Standard Model (SM) is the most successful theoretical framework to provide precise predictions. It was developed during the 60s and it has been tested continuously since then, using several experimental devices. Currently, the state-of-the-art facility to study the SM (and its extensions) is the Large Hadron Collider (LHC), a proton-proton accelerator providing the highest center-of-mass energy up to date, and the experiments in which the pp collisions are recorded: ATLAS, CMS, ALICE and LHCb.

It is usually stated that the study of the SM was culminated in 2012, when the Higgs boson, predicted 50 years before [1–6], was found by ATLAS and CMS [7, 8]. Nevertheless, this was only the starting point for an extensive experimental program constructed to fully understand the properties of the newly discovered particle. In this thesis, pp collision data collected with the CMS experiment are used to study the Higgs boson coupling to the top quark. This quark is the most massive elementary particle known up to date. As a consequence of its large mass, the top quark presents a unique feature: it decays before hadronization, being the only quark that can do so.

Run-1 of the LHC (2009-2012) allowed to discover the Higgs boson and measure its coupling to vector bosons. In addition, precision measurements related to top quark production were also carried out.

Run-2 (2015-2018) provided an unprecedented amount of data, collected at a center-of-mass energy of 13 TeV. The data available at the end of this data-taking period corresponded to an integrated luminosity of 138 fb^{-1}, allowing to study low cross section processes not measured before. With the full Run-2 dataset, the coupling of the Higgs boson to fermions, called Yukawa coupling, was investigated. The coupling of the Higgs boson to the third generation of fermions was proven and evidence for the coupling to the second generation was established.

In this thesis the dataset recorded during Run-2 is exploited to study the Higgs boson coupling to the top quark. This is the largest Yukawa coupling and deviations from

the SM prediction would indicate the presence of physics beyond the SM (BSM). The top-Higgs interaction can be directly measured at tree level only by studying the associated production of a Higgs boson with a top quark (tH) or a top-antitop quark pair (t$\bar{\text{t}}$H). I have worked on the measurement of the t$\bar{\text{t}}$H and tH inclusive production cross sections using final states with multiple leptons. The improvements in the experimental techniques applied on Run-2 data, some of which were developed in this work, allowed to observe the t$\bar{\text{t}}$H process in final states with leptons, as reported in Ref. [9]. Regarding the top-Higgs coupling, I have performed the search for charge-parity (CP) violation in the top-Higgs boson coupling. Results are published in Ref. [10], showing good agreement with the SM. A combination with other Higgs boson final states is included in the study, excluding the pure CP-odd scenario with 3.7 standard deviations (s.d.).

When studying t$\bar{\text{t}}$H, other t$\bar{\text{t}}$X processes arise as background. The t$\bar{\text{t}}$Z process was measured during Run-2 with a good precision of the order of 10%. However, the comprehension of t$\bar{\text{t}}$W production is more challenging. This process was studied firstly with part of the Run-2 dataset, showing a value for the cross section higher than the SM prediction, but consistent within uncertainties, both by ATLAS and CMS [11, 12]. It was also investigated in dedicated control regions in the context of t$\bar{\text{t}}$H dedicated measurements [9, 13, 14] and, later in t$\bar{\text{t}}$t$\bar{\text{t}}$ dedicated studies [15, 16] by both collaborations, as well. Measurements using the full Run-2 dataset showed tension between the production rate measured and the theory predictions [17–19]. Such prediction have shown that higher order electroweak contributions have a non-negligible contribution to the cross section. As a result, the modelling of t$\bar{\text{t}}$W process is an open question under discussion within the community, that would benefit from dedicated measurements of the inclusive and differential cross sections. The t$\bar{\text{t}}$W process also allows to study the top-electroweak coupling and has the unique feature that diagrams presenting gluon-gluon initial states are suppressed. During this thesis I have measured the t$\bar{\text{t}}$W inclusive production cross section using final states with two or three leptons. The result, published in Ref. [20], reduces the total uncertainty on the cross section by a factor of two with respect to the previous CMS result [12] and is in agreement within 2 s.d. with the latest NNLO (QCD) + NLO (EWK) SM prediction [21].

All studies reported in this thesis use final states with leptons, hence lepton reconstruction, identification, isolation, and selection are a key elements of the analysis. The performance of the lepton reconstruction and selection must be studied in each data-taking period, as it is crucial for the measurements performed by the CMS experiment. Additionally, identification and selection techniques need to be constantly improved, with the goal of achieving better efficiency and reducing the background rate. These improvements will ultimately result in the reduction of the systematic

uncertainties in measurements using leptons in the final state. This thesis was carried out during the LHC long shutdown after the end of the Run-2 and at the beginning of Run-3. In order to maintain or improve the performance obtained during Run-2, given the harsher conditions of the new data-taking period, with a higher number of simultaneous collisions, I have developed a new identification algorithm for muons based on machine learning techniques. The results of this new algorithm are published in Ref. [22].

Final states with multiple leptons in combination with missing transverse energy can be used as signature to search for new physics. During this thesis I have been involved in searches using events with two opposite-sign same-flavor (OSSF) leptons and missing transverse momentum in the final state. The results and limits obtained were published in Ref. [23]. Additionally, during the development of this thesis, the Run-3 of the LHC started; giving me the opportunity of performing the first CMS measurement of the quark anti-quark pair production cross section at the new energy of 13.6 TeV [24].

The results obtained in this thesis have been published in various articles. These publications and the responsibilities I had in the different CMS groups are summarized in the diagram displayed in figure 1.

The thesis is organized as explained in the following lines. Chapter 1 gives an overview of the theoretical framework on which the results of this thesis are based: the SM. It includes a brief discussion of its limitations and a summary of the Higgs boson and top quark sectors of the SM at the energy scale of the LHC. Chapter 2 describes the complex instruments used to produce and collect the data analyzed in this thesis: the LHC collider and the CMS experiment, and details about how particles are reconstructed in the detector are provided. Chapter 3 focuses on lepton reconstruction, selection and identification techniques in CMS, as this is one of the core elements of the studies presented in the thesis. Chapter 4 discusses the production of a pair of top quarks in association with a W boson or a Higgs boson. Common items in the measurement of both processes are given in this chapter. Chapter 5 describes the measurement of $t\bar{t}W$ cross section, while chapter 6 describes the measurement of the $t\bar{t}H + tH$ cross section in final states with multiple leptons, including an interpretation of the results of the top-Higgs coupling in terms of the CP symmetry. Finally, the conclusions are presented.
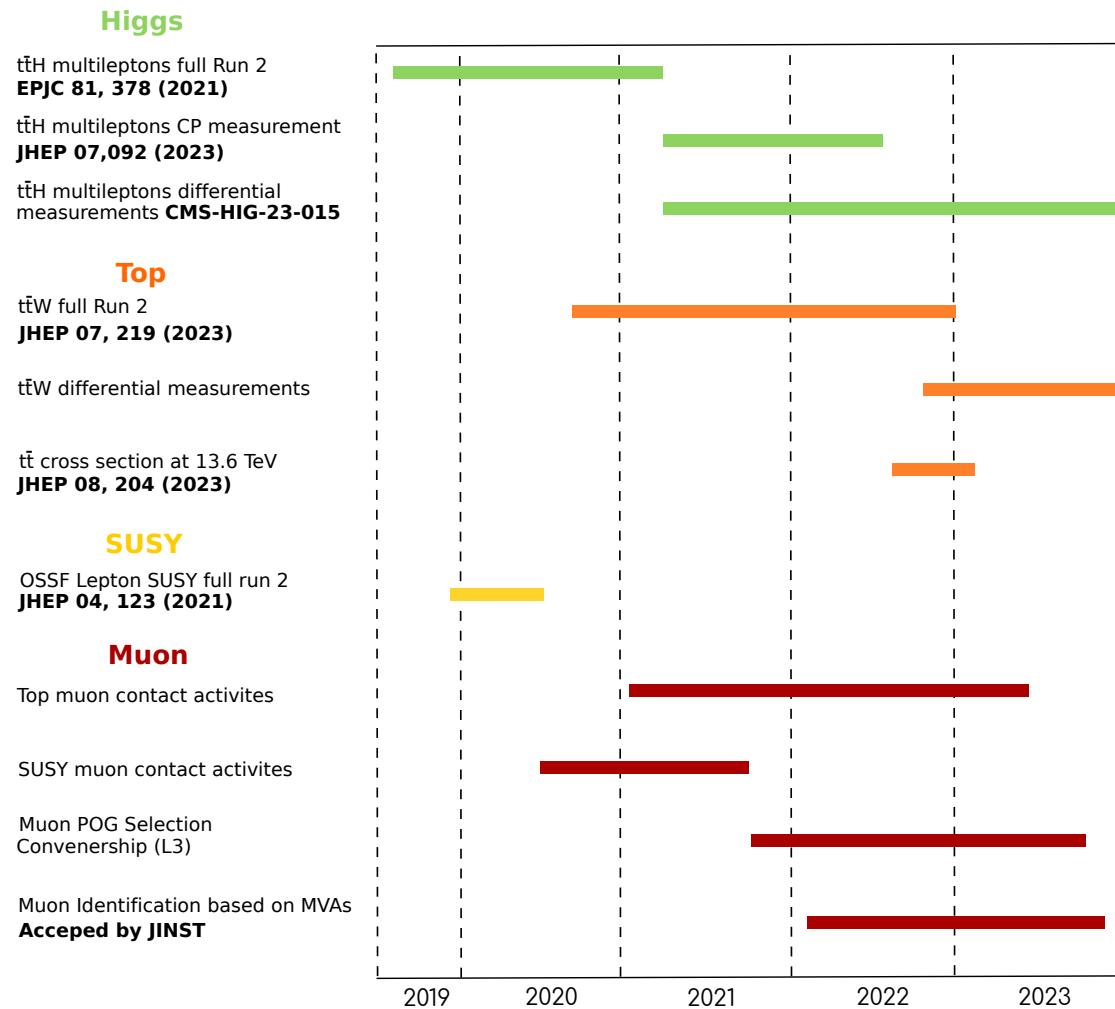
FIGURE 1: Timeline diagram summarizing the activities and publications carried out during this thesis.

1

Theoretical framework

In this chapter a brief review of the theoretical framework relevant in high energy physics, the standard model (SM), is given. Afterwards, the limitations of such framework are briefly discussed and some extensions of the SM are presented. Finally, I explain the phenomenology of proton-proton (pp) collisions.

1.1 Standard Model

Currently, the SM is the theoretical framework that better describes the elementary particles and their interactions [25]. It accurately portrays how the fundamental constituents of matter (and antimatter) interact among themselves across a wide range of energies, spanning from few eV to the TeV scale.

Within the SM, elementary particles are described as excited states of quantum fields, which are defined using the mathematical framework of Quantum Field Theories (QFTs). The final objective of the SM is to provide quantitative predictions about the real world; in order to do so, some conditions need to be fulfilled by the QFTs: they must be both renormalizable and gauge invariant. Additionally, the model must incorporate the principles of quantum mechanics and special relativity. The fields described in the SM can be fermionic (non-integer spin) or bosonic (integer spin), this allows to classify particles as fermions or bosons.

Gauge (or vector) bosons have spin 1 and are responsible for the interactions, i.e. they act as mediator of a certain force. The gauge bosons in the SM are: the photon (γ), the $W^\pm$ and Z bosons and the gluon (g). The photon is responsible for the electromagnetic force and it is massless; therefore, the range of action of this force is infinite. $W^\pm$ and Z bosons are the mediators of the weak interaction, both have mass and the $W^\pm$ boson also bears electric charge. Finally, the eight double-coloured massless gluons act as mediators of the strong force. The range of interaction for the gluon, the W and Z
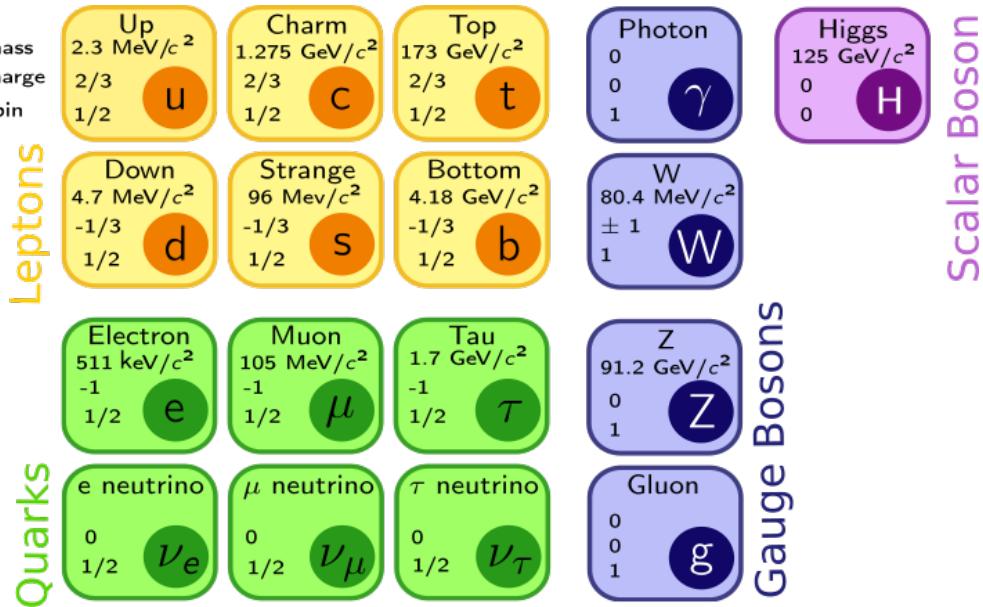
FIGURE 1.1: Summary of SM particles including its properties: spin, charge and masses. Information extracted from [26].

bosons is finite. Additionally, gluons self-interact, which leads to the phenomenon of colour confinement, explained in detail later. Gravity is not included in the SM description, but its effect is insignificant at the energy scales used in particle physics. The Higgs boson (H), with spin 0, is the only scalar boson in the SM. It was introduced in the Higgs-Brout-Englert mechanism, that explains how the $W^{\pm}$ and Z bosons, as well as fermions, acquire mass.

Fermions engage in interactions mediated by bosons and have spin 1/2. These particles can be categorized into two groups depending on whether they possess colour charge or not: quarks and leptons. Fermions are also classified in three distinct generations, with those within the first family being the constituents of visible and stable matter. Quarks, being colour-charged fermions, are sensitive to the strong interaction. They also have electric charges, that may take values of either +2/3 (for up, charm, and top quarks — u, c, t) or -1/3 (for down, strange, and bottom quarks — d, s, b). There are three electrically charged leptons: the electron, muon and tau (e, μ, τ) and three neutral leptons known as neutrinos (ν_e, ν_μ, ν_τ), which only interact weakly.

Each particle has a corresponding antiparticle, with the same mass and opposite charges. A illustrative representation of the particle postulated in the SM can be seen in Fig. 1.1.

The interactions and propagation of fields in the SM (and in any QFT) is fully described by a Lagrangian density, $\mathcal{L}(\psi_i, \partial\psi_i, x^\mu)$, where ψ_i are the fields, defined as the generators of a suitable representation of the Lorentz group, and x^μ the spacetime

coordinates. In particular, fermion fields are represented by spinors, gauge bosons by vectors and the Higgs boson is represented by a scalar field. The crucial items of any QFT —not discussed so far— are the symmetries postulated in the model: the SM Lagrangian must satisfy local gauge symmetries. These symmetries are responsible for the conservation of charges and imply the appearance of fields that corresponds to bosons which mediate the interactions. In particular, two interactions or sectors are present in the SM Lagrangian: the Electroweak (EWK) and the strong interaction, described by quantum chromodynamics (QCD). Each of them is represented by a local gauge symmetry. EWK interaction is invariant under transformation of the $SU(2)_L \times U(1)_Y$[1] group, while the QCD is invariant under the 3-dimensional special unitary group $SU(3)_C$[2]. In the following, I will categorize different portions (sectors) of the SM lagrangian based on the interaction they describe.

1.1.1 Electroweak sector

The electromagnetic and weak forces are described together in this sector of the Lagrangian. The electroweak theory was developed by Glashow, Weinberg, and Salam [27–29], and it is a unification of the electromagnetic and weak interactions. The electromagnetic interaction is described by the abelian symmetry group $U(1)_Y$, while the weak part of the interaction is represented by the non-abelian gauge group $SU(2)_L$. The weak interaction possesses the distinctive characteristic of violating parity, this fact is accommodated in the theory by using the chirality property of the fermion fields. Chirality is a Lorentz-invariant quantity corresponding to the eigenvalues of the operator: $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$, where γ^i are the Dirac matrices. The spinors representing the fermions, can be projected onto left (ψ_L) and right (ψ_R) handed spinors as follows:

$$\psi_L = \frac{1}{2}(1 - \gamma^5)\psi, \qquad \psi_R = \frac{1}{2}(1 + \gamma^5)\psi. \tag{1.1}$$

The $SU(2)_L$ part of the theory acts only on left-handed fermions. As a result, left-handed fermion fields transform as doublets under $SU(2)_L$, while right-handed fermion fields do it as singlets of the $SU(2)_L$ group. For first generation quarks, it can be written:

$$L = \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \quad u_R, \ d_R. \tag{1.2}$$

---

[1] The "L" refers to left and the "Y" to the hypercharge that we will describe later
[2] The "C" refers to the colour charge, which is the charge associated to the strong force

For other fermions it would be equivalent, except for neutrinos, which do not have a right-handed component.

The $SU(2)_L$ symmetry implies the existence of three gauge bosons, whose vector fields can be denoted as W_μ^i with $i = 1, 2, 3$ and μ the index going over spacetime components. By introducing them, the gauge transformation under which the Lagrangian is invariant can be defined, and the associated conserved charge is called weak isospin: I_3. The $SU(2)_L$ invariant Lagrangian can be written as:

$$\mathcal{L} = -\frac{1}{4} W_{\mu\nu}^i W_i^{\mu\nu} + i\bar{u}_R \gamma^\mu D_\mu u_R + i\bar{d}_R \gamma^\mu D_\mu d_R + i\bar{L}\gamma^\mu D_\mu L \tag{1.3}$$

$W_{\mu\nu}^i$ and D_μ are defined as:

$$W_{\mu\nu}^i = \partial_\mu W_\nu^i - \partial_\nu W_\mu^i - g_W \epsilon^{ijk} W_\mu^j W_\nu^k \tag{1.4}$$

$$D_\mu = \partial_\mu - ig_W \frac{\sigma_i}{2} W_\mu^i \tag{1.5}$$

where σ_i are the Pauli matrices and g_W is the coupling constant in the $SU(2)_L$ group.

An analogous development is performed in the $U(1)_Y$ group, yielding a vector field B^μ. The corresponding lagrangian density is:

$$\mathcal{L} = -\frac{1}{4} B^{\mu\nu} B_{\mu\nu} + \sum_{\psi = u_R, d_R, L} \bar{\psi}(i\gamma_\mu \partial^\mu - Y g_Y B_\mu)\psi \tag{1.6}$$

Where Y is the hypercharge, which is the current conserved by this symmetry. It is related to the isospin and the electric charge by the expresion: $Y = (Q - I_3)$. g_Y is the coupling constant associated to the vector field, and $B^{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$.

The $\mathcal{L}_{EWK}$ is constructed combining both Lagrangian densities. It is worth noting that, at this point, the fields present are massless, as any mass term of the form $m\psi\bar{\psi}$ will not be invariant under $SU(2)$ transformations. This will be solved later by the introduction of the electroweak spontaneous symmetry breaking, explained further in this chapter.

1.1.2   Strong sector

The theory in the SM that describes the strong interaction is Quantum Chromodynamics (QCD) and was developed by Fritzsch, Leutwyler and Gell-Mann [30], based on the

previous work of Yang and Mills [31]. It is a gauge theory symmetric under $SU(3)_C$ Lie group, which is the 3-dimensional special unitary group. The conserved quantity associated is the colour charge. Quarks are the only fermions presenting this charge, hence, they are represented as triplets; while leptons have the trivial representation.

$SU(3)_C$ group is non-abelian and has 8 generators, which are represented by the Glenn-Mann matrices (T^a). These generators are the eight vector bosons G_μ^a, called gluons, and defined as $G_\mu^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + g_S f^{abc} G_\mu^b G_\nu^c$. In the previous expression, f^{abc} are the structure constants of the $SU(3)_C$ group, where the a, b, c index run over the 8 types of gluons. It is worth noting that the last term of G_μ^a —introduced due to the fact that the group is non-abelian— is responsible for the self-coupling of the gluons.

Using this notation, the QCD Lagrangian can be written as:

$$\mathcal{L}_{QCD} = \bar{\psi}(i\gamma^\mu \partial_\mu - g_S \gamma^\mu T_a G_\mu^a - m)\psi - \frac{1}{4}G_{\mu\nu}^a G_a^{\mu\nu} \tag{1.7}$$

Where g_S is the strong coupling constant, later on in this thesis the used notation for this constant will be: $\alpha_S = g_S^2/4\pi$. One consequence of the $SU(3)_C$ group is that α_S runs with the energy scale of the interaction (Q) as:

$$\alpha_S(Q^2) \approx -\frac{1}{ln(Q^2/\Lambda)_{QCD}^2}. \tag{1.8}$$

The previous expression implies that the strong interaction will become stronger at low energy scales. As a consequence, quarks and glouns will not be observed as free particles; instead, they will form bounded colourless states called hadrons. This phenomenon is known as quantum colour confinement. The top quark is the exception: given its high mass it decays before hadronizing. In the high energy regime, the strength of the strong interaction is reduced, and the opposite phenomenon, asymptotic freedom, is predicted.

1.1.3 Symmetry breaking

As already stated, imposing invariance under local gauge transformations of the $SU(3)_C \times SU(2)_L \times U(1)_Y$ group implies that no mass terms are allowed in the Lagrangian. Hence, all fermions and bosons should be massless. This contradicts the experimental observations, therefore there is a clear need to modify the theory so it can accommodate massive fermions and bosons.

This is achieved by introducing the Higgs-Brout-Englert mechanism [1–6], which describes the Spontaneous Symmetry Breaking (SSB) of the EWK symmetry.

We consider a SU(2)$_Y$ doublet of scalar field:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tag{1.9}$$

where ϕ^+ is a positively-charged scalar field and ϕ^0 is a neutral scalar field.

Whit this we construct the Lagrangian:

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

Where the potential $V(\phi) = \zeta^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2$ with $\zeta^2 < 0$ and $\lambda > 0$, this potential is represented in Fig. 1.2. From the potential shape it can be observed that the symmetric solution is unstable and there are a set of indistinguishable minima at the same potential energy to which the system can evolve. This evolution will spontaneously break the symmetry and establish one of the minima as the physical one. As a result the unitary gauge is assumed, and can be written as:

$$\phi = \begin{pmatrix} 0 \\ \sqrt{\frac{-\zeta^2}{\sqrt{2}\lambda}} + \phi(x) \end{pmatrix} \tag{1.11}$$

where $\phi(x)$ represents the degree of freedom that remains after the symmetry breaking. It can be interpreted as a scalar field with mass: $\sqrt{2\lambda}v$, i.e. the Higgs boson. After adding this field to the Lagrangian, it can be shown that new combinations of the W_μ and B_μ fields with the components of the scalar field yields three massive and one massless new fields. These fields correspond to the bosons introduced at the beginning of this chapter. In particular, W_μ^3 and B_μ are mixed, they can be disentangled by applying a rotation:

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos \theta_W & -\sin \theta_W \\ \sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} W_\mu^{(3)} \\ B_\mu \end{pmatrix} \tag{1.12}$$

Where θ_W is the Weinberg angle that is defined by the following relationship: $\tan \theta_W = g_Y / g_W$.
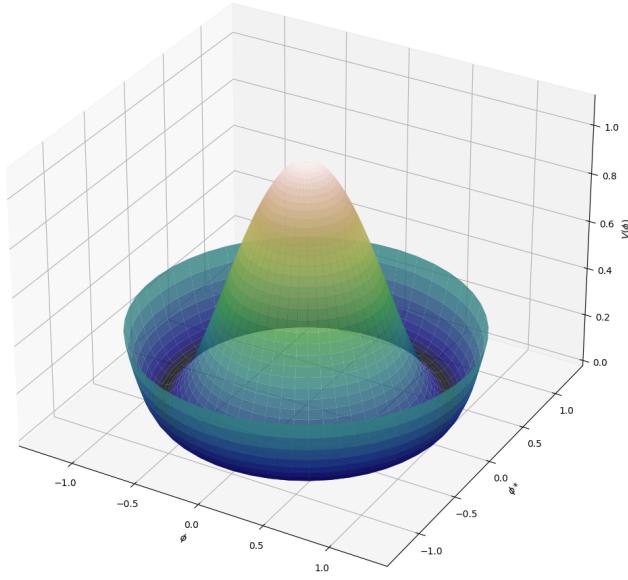
FIGURE 1.2: Representation of Higgs potential $V(\phi) = \zeta^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2$ with $\zeta^2 < 0$ and $\lambda > 0$.

The A_μ field defined is the photon (γ) field while the Z_μ corresponds to the Z^0. Both of them are neutral and the Z_μ field has an associated mass: $m_Z = \frac{v}{2}\sqrt{g_Y^2 + g_W^2}$.

It can also be shown that the W_μ^1 and W_μ^2 are mixed following the linear combination:

$$W^\pm = \frac{1}{\sqrt{2}}(W_\mu^{(1)} \mp W_\mu^{(2)}), \tag{1.13}$$

which describes the observed $W^\pm$ with an associated mass: $m_W = \frac{v}{2}g_W$.

1.1.4   Yukawa coupling

Fermion masses can also be derived from the interaction between the scalar doublet (Higgs) and the fermions. The most general Lagrangian density describing that interaction for a single generation of quarks is the following:

$$\mathcal{L} = -y_u \bar{L}\phi^c u_R - y_d \bar{L}\phi d_R + \text{h.c.}, \tag{1.14}$$

where y_u and y_d are called Yukawa couplings and are free parameters of the model. And $\phi^c = i\sigma^2 \phi^\dagger$. Expanding again around the vacuum expectation value of the scalar field defined in Eq. 1.11, a mass term is found with the following form:

$$\frac{y_\psi v}{\sqrt{2}} \psi \bar{\psi}, \tag{1.15}$$

hence, the mass of fermions equals $m_f = y_f v / \sqrt{2}$ for charged leptons and quarks. The non existence of right handed neutrinos implies that, according to this mechanism, neutrinos remain massless.

1.1.5 The CKM Matrix

So far we have only considered one generation of fermions, while the SM has three flavours of fermions. Nevertheless, the description used up to this point can be extend to account for all quarks, by using the following notation: Q_{Li}, u_{Ri} and d_{Ri}, where i is an index running over the 3 flavours. Then, equation 1.14 can be rewritten as follows:

$$\mathcal{L} = -Y_{ij}^u \bar{Q}_{Li} \phi^c u_{Rj} - Y_{ij}^d \bar{Q}_{Li} \phi d_{Rj} + \text{h.c.}, \tag{1.16}$$

where Y_{ij}^u and Y_{ij}^d are 3x3 complex matrices and the indices i and j run over the three flavours. The Y^u and Y^d matrices are not diagonal, leading to possible flavour mixing between generations. Consequently, fermion fields do not have well defined masses. To address this issue, the fermion fields need to be redefined as eigenstates of the Yukawa matrices. For quarks, a basis of fermion fields that are eigenstates of both Yukawa matrices and of electroweak interactions is not achievable. Therefore, the interaction of the W bosons does not happen among same flavour fermions; instead, it connects the quark mass eigenstates from different families. This phenomena is parameterized through the Cabibbo-Kobayashi-Maskawa (CKM) matrix [32]:

$$V_{CKM} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb}. \end{pmatrix} \tag{1.17}$$

This matrix is unitary and its elements are free parameters of the theory that need to be measured. Elements in the diagonal are found to be larger than their non-diagonal counterparts. In addition, the mixing in the first and second generation is larger than for the third generation, where the term is large enough to assure that most of the top quarks decays produce a b quark [26].

For leptons, as in the SM neutrinos are massless, this effect can be handled by redefining the three flavours of neutrinos with no additional criteria needed. In models

where neutrinos are not massless, a parametrization analogous to the CKM matrix is used: the PMNS (Pontecorvo-Maki-Nakagawa-Sakata) matrix [33].

1.2 Limitations of the SM

The SM has been tested and confirmed via observations since its formulation 5 decades ago, this makes the SM one of the most successful theories in physics. The culmination of that effort was the discovery of the Higgs Boson in 2012 by ATLAS [8] and CMS [7]. Nevertheless, there are still some topics for which the SM is not able to provide a description that explains satisfactorily the observations. In this section, some of these points are discussed along with few proposed extensions of the SM — Beyond the Standard Model theories (BSM) — that aim to accommodate these phenomena.

Hierarchy problem

In the SM, loop contributions from fermions and bosons to the Higgs propagator are found to be significant. These corrections, dominated by the top quark, can be summed up into a correction to the Higgs bare mass which is of the order of the plank scale ($\approx 10^{19}$ GeV) and proportional to the Yukawa couplings of the SM fermions: $\Delta(m_H)^2 \sim Y_x^2 \Lambda_P^2$. It is possible that such large corrections cancel out with other corrections, leading to the observed Higgs mass value of around 125 GeV. In particular, corrections from bosons contribute with the opposite sign to that of fermions. Nevertheless, a complete cancellation is unexpected, and this might be an artifact of the model, an indication that there should be a simpler explanation such as a yet undisclosed symmetry of the model.

Gravity

As stated in Section 1.1, the SM does not account for a description of gravity as a QFT. Nevertheless, at the energy scales used in the current colliders (of the order of TeV), this is not really a limitation, because gravity effects to the processes of study are negligible, and only become significant at the Planck scale.

Neutrino masses

In the SM formulation given above, neutrinos are massless. Nevertheless, experimental observations show that neutrinos change flavour (oscillate) [34]. This phenomena

can be explained if different generation neutrinos have mass. Their masses are currently unknown, but there is enough evidence to prove that they are non-zero. In the SM, neutrinos are right handed particles, and as a consequence, they are unable to acquire mass through the same mechanism employed by other fermions. There are several formulations to explain neutrino masses such as the see-saw [35] mechanism, but there is no consensus on which is the best extension to accommodate neutrino masses in the SM [36].

Matter antimatter asymmetry

Our Universe is predominantly composed by matter and the amount of antimatter is a small fraction, that usually can be explained by secondary origin in energetic particle collision, for example arising from cosmic rays. If charge conjugate parity (*CP*) symmetry was exact, the laws of Nature would be the same for matter and for antimatter. Consequently, in order to explain this asymmetry, *CP* violation is postulated as one of the necessary conditions for baryogenesis by Sakharov [37]. It could also be argued that the unbalance could be local, but there is still a need for a mechanism that generates it. Additionally, parts of the universe dominated by antimatter would have effects on our observations of the isotropy of cosmic microwave radiation.

There is also direct evidence of *CP* violation: it was first discovered in neutral K decays [38] in 1964, later in 2013 and 2019 it was also observed in B [39] and D [40] mesons decays by LHCb. These observations can be explained by the *CP* violation in the mixing of the neutral mesons. Such mixing can be introduced in the SM by a *CP*-violating complex phase in the CKM matrix described in Section 1.1.5. Nevertheless, this single *CP*-violating SM parameter is not able to explain the full matter-antimatter asymmetry in the Universe, failing by several orders of magnitude.

In the SM there are three sources of *CP* violation: the already mentioned complex phase in the CKM matrix, a complex phase in the PMNS matrix, and *CP* violation in the strong interaction. An experimental indication of the latest would be to observe the electric dipole moment of the neutron. This has not been accomplished, suggesting that the *CP* violation in the strong sector is also too small. In 2020, indications of *CP* violation in the lepton sector where reported by T2K Collaboration [41]. This last source of *CP* violation would allow for leptogenesis [42], an early universe scenario in which heavy neutral leptons —that were their own antiparticles— underwent *CP*-asymmetric decays, resulting in a Universe with unequal numbers of particles and antiparticles. Leptogenesis can successfully account for the present matter–antimatter asymmetry. These hypothesis can be tested in neutrino oscillation experiments.

Additionally, there are several BSM theories that would introduce *CP*-violating processes in different sectors that may explain the large matter-antimatter asymmetry produced in the early universe. Some BSM theories predict that *CP* violation in the top quark sector and the Higgs section can be particularly large. As a result, indications of *CP* violation might be found by studying the top decay, the top dipole moments, the production rate for $t\bar{t}$, single top and $t\bar{t}X$, the Higgs couplings and certain production modes of the Higgs boson [43]. These BSM proposals can be tested in the collider experiments.

Dark matter

Cosmological observations, such as the rotation of stars around the centre of the galaxy, are dependent on the gravitational effect of the massive objects in their surroundings. These observations show a large discrepancy between the amount of "visible" or baryonic matter and the gravitational effect measured [44]. This suggests the existence an "invisible"[3] type of matter, that would account for a large fraction of the mass forming these astrophysical systems. Currently, no particle in the SM is suitable to explain such kind of matter, referred to as dark matter (DM). Some BSM theories include the addition of new particles that can be DM candidates, such as the so-called weakly interacting massive particles (WIMPs) [45].

1.2.1 Beyond the SM

In order to cope with the limitations of the SM described above, as well as other not mentioned, BSM theories are developed. These are extensions of the SM that aim to improve the description of the universe given by the SM.

The most famous BSM theory is supersymmetry (SUSY) [46], and was developed in 1974. This theory is founded on the idea of introducing a new symmetry that transforms fermions into bosons by extending the Lie algebra. In these theories, particles are represented as multiplets of the associated algebra, called supermultiplets. The degrees of freedom of each supermultiplet are equal for fermions and bosons, hence, each particle has associated a superpartner. If the particle is a fermion, its superpartner should be a boson and viceversa. SUSY theories can solve the hierarchy problem, besides, they introduce new particles that can be DM candidates.

There are, in fact, many exotic ideas to extend the SM, some of them including an extra particle or field. Currently, an approach to look for new physics in a model

---

[3]As it would not interact via the electromagnetic force

independent manner is using the formalism of Effective Field Theories (EFTs) [47]. This is based on the idea that new physics might appear at energies that are too high to directly produce the resonance associated to the BSM particles with the current experiments. Nevertheless, at low energies, these new physics would be accessible as an integrated out theory, with a limited energy range of applicability. The effect can be the appearance of a new interaction vertex or a modification of an existing coupling. The EFT approach is sensitive to most of the SM issues proposed in this section. In particular, CP violation can be constrain by certain EFT parametrizations as the one reported in Ref. [48].

1.3   Phenomenology of the SM in colliders

Particle colliders serve as the primary experimental infrastructure designed to test the SM and search for new physic phenomena. In particular, this thesis is developed using data from collisions that took place at the Large Hadron Collider (LHC), a hadronic collider that will be described in Chapter 2. When modelling a pp collision, the first thing to bear in mind is that protons are not elementary particles, but composite states bounded by the strong interaction. At the LHC working energies, which are of the order of 10 TeV, the constituents of the protons (partons) can interact, leading to an inelastic collision where the proton breaks. These collisions are the ones that yield the most relevant physical processes. Additionally, QCD plays a significant role on the description of hadronic collisions. As a consequence, two energy regimes can be identified, depending on whether QCD is perturbative (for scales over Λ_{QCD}), or it is not (scales under Λ_{QCD}). For the former case, QCD can be treated as a weakly interacting theory in which free quarks exist, while for the later, coloured particles are force to hadronice.

The description of an inelastic collision in the LHC can be summarized in the following items:

- Description of the incoming proton and, in particular, the kinematic properties of the interacting partons.

- The hard scattering that describes the interaction between the two partons to produce a new state.

- Radiation coming from coloured states generated in the two previous items. At some point those coloured products of the collision will reach the non-perturbative regimen and hadronize. This includes the lower energy remnants of the incoming protons and their evolution after the hard scattering.

One of the most important physical quantities in the context of high energy physics is the cross section, which represents the probability that a certain process takes place. It depends on the centre-of-mass energy of the collision and is given in surface units. The inclusive cross section can be derived from any observable of interest, allowing to confront the measured data with the prediction. The differential cross section, which is the cross section as a function of a certain observable, can also be extracted to study how the prediction and the data agree, as a function of the selected observable.

In fact, any observable of interest can be computed from an inclusive or differential cross section. Cross section for the production of a given final state in pp collisions can be computed at fixed order in perturbation theory as:

$$\sigma = \sum_{i,j} \int dx_i dx_j f_i(x_i, \mu_F) f_j(x_j, \mu_F) \sigma_{ij}(x_1 x_2 s; \mu_R \mu_F) \tag{1.18}$$

The sum runs over all possible parton flavours that can produce the interaction. The expression is integrated over all the possible initial-state momenta of the considered partons. Those initial-state momenta are parameterized by the Bjorken variables: $x_i = \frac{p_{parton}}{p_{proton}}$. $f_i(x)$ are the parton distribution functions (PDFs), that give the probability of finding a parton of a given flavour (i) with a fraction of the proton's momenta x when interacting. Finally, σ_{ij} is the cross section of the partons with flavours i and j at the desired order in perturbation theory, that depends on the factorisation (μ_F) and renormalisation scales (μ_R) and the energy of the interaction.

Below, a more detailed description of each of these items is provided. Due to the complexity of the computations in practice, some of the steps of the computation are simulated using Monte Carlo (MC) methods, which are computationally affordable. Each of the steps are treated by different tools.

Parton distribution functions

As already stated, protons are composit states. Quantum numbers of a proton are given by its main constituents or valence quarks, which are of flavours u, u and d. Given the nature of strong interactions, the structure of the proton should also include other coloured particles, radiated from the valence quark interaction: these are gluons and other quarks, known as sea quarks.

In order to model the complex internal structure of the proton PDFs —denoted as $f_i(x, q^2)$— are used. They provide a parametrization of that structure for an external particle interacting at a certain energy regimen: q^2. In this use case, that energy
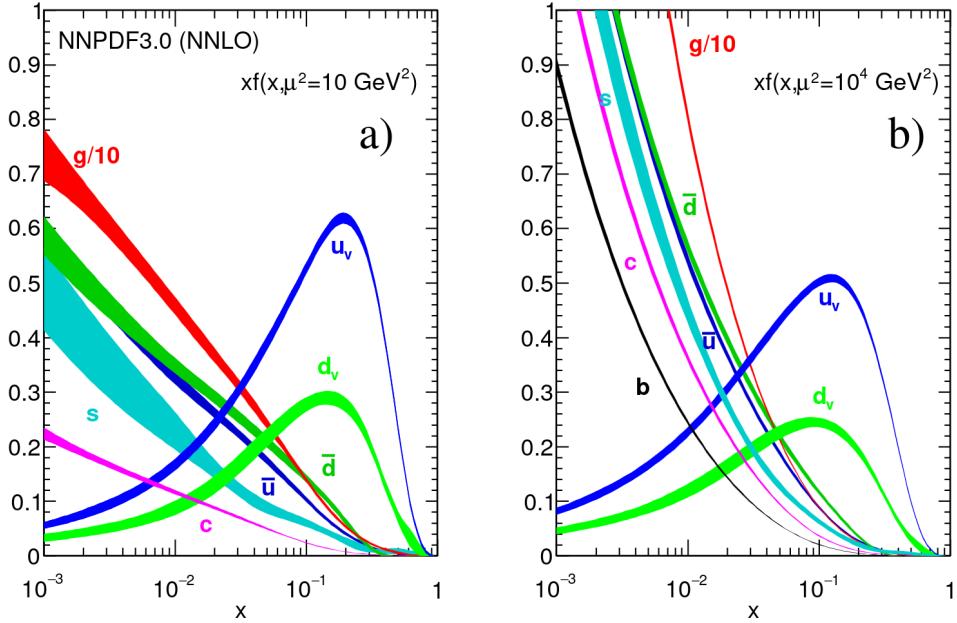
FIGURE 1.3: Parton distribution functions $f(x)$ (where $f = u_v$, d_v, $\bar{u}$, $\bar{d}$, $s \approx \bar{s}$, $c = \bar{c}$, $b = \bar{b}$, g) obtained in NNLO NNPDF3.0 global analysis [50] at scales $\mu^2 = 10$ GeV^2 (left) and $\mu^2 = 10^4$ GeV^2 (right), with $\alpha_s(M_Z^2) = 0.118$. Figure extracted from Ref. [26].

regime is given by the factorization scale μ_F, which is selected to regularize the effect of infrared divergences, separating the soft and hard processes. Hence, PDFs are used to summarize the structure of the proton that cannot be computed by perturbation QCD calculations.

PDFs have been experimentally obtained in deep inelastic scattering measurements at certain energy scales. These functions can then be propagated to the energy scale under study using the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) [49] equations. As a consequence, PDFs can be measured in the "clearer" environment provided by electron-proton collisions at lower energies, and then, extrapolated to LHC energies. In fact, the foundation of the modern PDFs, used in this thesis, are the ones provided by ZEUS and H1 collaborations at the HERA collider.

Figure 1.3 shows an example of the PDFs obtained in he NNLO NNPDF analysis [50] at two different energy scales of 10 and 10^4 GeV^2. The PDF of the valence quarks get closer to the sea quark ones as energy increases. At higher energy, the gluons take the most part of proton momentum.

Flavour schemes

When modelling the production of third generation quarks, two different approaches can be followed to treat initial-state b quarks: the first one, called 4-flavour scheme

(4FS), assumes the b quark to be massive; while the second, called 5-flavour scheme (5FS), considers it to be massless. In the 4FS, the mass of the quark is taken into account, and the b quark is included in the initial state only through gluon splitting, because its mass is larger than that of the proton. As a result, there are only four quark flavours (up, down, charm and strange) allowed in the proton. In the 5FS, instead, the b quark is modelled via a PDF as any other parton.

Both approaches were found to provide reliable predictions. In the 4FS the b quark emission is simulated exactly at fixed order in perturbation theory. This should, in principle, provide a more accurate description than the 5FS, where the emission is modelled by the parton shower. Nevertheless, the gluon splitting introduces a collinear divergence, that may break the convergence of the perturbation series.

Hard scattering

Hard scattering refers to the inelastic interaction between the two incoming partons. Usually, at the energies provided by the LHC, the parton contains a large fraction of the proton energy, big enough to remain in the energy regimen where QCD is perturbative. The term $\sigma_{ij}(x_1x_2s; \mu_R\mu_F)$ in equation 1.18 can be written as follows:

$$\sigma_{ij \to n}(x_1x_2s; \mu_R\mu_F) = \frac{1}{2s} \int d\Phi_n |\mathcal{M}_{ij \to n}(s, \Phi_n, \mu_R, \mu_F)|^2 \qquad (1.19)$$

where s is a Mandelstam variable of the two partons, Φ_n accounts for the phase space of the final state (that might contain n particles), and $\mathcal{M}$ is the matrix element (ME). The ME encodes the probability that a scattering of initial partons (i, j) produce the final state particles n. The matrix element can be computed using QFT by applying Feynman diagrams that represents all possible transitions between the initial and the final state and their associated amplitudes.

Matrix element computation as well as phase space integration is handled using MC simulation. Such MC generator programs are know as matrix element generators. Usually in these computations the hard scattering is simulated at Leading Order (LO) or Next-to-leading Order (NLO) accuracy in QCD using the POWHEG [51], MADGRAPH or MADGRAPH AMC@NLO [52] programs with a given choice of PDFs.

Parton showers, underlaying event and hadronization

So far the description provided corresponds to the first and second items in the enumeration given at the beginning of this section. Fixed order calculations grant an

accurate prediction for the partonic final state. However, they do not provide a complete description of the experimental final state, as particles in the partonic final state may further decay or radiate extra particles.

All charged particles might radiate photons, while all coloured particles can radiate gluons. These gluons may additionally radiate other coloured particles that, subsequently, will interact between themselves, producing a shower. This process is known as parton fragmentation. Each step of the shower will decrease the energy, and, at some point, the transition between the perturbative to the non-perturbative QCD regime will be reached. For this reason, the process is described by phenomenological models known as Parton Showers (PS).

We have described the origin of radiation in the final state (FSR), but a similar phenomenon happens for the rest of proton constituents that are not part of the hard scattering. Those partons may interact among each other, producing secondary interactions called Underlying Event (UE). The UE is modelled using phenomenological models that take into account perturbative and non-perturbative effects.

Finally, the particles produced in the shower are still coloured states, hence they are not observable. Instead, they will undergo an hadronization process in order to produce bounded colour neutral states, denoted as hadrons. This is a fully non-perturbative process, that occurs below a given energy scale. In order to describe the hadronization phenomenological models such as PYTHIA [53] and HERWIG [54] (that also model PS) are used.

A schematic depiction of the modelling of a pp collision, including all steps described in this section, is shown in Fig. 1.4.

1.4 Higgs boson and top quark physics at LHC

1.4.1 Higgs boson physics at LHC

The discovery of the Higgs boson represents a significant milestone in the field of LHC physics, completing the Standard Model. This experimental achievement initiated an extensive experimental program focused on precisely measuring the properties of the newly discovered boson, with the objective of verifying its compatibility with the SM Higgs boson. These properties include its mass, electric charge, interactions with other particles, and spin. In particular, the Higgs boson mass m_H is a free parameter of the SM that needs to be measured. This parameter determines the coupling of the Higgs to other particles as well as its production and decay rates.
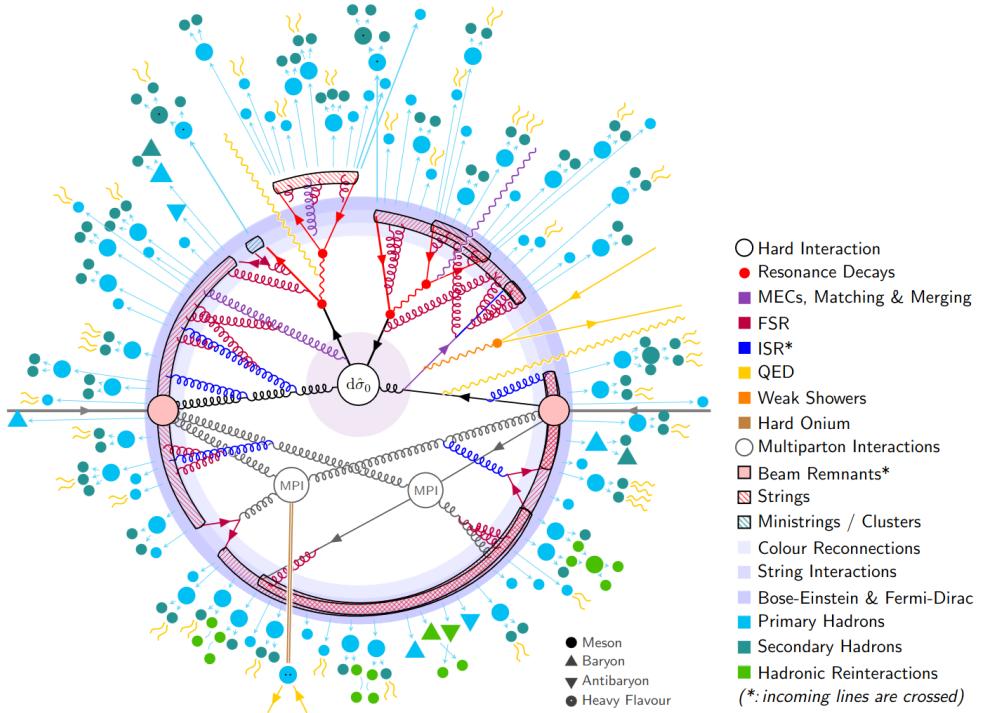
FIGURE 1.4: Simplified schematic representation of a proton-proton collision where a top-antitop quark pair is produced along with its subsequent PS. Extracted from Ref. [55].

The most precise measurement of the Higgs boson mass was performed combining data taken by the CMS experiment during the years 2011, 2012, and 2016 at $\sqrt{s} = 7$, 8, and 13 TeV and yields a value of $m_H = 125.38 \pm 0.14$ GeV [56].

Production modes

In pp collisions at the LHC the main production modes of the Higgs boson are: gluon-gluon fusion (ggH), Vector-boson fusion (VBF), Higgs-Strahlung (VH) and the associated production of a H with heavy quarks. The main leading-order Feynman diagrams of the four most common production modes of the Higgs boson in pp collisions are shown in Fig. 1.5. The cross sections for the different production modes spans over several orders of magnitude, as shown in Fig. 1.6 (left).

Gluon-gluon fusion is the dominant production mode in the LHC, with a cross section of 49 pb. As the Higgs boson does not couples directly to the massless gluon, it is produced via an intermediate loop of heavy quarks. This loop is dominated by top quarks and contributions from lighter quarks are suppressed proportionally to m_q^2. VBF production mode has a cross section of 3.8 pb, being the production mode with the second-largest cross section. This production mode proceeds as the scattering of two quarks-antiquarks mediated by the exchange of a W or Z boson, with the Higgs
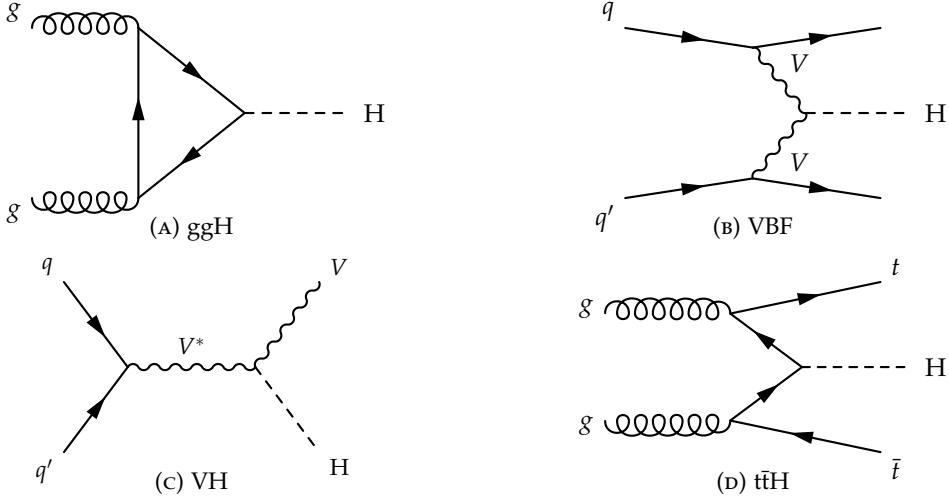
FIGURE 1.5: Main leading order Feynman diagrams contributing to the Higgs boson production in (a) gluon fusion, (b) Vector-boson fusion, (c) Higgs-Strahlung, and (d) associated production with a pair of top quarks.
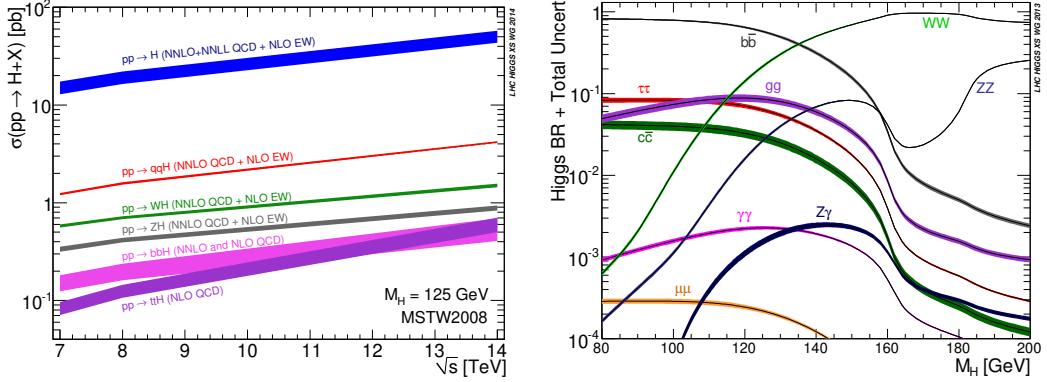
FIGURE 1.6: Left: SM Higgs production cross section in pp collisions as a function of the center-of-mass energy. Right: SM Higgs decay branching ratios as a function of the Higgs boson mass. Both figures extracted from Ref. [57].

boson radiated off the electroweak boson propagator. Therefore, Higgs production is purely electroweak, as any jet activity (from QCD) would happen around the forward outgoing quarks, so the experimental signature of this production mode is very clean. Higgs-Strahlung is the third most frequent mechanism, with a cross section of 2.3 pb. In this process, the Higgs boson is produced in association with an electroweak boson. As the VBF production mode, it can be used to probe the coupling of the H to the electroweak bosons.

The associated production with a pair of heavy quarks is the production mode with the next-to-the-lowest cross section and can be initiated by two incoming gluons (gg) or two incoming quarks (q$\bar{\text{q}}$). The heavy quarks can be a top-antitop quark pair (t$\bar{\text{t}}$H) or a pair of bottom-antibottom quarks (b$\bar{\text{b}}$H). The t$\bar{\text{t}}$H cross section is 507 fb, of the same order of b$\bar{\text{b}}$H cross section, nevertheless, the second one presents final states
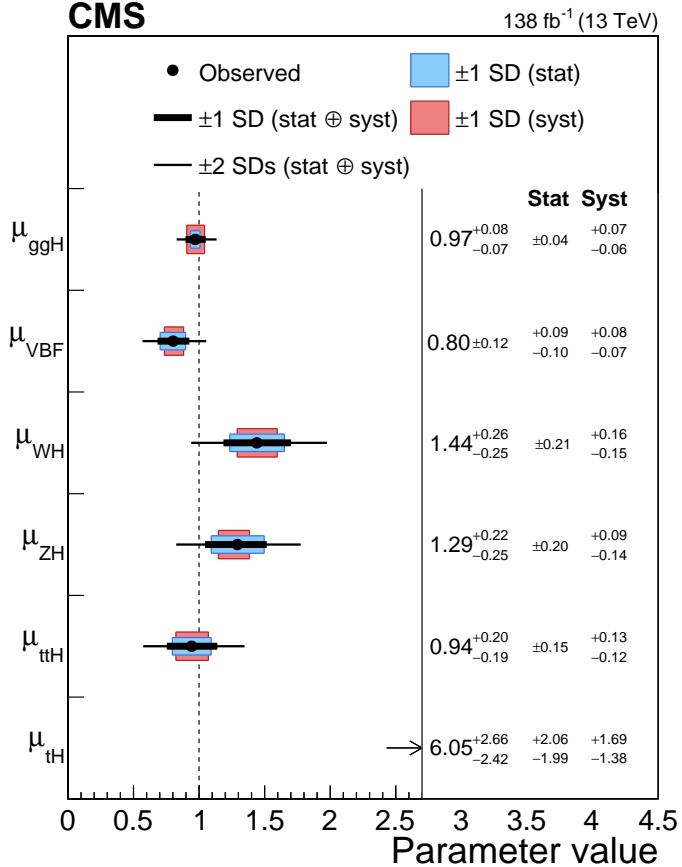
FIGURE 1.7: Signal strength parameters (σ/σ_{SM}) extracted for various production modes, assuming SM Higgs decay rates. The thick (thin) black lines indicate the 1 (2) s.d. confidence intervals, with the systematic and statistical components of the 1 s.d. interval indicated by the red and blue bands, respectively. Figure extracted from Ref. [58].

harder to distinguish from the backgrounds. The Higgs boson can also be produced with a single top, tH, being this the production mode with the lowest cross section: 89.5 fb.

All production modes of the Higgs except $b\bar{b}H$ and tH have been observed both by CMS and ATLAS [58, 59]. Figure 1.7 summarizes the values of the measured production cross sections using data collected during Run-2 by the CMS experiment. All the measurements are in good agreement with the SM.

Decay modes and couplings

The Higgs boson is not a stable particle, it decays before reaching the detector, therefore it must be studied via its decay products. It decays in several modes with different frequency, the branching ratio of the Higgs boson as a function of the mass of the Higgs is shown in Fig. 1.6 (right). At $m_H = 125\,\text{GeV}$, $H \to b\bar{b}$ is the most frequent

decay mode, followed by H → WW. Using Run-1 data, it was possible to observe the H decay to WW, ZZ and $\gamma\gamma$, separately [60, 61]. In fact, the two latter ones, despite having low branching ratios, were the channels that yielded to the H discovery. Both benefit from being resonant channels and from the high resolution of the detector to measure the invariant mass of the dilepton and diphoton systems. This probed the Higgs coupling to gauge bosons and showed that this coupling was proportional to m_V^2, as predicted by the SM.

Higgs coupling to fermions is directly proportional to the mass of the fermion, therefore, the largest value of the Yukawa coupling is given by the top-Higgs coupling. As the Higgs boson does not decay to top quarks, a pair of b quarks is the heaviest final state possible for the Higgs boson. Despite the large H branching ratio, this measurement is challenging given the large amount of QCD background. The H → b$\bar{\text{b}}$ was observed by CMS and ATLAS using Run-2 data [62, 63]. H → $\tau\tau$ has the second largest branching ratio, and a better signal to background ratio than the previous channel. This decay was observed by CMS and ATLAS during Run-1 [60, 61]. Using Run-2 data CMS reported the evidence of the coupling of the Higgs to the second generation of fermions, as the H → $\mu\mu$, with a significance of 3.0 s.d. [64].

In order to study the top-Higgs coupling, gluon gluon fusion production, or Higgs decay to $\gamma\gamma$ can be used. In both cases the top quark appears in loops. In order to study the Higgs-top coupling at tree level, t$\bar{\text{t}}$H and tH production modes are needed. These production modes will be covered in detail later in Chapters 4 and 6.

The current status of our understating of the Higgs coupling to electroweak bosons and fermions is summarized in fig 1.8, which was produced using data taken during Run-2 at $\sqrt{s} = 13\,\text{TeV}$.

Figures 1.7 and 1.8, along with the Refs. [58, 59] show the outstanding improvement in our understanding of the Higgs boson properties achieved during the ten years that followed its discovery, reporting the most up-to-date measurements of the properties of such boson. The results provided in the CMS publication [58] combine the studies presented in this thesis and published in Ref. [9].

1.4.2 Top quark physics at LHC

In order to test the SM and explore potential BSM physics models, it is imperative to investigate the properties of the top quark. This particle was initially postulated by Kobayashi and Maskawa in 1973, as part of the framework that required at least three
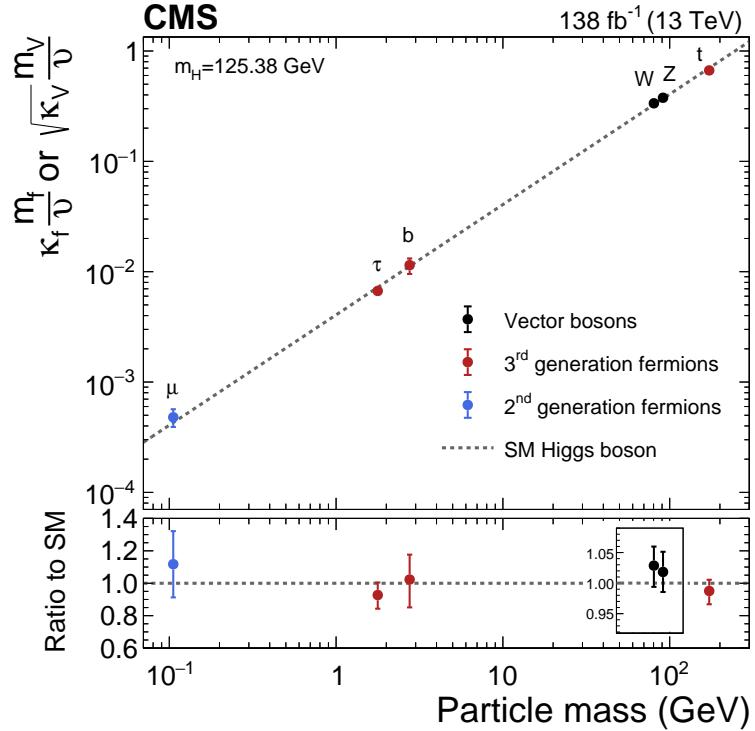
FIGURE 1.8: Measured coupling modifiers of the Higgs boson to fermions and heavy gauge bosons, as functions of fermion or gauge boson mass. ν is the vacuum expectation value. For electroweak bosons, the square root of the coupling modifier is plotted. Figure extracted from Ref. [58].

generations of quarks to explain *CP* violation through flavour mixing [65]. Nevertheless, the top quark was not discovered until 1995, when it was observed by the D0 and CDF experiments at Tevatron collider in p$\bar{\text{p}}$ collisions [66, 67]. These measurements allowed to measure its mass, showing that it was the most massive fermion. The top quark mass, as measured by CMS and ATLAS at the LHC and the experiments at Tevatron is $m_t = 172.69 \pm 0.3$ GeV [26], confirming it is the heaviest fundamental particle in the SM.

As a consequence of the large value of its mass, the top quark lifetime is remarkably small, of order $\tau \approx 10^{-25}$, which is one order of magnitude lower than the hadronisation time scale. As a result, the top quark undergoes weak force-mediated decays —most of the time into a W boson and a b quark— before hadronization. This is a special feature of the top quark, making it the only quark that can be studied as a free state, and one of the few available probes for perturbative QCD (pQCD).

A second implication of the top quark's mass is the strength of its coupling to the Higgs boson, which is approximately of order 1. It is commonly acknowledged that this Yukawa coupling is the only one in the SM considered "natural", as the others are orders of magnitude smaller. Given this fact, the top quark holds a unique position in

the spontaneous breaking of the electroweak symmetry. Furthermore, the top quark is responsible for adding sensible corrections to the Higgs boson propagator. Therefore, studying the top quark and its couplings is a good handle to solve the hierarchy issue in the SM.

Production modes

There are several production modes of the top quark in pp collisions at the LHC energy: production of quark anti-quark pairs (t$\bar{\text{t}}$), single top production and associated production with other particles.

Some of the leading order diagrams for t$\bar{\text{t}}$, which is the dominant top production mode at LHC, are shown in Fig. 1.9. At the LHC, the dominant production mechanism for the t$\bar{\text{t}}$ process is gluon-gluon fusion, corresponding to 85% of the total production. Instead, in p$\bar{\text{p}}$ collisions at the Tevatron energy scales, t$\bar{\text{t}}$ production was dominated by quark-anti quark annihilation. The next-to-next-to-leading order (NNLO) plus next-to-next-to-leading-logarithmic accuracy prediction of the t$\bar{\text{t}}$ cross section for a centre-of-mass energy of $\sqrt{s} = 13\,\text{TeV}$ yields $\sigma_{\text{t}\bar{\text{t}}} = 832^{+20}_{-29}(\text{Scale}) \pm 35(\text{PDF} + \alpha_S)$ pb [68].
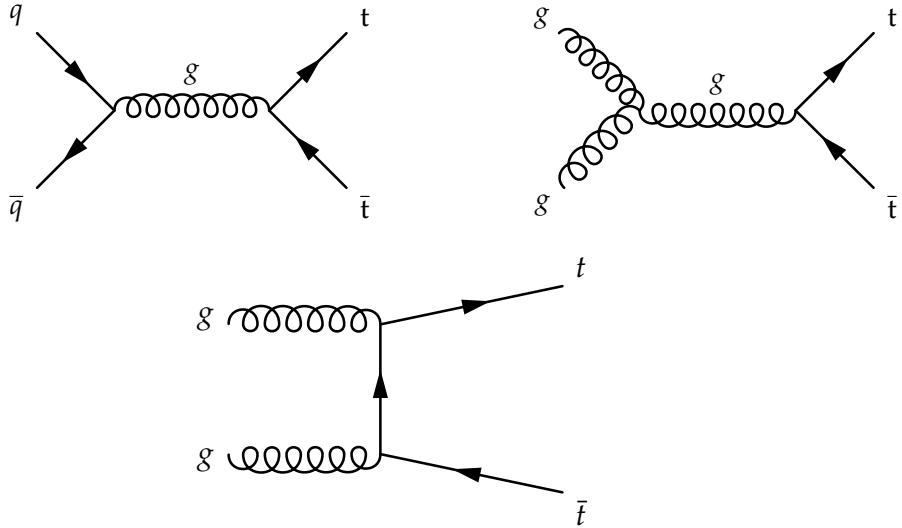
FIGURE 1.9: LO Feynman diagrams for t$\bar{\text{t}}$ production in pp and p$\bar{\text{p}}$ collisions. Top left diagrams shows the q$\bar{\text{q}}$ production mode, while the other two show the gluon-gluon fusion prodction mode.

This production mode has been studied both at Tevatron and at LHC, taking into account the different final states given by the top quark decay. Top quark can decay hadronically (q$\bar{\text{q}}$b) or leptonically, to a lepton, a neutrino, and a b quark ($\ell\nu$b). In terms of t$\bar{\text{t}}$ decay we can distinguish, then, three scenarios: fully hadronic, dileptonic and semileptonic, with a branching ratio of 46%, 10% and 44% respectively. The
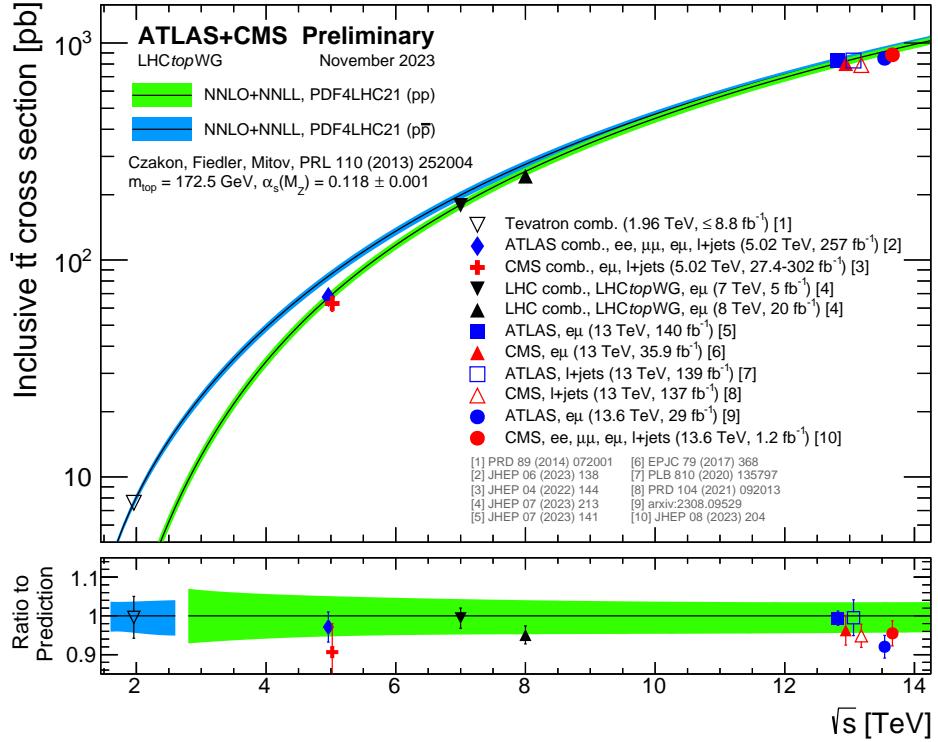
FIGURE 1.10: Summary of LHC and Tevatron measurements of the $t\bar{t}$ production cross-section as a function of the centre-of-mass energy compared to the NNLO QCD calculation complemented with NNLL resummation (top++2.0). The theory band represents uncertainties due to renormalisation and factorisation scale, parton density functions and the strong coupling. The measurements and the theory calculation are quoted at m_t=172.5 GeV. Measurements made at the same centre-of-mass energy are slightly offset for clarity. Figure extracted from Ref. [69].

dileptonic final state, despite its low branching ratio, provides a clean signature with two leptons in the final state and two b jets.

Currently the $t\bar{t}$ production cross section has been measured at center-of-mass energies from 1.96 TeV to 13.6 TeV, using data taken by D0 and CDF experiments at Tevatron, as well as, CMS and ATLAS at LHC. A summary of these experimental measurements is shown in Fig. 1.10. The measurements are in good agreement with the NNLO+NNLL predictions. During the development of this thesis, I contributed to the first measurement of $t\bar{t}$ production at $\sqrt{s}$ = 13.6 TeV by performing a cross check of the main analysis, using an event-counting method restricted to events with an opposite-sign eμ pair and at least one jet [24].

The single top production is the subleading top quark production mode at the LHC, with a cross section one order of magnitude lower than $t\bar{t}$ production. Several processes that yield only one top quark in the final state are included in this production mode, being the dominant ones the so-called t-channel, s-channel and tW-channel. LO

diagrams for those channels can be seen in Fig. 1.11. In these cases, the top production is mediated by the electroweak interaction, and a tWb vertex is present.
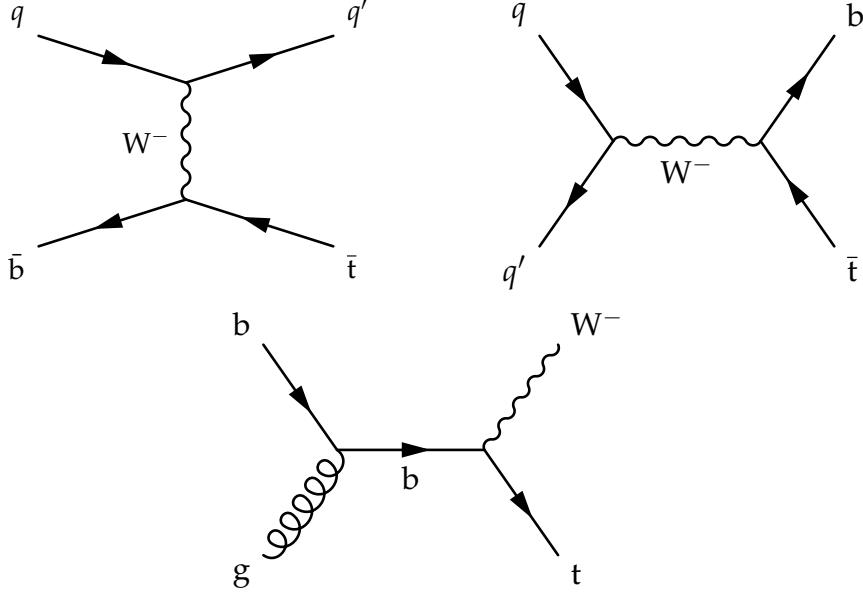
FIGURE 1.11: LO Feynman diagrams for the main modes of single top production in pp collisions: *t*-channel (top left), *s*-channel (top right) and tW-channel (bottom).

There are other top quark production modes with lower cross sections, in most of those cases the top quark is produced in association with other particles, therefore, those processes are probes to the couplings of the top quark to such particles. The amount of data delivered by the LHC allows to study those low cross section production modes.

These processes include the associated production of a top anti-top quark pair with bosons (t$\bar{\text{t}}$H, t$\bar{\text{t}}$W, t$\bar{\text{t}}$Z and t$\bar{\text{t}}\gamma$), as well as, single top associated production (tZq, tH, tHW); finally the t$\bar{\text{t}}$t$\bar{\text{t}}$ production is also included in this category.

The t$\bar{\text{t}}$H, tH and tHW processes had been already covered as Higgs production modes and, along with t$\bar{\text{t}}$W, will be studied in detail in this thesis; t$\bar{\text{t}}$Z and t$\bar{\text{t}}\gamma$ productions will also be covered later in Chapter 4.

The tZq production has been observed both by ATLAS [70] and CMS [71] using Run-2 data. This process allows to study the top-Z coupling but also the WWZ coupling. Besides, this production is sensitive to flavour chaining neutral currents (FCNC) that may appear in BSM scenarios.

The t$\bar{\text{t}}$t$\bar{\text{t}}$ production, with a SM predicted cross section of only 12 fb, has been observed by ATLAS with an observed (expected) significance of 6.1 (4.3) s.d. [72], while CMS

has reported the evidence of $t\bar{t}t\bar{t}$ production with an observed (expected) signal significance of 4.0 (3.2) s.d. [73]. This process can be used to test the top-Higgs Yukawa coupling, as one of the $t\bar{t}t\bar{t}$ production diagrams in the SM is Higgs-induced [74].

2

The LHC and the CMS experiment

The measurements described in this thesis have been performed using data from pp collisions at the LHC, recorded by the Compact Muon Solenoid (CMS) experiment. This is one of the four main experiments of the LHC at CERN. In this chapter a description of the experimental framework is provided.

2.1 The Large Hadron Collider

The LHC is a hadron circular accelerator and collider designed to achieve a center-of-mass energy of the order of 10 TeV. The LHC is capable of accelerating several types of nuclei: hydrogen nuclei (protons) are most commonly accelerated, but heavier nucleai, such as lead, are also accelerated in special data taking runs. At the date of writing, the LHC has provided pp collisions at a center-of-mass energy of 13.6 TeV, the highest ever achieved. The LHC is composed of two superconducting rings hosted in a 27 km circumference tunnel 100 m below the surface, which is inherited from the Large Electron Positron collider (LEP) [75]. In each of the rings a beam of protons is accelerated in a direction. Both beams are allowed to collide in specific points of the LHC circumference, where the four multi-purpose particle detectors ALICE, ATLAS, LHCb and CMS are located. In this section, a brief summary of LHC description is provided, a more detailed description can be found in Ref. [76].

The beam within the LHC is supplied by a chain of colliders. Since 2020 the process starts with negative ionised hydrogen atoms (one proton with two electrons), which are accelerated by Liniac 4, a linear collider, up to 160 MeV. Then, they are separated into protons and electrons, and the resulting protons are injected to a circular synchrotron, the Proton Synchrotron Booster (PSB). Before 2020, protons were introduced into the Linac 2 which accelerated them to an energy of 50 GeV before injecting them to the PSB. The PSB is capable of arranging clusters of protons, called bunches, and

further accelerating them to an energy of 1.4 GeV. Subsequently, these bunches are injected into the Proton Synchrotron, which raises their energy to 25 GeV and establishes a separation of 25 ns between them. Finally, the bunches are injected into the Super Proton Synchrotron, which serves as the ultimate stage before they enter the LHC, with an energy level of 450 GeV [77].

The LHC is able to accelerate each of the beams circulating in opposite directions by using 16 radio frequency cavities distributed along the circumference of the collider. The cavities operate at a frequency of 400 MHz, each one providing a 2 MV acceleration voltage. The voltage in the cavities is synchronized with the arrival time of the protons. Consequently, protons arriving earlier are accelerated less than the ones arriving later. This helps to further collimate the bunches.

The beams are bent to be inside the LHC circumference by applying a magnetic field of up to 8.3 T. This field is created by superconducting magnets located in many sectors along the circumference. In order to operate properly, these magnets have to be kept at a temperature of 1.9 K: this is achieved by a cooling system using liquid helium. Between the bending magnets, there are sets of quadrupoles, sextupoles, and octupoles magnets used to focus the beam and keep the bunch coherence. Once the beams have achieved the target collision energy (i.e. 6500 GeV for the data used in this thesis), they are deviated slightly from its trajectory, in order to make both proton bunches collide at the collision points. This is called bunch crossing (BX), and it is worth noting that in each BX more than one proton pair can interact simultaneously. This effect, called pileup (PU), leads to the presence of more than one interaction vertices in the registered event.

Other than the beam energy, the instantaneous luminosity ($\mathcal{L}$) is the other relevant parameter related to the collision setup. The luminosity is defined as the quotient between the event rate and the cross-section of a given process (σ).

$$\frac{dN}{dt} = \mathcal{L}\sigma \tag{2.1}$$

This parameter depends only on the conditions of the accelerator and can be expressed as follows:

$$\mathcal{L} = \frac{n_b N_p^1 N_p^2 f}{4\pi\sigma_x\sigma_y}R \tag{2.2}$$

Where n_b is the number of colliding bunches, N_p^1 and N_p^2 are the number of protons per bunch in each colliding beam (around 10^{11}), f is the beam revolution frequency of the LHC (40 MHz), σ_x and σ_y are the effective sizes of the bunches in the transversal
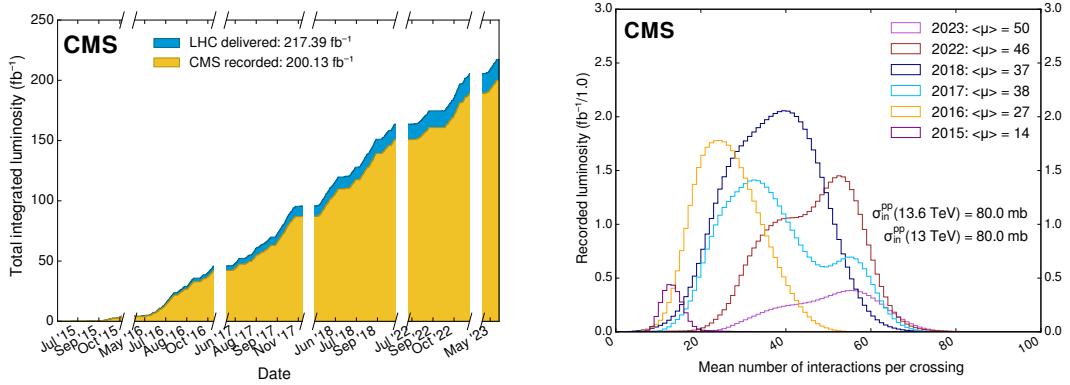
FIGURE 2.1: Left: cumulative delivered and recorded luminosity versus time for 2015-2018 and 2022-2023 data-taking periods using pp collisions data only, taken from [78]. Right: distribution of the average number of interactions per bunch crossing (pileup) for pp collisions in 2015 (purple), 2016 (orange), 2017 (light blue), 2018 (navy blue), and 2022 (brown); taken from [78].

directions, and R is a calibrated geometrical factor that takes into account the crossing angle at the interaction point.

The integrated luminosity (L) can also be defined by integrating the instantaneous luminosity over time. The integrated luminosity during the LHC data-taking periods related to this thesis are shown in the left depiction of Fig. 2.1. Each of these data-taking periods is known as a "Run", the figure shows the accumulated luminosity for Run-2 and the beginning of Run-3. Run-2 started in 2015 and lasted until 2018: during that period of time the LHC worked (most of the time) at an energy of 13 TeV. Run-3 started in 2022 and is currently on going. LHC was initially designed to provide an instantaneous luminosity of about 10^{34}cm^{-2}s^{-1}, but during the running period between 2017 and 2018 twice this value was reached by collimating a higher number of protons per bunch, yielding to simultaneous collisions. The mean number of interactions per BX is shown in Fig. 2.1 (right) for each year.

2.2 The CMS experiment

CMS stands for Compact Muon Solenoid. This apparatus is, along with ATLAS, one of the two general-purpose experiments at the LHC. It has been designed to be able to measure a broad type of physically interesting processes that can happen when protons collide at the LHC energy scale. CMS is a cylinder-shaped detector, composed by several layers of subdetectors, built to enclose the LHC beam pipe around one of the LHC interaction points. The subdetectors that constitute CMS are: a high granularity tracker, two hermetic calorimeters, and a highly redundant muon system. On top of

CMS DETECTOR

Total weight   : 14,000 tonnes
Overall diameter : 15.0 m
Overall length   : 28.7 m
Magnetic field   : 3.8 T

STEEL RETURN YOKE
12,500 tonnes

SILICON TRACKERS
Pixel (100x150 μm²) ~1 m² ~66M channels
Microstrips (80–180 μm) ~200 m² ~9.6M channels

SUPERCONDUCTING SOLENOID
Niobium titanium coil carrying ~18,000 A

MUON CHAMBERS
Barrel: 250 Drift Tube, 480 Resistive Plate Chambers
Endcaps: 540 Cathode Strip, 576 Resistive Plate Chambers

PRESHOWER
Silicon strips ~16 m² ~137,000 channels

FORWARD CALORIMETER
Steel + Quartz fibres ~2,000 Channels

CRYSTAL
ELECTROMAGNETIC
CALORIMETER (ECAL)
~76,000 scintillating PbWO₄ crystals

HADRON CALORIMETER (HCAL)
Brass + Plastic scintillator ~7,000 channels

FIGURE 2.2: Cutaway diagrams of CMS detector. The different subdetectors are labeled in the figure. Figure taken from Ref. [81].

that, a superconducting solenoid provides a magnetic field of 3.8 T. A sketch of the detector is shown in Fig. 2.2. The CMS Experiment is the experimental set-up of this thesis: in this section some details about the detector are given, for a more exhaustive discussion of its design see Ref. [79]. For an overview of the detector performance during Run-2, and a description of the upgrades in place for Run-3, see Ref. [80].

2.2.1   Coordinate system

In order to define the physical observables used in this thesis, as well as describe the detector components, a coordinate reference system is defined. The z axis is fixed along the beam direction with the positive semi-axis pointing west, toward the Jura mountains. The $x - y$ plane is the transverse plane, perpendicular to the beam direction. The x positive semi-axis points to the centre of the LHC while the y positive semi-axis points to the surface. Given the cylindrical shape of the detector, the azimuthal angle (φ), and the polar angle (θ), are also used. φ describes angles between the y and x axis, while θ is defined in the z-x plane. θ is used to define the pseudorapidity as: $\eta = -\log(\tan(\theta/2))$. It can be proven that the difference in η between two massless particles is invariant under Lorentz boosts along the z axis.

Taking into account this geometry, we define the projection of the momentum of a certain particle in the transverse plane as the transverse momentum, p_T. In order to measure angular distances between particles we define the quantity:

$$\Delta R = \sqrt{\Delta \phi^2 + \Delta \eta^2} \qquad (2.3)$$

2.2.2 Tracking system

This is the innermost layer of the CMS detector; it is based on silicon sensors and its propose is to measure the trajectory of charged particles (tracks). The precise determination of the tracks is crucial to obtain a high resolution estimation of the particle track parameters, as well as, to associate the track with one of the interaction vertex. This avoids the wrong assignation of a particle to one vertex originated due to pileup interactions. The impact parameter determination is also a key element when tagging jets coming from a b meson.

The tracking system is divided in two parts: the pixel detector and the silicon strip detector.

The pixel subdetector is the closest one to the beam axis. After the addition of an extra layer in 2017, the innermost barrel pixel layer is located around 2.9 cm away from the beam axis, while the outermost is at 16 cm. In the endcap, the pixel detector covers up to approximately 50 cm in the Z axis. Given the high occupancy expected in this region and the need of high spacial resolution for a precise measurement of the impact parameter, the pixel detector has been designed to provide a high granularity, yielding a spatial resolution of between 10 and 15 μm.

The strips detector is located between 20 to 116 cm away of the beam axis. Silicon strips are located following the z direction in the barrel, and the radial component in the endcap. The obtained resolution along the Z axis is around 200 μm; in the case of the transverse plane the resolution is typically around 20 μm.

2.2.3 Calorimeters

CMS has two calorimeters: the electromagnetic calorimeter (ECAL) devoted to measure the energy of electrons and photons, and the hadronic calorimeter (HCAL) which is designed to measure the energy of hadrons. The ECAL is the innermost of the two, it is located surrounding the tracker. Then, the HCAL is constructed between the ECAL and the solenoid magnet. Hermeticity of the calorimeters is crucial in order to provide a good estimation of the momentum imbalance of the event.

The ECAL is a scintillation calorimeter composed by tenths of thousand's of lead tungstate (PbWO$_4$) crystals. This material is very dense (8.3 g/cm^3) and has both small radiation length (0.89 cm) and Moliere radius (2.2 cm). The light generated within the crystals is converted into an electric signal that can be read out by photo detectors based on silicon. The ECAL resolution can be parameterized as:

$$\frac{\sigma(\mathrm{E})}{\mathrm{E}} = \frac{2.8\%}{\sqrt{\mathrm{E/GeV}}} \oplus \frac{12\%}{\mathrm{E/GeV}} \oplus 0.3\%, \tag{2.4}$$

where the first term is the stochastic one, the second takes into account to the electronic noise and the last one is from the non-uniformity of the detector and calibration uncertainties.

The HCAL is a sampling calorimeter composed by alternating layers of brass and a scintillator material. It is devoted to measure the energy of neutral hadrons and gives an alternative measurement of charged ones. A part of the subdetector is located outside the magnetic solenoid in order to provide extra absorption and increase the hermeticity of the detector. In addition, the very forward region ($3.0 < |\eta| < 5.0$) is covered by the forward calorimeter, located at 11 cm from the center of the detector and composed of steel.

The energy resolution of the HCAL is given by:

$$\frac{\sigma(\mathrm{E})}{\mathrm{E}} = \frac{85\%}{\sqrt{\mathrm{E/GeV}}} \oplus 7.4\%, \tag{2.5}$$

being the first uncertainty the one corresponding to the stochastic term and the second one accounting for calibration uncertainties.

2.2.4 Muon system

The muon system is located in the outermost part of the detector and it provides efficient and robust measurement of muons. As the only SM particles reaching this part of the CMS detector are muons and neutrinos, the muon system is able to identify the presence of a muon. The muon system also tracks the muon trajectory, reconstructs its momenta and triggers based on the presence of muons. The precise reconstruction and identification of muons is a key element of the CMS physics program, as muons are particles more efficient to select and less likely to be misidentified than electrons. Hence, the presence of such particles is a clear signature of interesting processes.
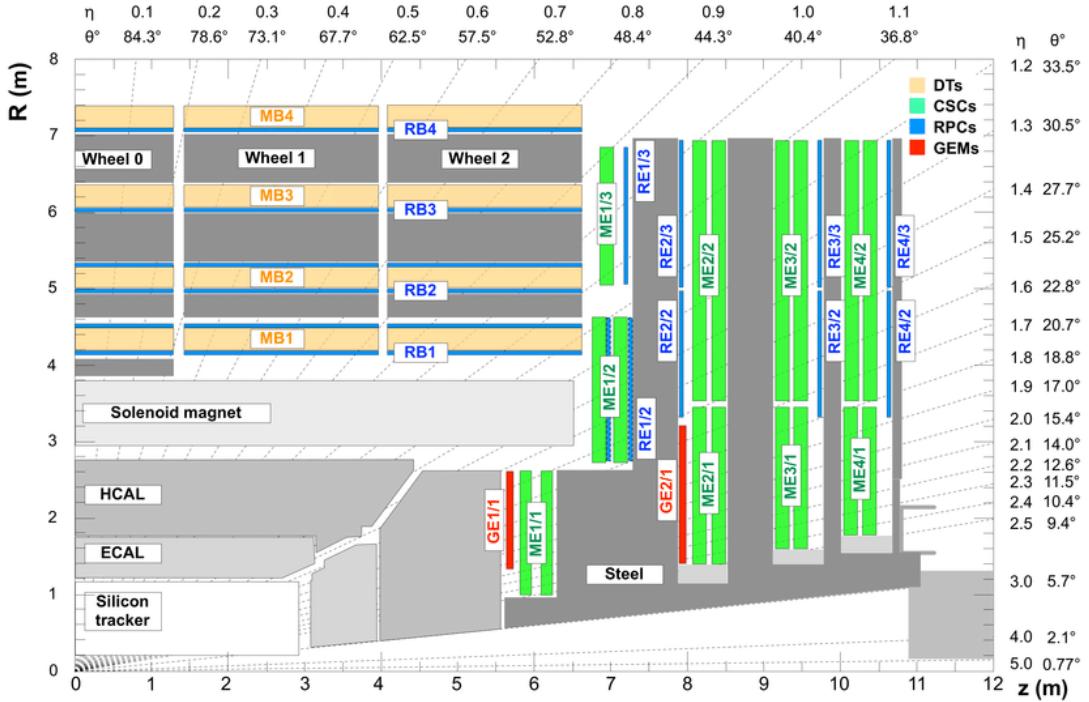
FIGURE 2.3: Scheme of the plane z-R of the CMS experiment. The muon chambers are detailed and the four types are shown in different colours [82].

The muon system was composed by three types of gaseous detectors during Run-2: Drift Tubes (DTs), Cathode Strip Chambers (CSCs) and Resistive Plate Chambers (RPCs). During 2020 the installation of a fourth type of chamber called Gas Electron Multipliers (GEMs) started in preparation of Phase-II of the LHC. Each type of chamber is designed with different features, depending on its geometrical location. Therefore, the four type of chambers are used in a complementary way. In Fig. 2.3 the location of each type of chamber is depicted.

DTs are located in the barrel region, covering an $|\eta| < 1.2$. In this part of the detector the muon rate is low and small contribution from neutron induced background is expected. DTs are organized in 5 wheels located at different positions along the z axis. Each wheel contains 4 stations situated at a different distance from the collision point and alternated with the layers of the magnet return yoke. Stations are composed by superlayers, each superlayer can be divided in layers, which is composed by cells. Each DT cell is a cavity containing gas, an anode wire of 50 μm of diameter and two electrode plates. When the muon passes through the cell its position can be obtained by measuring the drift time of the charged particles produced by the ionization of the gas to an anode wire. Depending on the orientation of the wires on each layer, the DTs are able to measure the z position and the $r - \phi$ bending angle in each chamber. The exception is the outermost chambers, which only measure the $r - \phi$ coordinate.

CSCs are placed in the endcap covering a $0.9 < |\eta| < 2.4$, in this part of the detector a large background contribution is expected, besides the magnetic field in this region is not uniform. The CSC chambers are organized in 4 vertical disks in each endcap. Each disk is divided in chambers split in ϕ sectors. CSC are multiwire proportional chambers designed to have high granularity, high response and tolerate high levels of radiation. The wires point in the radial direction, yielding precise measurement of the ϕ component.

RPCs are gaseous parallel-plate detector found both in the barrel and endcap, in the $|\eta| < 1.9$ region. They provide a high time resolution, of around 1 ns. Given this feature, the RPCs complement DTs and CSCs timing measurement and trigger capabilities.

Finally the GEM detectors are located in the endcaps close to the beam pipe, and will complement the CSCs in scenarios with high rate expected during LHC Phase-II.

2.2.5 Trigger system

The LHC collision rate can reach 40 MHz, nevertheless, the experimental setup is not able to process and store that amount of data. As a result, the need of a trigger system able to quickly select interesting collisions is patent. The CMS trigger system is divided in two steps or logical layers: Level-1 Trigger (L1) and High Level Trigger (HLT).

The L1 trigger is designed to have a latency of up to 4 μs, therefore it is hardware-based, because software triggers cannot reach these latencies. The L1 trigger takes information only from the muon system and the calorimeters. Tracking detectors are not used, as it is not possible to read out the information in each BX. Each subdetector has its own L1 algorithm. In the case of the calorimeter the first step is to promptly reconstruct the energy deposits, while for the muon system it is to reconstruct muon hits. Then, a two-layer trigger reconstructs electrons, jets, and hadronic τ candidates using the calorimeter information. For the muon system, this second step is done by three track finders. They combine the information from all the muon subdetectors in order to create muon candidates. Finally, the aforementioned information coming from the calorimeter and muon track finders is received and combined by the global trigger, that decides whether the event is worth to be kept or not. This decision is usually based on the presence of simple topologies on the event, such as the presence of single and double objects, or combinations of objects passing minimum momentum

thresholds. The readout of the CMS detector limits the acceptance rate of the L1 trigger to be 100 kHz.

The second step, the HLT, is not so constrained by speed requirements; as a result it can run on commercial computers. After a certain event passes the L1, the full event information can be processed by the HLT, which performs a more complex and precise reconstruction.

The HLT is organized in several algorithms (or paths) that target specific topologies. The event reconstruction is seeded by a L1 positive decision with that specific topology. The paths run depend on the event: first, the less complex reconstructions are run, aiming to reject spurious events promptly. Then, the paths are run ordered depending on the complexity of the reconstruction. For this reason, the amount of time that it takes for an event to be processed by the HLT varies. The usual average time is of 150 ms. The HLT is able to take a rate of around 1 kHz, integrated over few hours.

2.2.6   Event offline processing

Once events are selected by the trigger they are stored to be analyzed offline. Given the amount of data registered, the computation resources needed to store and process it are not present in a single institution. Apart from data, simulations of the physical processes are needed. These simulations, which should include the detector response to the particles, are computationally intensive and require significant storage capabilities. That is why the computing model used by CMS and other large experiments at the LHC is based on the Worldwide LHC Computing Grid (WLCG) [83]. This is a network that connects computer farms in different institutions and parts of the world, in order to optimize usage of resources.

The network is organized in layers, each comprising distinct computing centers with different capabilities and playing diverse roles. The first layer is the single Tier-0 (T0) physically located at CERN and Budapest. T0 is used for data acquisition, triggering as well storage of raw data and its prompt reconstruction. The next stage is composed by tens of Tier-1 facilities, that share the responsibility of store and process raw data contained in CERN's T0. The following layer is composed by around 50 Tier-2, where the reconstructed data is stored, simulations of events are generated and users can perform the first steps of the data analysis tasks. Finally, there is a last layer: Tier-3, which is composed by smaller computing facilities where local data analysis activities are carried out.

2.3   Event reconstruction

The information registered by each subdetector in CMS is combined in order to reconstruct the collision products. The goal is to identify these objects, measure their kinematic properties and, eventually, compute the physical observables needed for the analysis of the data. In CMS the algorithm developed to perform this task is the Particle-Flow (PF) algorithm [84], based on the concept of combining the information from all subdetectors into a global event reconstruction. The reconstruction is performed first locally in each subdetector, in order to reconstruct tracks in the tracker and muon system, and energy deposits in the calorimeters. The PF algorithm takes these information as input.

Tracks in the tracker are reconstructed using an iterative approach. They are seeded by two hits in consecutive pixel layers. Then, in each successive layer, the algorithm searches for compatible hits in order to propagate the trajectory. The compatibility between the hit and the hit pattern expected for a charged-particle is evaluated using a Kalman-filter [85]. When a hit is considered to be compatible with the track, it is added to the track and its parameters are updated for the next extrapolation. The process is repeated in all tracker layers until the algorithm reaches the end of the subdetector. Then, the track is fitted to extract its parameters, and the hits associated to the track are removed for subsequent iterations. This process is, then, repeated with a new seed. In order to keep tracking efficiency high, the seeding and quality requirements applied to reconstruct the track are relaxed in each iteration. Once tracks have been reconstructed, primary vertices, that may correspond to several interactions in the same event, can be identified. The primary vertex is defined as the one with the highest quadratic p_T sum of their tracks.

Energy deposits in the calorimeters are also reconstructed using an iterative approach. Clustering of the deposits is performed independently in the ECAL barrel and endcaps, as well as, on the HCAL barrel and endcaps. Cells with energy higher than a certain threshold, and higher than the neighbouring cells are used as seeds. Clusters are created by aggregating cells which are adjacent to the clusters and have an energy excess.

These basic elements reconstructed in each subdetector are linked by the PF algorithm using geometrical matching. Reconstructed tracks are extrapolated to the calorimeters, deposits in the ECAL and HCAL are compared in order to check for their compatibility. Tracker tracks are extrapolated to the muon system and its compatibility with muon hits is checked. This linkage is used to identify the particle, as each type will

interact with a certain subdetector or subdetectors. Tracker tracks that can be associated to deposition in the ECAL cluster are good electron candidates. Muons can be reconstructed by associating tracker tracks with segments in the muon system. Electron and muon reconstruction is treated in detail in Chapter 3. Tracks reconstructed in the tracker and matched to depositions in the calorimeters are used to reconstruct charged hadrons, while photons and neutral hadrons will leave only energy deposits in the ECAL and HCAL, respectively. With this information it is possible to reconstruct the physical objects used in the analysis, namely jets, b-tagged jets, missing transverse energy, electrons, muons, taus, and photons. In the following sections, a summary on how these objects are reconstructed is given, except for leptons and taus, which are covered in Chapter 3.

2.3.1 Jets

As discussed in Chapter 1.1, the effect of colour confinement implies that quarks and gluons produced in the collision are not allowed to remain as singlet states. They therefore radiate partons until they form hadrons, which are bounded, colourless states. This process is called hadronization and generates a spray of particles collimated in a given direction that interacts with the detector. This spray of particles is know as a jet. It is needed, then, to cluster the multiple particles coming from a certain hadron into a jet. This is a difficult task that in CMS is performed by the anti-k_t algorithm [86]. This algorithm uses all the particle candidates constructed by PF algorithm to run an iterative grouping. The criteria is to group together near particle candidates using the p_T and angular distance. The process is repeated taking into account all particle candidates until they are clustered into jets for a given radius parameter R. In CMS a distance parameter of 0.4 is usually used.

2.3.2 b tagging

It is useful to determine the flavour of the partons that have produced a certain jet. In the case of b quarks and c quarks it is possible, as the hadrons containing these types of quarks present certain features. These hadrons present a longer lifetime than other light flavour hadrons, such characteristic allows them to fly several millimetres before decaying. Consequently, a secondary vertex is generated within the jet cone section, as depicted in Fig. 2.4. Besides, b (and c) quarks mass is larger than the one of light flavour quarks and gluons; therefore, the particles produced in the decay will have a larger momentum relative to the jet axis.
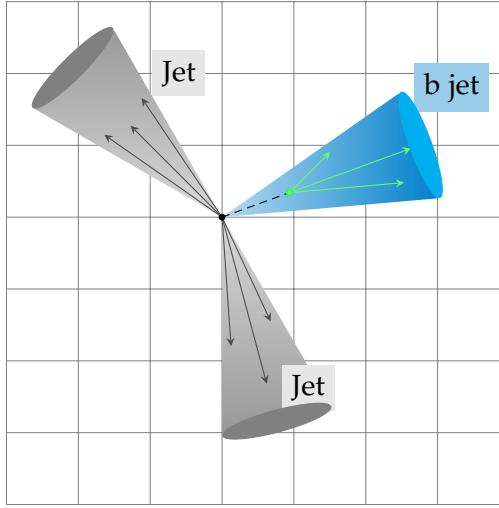
FIGURE 2.4: Depiction of three jets, the blue one is a b jet. A secondary vertex is visible within the b jet cone.
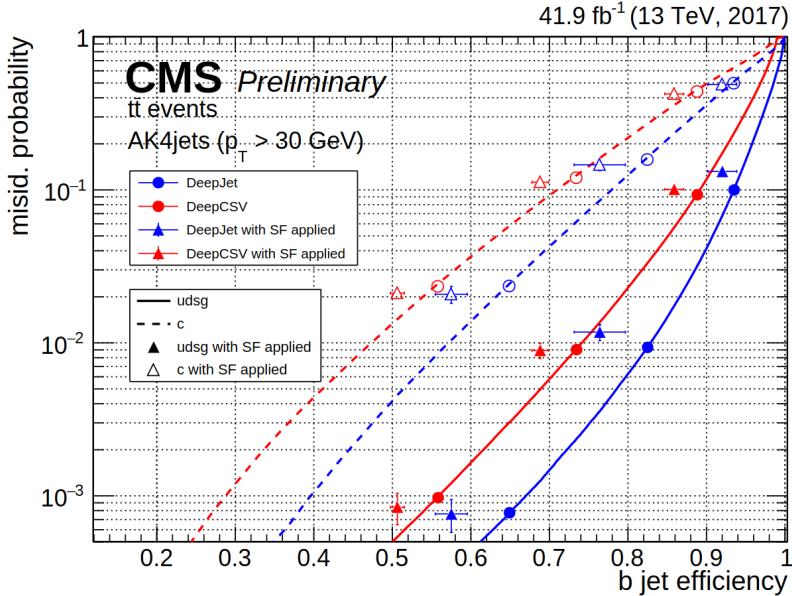
FIGURE 2.5: Performance of the `DeepJet` and `DeepCSV` b-tagging algorithms on $t\bar{t}$ events with both top quarks decaying hadronically. The jets are required to have $p_T > 30$ GeV . The performance is shown for both b vs. c classification (dashed lines), and b vs. light (solid lines) [88].

Several algorithms have been developed within CMS for b jet identification. In this thesis the `DeepJet` [87] algorithm is used. This tagger is based on a neural network that makes use of full information of all jet constituents, charged and neutral particles, secondary vertices, and global event variables simultaneously. Depending on the light jet misidentification probability allowed, three working points (WPs) are defined to tag b jets. WPs loose, medium and tight, target a misidentification probability of 10, 1, and 0.1%, respectively. Performance of the `DeepJet` discriminant in comparison with the previous, widely used discriminant (`DeepCSV`) is shown in Fig. 2.5 .

2.3.3 Photons

Photons usually pass the tracker without interacting and deposit more than 95 % of its energy in the ECAL, producing a shower. Therefore, photon reconstruction is seeded from ECAL energy clusters with $E_T > 10$ GeV which are not related to any good quality tracker track; the energy cluster should be isolated and be compatible with a photon shower.

In some occasions, photons interact with the tracker materials and produce an electron-positron pair before reaching the ECAL. This is the so-called *photon conversion*. These type of photons present a different signature: the produced electron and positron have bent trajectories in the tracker with $\vec{p}_T$ almost parallel and energy spread in the ϕ plane. An alternative reconstruction method is applied to also reconstruct this kind of events.

2.3.4 Missing transverse energy

Precise reconstruction of particles in the event, as well as, the correct association of this particles to a specific vertex is crucial in order to infer the presence of particles not interacting with the detector, i.e. neutrinos, but also new particles postulated in BSM theories. The missing transverse momentum, $\vec{p}_T^{\text{miss}}$, is defined as follows:

$$\vec{p}_T^{\text{miss}} = -\sum_i \vec{p}_T(i) \qquad (2.6)$$

Where the index i runs over all reconstructed particles associated to a certain vertex. Give momentum conservation, in case no invisible particles are present the p_T^{miss} in the event should be zero. Nevertheless, there are detector effects that prevent this from being true, such as energy resolution, detector noise, misreconstructed particles, primary vertex determination and underlying event.

In order to mitigate these effects, a set of filters [89] had been developed to identify events where the p_T^{miss} is expected to be poorly reconstructed. The impact of applying these filters is depicted in Fig. 2.6.

Other relevant related quantity is H_T^{miss}, which is defined as the p_T^{miss} but considering only jets, leptons and τ_s with certain selection criteria. This definition is less sensitive to invisible particles, but it is also less dependent of energy mismeasurements and spurious signals.
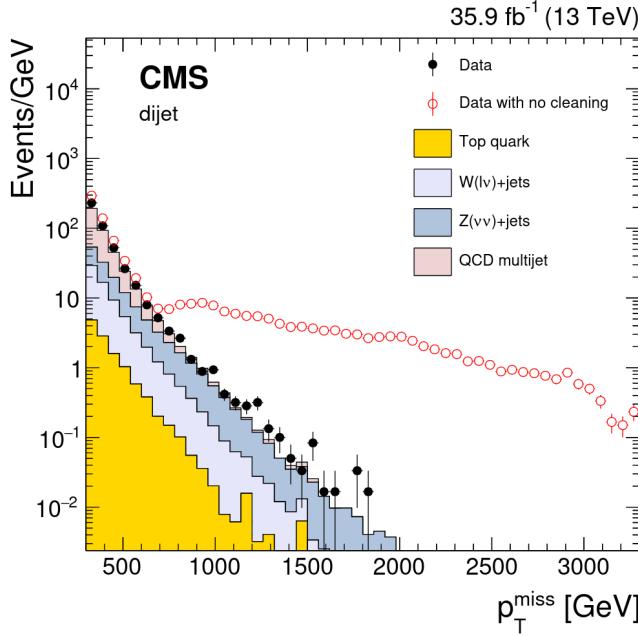
FIGURE 2.6: The p_T^{miss} distributions for events in a multi-jet selection with the event filtering algorithms applied (filled markers), without the event filtering algorithms applied (open markers), and from simulation (solid histograms). Taken from [89].

2.4 Data taking conditions

The LHC collider and the CMS experiment are very complex experimental setups that are not completely stable with time. The working conditions of the LHC changes, while CMS undergoes upgrades and suffers degradation in some of its subdetectors as a consequence of the collisions. These factors collectively introduce non-uniform data acquisition conditions over time. These effects are carefully considered during the data analysis by applying corrections or dedicated simulations. The main features of Run-2 data taking period are described below.

2.4.1 Pileup conditions

As described in Section 2.1, the pileup conditions are a direct consequence of the LHC collision parameters selected. Figure 2.1 shows that the number of simultaneous interactions per BX increased significantly for 2017 and 2018 with respect to 2016 data taking period. This directly affects the performance of p_T^{miss}, jet energy and lepton isolation measurements. Additionally, triggers are affected, as the increase in instantaneous luminosity creates the need to tighten the HLT requirements to keep the rate under control.

2.4.2   Detector upgrades

A partial update of the tracker was performed during 2017. The upgrade introduced an additional layer located in the innermost layer of the pixel tracker, both in the endcap and the barrel. This would allow at least 3-hit coverage across all the tracker acceptance as well as improve the resolution on the impact parameter measurement.

Nevertheless, during 2017 data taking some of the pixel modules failed. The failure was caused by radiation-induced damage in the components of the power supply. Despite this, the degradation in performance was not dramatic given the redundancy of the system. Before 2018 data taking period all the modules where repaired.

2.4.3   Degradation effects on the detector

During data taking the detector is exposed to high radiation environments, specially the innermost detectors. The tracking system is the subdetector subjected to the highest radiation densities, but it is also built using materials resilient to the radiation. For this reason, the subdetectors suffering more significant degradation effects due to radiation are the calorimeters.

ECAL crystals suffer a gradual lose of transparency as the radiation dose accumulates. As a result, the response of the crystals is delayed, shifting the ECAL timing. This shift was not propagated to L1 calibrations during 2016 and 2017 data taking years, meaning that some of the objects reconstructed in the ECAL (at high η) and firing the trigger were associated to the wrong bunch crossing. The L1 trigger logic does not allow to trigger in the two BX after the triggered one, as a consequence, if this effects takes place, the collision to be measured would be vetoed and the wrong event would be saved, this effect is know as L1 preferring. It is possible to measure the probability of this effect to happen by selecting events 3 BX after a trigger has been fired. These events are not affected by the prefiring by definition of the trigger rules. Corrections due to the prefiring are applyed when needed, and are found to be small, of the order of a few percent. In the case of data taken during 2018, the effect was accounted for in the L1 calibrations, so no correction is needed for that year.

In addition, in 2018, a section of the HCAL in the endcap region experienced a failure and could not be recovered during the data taking period. As a consequence, the energy estimation of jets, photons and electrons in the $-1.57 < \phi < -0.87$ and $-3.0 < \eta < -1.3$ region suffered a miscalibration. In the analysis presented in this thesis, an extra uncertainty is applied to account for this miscalibration.

2.5   Corrections to simulation

Monte Carlo simulations reproduce quite reliably most physical processes that take place in a pp collision; in Section 1.3 the generators used to model the collision are described. Nevertheless, one needs also to model how the final state particles interact with the detector, as well as, the detector response. `Geant4` [90] is used to model detector interaction, but there are certain effects such as complex interaction processes, magnetic field anisotropies or alignment that are not straight forward to include in the simulation. As a result, part of the data taken by the detector must be used to extract calibrations for the objects reconstructed and identified and use those calibrations to correct the simulation. These correction have associated uncertainties that must be taken into account in the data analysis as systematic uncertainties. In this section the methods used to derive these calibrations are discussed.

2.5.1   Trigger efficiency

There are several strategies to estimate the trigger efficiency depending on the nature of the trigger used. In this thesis a combination of single, double and triple lepton triggers is used. The method used to measure the trigger efficiency is called orthogonal trigger method, and is described in the following lines.

Events used to measure the trigger efficiency are collected using a set of triggers, known as reference triggers, that are independent to the ones required at analysis level. In particular, the reference triggers used in this thesis are based on the momentum imbalance and the jet activity of the event. Furthermore, events should pass a minimal offline selection (Off. Selection) related to the number of leptons in the final state and their minimum p_T. Then, the efficiency for a given leptonic trigger can be computed as:

$$\epsilon(\text{Trigger}) = \frac{\text{N(Off. Seletion \& Ref.trigger \& Trigger)}}{\text{N(Off. Seletion \& Ref. Trigger)}}. \qquad (2.7)$$

Trigger efficiencies are highly dependent of the lepton kinematics, hence the measurement is usually performed as a function of the p_T and η of the leptons. Uncertainties on the measurement are considered: the statically uncertainty of the selected samples and the comparison between both sets of reference triggers. The correlation between the reference triggers and the leptonic triggers is not zero, as the mismeasurement of a lepton can yield to instrumental missing transverse energy. Nevertheless, the correlation is small and covered by the uncertainties.

2.5.2 Lepton efficiency

Efficiency of lepton reconstruction and selection is evaluated using a "tag-and-probe method" [91]. This will be covered in detail in Chapter 3, where a dedicated description of lepton reconstruction and selection is given.

2.5.3 Jet energy and resolution

Jet reconstruction is described in Section 2.3.1. As jets are composite objects, their reconstruction is particularly complex. To ensure an accurate estimation of the jet energy scale and p_T^{jet} resolution, a series of corrections are applied in both data and simulation. Corrections on the jet energy resolution ensure that the simulation describes well the data.

The correction procedure is divided into several sequential steps, with a comprehensive description of each step provided in Ref. [92]. Firstly, contributions from pileup and underlying event particles are subtracted in both data and simulation. Subsequently, a comparison is made between the momentum of the simulated jet at the reconstruction level and that at the generation level. Corrections are then derived for both data and simulation to ensure a consistent response as a function of p_T and η. Finally, residual corrections are applied only on data, to account for discrepancies between data and MC simulations in various topological scenarios.

All these jet corrections are applied individually to each jet within each event, taking into account their p_T, η, and the detector region. Moreover, these corrections are propagated to certain sensitive variables, such as $\vec{p}_T^{\text{miss}}$. Additionally, each correction comes with a corresponding set of systematic uncertainties, which are appropriately propagated into the estimation of the p_T and other affected observables.

2.5.4 b tagging

The b-tagging algorithms described in Section 2.3.2 are susceptible to missmodelling effects in simulation. It is needed to evaluate the b-tagging efficiency, as well as the mistagging efficiency, the techniques used are fully described in Ref. [93], below a summary is presented.

Measuring efficiency in MC simulations is straightforward, as the generated information allows to determine if a jet originated from a b hadron. Therefore, efficiency is defined as the number of b jets correctly tagged by the algorithm (at a specific WP)

divided by the total number of actual b jets. Evaluating efficiency in data demands a more difficult approach. This process involves selecting a sample enriched in b quarks, achieved via two methods: employing a multijet sample where events with non-isolated muons in the low-p_{T} range are chosen, or utilizing a $t\bar{t}$ sample focused on final states with two isolated leptons. In both cases, several techniques exploiting kinematic differences between light jets and b jets, are applied to accurately determine the proportion of jets and b jets that either pass or fail the b-tagging selection.

In order to estimate the mistagging efficiency, corresponding to jets originating from gluons or light quarks a multijet enriched sample is used and the "negative-tag method" is applied. This technique is characterized by the definition of positive and negative taggers. These taggers are constructed using the nominal tagging algorithms, although restricted to jets with tracks exhibiting positive or negative values of the impact parameter and secondary vertices with either positive or negative flight distances. The algorithms are expected to be symmetric for light jets, while an asymmetry is shown for b and c jets, making it possible to measure the proportion of light jets classified as b jets by the algorithm.

The uncertainty associated with the measurement of b-tagging efficiency includes factors such as lepton and jet identification efficiency, energy scale corrections, and the effects arising from gluon splitting, quark fragmentation, and the methods employed to differentiate light jets from b jets.

Additionally, quark- and gluon-inducedf jets can be distinguished in CMS, using the Quark-Gluon Likelihood algorithm [93]. Gluon-induced jets are wider, present higher particle multiplicities and have a more uniform energy fragmentation when compared to a quark-induced one. These characteristics are exploited by the algorithm in order to classify jets as quark- or gluon-induced.

2.6   Statistical treatment of the data

When analyzing the data taken by CMS, it is usual to make use of statistical techniques that maximise the separation power between two classes of entities, namely two classes of events, or two classes of physics objects. Multivariate analysis techniques (MVAs) are used to classify events in two or more categories using several physical observables as input variables.

Three types of supervised learning algorithms are used in this thesis: Random Forests (RFs), Boosted Decision Trees (BDTs) and Neural Networks (NNs). They are trained on a simulated sample where it is possible to know, by construction, the class each

event or object belongs to (e.g. signal vs background, muon vs jet misreconstructed as a muon, etc); then, the algorithm *learns* to perform the classification based on the input information provided and can be applied to other (not tagged) simulation events or to data.

RFs are classifiers combining several decision trees to produce an unique output. A decision tree is a classifier that applies several sequential criteria. Hence, a first selection is done in a certain variable and, according to it, the data set is divided in two subsets or nodes. The process is then repeated with other selection criteria until a certain predetermined level of signal purity is achieved in each node. The maximum number of branching operations needed to achieve the target purity is called the depth of the tree. In a RF several decision trees are used: each of them is constructed based on a random subset of the training data and a random subset of features at each node. This randomness introduces diversity among the trees, and the final prediction is obtained by aggregating the predictions of all individual trees. For classification tasks, the most frequent class among the trees is selected, while for regression tasks, the average prediction is taken. A extended description of random forest algorithms can be found in Ref. [94].

BDTs are binary classifierss also built using decision trees. In this case, in order to combine the trees, a technique called *boosting* is used. This is based on training a first layer of *weak learners*. Then, each subsequent tree is trained to classify samples for which the previous classifier has failed. There are several boosting algorithms available. More information about this kind of discriminants can be found in Ref.[95].

Many types of NNs have been developed recently [96]. The NNs used in this thesis are feed-forward NNs. A neural network is an algorithm that creates a mapping or correspondence between a set of *neurons* in the input layer, as many as the input variables, and a certain number of neurons in the output layer, that provide the desired classification. A certain activation function (f) is applied to create each neuron as $f(\sum a_i x_i + b_i)$ where x_i are the input variables, while a_i are the weights and b_i the biases, parameters of the model. Such parameters are obtained during the training and are chosen to minimize the cross-entropy variable. Several activation functions and minimization techniques to obtain the parameters can be used.

3

Lepton reconstruction, selection, identification and isolation

In this thesis, final states with multiple leptons are used to study the physical processes of interest. For this reason, efficient lepton reconstruction, identification, and selection are key elements and are covered in this chapter. Firstly, muon and electron reconstruction algorithms within CMS are explained. Afterwards, muon and electron identification is discussed. As muon selection was one of the tasks developed during my PhD, in this section special focus is given to the performance of muon identification during Run-2. A new identification technique developed for Run-3 and based on an MVA is also presented. Later, how to select isolated leptons is discussed, including the description of an MVA trained to select prompt isolated leptons as the ones targeted in the final states of this thesis. Finally, hadronic τ reconstruction is briefly discussed.

3.1 Muon reconstruction

The tracker and the muon system are the two CMS subdetectors that play a role in muon reconstruction. Three algorithms are used to construct muon tracks: *standalone muon*, *tracker muon* and *global muon*. In this section they are briefly discussed, a complete description of the algorithms can be found in Ref. [97]. The first step of the reconstruction is performed locally in the muon system and the tracker separately.

Local reconstruction in the muon system uses information from the several types of detectors present in this part of CMS, as explained in Section 2.2.4. The precise location of each hit is reconstructed from electronic signals in each chamber. The algorithm used for this task is different depending on the detector technology. In addition, DTs and CSCs are multilayer detectors, therefore, hits in each layer can be associated

following a straight-line track within each CSC or DT chamber to create the so-called segments. On the other hand, local reconstruction on the tracker has already been discussed in Section 2.3.

standalone muon reconstruction is seeded by groups of DT or CSC segments. Then, the track candidate is propagated and updated adding also information from RPCs and using a Kalman-filter algorithm [98]. Tracker muon tracks are reconstructed by extrapolating all tracks in the tracker to the muon system. A geometrical matching is performed between the extrapolated track and DT or CSC segments. In order to consider the track as muon tracker track, it should match at least one muon segment. Finally, global muon tracks are built by matching standalone tracks with tracker tracks. Both tracks are extrapolated to a common surface, where its parameters are compared by performing a combined fit using a Kalman-filter.

For muons produced within the muon system acceptance the probability of being reconstructed either as global or tracker muons is of 99%. Tracker muon algorithm is designed to be highly efficient, specially in zones where the muon system is less redundant. The global muon algorithm reconstructs muons passing the muon system with high purity. These reconstructed muon tracks are fed into the PF algorithm applying additional criteria from other CMS subdetectors.

Muon momentum is estimated using the inner track information, as for muons with $p_T < 200$ GeV, the contribution from the muon system to the momentum measurement is marginal. For high p_T muons, dedicated techniques are used [99].

3.2   Electron reconstruction

Electrons are reconstructed by using information form the tracker and the calorimeters, a complete description of electron reconstruction in CMS can be found in Ref. [100]. As electrons pass through the tracker, they will suffer bremsstrahlung and, as a consequence, they will radiate photons. This implies that the trajectory of the electron will change. Furthermore, when the electron reaches the ECAL, it will probably be composed by a combination of electrons and photons. Reconstruction algorithms take this into account in order to create a single electron candidate containing all the associated objects.

In the ECAL, dedicated electron clustering algorithms are used. These take into account that the energy of the radiated photons mainly spreads along the ϕ direction. This is a consequence of the bending of the electron trajectory in the magnetic field.

Two algorithms, applying different criteria for the barrel and endcap, are used to associate the deposits corresponding to electrons and photons into a supercluster (SC).

Three dedicated electron seedings are used. The ECAL-based seeding uses ECAL deposits corresponding to a electron candidate and associates related electrons and photons into a SC. The second type of seeding uses doublets of tracker hits matched geometrically to the ECAL SC. The tracker-based seeding uses tracks reconstructed using the general algorithm for charged particles, that are compatible with an electron. To recover efficiency, the two later methods are combined using dedicated tracking algorithms to model the energy loss due to bremsstrahlung; these algorithms use Gaussian Sum Filtering (GSF) [101]. Electron tracks reconstructed using these techniques are called GSF tracks.

ECAL clusters, GSF tracks, as well as SC associated to tracker tracks are passed as input for the PF algorithm, which finally produces electron and photon candidates.

3.3   Muon identification

Several quality criteria on reconstructed muons are applied at analysis level to assure high purity. Three working points are defined for muons of p_T between 15 and 200 GeV, depending on the efficiency and purity level needed for a specific study: tight, medium, and loose identification (ID) criteria. For low-p_T and high p_T muons, dedicated selections are developed, but are not described in this thesis.

Loose ID aims to be highly efficient, but maintaining a low rate on charged hadrons misidentified as muons. It is able to identify prompt muons, which are those originating at the primary vertex, as well as, muons from light and heavy flavour decays. A loose muon is defined as a muon selected by the PF algorithm that is also a tracker or a global muon.

Medium and tight ID muons are defined as muons passing the loose ID criteria and adding extra sequential quality criteria on several variables such as[1]:

  − The muon track fit χ^2.

  − The number of muon stations with hits included in the muon track fit.

  − The number of hits in the pixel detector included in the muon track fit.

  ---
  [1]Loose, medium and Tight ID are know as cut-based IDs

— The χ^2 of a kink-finding algorithm on the inner track. The algorithm splits the inner track in two parts in several places along the trajectory and compares them. A large χ^2 indicates that the two parts are not compatible with being a single track.

— For global muons, the χ^2 of the position matching between the inner and standalone tracks.

— The muon segment compatibility, which is evaluated by using both the number of matched segments in all stations and the closeness of the matching in position and direction with the pattern expected for a minimum ionizing particle.

— The number of tracker layers used in the muon track fit.

— The fraction of tracker hits used for the fit of the inner track.

— The number of matched stations: number of segments (one per station) reconstructed in the muon chambers and used in the global muon track fit.

— The compatibility of the track with the PV both in the xy and z plane.

Medium ID is created in order to select prompt muons and muons from heavy flavour decay [97], the corresponding selection is defined in the following lines. The tracker track of a medium muon is required to use hits from more than 80% of the inner tracker layers that it traverses. If the muon is not reconstructed as global, the muon segment compatibility must be greater than 0.451. In any other case, when the muon is reconstructed as both a tracker muon and a global muon, the muon segment compatibility is required to be > 0.303; In that scenario other requirements are applied: the muon track fit must have a $\chi^2 < 3$, the χ^2 of the position matching between the inner and standalone tracks must be lower than 12, and the maximum χ^2 of the kink-finder must be less than 20.

Tight ID is defined to select prompt muons, rejecting both muons from decays in flight, as well as, charged hadrons misidentified as muons [97]. The selection is described in the following lines. A tight muon is a global muon whose tracker track uses hits from more than 5 layers of the inner tracker and includes in the tracker fit at least one pixel hit. Additionally, the tracker muon track must match hits in more than one muon stations. The global muon fit must include at least one hit in the muon system and have a $\chi^2 < 10$. Compatibility criteria with the PV is also imposed: requiring a transverse impact parameter $|d_{xy}| < 0.2$ cm and a longitudinal impact parameter $|d_z| < 0.5$ cm.

3.3.1   Muon performance during Run-2

Muon selection performance can be evaluated by computing the efficiency as a function of the muon p_T and η, using the "tag-and-probe method". This technique selects muons coming from the Z boson decay, as described in Ref. [91]. Tag muons are required to pass the tight ID criteria and, to avoid any bias in the efficiency measurement, to be geometrically matched with the muon that triggered the event. Probe muons are required to be tracker muons and have a $p_T > 20$ GeV. Simulated samples of Z/γ events, also known as Drell-Yan (DY), are usually used.

Muon performance during Run-2 have been documented in Refs. [102] and [103]. The efficiency is shown to be stable between the across the data taking years, even with different detector geometry, which show the robustness of muon identification. As an example, 2018 efficiencies for loose ID and tight ID muons are shown in Fig. 3.1. The efficiency of the loose selection is above 98% for all η and p_T regions. Tight muon efficiency is above 95% except in η regions where the muon system is less instrumented.

3.3.2   Developments on muon selection for Run-3

As shown in the previous section, the muon identification performance during Run-2 was outstanding. Nevertheless, in cases where the purity of the selected sample is critical, or when the muon identification performance is degraded due to PU, the use of more advanced techniques is crucial to maintain an acceptable compromise between efficiency and purity. Studies on the medium and tight identification criteria showed that these selections could be optimized, specially in the context of higher number of simultaneous collisions.

In order to do so, an MVA technique was used. In particular a RF classifier trained in a $t\bar{t}$ sample. The training variables are the ones used to define the cut-based muon selection used during Run-2, and described in Section 3.3; with the exception of the impact parameters. This discriminant, that from now on will be called muon MVA ID, is trained to select genuine muons.

A $t\bar{t}$ sample was chosen for the training as it provides genuine muons from diverse sources, such as muons from prompt decays and from heavy-flavour hadron decays. The MVA is trained and tested using 2018 simulation, as this year is consider to be representative of Run-2 conditions. Besides, the PU profile was higher than in other Run-2 data taking years, and closer to the one in Run-3.
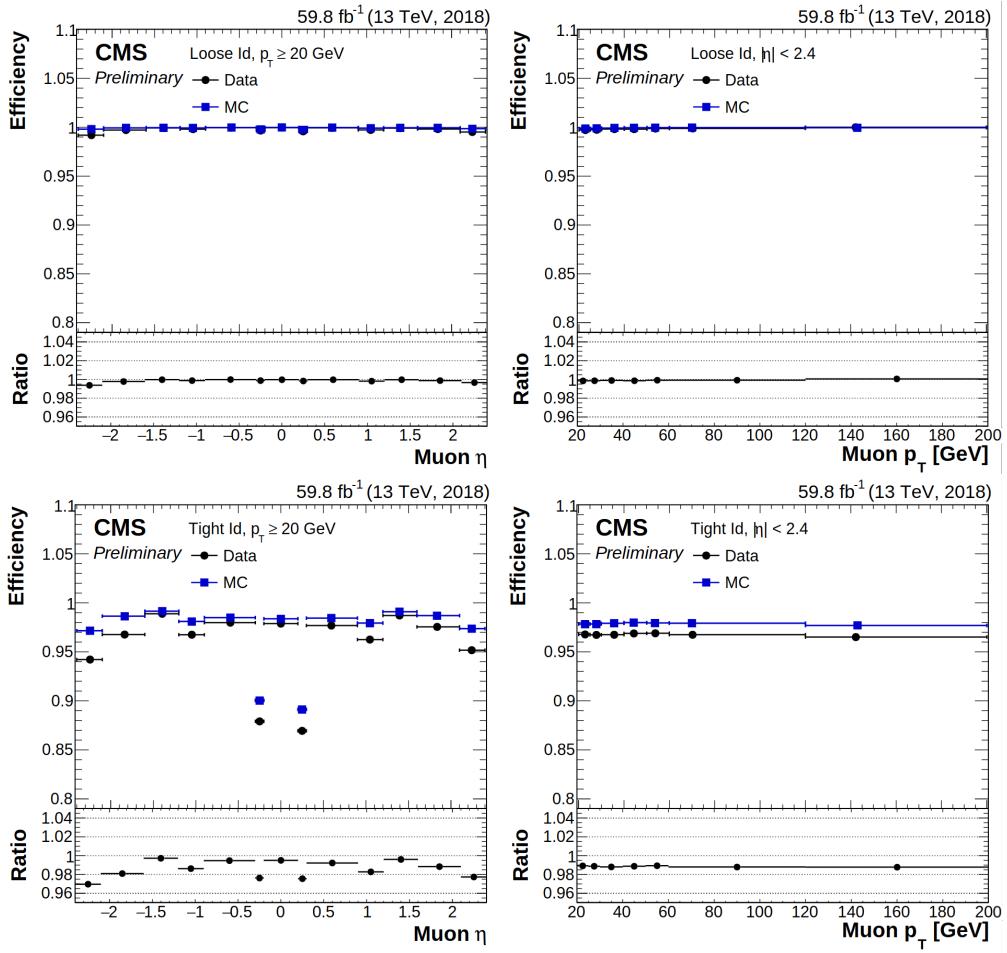
FIGURE 3.1: Muon efficiency for the loose ID (upper row) and tight ID (lower row) selection in 2018. Efficiency is measured using data and simulation. Left plots show the efficiency as a function of η while right plot shows it as a function of p_T. The lower plot shows the data to simulation agreement. Figures extracted from Ref. [103].

In order to determine the source of the muon in simulation, generation information is used. A geometrical matching between the hits of the simulated muon track and those of the muon track reconstructed in the muon system is performed, then the simulated muon track is used to trace the generated parent particle. Signal muons are those produced promptly in the decay of W, Z, and Higgs bosons, isolated muons from τ leptons decays, and nonisolated muons from heavy-flavour hadron decays. Muons from the decay of light hadrons (such as kaons and pions), and spurious signatures in the detector misreconstructed as muons, i.e. hadrons reaching the muon system, are classified as background.

The optimization was performed using the area under the receiver operating characteristic (ROC) curve, as well as the efficiency and the background rejection as figures of merit. The ROC curve for this MVA is shown in Fig. 3.2, which also shows the signal efficiency and the background rejection for the cut-based ID working points medium
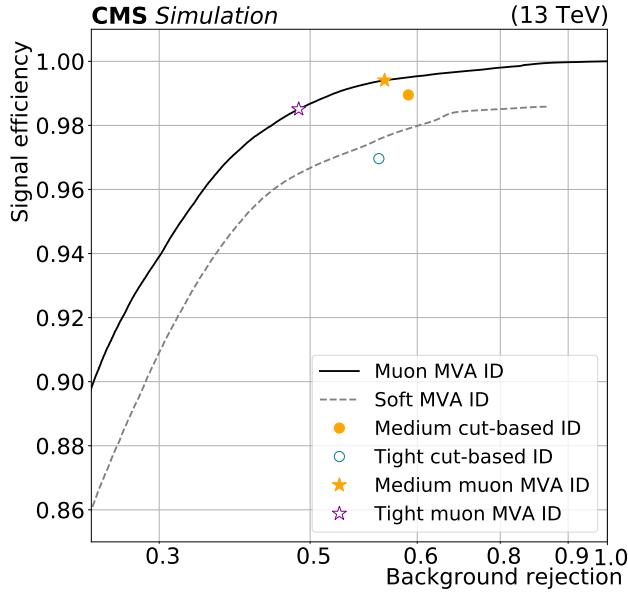
FIGURE 3.2: ROC curve for muons with $p_T > 10$ GeV for the muon MVA ID discriminant (black solid line) its medium and tight WPs shown as orange solid and purple open stars, respectively. Orange solid and blue open points show the medium and tight WPs of the cut-based ID. The ROC curve of the soft MVA ID [104] is also shown (grey dashed line).

and tight. Additionally, the ROC curve for the soft MVA [104] is shown: this MVA was developed during Run-2 to select low-p_T muons specially arising from heavy-flavour decays.

Two working points for the muon MVA ID are provided. The Medium WP is designed to provide the same background rejection as the cut-based selection medium WP, calculated for muons with $p_T > 20$ GeV. The tight WP is designed to reduce the background contribution a 10% more than the medium WP, but still keeps high efficiency.

Performance of the MVA is checked in data and simulation using the tag-and-probe method introduced in the previous section; in this case, probe muons are required to be loose muons. Efficiency is shown for the medium WP of the muon MVA ID and the cut-based ID as a function of p_T, η and number of primary vertices in Fig. 3.3. The efficiency of the medium MVA ID is higher than 99.5%, and about 0.5% higher than the one achieved by the cut-based ID for a similar background contamination. MVA ID efficiency remains above 99.5% even for events with up to 60 PU vertices, showing great stability as a function of PU. Instead, for the medium ID cut-based selection a decrease in efficiency as a function of PU is shown. This is an important feature of the MVA ID, as it will behave better than the cut-based one in Run-3 PU conditions.
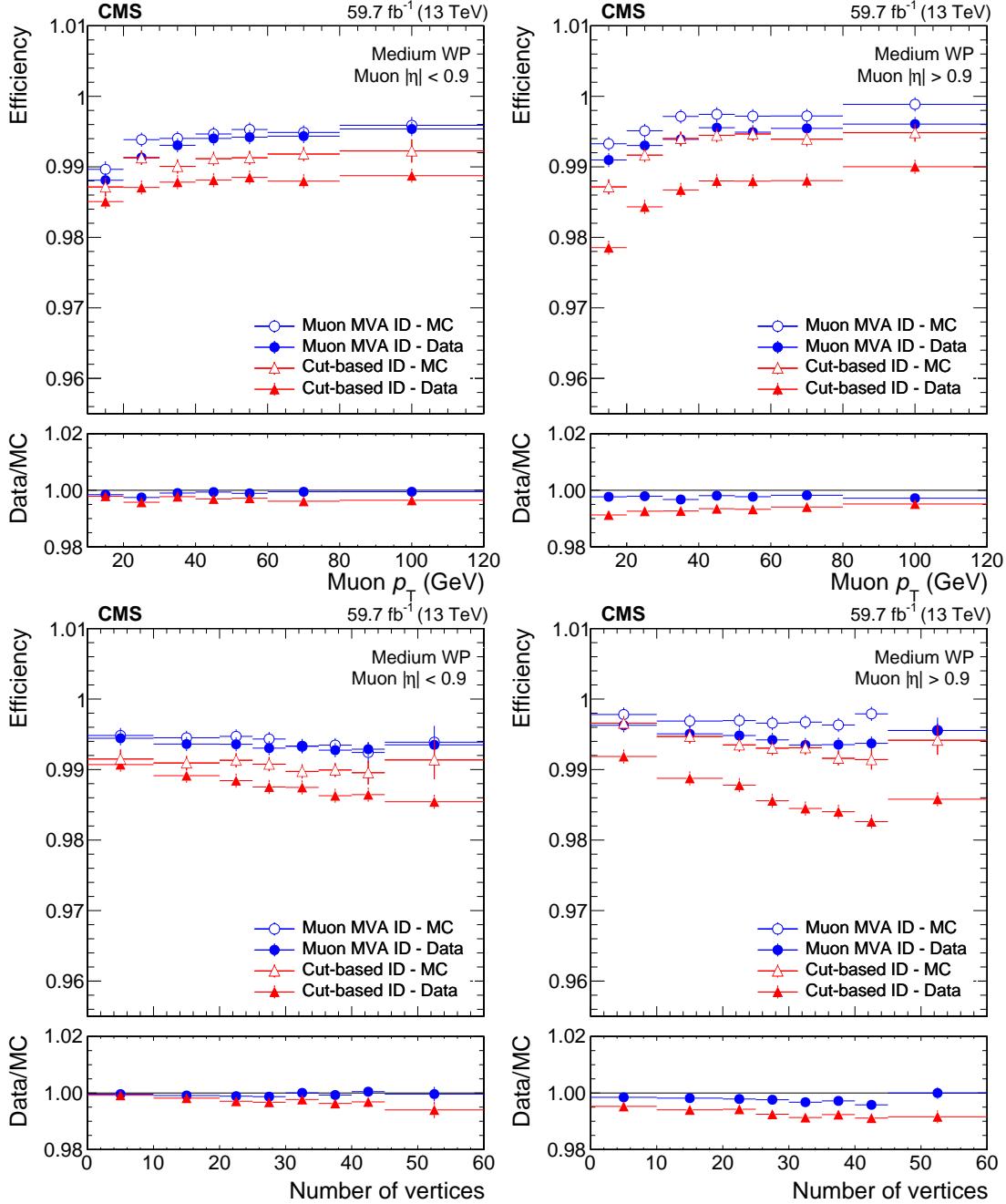
FIGURE 3.3: Muon identification efficiency for the medium WP as a function of p_T (upper row) and the number of vertices (lower row) for muons with $|\eta| < 0.9$ (left) and $|\eta| > 0.9$ (right). Blue dots show the muon MVA ID performance both for the 2018 data set and DY simulation, while red triangles show the efficiency of the medium cut-based ID used during Run-2. Data to MC ratio is also shown.

3.4   Electron identification

Additional criteria are applied over reconstructed electrons in order to select prompt isolated electrons and distinguish them from photon conversions, jets misidentified as electrons, or electrons from semileptonic decays of b and c quarks. Several discriminating variables can be used in order to select prompt electrons. More complex discriminants based on MVAs are also developed. Here a brief review is given, for a more comprehensive description see Ref. [100]

In order to discriminate electrons from jets misidentified as electrons, the lateral extension of the shower in the η direction can be defined as follows:

$$\sigma_{\eta\eta}^2 = \frac{\sum(\eta_i - \eta)^2 w_i}{\sum w_i} \tag{3.1}$$

The sum runs over the 5×5 matrix of ECAL crystals around the crystal with highest energy in the SC, w_i is a weight depending logarithmically on the energy, and the positions η_i are expressed in units of crystals. Variables related to the energy deposits are also used. For example, the ratio between the energy deposited in the ECAL and HCAL, H/E; and the difference between the inverse of the SC energy and the inverse of the track momentum, $1/E_{SC} - 1/p$.

In order to discriminate signal electrons from conversion ones, the pattern of track hits can be used. When the photon conversions happens inside the tracker, the first hit on the track will not be located in the innermost layer of the tracker. Therefore, in order to reject photon conversions, tracks with not missing hits are selected. Topological features of the track can also be used to decide if the electron is likely to come from a photon. Such information can be used to reject events compatible with a conversion.

More advance selection criteria were developed in order to improve electron selection. A BDT is trained in a DY+jets simulation, using prompt electrons as signal and jets as background. The input variables to the BDT are the variables defined above, along with the SC position, relative position between the SC and the tracker track and tracker parameters. Several WPs are defined, in this thesis, *WP-loose* and *WP-80*, corresponding to a 98% and 80% signal efficiency, respectively, are used.

3.5 Lepton isolation and the prompt lepton MVA

3.5.1 Lepton isolation

In order to select prompt leptons coming from the decay of a Z, W or H bosons; or leptons coming from a τ decay, extra criteria is needed. Those leptons need to be distinguished from other genuine leptons, arising from heavy flavour decays. Non-prompt leptons are usually produced within a jet cone; therefore, the detection of other objects around the lepton can be used to discriminate them. Isolation variables measure the energy in the vicinity of the lepton and provide discrimination between prompt and nonprompt leptons.

PF relative isolation is computed as the sum of the p_T of all charged hadrons, the transverse energies of all photons and of all neutral hadrons reconstructed by the PF algorithm within a ΔR region surrounding the lepton, and divided by the p_T of the lepton. Charged PF candidates not associated with the PV are not considered, and contributions from neutral PU particles are taken into account with two alternative estimations: the $\Delta\beta$ corrections or the effective area corrections. The isolation can then be expressed as follows:

$$I_{PF} = \left(\sum p_T^{\text{ch. hadrons}} + \max\left(0, \sum p_T^{\text{neutral hadrons}} + \sum p_T^{\gamma} - p_T^{\text{PU}} \right) \right) / p_T^{lepton} \qquad (3.2)$$

When using the $\Delta\beta$ corrections, the PU term corresponds to the sum of charged-hadron deposits originating from PU vertices scaled by a factor of 0.5. This factor, computed in simulation, is the ratio between charged and neutral particle production rate at inelastic pp collisions. For the effective areas method, the PU term is expressed as: $\rho \mathcal{A}_{eff}$, where ρ is the average energy density in the event, and $\mathcal{A}_{eff}$ (the effective area) express the dependence of isolation as a function of the number of reconstructed vertices in the collision.

The size of the cone depends on the desired selection and on the type of lepton. The standard size used in CMS for electrons is $\Delta R = 0.3$, while for muons, $\Delta R = 0.4$ is used. In boosted topologies, varying the size of the cone with the p_T is desirable, as this allows to recover efficiency when leptons are emitted colinearly with the rest of the decay products. This type of isolation is called mini-isolation, and the cone size is reduced as a function of p_T, taking values between 0.05 and 0.2.

3.5.2 Prompt lepton MVA

In this thesis, the tt̄H, tH, and tt̄W processes are studied in final states with multiple leptons. One of the leading backgrounds arises from the misidentification of non-prompt leptons. Given the low cross section of these processes, especially tH and tt̄H, a very efficient selection of prompt leptons and leptons coming from taus is a key element. In order to enhance the discriminating power between prompt and non-prompt leptons, a multivariate discriminant, using both muon identification and isolation variables, was trained. This method has been crucial in several important CMS publications using Run-2 luminosity, such as the observation of tt̄H [9, 105], tt̄tt̄ [106] and tZq [107] processes, the precision measurement of WZ [108–110] process and the increase in sensitivity in electroweak SUSY production searches [111]. The technique has been documented for muons in Ref. [112]. The description provided in this section is done for the MVA trained and optimized in the context of tt̄H analysis. A retraining of this method is also available and was used in tt̄Z and tZq measurements. In the context of this thesis, this alternative training is used in the tt̄W measurement.

The discriminant is a BDT with gradient boosting, trained with the TMVA tool [113], using tt̄H simulation as signal and tt̄ simulation as background. Signal leptons are those reconstructed leptons in the tt̄H simulated sample that are matched at generator-level to the decay of a Z, or W boson, or a τ. Background leptons are those present in the tt̄ semileptonic sample that are not matched to a prompt W or Z boson decay or a τ decay. A very loose preselection is applied before training, requiring leptons to pass the requirements described in Table 3.1.

	Electrons	Muons
p_T	5 GeV	5 GeV
$|\eta|$	<2.5	<2.4
ID	WP-loose of the MVA and missing inner hits < 2	loose ID
Relative mini-isolation	< 40% p_T	< 40% p_T
$|d_\mathrm{xy}|$	< 0.05	< 0.05
$|d_\mathrm{z}|$	< 0.1	< 0.1
d/σ_d	< 8	< 8

TABLE 3.1: Preselection applied before the prompt lepton MVA training.

Two models are trained independently for electrons and muons. Besides, in order to account for differences in the detector and PU conditions between 2016 and 2017, two different training are applied. For 2018, the training performed in 2017 simulation is used, as no significant changes were present between the two years.

Input variable considered are kinematic variables of the leptons, isolation of leptons, impact parameter variables and, if present, properties of the jet associated with the

lepton. This jet is the one in which the PF candidate that conforms the lepton is included. Only jets with $p_T > 10$ GeV are considered. The training variables are:

- Lepton p_T and $|\eta|$.

- Charged component of the mini-isolation variable, defined as:

$$I_\ell^{\text{charged}} = \sum p_T^{\text{charged}} / p_T^\ell. \tag{3.3}$$

- Neutral component of the mini-isolation variable, using the effective areas method [97] to corrected for PU effects. The variable is defined as:

$$I_\ell^{\text{neutrals}} = \max\left(0, \sum p_T^{\text{neutral hadrons}} + \sum p_T^\gamma - \rho\,\mathcal{A}\left(\frac{R}{0.3}\right)^2\right) / p_T^\ell. \tag{3.4}$$

- Lepton-to-jet p_T ratio variable. It is the ratio of the transverse momentum of the lepton to the transverse momentum of the nearest jet: $p_T^{\text{ratio}} := p_T^\ell / p_T^{\text{jet}}$. If no jet associated to the lepton is present, this variable is set to $1/(1 + I_{\text{rel}})$, where I_{rel} is the relative isolation. Related to this variable is the jet relative isolation variable, defined as $(1/p_T^{\text{ratio}} - 1)$

- Lepton relative p_T variable. Defined as the projection of the lepton momentum in the transverse direction to the jet, $p_T^{\text{rel}} := p^\ell \sin\theta$, where θ denotes the angle between the lepton and jet momentum vectors and p^ℓ, the magnitude of the lepton momentum. If no jet associated with the lepton exists, this variable is set to zero.

- Jet b-tagging score: the value of the `DeepJet` b-tagging algorithm discriminant (defined in Section 2.3.2) of the associated jet. The variable is set to zero when there is no associated jet to the lepton.

- Jet charged constituents: the number, N_{charged}, of charged PF candidates within the associated jet. Tracks entering in the counting must be within a $R < 0.4$ of the lepton and are required to come from the PV. Minimal track quality, p_T, and impact parameter criteria are also required. If no associated jet is present, this variable takes the zero value.

- Longitudinal (d_z) and transverse (d_{xy}) impact parameters. Since these variables span over several orders of magnitude, their logarithmic value, is used.

- Significance of the three-dimensional impact parameter. It is defined as the ratio of the impact parameter of the lepton track in three dimensions with respect to the PV and its uncertainty. It is denoted as d/σ_d.
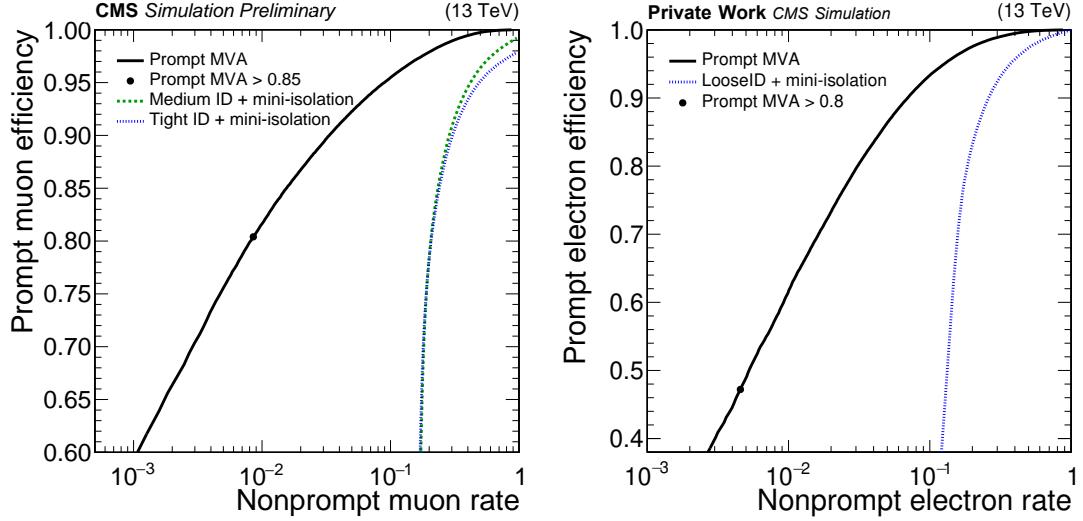
FIGURE 3.4: Black solid lines shows the ROC curve for the prompt-lepton MVA for electrons (left) and muons (right). Dashed lines show the performance for a mini-isolation based selection.

— Output of the Electron MVA ID: BDT trained to separate electrons from jets, described in Section 3.4. This variable is only used for electrons.

— Muon segment compatibility variable as defined in Section 3.3. This variable is only used for muons.

Performance is tested in a $t\bar{t}$ sample using generation information. The area under the ROC curve is used as figure of merit to tune the MVA hyperparameters. In Fig. 3.4, the performance of the prompt lepton MVA is shown along with the performance for other, simpler, isolation selections. For muons, two selections are tested; in both cases a range of mini-isolation criteria are applied on top of the tight ID and the medium ID criteria, respectively. For electrons, the selection requires the electron to pass the loose ID criteria provided by the MVA described in Section 3.4 and mini-isolation. A reduction in the nonprompt rate is observed in any point of the curve, when comparing the prompt-lepton MVA selection versus a selection based on the cut-based ID and mini-isolation. For $t\bar{t}H$ the optimal WP was found by selecting an output on the MVA score greater than 0.85 for muons and 0.8 for electrons.

3.5.3 Nonprompt muon rate measurement

For muons, the nonprompt rate is evaluated in simulations and data as a function of p_T and η. In order to perform this measurement, a sample of multijet data events enriched in nonprompt leptons is used. Events in the multijet enriched sample must contain one muon passing the single muon triggers in Table 3.2 and the selection

criteria described in Table 3.1. Additionally, the muon is required to pass the medium ID selection to suppress contributions from light flavour decays, which are poorly modelled in the used simulation. Besides, at least one jet recoiling against the muon with a $\Delta R > 0.7$ should be present in the event.

HLT path	Prescale	Lepton reco-p_T	Jet p_T
`HLT_Mu3_PFJet40`	8.99×10^3	> 3 GeV	> 45 GeV
`HLT_Mu8`	1.59×10^4	> 8 GeV	> 30 GeV
`HLT_Mu17`	5.94×10^2	> 17 GeV	> 30 GeV
`HLT_Mu20`	2.25×10^2	> 20 GeV	> 30 GeV
`HLT_Mu27`	2.25×10^2	> 27 GeV	> 30 GeV

TABLE 3.2: Triggers used to record events for the measurement of the nonprompt muon rate.

The selection described above is dominated by multijet events, nevertheless contributions due to electroweak-induced processes such as $t\bar{t}$, $W + \text{jets}$ and $Z + \text{jets}$ are expected. In order to provide an accurate measurement of the nonprompt rate, the prompt muon contribution from those processes should be extracted. The variable m_T^{fix} provides a handle to distinguish prompt and nonprompt contributions. This variable is defined as:

$$m_T^{\text{fix}} = \sqrt{2\, p_T^{\text{fix}}\, p_T^{\text{miss}}\, (1 - \cos \Delta\phi)} \qquad (3.5)$$

where $p_T^{\text{fix}} = 35\,\text{GeV}$ and $\Delta\phi$ is the azimuthal angle between the muon momentum and the $\vec{p}_T^{\text{miss}}$. This quantity is a variation of the transverse mass of the lepton and p_T^{miss}. The m_T^{fix} variable shows a kinematic endpoint around the W boson mass, similar to the m_T observable. Nevertheless, m_T is correlated with the p_T of the muon, while m_T^{fix} is not, avoiding potential biases in the measurement.

Two categories of events are created, depending on whether they pass the muon selection criteria that we what to study. In this case, the criteria applied is that the muon has an MVA score larger than 0.85. The nonprompt rate, f, is then defined as:

$$f = \frac{N_{\text{pass}}}{N_{\text{pass}} + N_{\text{fail}}} \qquad (3.6)$$

The contributions for nonprompt muons in both categories (N_{pass} and N_{fail}) is extracted from a fit to the m_T^{fix} shape. This fit is preformed simultaneously in the passing and failing categories, using templates obtained from multijet simulations and simulations from the processes leading to prompt-leptons as fitting distributions. The normalizations of the templates are free parameters in the fit. Modification on the

shape of these templates are allowed by adding nuisance parameters to the fit. Statistical uncertainties of the templates are also nuisance parameters in the fit. An example of m_T^{fix} distribution after the fit is shown in Fig. 3.5.
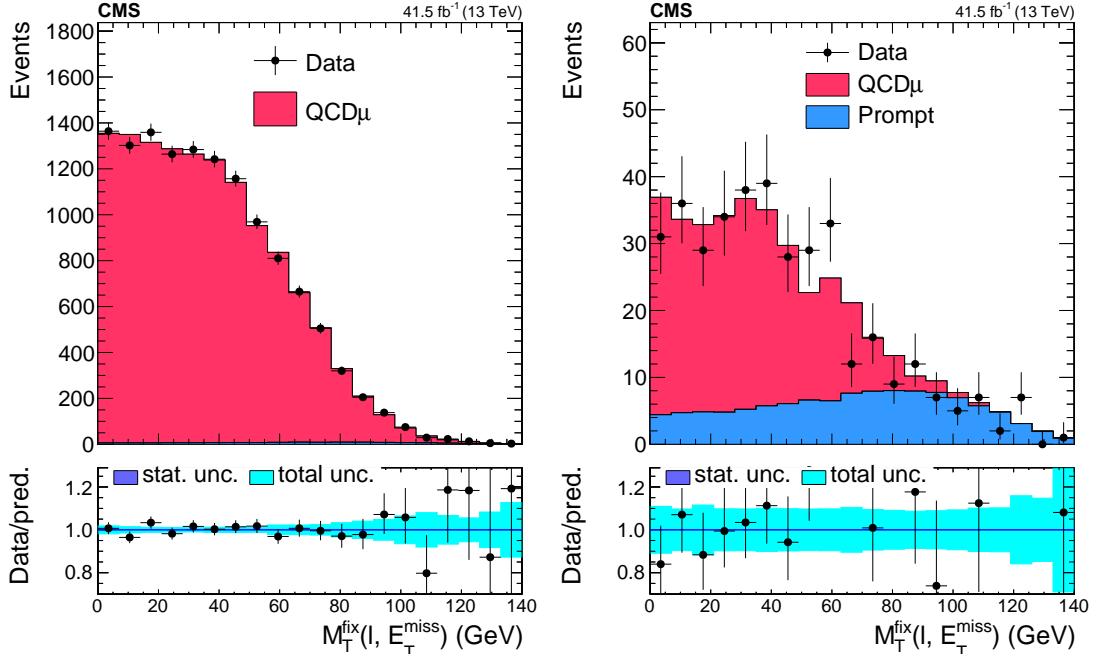
FIGURE 3.5: Distribution of the m_T^{fix} in the multijet sample after the fit has been performed for muons with $15 < p_T < 20$ GeV and $|\eta| > 1.2$ passing (left) and failing (right) the muon MVA selection. Black dots show the data while the solid histogram is the prediction from simulation. In the lower part of both plots the ratio between data and prediction is shown.

The nonprompt muon rate computation is also performed using a cut-based selection based on ID an mini-isolation requirements to provide a comparison. The performance of mini-isolation shows a dependence with the p_T, hence, the requirement on mini-isolation applied is selected for each p_T range to give the same efficiency as the prompt MVA, in order to obtain a fair comparison. The nonprompt rate for the prompt MVA selection and the cut-based selection is shown in Fig. 3.6. The nonprompt rate is reduced a factor of 2 when using the prompt MVA with respect to the mini-isolation selection for muons in the barrel. The reduction is even bigger, a factor of 3, for the $|\eta| > 1.2$ region. The agreement between the rate computed in data and simulation is also good.

3.5.3.1 Retraining of the prompt MVA

As already stated, the technique described above was initially developed in the context of the $t\bar{t}H$ multilepton analysis. Afterwards, it has been retrained and optimized for other analysis such as $t\bar{t}Z$ or tZq measurements. The alternative version of the MVA
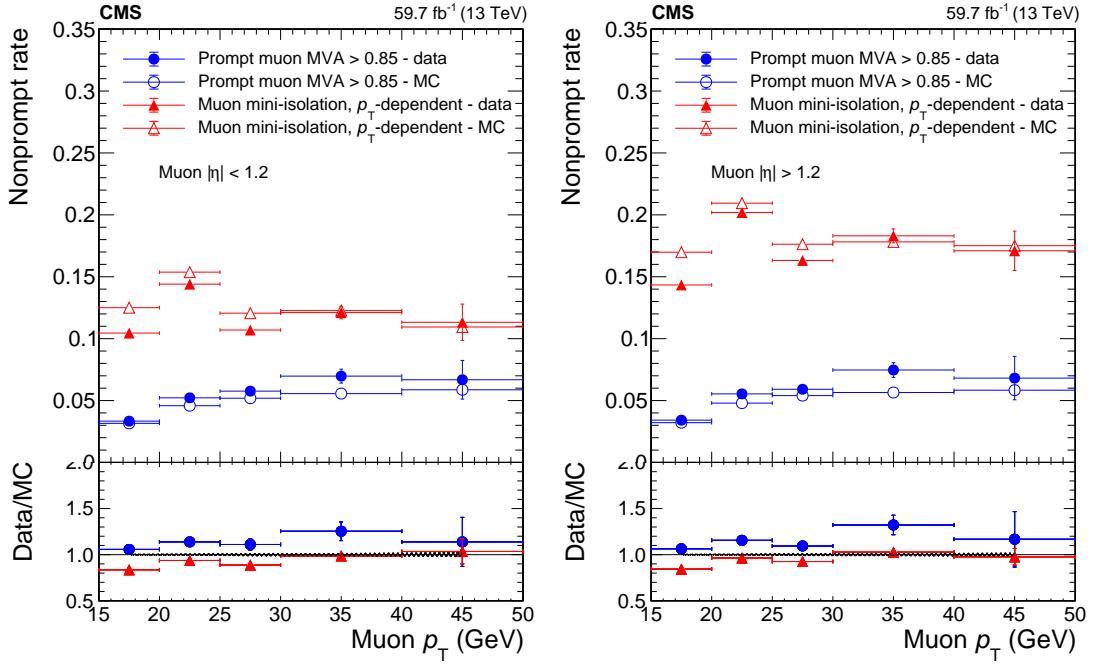
FIGURE 3.6: Measurement of the nonprompt-muon rate of a prompt muon MVA (blue dots) and mini-isolation (red triangles) selection as a function of p_T for muons with $|\eta| < 1.2$ (left) and $|\eta| > 1.2$ (right).

was trained simultaneously using $t\bar{t}W$, tZq, and $t\bar{t}Z$ simulation as signal, $t\bar{t}$ simulation was used as background. The preselection applied is also different. For electrons, it is not required that electrons pass the WP-loose of the identification MVA. For muons, the ID selection is tightened and requires muons to pass the medium identification criteria. Both for muons and electrons, the p_T threshold used for the training was raised from 5 GeV to 10 GeV. When comparing the performance for both MVAs, no significant difference where found.

3.6 Hadronic tau reconstruction

Taus are the only leptons with a mass large enough to decay into hadrons, accompanied by a single neutrino. Approximately 36% of the time, the taus decay into either an electron or a muon. In those cases, the reconstruction is performed using algorithms designed for electrons or muons, as previously outlined. In the remaining scenarios, the taus undergo decay into charged hadrons (mostly $\pi^{\pm}$) and neutral hadrons (π^0), plus a neutrino. The hadronically decaying taus (τ_h) are reconstructed using the hadrons-plus-strips (HPS) algorithm [114].

This algorithm is seeded by the PF jets constituents and it is designed to reconstruct π^0 (which are present in most of τ decays) within the jet. The π^0 will decay as $\pi^0 \rightarrow \gamma\gamma$, where photons will very likely convert into a electron pair. Hence, the π^0 manifests

in the jet as a photon PF candidates or as electron candidates. The deposits of these candidates in the ECAL are clustered in the so called *strips*. The τ_h candidates are then created by combining the charged-particle constituents of the jet (*prongs*) with the strips. Depending on the number of strips and charged particles observed, different decay modes are reconstructed: a single charged particle without strips ($h^\pm$); a single charged particle and one strip ($h^\pm \pi^0$); a charged particle and two strips ($h^\pm \pi^0 \pi^0$); and the combination of three charged particles ($h^\pm h^\mp h^\pm$). Where h denotes a charged pion or kaon. Extra quality criteria are applied to the τ candidates: for candidates containing more than one charged hadron, tracks should originate from the same vertex, the τ constituents should be contained within a cone which size depends on the p_T of the candidates and the total electric charge should be ± 1.

The τ_h reconstruction is completed by discriminating genuine τ_h from QCD-induced jets, muons and electrons. Several algorithms had been designed with that propose. In this thesis, a Deep NN (DNN) called *DeepTau* [115] is used. The algorithm takes as input both low level information about the PF candidates within the τ cone, along with high-level information of the reconstructed τ_h. The DNN was trained using $t\bar{t}$ and $W + \text{jets}$ simulation. The network offers three distinct output nodes, specialized in distinguishing τ_h from jets, electrons, and muons, respectively. Based on the score from each node, several working points are established: eight for the node discriminating against jets, four for the node targeting muons, and eight for the node trained to identify electrons.
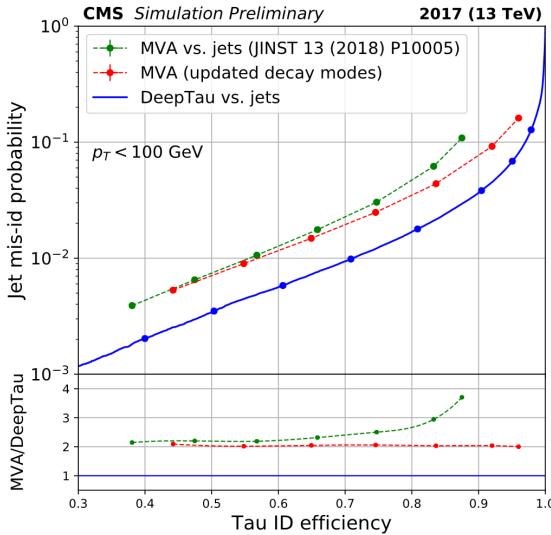
FIGURE 3.7: Jet misidentification probability as a function of τ efficiency. The performance of the `DeepTau` algorithm in simulated events (blue line) is compared to the performance for other discriminants (red and green lines). Figure taken from [115].

The performance of the `DeepTau` algorithm improves significantly with respect to the previous MVA-based algorithms used [116]. This improvement is shown in Fig. 3.7

where the eight working points defined to discriminates against jets are depicted. The τ efficiency is measured using $H \rightarrow \tau\tau$ simulation, while the misidentification probability is estimated using $t\bar{t}$ simulations.

<div align="center">

4

</div>

Analysis methodology for $t\bar{t}X$ production in multilepton final states

The main topic of this thesis is the measurement of processes where top quarks are produced in association with a boson: $t\bar{t}X$ (tX). In particular, the associated production of a top quark-antiquark pair with a Higgs boson or a W boson are studied. In Chapter 1.4, the unique features of Higgs production in association with top quarks were already introduced. Here, $t\bar{t}H$ and tH production are explained in detail, including how such production modes are sensitive to BSM scenarios. Other processes where top quarks are associated with bosons present very similar signatures to the aforementioned Higgs boson production mode. Hence, a good understanding of processes such as $t\bar{t}W$ or $t\bar{t}Z$ is mandatory.

In this chapter, an overview of the relevant characteristics of $t\bar{t}X$ production, as well as, the experimental status of such measurements is given. Afterwards, a description of the analysis strategy followed in this thesis to measure $t\bar{t}H$ and $t\bar{t}W$ cross sections in final states with multiple leptons is presented, in particular some common items of both analyses are explained.

4.1 $t\bar{t}H$ and tH production

4.1.1 SM production

In pp collisions $t\bar{t}H$ can be produced in a gg or q$\bar{q}$ initials state, Fig. 4.1 shows two LO diagrams for this process. From the diagrams it is clear that the Higgs boson couples at tree-level to the top quark, as a result $t\bar{t}H$ production cross section is proportional to y_t^2. $t\bar{t}H$ cross section has been computed at NLO in QCD with NLO electroweak corrections [117], yielding a value of $\sigma_{\text{SM}}^{t\bar{t}H} = 506.5$ fb, with a renormalization and factorization scale uncertainty of +5.8/-9.2 % and a PDF+α_s uncertainty of 3.6%. This

computation is used as reference in the studies reported in this thesis. Latest calcu-
lation at NNLO in QCD reports a value of $\sigma_{\text{SM}}^{\text{t\bar{t}H}} = 507.0\ (3.1)\ ^{+0.9\%}_{-3.0\%}$ fb [118], where
the uncertainty in brackets is the combination of the the uncertainty from the soft
Higgs boson approximation and the systematic uncertainty from the subtraction pro-
cedure; and the uncertainty given as a percentage accounts for the renormalisation
and factorisation scales variations.
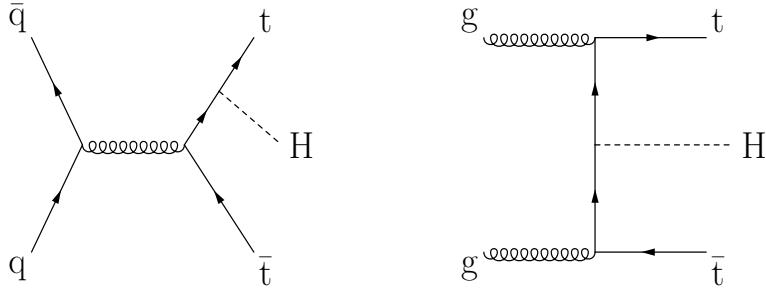
FIGURE 4.1: Representative LO Feynman diagrams for the t̄tH process in pp collisions.

The measurement of the t̄tH cross section can be used to study the module of $|y_t|$, but
it is not sensitive to any information about its sign. The relative sign of y_t and the
Higgs to W coupling (λ_W) can be inferred from the Higgs production in association
with a single top. This production occurs mainly via two channels: t-channel, usually
denoted as tHq, and tW-channel, or tHW. The s-channel has a negligible cross section
and it is usually not considered. At LO in the 5FS both production modes are well
defined; for higher orders, or in the 4FS, some processes interfere. For example, when
the W boson decays hadronically, the tHq interferes with both the tHW and the s-
channel at higher orders. Nevertheless, computations at NLO in the 4FS and 5FS
exists for the t-channel. Similarly, the tW-channel interferes with t̄tH production at
NLO, but there are calculations available to account for such interference as the ones
reported in Ref. [119]. Representative LO diagrams for the tHq and tHW production
modes are shown in Fig. 4.2. Both production modes are directly proportional to the
coupling of the Higgs to the top and of the Higgs to the W. Additionally, as the initial
state is the same in diagrams with a top-H vertex (left) and a H-W vertex (right), there
is an interference between both diagrams of each channel. In the SM such interference
is destructive, yielding values for the cross section of tHq and tHW of $\sigma_{\text{SM}}^{\text{tHq}} = 74.3$
fb and $\sigma_{\text{SM}}^{\text{tHW}} = 15.1$ fb, respectively. These values are computed at NLO in QCD and
evaluated in the 5FS [117], the corresponding uncertainties are reported in Table 4.1.
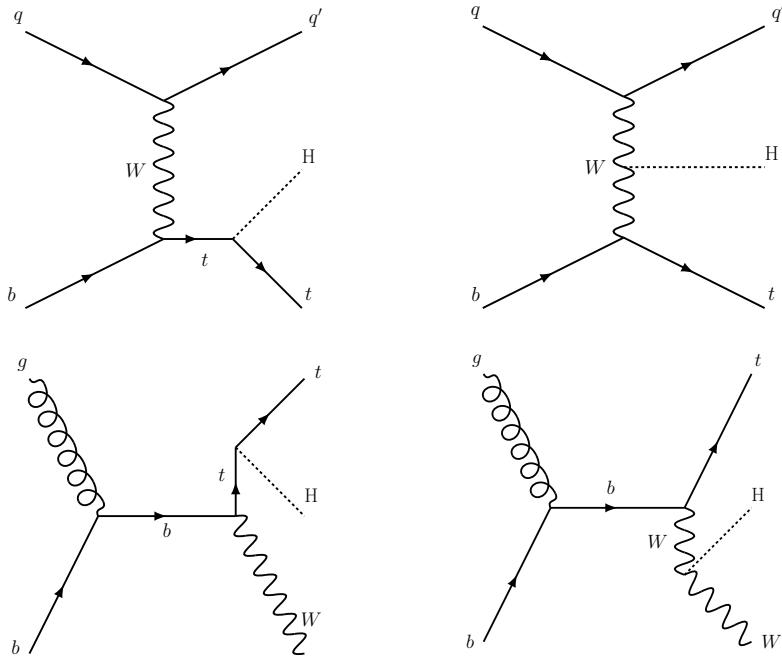
FIGURE 4.2: Representative LO Feynman diagrams for the tH process in the *t*-channel (upper row) and tW-associated production mode (lower row).

	σ_{QCD}^{NLO} [fb]	Scale [%]	PDF $+\alpha_s$ [%]
tH tW-channel	15.17	+4.9/-6.7	$\pm\,6.3$
tH t-channel	74.25	+6.5/-14.9	$\pm\,3.7$

TABLE 4.1: Cross section for tH production computed at NLO QCD for $\sqrt{s} = 13$ TeV and $m_{\mathrm{H}} = 125.09$ GeV. Scale stands for the relative renormalization and factorization scale uncertainties, including uncertainties on the flavour scheme. PDF+α_s stands for the relative uncertainty from the PDF choice and α_s value. [117].

4.1.2 Yukawa sector and BSM interpretations

As discussed in Chapter 1 the top Yukawa coupling cannot be studied via Higgs boson decays, since the decay of the Higgs boson in two top quarks is not allowed. Instead, such coupling can be tested using the gluon fusion production mode or the Higgs decay to $\gamma\gamma$. In both cases the top quark contributions appear in loops, as shown in Fig. 4.3. Alternatively, the coupling can be studied using tt̄H and tH production modes. This allows to perform the measurement at tree level, avoiding BSM contributions that may appear in the loops.

In that sense, the measurements of the top Yukawa coupling in the SM can be extended to search for new physics. To perform this kind of studies tt̄H and tH measurements can be interpreted in the context of the Kappa-framework [120]. This framework allows to introduced BSM modifications only on the couplings strengths of the Higgs to other particles. The coupling modifiers are denoted as κ_i and defined in such way
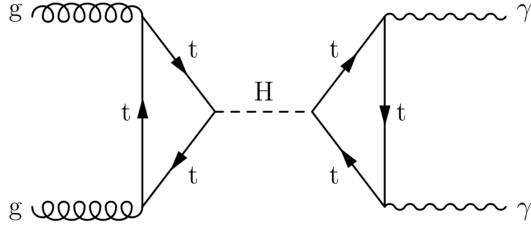
FIGURE 4.3: Feynman diagram showing Higgs boson production via gluon fusion mode and its decay into a photon pair.

that the cross section and partial decay widths of the Higgs scale as κ_i^2 at leading order. For the loop-induced vertices, the interaction can be effectively parameterize through κ_γ and κ_g, to scale the couplings of the Higgs to photons or gluons.

Two BSM interpretations of the coupling are studied in this thesis. The first one corresponds to the relative sign of the top Yukawa coupling and the H-W coupling. The second is related to the existence of CP violation in the Higgs sector.

Relative sign of the top Yukawa coupling

The tH cross section can be parameterized as a function of the coupling modifiers κ_W and κ_t, at LO and $\sqrt{s} = 13$ TeV [121]:

$$\sigma_{\text{tHq}} = (2.63\kappa_t + 3.58\kappa_W - 5.21\kappa_t\kappa_W) \times \sigma_{\text{tHq}}^{SM} \tag{4.1}$$

$$\sigma_{\text{tHW}} = (2.91\kappa_t + 2.31\kappa_W - 4.22\kappa_t\kappa_W) \times \sigma_{\text{tHW}}^{SM}. \tag{4.2}$$

As stated in Section 4.1.1, the interference in the SM is destructive, in that case $|\kappa_t| = |\kappa_W| = 1$, and $\frac{\kappa_t}{\kappa_W} = -1$. Anomalous couplings would affect the inference between the diagrams, modifying the cross section of tH production mode. The particular case where $|\kappa_t| = |\kappa_W| = 1$, and $\frac{\kappa_t}{\kappa_W} = 1$ is known as the Inverted Top Coupling (ITC) scenario. Under that assumption, the cross sections for tHq and tHW would increase by a factor of 11 and 9, respectively. This feature makes the study of tH production particularly relevant, as tt̄H cross section remains unchanged between the the SM scenario and the ITC scenario.

Charge-Parity violation in the Higgs sector

In the SM the Higgs boson is a pseudo-scalar invariant under CP transformations. This hypothesis was proven in data by studying the Higgs decay to vector boson

pairs [122]. Those studies, however, do not rule out *CP* violation in the Yukawa sector, as the Higgs boson could be a mixture of *CP*-eigenstates with a *CP*-odd component that does not couple to vector bosons. In particular, outside the kappa framework approach, one can write the Lagrangian of the top Yukawa sector as [123]:

$$\mathcal{L} = \bar{\psi}_t(\cos(\alpha)\kappa_{Htt} + i\sin(\alpha)\kappa_{Att}\gamma_5)\frac{y_t}{\sqrt{2}}\psi_t\phi, \tag{4.3}$$

where ψ_t represents the top quark field, ϕ is a scalar representing the Higgs field, α is the *CP*-mixing angle and κ_{Att} and κ_{Htt} are real dimensionless rescaling parameters for the *CP*-odd and *CP*-even components respectively. Depending on the choice of α, three scenarios are identified, as detailed in Table 4.2.

Scenario	α
Purely *CP*-even	0° or 180°
Purely *CP*-odd	90° or 270°
Mixed	$\neq 0°, \neq 90°, \neq 180°$

TABLE 4.2: *CP* scenarios

As already stated, the nature of the top-Higgs coupling affects ggH production mode, as well as, Higgs decay to $\gamma\gamma$ via loops. The Higgs coupling to gluons and photons has been measured, providing constrains on this Lagrangian. Such constrains in the above parametrization imply: $\kappa_{Att} = 2/3$ and $\kappa_{Htt} = 1$. This selection of parameters allows to recover the SM coupling for the Hgg and H$\gamma\gamma$. Nevertheless, tt̄H and tH production are still affected by the *CP*-phase as shown in Fig. 4.4, where the cross sections for both processes are represented as a function of α.

Trilinear Higgs coupling

The tt̄H and tH production are sensitive to some terms of the Higgs potential. In particular, the trilinear coupling (κ_λ) affects these processes in diagrams including higher order electroweak corrections, as the ones in Fig 4.5. At low energy, new physics could modify the value of this coupling, changing the Higgs potential. The tt̄H production is one of the Higgs production modes more sensitive to these higher order corrections, differences are expected in kinematic distributions as reported in Ref. [124].

4.1.3 Experimental status

The tt̄H production was observed both by ATLAS and CMS in 2018 by combining data at $\sqrt{s} = 7, 8$ and 13 TeV, amounting to an integrated luminosity of 61 fb^{-1} used in the
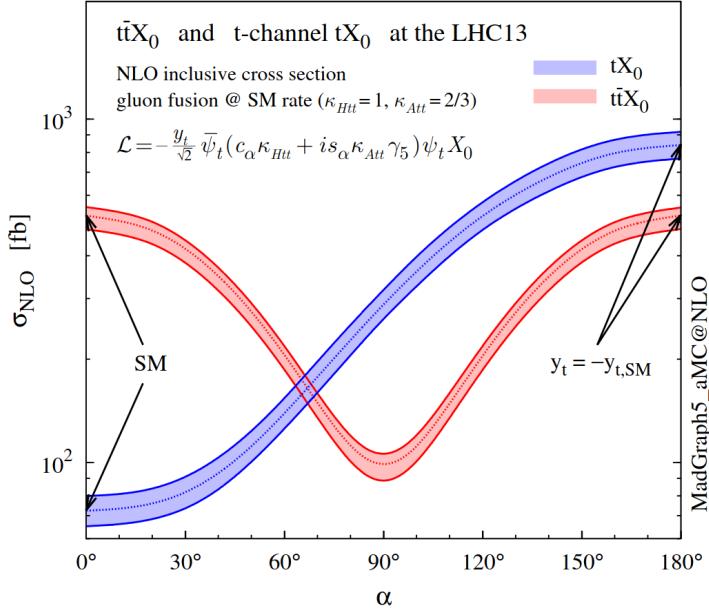
FIGURE 4.4: NLO cross sections for t̄tX and t-channel tX productions at the 13 TeV as a function of the *CP*-mixing angle (α), coloured band represents the scale uncertainties. κ_{Htt} and κ_{Att} are fixed to 1 and 2/3 respectively. Figure extracted from Ref. [123].
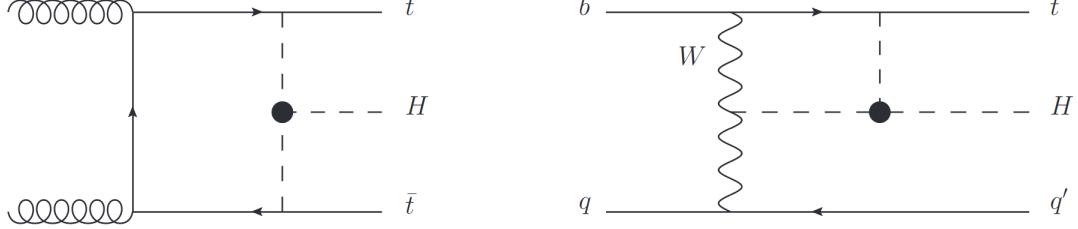
FIGURE 4.5: One-loop Feynman diagrams in t̄tH (left) and tH production with anomalous trilinear coupling.

CMS measurement [105] and to a 105 fb^{-1}, for the ATLAS one [125]. In the search, final states targeting the Higgs decay to b̄b, WW, ZZ, $\gamma\gamma$, and $\tau\tau$ were used, in order to maximize the sensitivity. The signal strength parameter ($\mu_{t\bar{t}H}$), which is the quotient between the measured cross section and the predicted by the SM, obtained by both experiments is in good agreement with the SM. ATLAS reported a $\mu_{t\bar{t}H} = 1.32^{+0.28}_{-0.26}$, while CMS measured a $\mu_{t\bar{t}H} = 1.26^{+0.31}_{-0.26}$, yielding a significance over the background only hypothesis of 6.3 s.d. and 5.2 s.d., respectively. Figures 4.6 and 4.7 show the signal strength parameters in each of the Higgs decay modes.

Once t̄tH production was observed, both experiments continued measuring this production mode in each of the channels individually. The $\gamma\gamma$ decay mode, despite presenting a low branching ratio, provides a clear signal with a low background setting. Furthermore, in this final state the Higgs system can be resolved. The CMS [126] and ATLAS [127] collaborations have measured t̄tH production cross section in this final
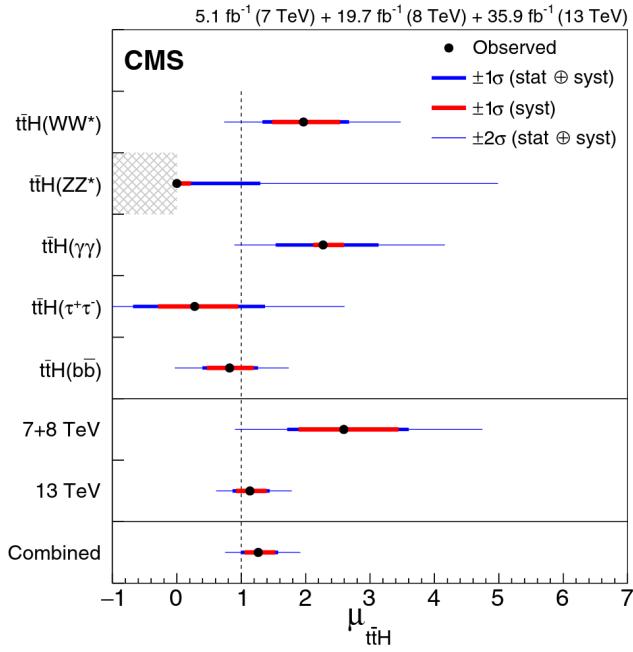
FIGURE 4.6: Signal strength of t$\bar{\text{t}}$H process in each of the final states used for the CMS measurements at $\sqrt{s} = 7$, 8 and 13 TeV. The combined result is also shown. Figure extracted from Ref. [105].
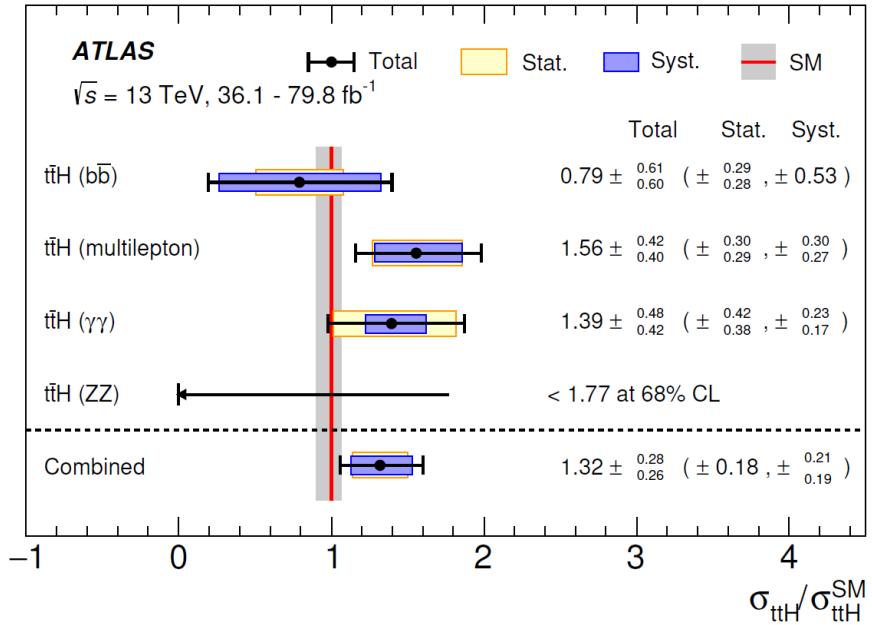
FIGURE 4.7: Signal strength of t$\bar{\text{t}}$H process in each of the final states used for the ATLAS measurements at $\sqrt{s} = 7$, 8 and 13 TeV. The combined result is also shown. Figure extracted from Ref. [125].

state using the full Run-2 dataset achieving an observed (expected) significance of 6.6 (4.7) and 5.2 (4.4) s.d., respectively. ATLAS analysis also studies tH production, but sensitivity is not high enough to claim evidence of this H production mode. Both analyses include a *CP* interpretation of the top-Higgs coupling, excluding the pure *CP*-odd scenario with 3.2 s.d and 3.9 s.d., respectively.

The b͞b final state is the dominant one in terms of branching ratio, yet it is affected by large background in the context of LHC measurements. This analysis can be performed by combining several channels depending on the top quark decay: fully hadronic, semileptonic, and dileptonic. The latest measurements of t͞tH in this channels reported by CMS [128] and ATLAS [129] uses 138 and 36.1 fb^{-1}, respectively. The ATLAS measurement yields an observed (expected) sensitivity of 1.4 (1.6) s.d. using only final states with one or two leptons. The CMS measurement, using the full Run-2 dataset and all channels, achieves an observed (expected) sensitivity of 1.3 (4.1) s.d. CMS analysis also aims to measure tH production, setting limits to its production rate. Both experiments have studied the *CP* structure of the top-Higgs coupling in this decay mode of the Higgs. Particularly, ATLAS excludes the pure *CP*-odd hypothesis with 1.2 s.d. [130].

This thesis aims to study t͞tH and tH production in final states characterized by the presence of multiple leptons. This targets cases where the Higgs boson decays predominantly into W^+W^- and $\tau^+\tau^-$ final states, although a minor contribution from ZZ decays is also included. It is worth noting that H $\to$ ZZ $\to 4\ell$ is usually covered by other Higgs analyses [131, 132] that includes t͞tH production. Therefore, this specific decay channel is excluded from consideration through a veto in the multileptonic selection process. H $\to$ ZZ $\to 4\ell$ holds the characteristic that the Higgs system can be reconstructed. However, due to the low branching ratio of this decay mode accurate measurements in this channel are severely limited by sensitivity constraints.

Multileptonic final states feature an intermediate branching ratio and certain final states, such as final states with two leptons with the same electric charge in high jet multiplicity topologies, allow to reduce the backgrounds. Prior to the beginning of this thesis, t͞tH was studied in final states with multiple leptons both by CMS and ATLAS in multileptonic final states using a part of the Run-2 dataset. CMS reported a signal strength of $\mu_{t\bar{t}H} = 0.96^{+0.34}_{-0.31}$, corresponding to the first evidence of t͞tH production in multileptonic final states [133]. Latest measurements obtained by ATLAS with a total integrated luminosity of 80 fb^{-1}, yielded $\mu_{t\bar{t}H} = 0.58^{+0.36}_{-0.33}$ [14].

4.2   t\bar{t}Z, t\bar{t}γ and t\bar{t}W production

The associated productions of a top-antitop quark pair with a W or Z/γ boson represent one of the leading irreducible background for t\bar{t}H measurement. Their production rate is of the same order as t\bar{t}H production and, given the decay of the W and Z bosons, its final states will provide signatures similar to the ones expected in t\bar{t}H when the Higgs decays to W^+W^- and $\tau^+\tau^-$. These processes also serve as valuable tools for testing the SM and exhibit sensitivity to BSM effects.

The t\bar{t}Z and t\bar{t}W production allow to measure directly the electroweak coupling of the top quark. Moreover, these processes are sensitive to various EFT operators. Therefore, conducting differential measurements of the t\bar{t}Z and t\bar{t}W cross sections enables the precise constraint of anomalous couplings in these production modes.

4.2.1   t\bar{t}Z

The cross section for the t\bar{t}Z production was computed at NLO accuracy plus electroweak corrections [117]. Corrections from hard, non-logarithmically enhanced radiation are reported to be large, resulting in a strong dependence of the renormalization scale choice. Calculations are also performed up to the next-to-next-leading logarithmic (NNLL) accuracy and matched to the complete NLO results including QCD and electroweak effects [19]. These computations yield a value of $\sigma_{t\bar{t}Z}^{\text{NLO+NNLL}} = 859^{+8.6\%+2.3\%}_{-9.5\%-2.3\%}$ fb and stabilize the scale dependency as depicted in Fig. 4.8 (left).

Regarding the experimental status for this process, it has been measured both by CMS and ATLAS [134, 135] using Run-2 data and yielding cross section values in good agreement with the theory predictions and a high level of precision, with a total uncertainty around 7-8%, these results are shown in Fig. 4.10. Recently ATLAS presented a t\bar{t}Z measurement including EFT and spin correlations interpretations [136].

4.2.2   t\bar{t}γ

It has been measured both by CMS and ATLAS using different final states and the full Run-2 dataset. CMS reported the production in final states with two leptons [137] and a lepton and jets [138], while ATLAS performed a combined measurement of t\bar{t}γ +tWγ in final state with an electron and a muon [139]. All measurements are shown in Fig. 4.10 and are in good agreement with the prediction.
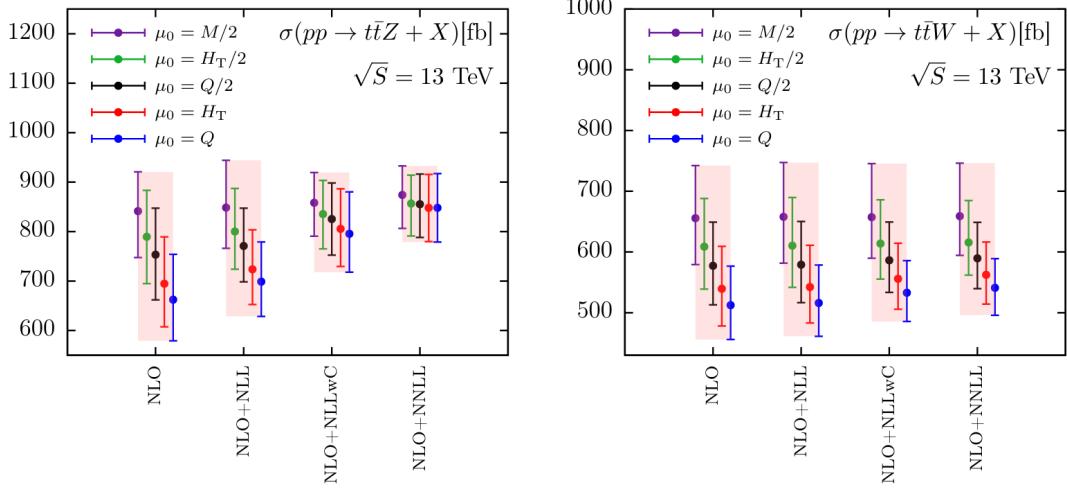
FIGURE 4.8: Cross section calculation for tt̄Z (left) and tt̄W (right) at $\sqrt{s} = 13$ TeV, different scale choices are used. Figures extracted from Ref. [19].

In this thesis, the contributions arising from off-shell $t\bar{t}\gamma^*$ and $t\gamma^*$ production are taken into account in two separated contributions. Those events where at least one of the leptons is matched to a photon at generator level, are considered as photon conversions, as explained in detail later in this chapter. In any other case, this production mode is included within the tt̄Z background.

4.2.3 tt̄W

The tt̄W production exhibits some differences with respect to tt̄Z production. Firstly, the W boson cannot be radiated from the top quarks in the final state. Secondly, at LO the production of tt̄W is $q\bar{q}$ induced, and gluon-gluon initial states are suppressed. In pp collisions this yields to a sizable difference between tt̄W$^+$ and tt̄W$^-$ production rates, motivated by the differences in the PDFs of the valence and sea quarks in the proton. In particular, the main production modes at LO are the channels: $u\bar{d} \rightarrow t\bar{t}W^+$ and $d\bar{u} \rightarrow t\bar{t}W^-$. It is worth noting that the $u\bar{d}$ initial state is two times more probable than the $d\bar{u}$ and has larger momentum of the partons. In Fig. 4.9 four representative diagrams of tt̄W production are shown at LO and NLO in QCD. The fact that gg initial states are suppressed, makes tt̄W process a good handle to study the parton PDFs and the tt̄ central-peripheral charge asymmetry in pp collisions. In tt̄ production, the top quark (antiquark) is preferentially emitted in the direction of the incoming quark (antiquark). At the LHC, in pp collisions, this result in a central-forward rapidity charge asymmetry, where top quarks (antiquarks) are produced with more forward (central) rapidities. Additionally, the fact that tt̄ production is dominated by gg initial state —which is charge-symmetric— makes this effect very small, of the order of 1%.
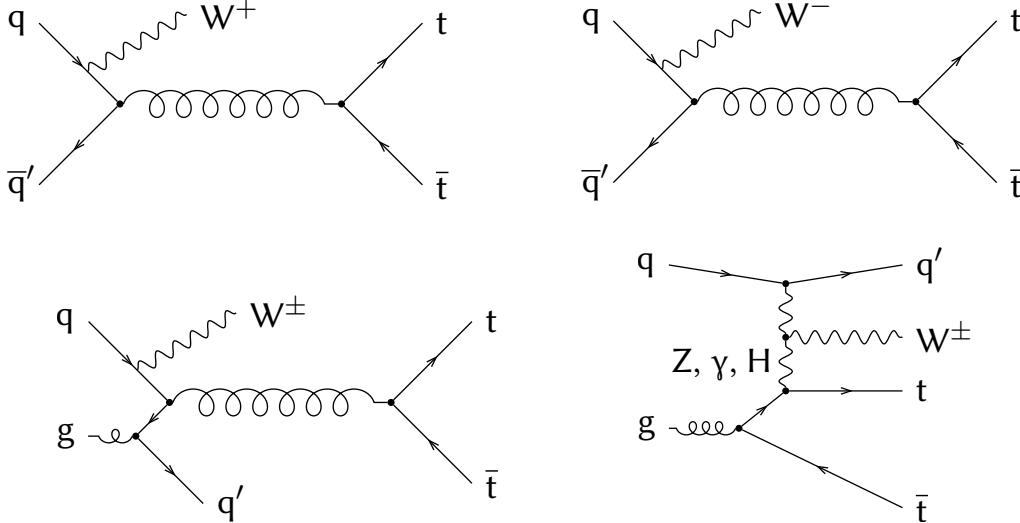
FIGURE 4.9: Representative Feynman diagrams for tt̄W production at LO (upper row) and NLO (lower row).

In tt̄W the dominance of $q\bar{q}$ initial state and the polarization of the initial state by the emission of the W enhance the asymmetry between the decay products of the top quarks and top antiquarks. This asymmetry can then be defined using the lepton decay of the top quark as:

$$A_c^\ell = \frac{N(\Delta\eta^\ell > 0) - N(\Delta\eta^\ell < 0)}{N(\Delta\eta^\ell > 0) + N(\Delta\eta^\ell < 0)} \tag{4.4}$$

Where $\Delta\eta^\ell = |\eta_\ell| - |\eta_{\bar{\ell}}|$ is the difference between the absolute value of the psedorapidities of the lepton coming from the top quark and the antiquark, respecitvely. The A_c^ℓ has been measured by ATLAS [140].

As already stated, tt̄W is a key background when measuring tt̄H production in final states with leptons. tt̄W process is also an important background in tt̄tt̄ production measurements, as well as, searches in final states with two leptons with the same electric charge (same-sign, *ss*) and three leptons.

Regarding the experimental status of tt̄W measurement, both the CMS and ATLAS Collaborations have observed this process at 8 [141, 142] and 13 TeV [11, 12]. Prior to the initiation of this thesis, the measurements at 13 TeV, conducted using data collected in 2016, were the most precise tt̄W measurements. The cross section values obtained are depicted in Fig. 4.10, showing a central value slightly higher than the SM prediction, but remaining well within the total uncertainties, which are of the order of 20%.
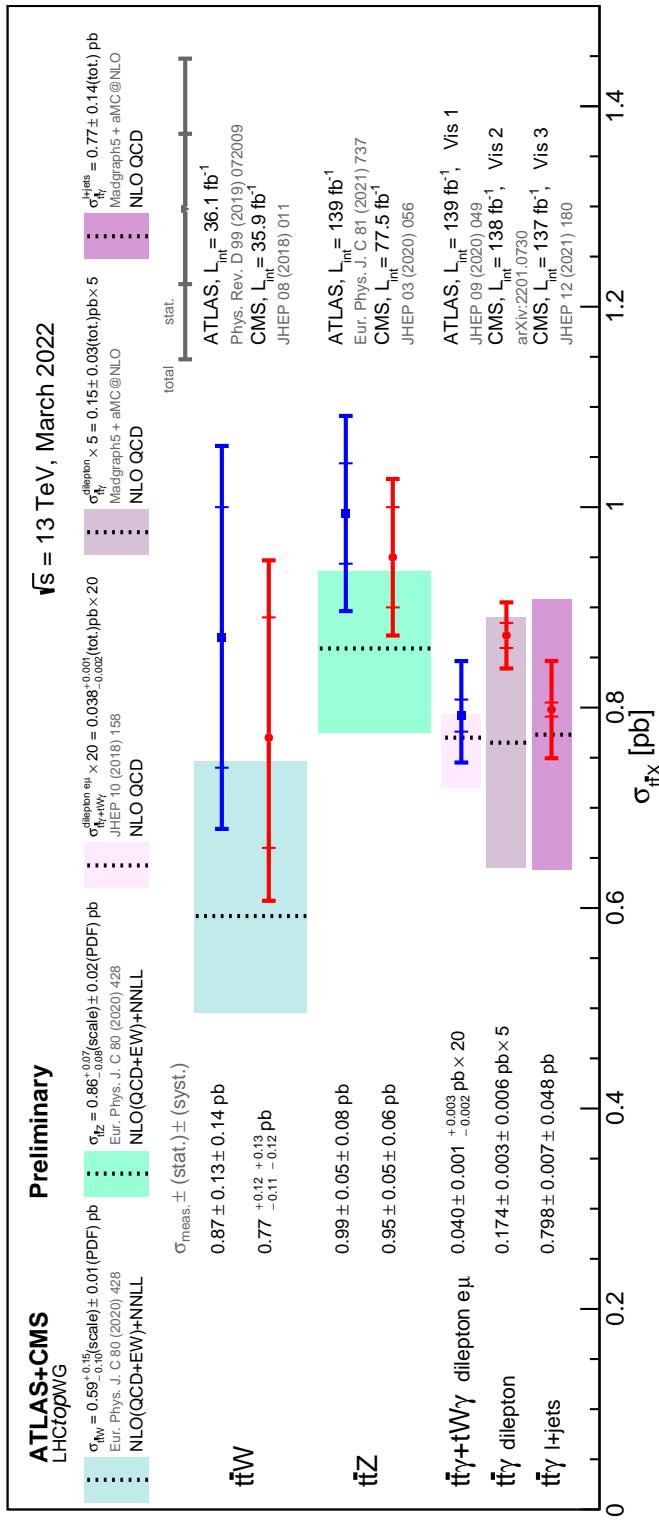
FIGURE 4.10: Summary of ATLAS and CMS measurements of t$\bar{\text{t}}$X (X = W, Z or γ) cross sections at 13 TeV, this shows the experimental status prior to the measurement presented in this thesis. The measurements of the t$\bar{\text{t}}$W and t$\bar{\text{t}}$Z cross sections are compared to the NLO QCD and EW theoretical calculation complemented with NNLL resummation. The theory band represents uncertainties due to renormalisation and factorisation scales and parton density functions. Figure extracted from Ref. [69].
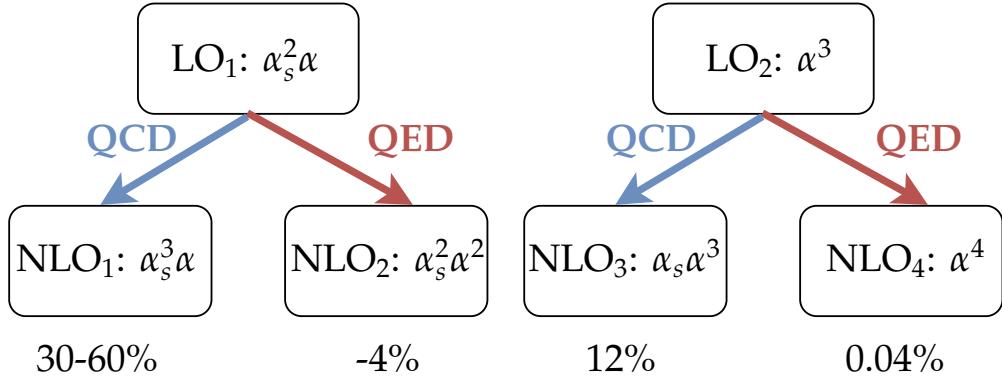
FIGURE 4.11: Diagram summarizing the LO and NLO QCD and EW corrections. Figure created by S. Sánchez.

During the t̄tH measurement reported in this thesis and the recent t̄tt̄t measurement [106, 143] —performed using the full Run-2 luminosity— the t̄tW production rate was checked in dedicated control regions. A tension between the theory and the prediction was reported in these background studies.

All the above motivates a dedicated inclusive t̄tW measurement using the full Run-2 dataset. Differential measurements of the process should follow in order to improve the understanding of t̄tW modelling.

4.2.3.1 t̄tW **modelling**

The t̄tW production is largely affected by contributions from electroweak diagrams at LO and NLO proportional to α^3 and $\alpha_S\alpha^3$ terms, respectively. These corrections were naively expected to be much smaller than the NLO (QCD) corrections, especially the $\alpha_S\alpha^3$ (NLO$_3$) term, which was considered negligible. Nevertheless, the complete NLO calculation reported in Ref. [17], showed that the NLO$_3$ term contributes with a 12% increase in the cross section with respect to the LO calculation. In Fig. 4.11 a diagram summarizing the LO and NLO QCD and EW corrections is shown. The percentage in the bottom represents the size of the correction with respect to the LO calculation. The large value of the NLO$_3$ correction is explained by the fact that it contains $gq \rightarrow$ t̄tWq' real emission diagrams, as shown in Fig. 4.9 bottom right, where tW $\rightarrow$ tW scattering is present. It is also worth noting that the LO mixed term $\alpha_S\alpha^2$, resulting from the interference of QCD and EWK diagrams, does not contribute to t̄tW production. The reason is that in q̄q channels the interference vanishes due to the colour structure of the initial state.

Aside from large electroweak corrections, t̄tW is affected by large scale uncertainties. The complete NLO+NNLL calculation reported in Ref. [19] allows to stabilize the cross
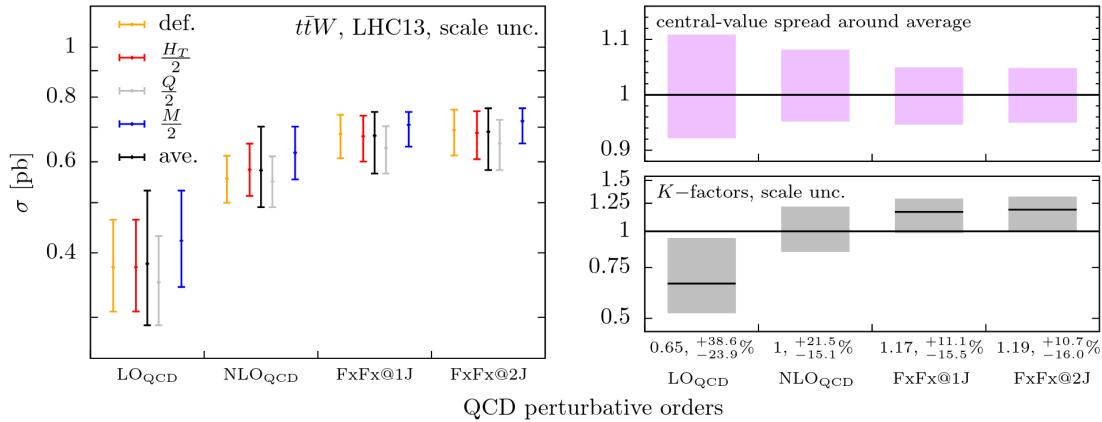
FIGURE 4.12: The t̄tW production cross sections at various QCD orders (left plot). Deviation of the different central values with respect to their average in percentage (upper right ratio plot). K-factors of the averaged central values with the combined scale uncertainties (lower right ratio plot). Figure extracted from Ref. [18].

section value as a function of the scale choice for t̄tH and t̄tZ, reducing the associated systematic uncertainty by up to a factor of two. Nevertheless, the stabilization is not that strong for t̄tW. This effect is clearly depicted in Fig. 4.8.

In order to improve the scale dependency a new computation was reported in Ref. [18]. This calculation aims to incorporate part of the corrections arising at NNLO —which are expected to be non-negligible— by applying multi-jet merging at NLO computation. This calculation features NLO QCD merging for t̄tW up to two jets, NLO_3 corrections are also included. At NLO in QCD, real-emission radiation can come from a QCD (QCD-jet) or an EWK (weak-jet) vertex. At LO in the standard merging scheme, the contributions from t̄tW + weak-jet are ignored below the merging scale. Analogously, at NLO merging, contributions are evaluated at LO below the selected merging scale and at NLO above it. In the computation presented in Ref. [18], the merging scheme is modified in order to take into account correctly the jets from electroweak vertices. Results are shown in Fig. 4.12, reporting a reduction in the scale dependency and a central value for the cross section slightly higher than the full NLO computation.

The state-of-the-art t̄tW cross section calculation is performed at NNLO in QCD and including NLO EWK corrections [21]. The cross section values obtained for t̄tW production as well as t̄tW$^-$ and t̄tW$^+$ are shown in Table 4.3, it is worth noting that t̄tW$^+$ production is two times more likely than t̄tW$^-$. In Fig. 4.13 the inclusive t̄tW cross section is shown at different orders in QCD perturbation. This result significantly reduces the uncertainty with respect to previous theoretical calculations.

	$\sigma_{tt̄W^+}$ [fb]	$\sigma_{tt̄W^-}$ [fb]	$\sigma_{tt̄W}$ [fb]	$\sigma_{tt̄W^+}/\sigma_{tt̄W^-}$
NNLO$_{QCD}$ + NLO$_{EW}$	$497.5 \pm 6.6\% \pm 1.8\%$	$247.9 \pm 7.0\% \pm 1.8\%$	$745.3 \pm 6.7\% \pm 1.8\%$	$2.007 \pm 2.1\%$

TABLE 4.3: State-of-the-art cross section prediction for tt̄W$^+$ and tt̄W$^-$ production, together with its sum and ratio. Uncertainties are computed through scale variations and are symmetrised. Taken from [21].
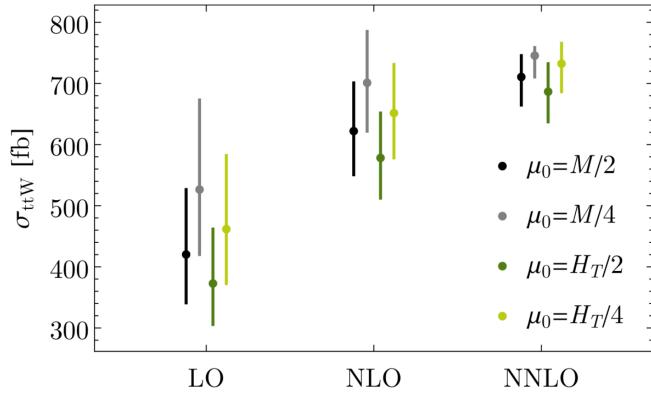
FIGURE 4.13: tt̄W production cross sections at various QCD orders. Figure extracted from Ref. [21].

4.3 Analysis strategy for tt̄X measurements in final states with leptons

In the following sections, common items of tt̄H and tt̄W measurements are discussed. In both cases the presence of multiple leptons is required in the final state and events are categorized depending on the number of leptons. For tt̄H, final states with hadronic taus (τ_h) are also considered in order to target the H $\to \tau^+\tau^-$ decay mode. In the tt̄H analysis a total of 10 signal region categories are examined, each characterized by relatively high lepton and τ_h multiplicities. Among these categories, the most sensitive ones include the one with two same-sign leptons and zero τ_h, the category with three leptons and zero τ_h and the one with two same-sign leptons and one hadronic tau. For the tt̄W analysis, two signal region categories with two same-sign leptons and three leptons, respectively, are used.

The selected topologies aim to reject as much as possible SM backgrounds with large cross sections, which are not expected to populate those final states. Nevertheless, given the low cross section of the signal —several orders of magnitude lower than the backgrounds— those backgrounds can still have a sizable impact arising from events where the leptons have been misidentified or come from nonprompt lepton decays. Consequently, lepton selection is key in both analyses. In order to efficiently select muons and electrons from bosons and τ decays, while rejecting those from nonprompt decays, the MVA described in Section 3.5.2 is used. Besides, for τ_h reconstruction and

identification state-of-the-art identification techniques based on ML algorithms are used as described in Section 3.6.

In each lepton category, a dedicated selection is imposed based on: lepton charge, jet and b tag multiplicities, as well as, kinematic variables related to the signal expected topology. In many cases the amount of background after the selection is still sizable, so multivariate techniques are developed in order to discriminate signal and background.

The backgrounds affecting the measurements can be classified in two groups: irreducible and reducible.

Irreducible backgrounds are those where reconstructed electrons and muons are, indeed, prompt leptons with the charge correctly reconstructed and all reconstructed τ_h are genuine hadronic τ decays. This background category is dominated by tt̄W and tt̄Z production in the context of the tt̄H analysis, and by tt̄H and tt̄Z processes for the tt̄W analysis. Additionally, contributions from tZq and WZ+jet production are also significant. These processes are estimated using the latest MC simulations available and dedicated control regions are defined in order to check the data to simulation agreement.

Reducible backgrounds are those entering the signal region due to the misreconstruction of leptons or τ_h. Three sources are considered: *nonprompt* lepton background, coming from misidentification of prompt leptons and τ_h; *conversions*, originated from the asymmetric conversions of a photon into electrons; and the so called *charge misidentification*, arising from events where the lepton charge was wrongly assigned. The nonprmopt lepton and charge misidentification backgrounds are estimated by using dedicated data-driven techniques, while conversions are estimated using simulation.

4.4   Object selection

4.4.1   Electrons and muons

Three levels of lepton selection are defined for electrons and muons with tightening criteria: loose, fakeable and tight. Loose leptons are required to have a minimum p_T and to be within the acceptance: $|\eta| < 2.4$ for muons and $|\eta| < 2.5$ for electrons. Additional requirements are applied on the impact parameter with respect to the primary vertex, on the significance of this impact parameter and on the isolation of the lepton. In the case of electrons, an additional criterion mandates that no more than one missing hit is allowed from its expected trajectory in the tracker. Certain restrictions on the muon cut-based ID defined in Section 3.3 and the electron MVA-ID

selection defined in Section 3.4 are also applied. This minimal loose selection aims to assure no duplicated nearby leptons are included in the analysis and to reject leptons originating from low mass resonances.

The fakeable lepton selection is constructed on top of the loose selection, and is used to estimate the nonprompt contribution in the signal region from data samples. Consequently, this selection is designed to accept nonprompt leptons, but it is also optimize to keep a good closure for the estimation in data and simulation. When selecting nonprompt leptons, the cone-p_{T} ($p_{\text{T}}^{\text{cone}}$) is defined. For nonprompt leptons, this is a proxy variable for the p_{T} of the parton originating the jet from which the lepton is coming. For prompt leptons, instead, the $p_{\text{T}}^{\text{cone}}$ is the reconstructed lepton p_{T}. For this reason, the variable is defined as follows:

$$p_{\text{T}}^{\text{cone}} = \begin{cases} p_{\text{T}} & \text{if lepton passes the tight selection} \\ xp_{\text{T}}(1 + I_{\text{Jet}}) & \text{otherwise} \end{cases}, \qquad (4.5)$$

where x is a factor that needs to be calculated applying the condition that the $p_{\text{T}}^{\text{cone}}$ is a continuous variable, and I_{Jet} is the jet relative isolation. Variables used at event-level are constructed using the $p_{\text{T}}^{\text{cone}}$; from now on in this thesis p_{T} stands for cone-p_{T} unless otherwise specified.

In the fakeable lepton selection, electron requirements are tightened by applying criteria on the ratio between the energy deposited in the ECAL and HCAL (H/E), the difference between the inverse of the cluster energy and the inverse of the track momentum ($1/E_{SC} - 1/p$) and on the spread of the cluster in the η region ($\sigma_{i\eta i\eta}$). This last criterion is applied as a function of the η of the supercluster. Events compatible with conversions are also rejected. Besides, for t$\bar{\text{t}}$H studies, the number of allowed missing hits in the tracker trajectory is reduced to zero.

Additionally, for t$\bar{\text{t}}$H studies, it is required that the jet overlapping with the lepton[1] fails the selected WP of the `DeepJet` b-tagging discriminant. Such requirement is also added to the selection for t$\bar{\text{t}}$W studies whenever the event does not pass the prompt-MVA selection used to define tight leptons. Furthermore, in both object selections, if the event does not pass the prompt-MVA requirement, some additional criteria are applied on the isolation of the overlapping jet, and the electron ID.

Finally, the tight lepton selection aims to provide high purity in prompt leptons, the leptons passing that selection are the ones used to define the analysis signal and control regions. Tight leptons are fakeable leptons that have a prompt-lepton MVA score

---

[1]matched to the lepton within a $\Delta R < 0.4$

higher than a certain value. This value was selected to be optimal for t̄tH and t̄tW analysis respectively, in a different preselection also aligned with the analysis features.

Detailed values for the three selections are provided in Tables 4.4 and 4.5 for electrons and muons respectively.

Observable	Loose	Fakeable	Tight
(Cone-)p_T	$> 7, 10$ GeV	> 10 GeV	> 10 GeV
$|\eta|$	< 2.5	< 2.5	< 2.5
$|d_{xy}|$	< 0.05 cm	< 0.05 cm	< 0.05 cm
$|d_z|$	< 0.1 cm	< 0.1 cm	< 0.1 cm
d/σ_d	< 8	< 8	< 8
I_e	$< 0.4 \times p_\mathrm{T}$	$< 0.4 \times p_\mathrm{T}$	$< 0.4 \times p_\mathrm{T}$
$\sigma_{i\eta i\eta}$	—	$< \{ 0.011 / 0.030 \}$[1]	$< \{ 0.011 / 0.030 \}$[1]
H/E	—	< 0.10	< 0.10
1/E - 1/p	—	> -0.04	> -0.04
Conversion rejection	—	✓	✓
Missing hits	≤ 1	$= 0, \leq 1$	$= 0, \leq 1$
Electron MVA ID	$>$WP-loose[2] —	$>$WP-80 ($>$WP-loose)[2] $>$WP-loose (—)[2]	$>$WP-loose[2] —
`DeepJet` of nearby jet	—	$<$WP-medium[3] $<$WP-interp.[3] (—)	$<$WP-medium[3] —
Jet relative isolation[4]	—	$< 0.7, 1.0$ (—)	—
Prompt-e MVA	—	$<0.8, 0.4$ ($>0.8, 0.4$)	$> 0.8, 0.4$

[1] Barrel / endcaps.
[2] WPs defined in Section 3.4.
[3] WPs of the `DeepJet` algorithm, defined in Section 2.3.2. WP-interp. denotes that the score is selected by defining a linear interpolation between the WP medium and loose of the `DeepJet` discriminant as a function of the cone-p_T.
[4] Defined as $1/p_T^{\mathrm{ratio}}$-1 if the muon is matched to a jet within $\Delta R < 0.4$ or as the PF relative isolation with ΔR=0.4 otherwise.

TABLE 4.4: Loose, fakeable and tight selection criteria for electrons. Requirements shown in black are applied when studying both t̄tH and t̄tW, while the ones in blue are only applied in the t̄tH measurements and the ones in green only in the t̄tW one. The cuts in parentheses are applied only when leptons fail the tight lepton MVA cut. A long dash (—) indicates selection criteria that are not applied.

4.4.2 Hadronic taus

The method used to identify τ_h was reported in Section 3.6, and is based on the `DeepTau` discriminant. τ_h are used only in the case of t̄tH measurements, in order to target the decay of the Higgs into this particle. As in the case of muons and electrons, three levels of τ_h selection criteria are used: loose, fakeable and tight.

Observable	Loose	Fakeable	Tight
(Cone-)p_T	$> 5, 10$ GeV	> 10 GeV	> 10 GeV
$|\eta|$	< 2.4	< 2.4	< 2.4
$|d_{xy}|$	< 0.05 cm	< 0.05 cm	< 0.05 cm
$|d_z|$	< 0.1 cm	< 0.1 cm	< 0.1 cm
d/σ_d	< 8	< 8	< 8
I_m	$< 0.4 \times p_T$	$< 0.4 \times p_T$	$< 0.4 \times p_T$
Muon ID	>WP-loose >WP-medium[1]	>WP-loose >WP-medium[1]	>WP-medium[1]
`DeepJet` of nearby jet	—	<WP-interp.[2] ($<$ WP-medium,—)	<WP-medium[2], —
Jet relative isolation[3]	—	$<0.5, 1.2$ (—)	—
Prompt-μ MVA	—	$<0.85, 0.4$ ($>0.85, 0.4$)	$> 0.85, 0.4$

[1] Muon ID WPs are defined in Section 3.3.
[2] The `DeepJet` score is selected by defining a linear interpolation between the WP medium and loose of the `DeepJet` discriminant (defined in Section 2.3.2) as a function of the cone-p_T.
[3] Defined as $1/p_T^{\text{ratio}}$-1 if the muon is matched to a jet within $\Delta R < 0.4$ or as the PF relative isolation with $\Delta R = 0.4$ otherwise.

TABLE 4.5: Loose, fakeable and tight selection criteria for muons. Requirements shown in black are applied when studying both tt̄H and tt̄W, while the ones in blue are only applied in the tt̄H measurements and the ones in green only in the tt̄W one. The cuts in parentheses are applied only when leptons fail the tight lepton MVA cut. A long dash (—) indicates selection criteria that are not applied.

All τ_h selected must have a p_T greater than 20 GeV, be within the acceptance ($|\eta| < 2.3$), and not overlay with a loose electron or muon within $\Delta R < 0.3$. Moreover, the longitudinal impact parameter is required to be less than 0.2 cm.

Loose τ_h selection is used to assure no τ_h enter the purely light lepton categories. On top of the previous selection τ_h should pass the Very-Very Loose working point of the `DeepTau` discriminant against jets.

The fakeable τ_h definition is used to contract the regions used to estimate the contribution of background from jets misidentified as τ_h. This selection adds, on top of the loose τ_h selection, criteria on the discriminant against muons and electrons.

Finally, the tight τ_h selection is used to define the τ_h that will populate the signal regions. In each of the categories a different background contribution from jets misidentified as τ_h is expected, as a result, different working points of the jet-τ_h discriminate are used in each category to maximize the signal efficiency. The used working points go from the very-loose to the very-tight one, corresponding to signal efficiencies from 90% to 50% and the corresponding jet misidentification probabilities from 4% to 0.3%.

Observable	Loose	Fakeable	Tight
p_{T}	> 20 GeV	> 20 GeV	> 20 GeV
$\lvert \eta \rvert$	< 2.3	< 2.3	< 2.3
$\lvert d_z \rvert$	< 0.2 cm	< 0.2 cm	< 0.2 cm
DeepTau vs. jets	$>$ WP-VVLoose	$>$ WP-VVLoose	Channel-dependent
DeepTau vs. muons	—	$>$ WP-VLoose	$>$ WP-VLoose
DeepTau vs. electrons	—	$>$ WP-VVVLoose	$>$ WP-VVVLoose

TABLE 4.6: Loose, fakeable and tight selection criteria for hadronic τ decays. A long dash (—) indicates selection criteria that are not applied.

Additionally, to discriminate against muons (electrons) the very-very-very loose (very loose) working points of the corresponding discriminants are used in all channels.

The criteria for each selection is described in detail in Table 4.6.

4.4.3 Jets and b tagging

Jet reconstruction and identification techniques were covered in Section 2.3.1. In this thesis jets are reconstructed from particle flow candidates by the anti-k_t algorithm, using a $\Delta R = 0.4$ as distance parameter. Jets need to have a $p_{\mathrm{T}} > 25$ GeV and $\lvert \eta \rvert <$ 2.4 to be considered in both tt̄W and tt̄H + tH studies. In the second case, in order to increase the sensitivity for tH production —where the spectator quark is expected to be emitted in the forward direction— the jet selection is enlarged by considering "forward jets". These forward jets can have an $\lvert \eta \rvert$ up to 5. Jets in the region $2.7 <$ $\lvert \eta \rvert < 3$ must have a $p_{\mathrm{T}} > 60$ GeV; this requirement is added due to the calorimeter noise that specially can affect low-p_{T} jet measurements. Finally, jets that have any lepton that passes the fakeable selection within a $\Delta R < 0.4$ cone are removed to select only jets produced in the parton shower.

In order to distinguish jets from light quarks and gluons from those coming from b quarks, the DeepJet discriminant is used, as described in Section 2.3.2. In this thesis both the loose and medium WP are used. In addition, some of the MVAs used at analysis level to classify events as signal or background make use of the score of the DeepJet discriminant.

4.4.4 Missing transverse energy

In Section 2.3.4 two quantities were defined to account for the momentum imbalance in the transverse plane: $p_{\mathrm{T}}^{\mathrm{miss}}$ and H_{T}. In the tt̄H studies presented in this thesis a

linear combination of both is used in order to select events:

$$L_D = 0.6 \times p_T^{\text{miss}} + 0.4 \times H_T \tag{4.6}$$

In events with genuine missing transverse energy both observables are correlated, while in events with instrumental missing transverse energy they are not. In tt̄H multileptonic final states we expect the observables to be correlated when compared with some of the backgrounds. The coefficients in Eq. 4.6 were optimized to increase the separation between tt̄H and the Z+jet background.

4.5 Signal and background estimation

4.5.1 Signals and irreducible backgrounds

Signals and irreducible backgrounds are described using Monte Carlo simulations.

The tt̄H samples used in tt̄H measurements are simulated with MADGRAPH_AMC@NLO at NLO accuracy in QCD; in tt̄W measurements the tt̄H simulation is computed with POWHEG at NLO. The tH process is simulated at LO in QCD using MADGRAPH_AMC@NLO. The 4FS is used to simulate tHq channel, in order to better simulate the additional quark from the gluon splitting. In the case of the tHW sample the 5FS is used in order to avoid interference with tt̄H at LO.

The tH samples include weights that model variations on the process kinematics as a function of the Higgs-W and Higgs-top couplings. These variations are parameterized using κ_t and κ_W and imply, as well, modifications in the normalization. In order to model tt̄H dependency of the Higgs-top coupling, an alternative LO sample simulated with MADGRAPH is used.

The tt̄W process is simulated at NLO accuracy using MADGRAPH_AMC@NLO. An additional sample, simulated with MADGRAPH_AMC@NLO, is used to account for the electroweak corrections of order $\mathcal{O}(\alpha_S \alpha^3)$. This represented the stat-of-the-art Monte Carlo for tt̄W production at the time when these studies were conducted.

The main irreducible backgrounds when studying both tt̄H and tt̄W production are tt̄Z and diboson production with additional jets. tt̄Z, tt̄γ, WZ and semileptonic ZZ contributions are simulated at NLO accuracy using MADGRAPH_AMC@NLO, for tt̄Z the selected phase space takes a Z mass higher than 1 GeV. For ZZ production to 4ℓ

POWHEG generator is used. Although less relevant, the $W^\pm W^\mp$ and $W^\pm W^\pm$ productions are considered as well. $W^\pm W^\mp$ events are simulated using POWHEG at NLO and $W^\pm W^\pm$ using MADGRAPH_AMC@NLO at LO accuracy.

Triboson production (WWW, WWZ, WZZ, ZZZ) as well as other low cross section processes involving the top quark: tZq and tt̄tt̄ are simulated at NLO using MADGRAPH_AMC@NLO, while tt̄WV (V = W, Z, and H) production is simulated at LO using MADGRAPH.

Non-tt̄H production modes account for a small contributions in the selected final states, they are modelled by NLO accuracy simulation computed in POWHEG, except VH production which is modelled using MADGRAPH_AMC@NLO.

A summary of the samples used in the analysis and the cross section values used to normalize the predictions is shown in Table 4.7.

4.5.2 Reducible backgrounds

Four sources of reducible backgrounds are considered: nonprompt lepton background and misidentified τ_h, charge misidentification and conversions.

4.5.2.1 Nonprompt leptons

Dedicated techniques, introduced in Section 3.5.2 are used to reject nonprompt leptons and are shown to be very efficient. Nevertheless, given the large cross section of some of the backgrounds (such as tt̄) when compared to the signal, there is a non-negligible contribution from processes with nonpromopt leptons populating the signal region. As an example, in final states with two *ss* leptons, events arising from tt̄ semileptonic decays can enter the selection if one of the b jets produces a *ss* lepton. Similarly, in a final state with three leptons, dileptonic tt̄ events can pass the selection if the third lepton is a nonprompt one. This background is estimated using data-driven techniques, in particular the so called nonprompt rate method. An Application Region (AR) is defined by loosening the signal region definition. The aim is to allow fakeable leptons, not passing the tight lepton selection described in Section 4.4.1, to enter this region. Therefore, the AR will be enriched in events with nonprompt leptons and can be used to estimate the contribution of such events in the signal region by applying a factor, called nonprompt rate (f). The factor f is measured in a Measurement Region (MR) and then applied on the AR to obtain the estimated contribution of nonprompt leptons in the signal region. Afterwards, the method is validated by performing a closure test using simulation.

Process	Generator	Order in QCD	Cross section [fb]
Signals			
t̄tH	MADGRAPH_aMC@NLO/ POWHEG	NLO	507 [117]
t̄tH alternative couplings	MADGRAPH_aMC@NLO	LO	507 [117]
tHq	MADGRAPH_aMC@NLO	LO	70.9 [144]
tHW	MADGRAPH_aMC@NLO	LO	15.6 [144]
t̄tW	MADGRAPH_aMC@NLO	NLO	601 [144], 592 [19]
Top-related backgrounds			
t̄tWW	MADGRAPH_aMC@NLO	LO	6.98 [145]
t̄tWZ [1]	MADGRAPH_aMC@NLO	LO	3.8 [144]
t̄tZ/t̄tγ *			839 [144] 859 [19]
$1 \leq m_{\ell\ell} < 10\,\mathrm{GeV}$	MADGRAPH_aMC@NLO	NLO	
$m_{\ell\ell} \geq 10\,\mathrm{GeV}$	MADGRAPH_aMC@NLO	NLO	
tZq	MADGRAPH_aMC@NLO	NLO	73.6
t̄ttt̄	MADGRAPH_aMC@NLO	NLO	8.21
Diboson production			
WW			1.19×10^5 [146]
$W^{\pm}W^{\mp}$	POWHEG	NLO	
$W^{\pm}W^{\pm}$	MADGRAPH_aMC@NLO	LO	
WZ			4.50×10^4 [146]
$WZ \to \ell\ell\ell\nu$	MADGRAPH_aMC@NLO	NLO	
$WZ \to \ell\ell qq$ [3]	MADGRAPH_aMC@NLO	NLO	
ZZ			1.69×10^4 [146]
$ZZ \to \ell\ell\nu\nu$ [1]	MADGRAPH_aMC@NLO	NLO	
$ZZ \to \ell\ell\ell\ell$	POWHEG	NLO	
$ZZ \to \ell\ell qq$ [3]	MADGRAPH_aMC@NLO	NLO	
Triboson production			
WWW	MADGRAPH_aMC@NLO	NLO	208.6
WWZ	MADGRAPH_aMC@NLO	NLO	165.2
WZZ	MADGRAPH_aMC@NLO	NLO	55.6
ZZZ	MADGRAPH_aMC@NLO	NLO	13.9
Other Higgs production modes			
ggH	POWHEG	NLO	4.86×10^4 [144]
VBF	POWHEG	NLO	3.78×10^3 [144]
ZH [2]	POWHEG	NLO	884 [144]
VH	MADGRAPH_aMC@NLO	NLO	1.37×10^3 [144]
t̄tWH [2]	MADGRAPH_aMC@NLO	LO	1.58 [144]
t̄tZH [2]	MADGRAPH_aMC@NLO	LO	1.54 [144]

TABLE 4.7: Summary of MC simulation used in this thesis to model the signal as well as the irreducible backgrounds. For each process the generator, accuracy in the QCD perturbative computation used to produce the sample, and the cross section used to normalize it are given. In some cases the normalization used when performing the t̄tH measurement and the dedicated t̄tW measurement are different: green numbers correspond to the normalisation applied in the t̄tW analysis, while the ones in blue correspond to the ones applied in t̄tH measurement.
[1] Sample only applied in t̄tW studies.
[2] Sample only applied in t̄tH studies.
[3] Background only considered as irreducible in $1\ell+1\,\tau_{\mathrm{h}}$ and $0\ell +2\,\tau_{\mathrm{h}}$ categories of t̄tH analysis.

Nonprompt rate measurement

The procedure used to measure the nonprompt rate was given in Section 3.5.2. Here a revisit of the selection including also electrons is provided.

Events in the measurement region have exactly one lepton passing the fakeable selection criteria. Additionally, at least one recoiling jet, not overlapping with the lepton

within a $\Delta R > 0.7$, is required in the event. Events are selected by triggers not including isolation requirements. Muon triggers were shown in Table 3.2, while electron triggers are detailed in Table 4.8. Electron triggers require the presence of an extra jet and impose certain quality criteria on the electrons. To avoid any possible bias these quality criteria are mimicked with the selection in Section 4.4.1.

As already anticipated in Section 3.5.2, this region will be enriched in QCD events, while the main contribution for nonprompt background is coming from tt̄. As a result, a different flavour composition and momentum spectrum of the nonprompt lepton is expected. To mitigate this effect, the fakeable selection was tuned to assure a similar flavour composition between both samples.

The nonprompt rate computation is explained in Section 3.5.2, in particular the definition of the nonprompt rate, f is given in Eq. 3.6.

HLT path	Prescale	Lepton reco-p_T	Jet p_T
HLT_Ele8_CaloIdM_TrackIdM_PFJet30	15-45 GeV	> 8 GeV	> 30 GeV
HLT_Ele17_CaloIdM_TrackIdM_PFJet30	25-100 GeV	> 17 GeV	> 30 GeV
HLT_Ele23_CaloIdM_TrackIdM_PFJet30	32-100 GeV	> 23 GeV	> 30 GeV

TABLE 4.8: Triggers used to record events for the measurement of the nonprompt electron rate.

Aplication of the nonprompt rate

After the measurement of the nonprompt rate, a transfer factor between the AR and SR is computed depending on the lepton multiplicity in the SR. If two tight leptons are required in the final state, the transfer factor can be denoted as N_{PP}, and defined in the equation:

$$N_{PP}^{fake} = \sum_{fp} F_1 + \sum_{pf} F_2 - \sum_{ff} F_1 F_2, \tag{4.7}$$

where F_i is constructed using the nonprompt-rate evaluated in the i-th lepton, sorted by cone-p_T (f_i), as $F_i = \frac{f_i}{1-f_i}$. The sums in the equation runs over the events in the AR for specific combinations of the leptons. The symbol f in the sum denotes that the lepton has failed the tight lepton selection; while p means that the lepton is a tight lepton. The first position indicates that the requirement is applied on the leading lepton, while the second position, that it applies to the subleading lepton. As an example, $\sum_{pf}$ runs over all events in the AR where the leading lepton is tight and the subleading lepton is a fakeable lepton.

Analogously, the expression can be extended to final states with three and four tight leptons. For three leptons the expression would be:

$$N_{\mathrm{PPP}}^{\mathrm{fake}} = \sum_{\mathrm{fpp}} F_1 + \sum_{\mathrm{pfp}} F_2 + \sum_{\mathrm{ppf}} F_3$$
$$- \sum_{\mathrm{ffp}} F_1 F_2 - \sum_{\mathrm{fpf}} F_1 F_3 - \sum_{\mathrm{pff}} F_2 F_3 + \sum_{\mathrm{fff}} F_1 F_2 F_3. \tag{4.8}$$

The contribution from prompt leptons in the AR is estimated using simulation and is removed before applying the transfer factor.

Closure

A closure test is performed in simulated events passing the requirements of the signal region in order to test the methodology of the data-driven estimation and show no biases are present. This ensures that the method is not introducing any bias due to the different flavour composition of the measurement region with respect to the application region, as well as due to the kinematic of the signal region.

The test is performed independently for muons and electrons using a t̄t sample in the signal region with two *ss* leptons in the final state. The number of nonprompt events in the SR estimated using semileptonic t̄t Monte Carlo (the so-called nominal selection) is compared with the estimation of nonprompt events computed by applying the nonprompt rate to events in the AR (relaxed selection). Two nonprompt rates are used for the estimation, in both cases events used for the computation of the f must pass the MR selection. In the first case, the calculation of f is done on a multijet simulation sample, while in the other a t̄t simulation sample is used. By comparing both estimations, the effect of the residual flavour composition and kinematics between the nonprompt leptons produced in multijet events and those produced in t̄t events is shown. In addition, the comparison between the t̄t simulation and nonprompt-rate estimation using a multijet sample gives an indication of the total bias expected in the method.

Figure 4.14 shows the closure test for muons and electrons as a function of the cone-p_{T}. Good closure of the order of 10% is shown for muons, while for electrons, a 30 to 40% non-closure is depicted.

4.5.2.2 Misidentified τ_h

The misidentification rate for τ_h can be defined as the number events in which a quark or gluon jet passes the tight τ_h selection over the number of such events passing the fakeable τ_h selection.
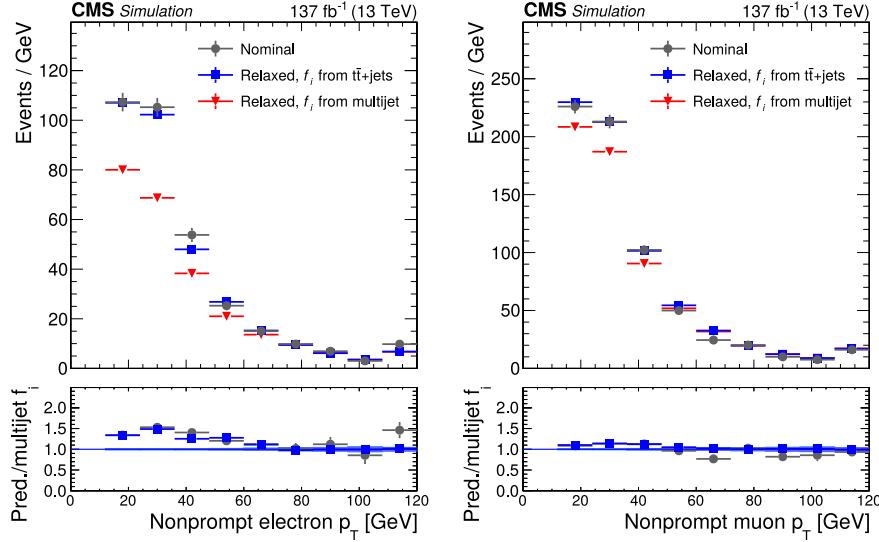
FIGURE 4.14: Transverse momentum distributions of nonprompt (left) electrons and (right) muons in simulated tt̄+jets events, for the three cases "nominal", "relaxed, f_i from tt̄+jets", and "relaxed, f_i from multijet" discussed in text. The figure illustrates that a nonclosure correction needs to be applied to the probabilities f_i measured for electrons in data, while no such correction is needed for muons.

The rate is measured using tt̄+jets events in which the two W bosons produced in the top quark decays decay to an electron and muon pair. Hence, the selection requires the presence of an opposite-sign eμ pair, as well as, the presence of a τ_h passing the fakeable selection. Additionally, at least two jets should be present in the event and the same b-tagging requirement as in the signal region is applied. Besides, events in the selection must have m$_{e\mu}$ > 12 GeV in order to reduce the contributions from low mass resonances. Contributions of genuine τ_h are modelled using the MC simulation and subtracted. In the above selection, the efficiency of the fakeable τ_h to pass also the tight τ_h criteria is computed following the same prescription as for nonprompt light leptons.

4.5.2.3   Charge misidentification

The charge of an electron or muon might be incorrectly determined, yielding to the appearance of this background in categories where the selection requires to have two leptons with equal charge. This incorrect determination is mainly a consequence of the Bremsstrahlung effect. As electrons traverse the detector, they radiate photons that may subsequently convert to electron-positron pairs. This makes it harder to correctly determine the curvature of the electron track. Given the larger mass of the muons their Bremsstrahlung effect is smaller, therefore the muon charge misidentification rate is negligible.

This background is estimated using data, by measuring the number of events with two electrons with equal charge and two electrons with opposite sign of the charge in the process $Z \rightarrow e^{\pm}e^{\mp}$. A measurement region is defined to select this kind of events by requiring two leptons with invariant mass within the Z peak. The events within the measurement region are further categorized into same-sign (*ss*) and opposite sign of the charge (*os*) samples, with N_{ss} and N_{os} the number of events in each of them. The charge misidentification rate can then be defined as: $r = \frac{N_{ss}}{N_{ss}+N_{os}}$. Moreover, the measurement is performed as a function of p_T and η of the leptons. The number of events in each category of the measurement region is obtained by performing a fit to the Z peak. To perform such fit, the shapes of the $m_{\ell\ell}$ distribution taken from simulation are used as templates. The simulation used for this propose is DY (Z/γ^*) simulation produced with MADGRAPH_AMC@NLO at NLO in QCD. The estimation of events with misidentified charge entering the signal region is estimated by applying the misidentification probability (r) to data events in the signal region as a function of p_T and η.

4.5.2.4   Conversions

This includes events in which an electron is produced in the conversion of an unidentified photon. The dominant contribution in the analysis presented in this thesis is from t̄tγ events and is estimated using simulation.

4.6   Statistical techniques for signal extraction

In the physics analyses reported in this thesis, statistical techniques are used in order to interpret the measured observables. Events are classified in different categories, where counting experiments are performed to measure cross sections of known processes and set confidence intervals on the presence of new processes.

To exploit all the information in the different categories, all counting experiments must be combined. This can be done by using a maximum-likelihood fit, following the approach in Ref. [147].

The likelihood $\mathcal{L}(\text{data}|\boldsymbol{\mu}, \boldsymbol{\theta})$ depends on the observed data, the vector $\boldsymbol{\mu}$ representing the signal strengths for the signal processes targeted by the analysis, and the nuisance parameters denoted by $\boldsymbol{\theta}$, which parameterize the systematic uncertainties. The expected signal and background yields are a function of those parameters, so the number of expected events can be written as $\nu_i = \boldsymbol{\mu} \cdot \boldsymbol{s}_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta})$ where $\boldsymbol{s}_i$ and b_i stands for

the number of expected signal and background events in each category i, respectively. The likelihood is defined as a product of probabilities:

$$\mathcal{L}(\text{data}|\boldsymbol{\mu}, \boldsymbol{\theta}) = \prod_i \mathcal{P}(n_i|\boldsymbol{\mu} \cdot \boldsymbol{s}_i(\boldsymbol{\theta}) + b_i(\boldsymbol{\theta})) \prod_k \rho_k(\bar{\theta}_k|\theta_k), \qquad (4.9)$$

where $\mathcal{P}$ denotes the probability function for the Poisson distribution, in particular, it is the probability of observing n_i events in that category, given that ν_i events are expected. The functions $\rho_k(\bar{\theta}_k|\theta_k)$ account for the prior probability of observing a value $\bar{\theta}_k$ for that nuisance parameter, given that the true value is θ_k. The uncertainties affecting the normalization of the processes are represented using a log-normal probability density function, uncertainties with statistical origin are modelled using a gamma function, and systematic uncertainties that also affect the shape of discriminating observables are incorporated into the fit as nuisance parameters [148] and are modelled by a Gaussian probability density function. In some specific cases, the normalization of certain backgrounds is left unconstrained in the fit.

The likelihood function can be maximised to find the best values for the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\theta}$. In order to remove the dependence on the nuisance parameters, the likelihood ratio test statistic is defined as a quotient where: the numerator is the likelihood where the nuisance parameters are substituted with their maximum likelihood estimate conditional on the current value of $\boldsymbol{\mu}$; the denominator is the likelihood where both $\boldsymbol{\mu}$ and $\boldsymbol{\theta}$ are set to their joint maximum likelihood estimate. This test statistic, when maximised, provides an estimate for $\boldsymbol{\mu}$ that accounts for, but not depends on, the nuisance parameters $\boldsymbol{\theta}$. For computational reasons, instead of maximising the likelihood ratio, the negative log likelihood ratio is minimised:

$$q_\mu = -2\log \frac{\mathcal{L}(\text{data}|\boldsymbol{\mu} \cdot \boldsymbol{s}(\hat{\hat{\boldsymbol{\theta}}}) + b(\hat{\hat{\boldsymbol{\theta}}}))}{\mathcal{L}(\text{data}|\hat{\boldsymbol{\mu}} \cdot \boldsymbol{s}(\hat{\boldsymbol{\theta}}) + b(\hat{\boldsymbol{\theta}}))}, \qquad (4.10)$$

where $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\theta}})$ is the maximum likelihood estimator and $\hat{\hat{\boldsymbol{\theta}}}$ maximizes the likelihood for a given $\boldsymbol{\mu}$ value. The minimization of the test statistic allows to obtain confidence intervals by applying Wilks' theorem [149]. This states that in the limit of large number of observations, q_μ follows a χ^2_n distribution if $\boldsymbol{\mu}$ is the true value, where n is the number of dimensions of $\boldsymbol{\mu}$. Under that circumstances, the crossing of $\boldsymbol{\mu}$ with the quantiles of χ^2 would indicate the confidence intervals.

4.7   Systematic uncertainties

The event yields and the shape of the distribution of any observable used in the analysis, may be altered by experimental or theory-related effects, referred to as systematic uncertainties.

Experimental sources of uncertainty include those related to the object selection, uncertainties in the data-driven estimates of the misidentified leptons and charge misidentification backgrounds and those derived from auxiliary measurements used to validate and correct the simulations. Theoretical or modelling uncertainties are those arising from missing higher-order corrections in the cross section calculations and simulation models, and from uncertainties in the PDFs.

Uncertainties are taken into account in the analyses as nuisance parameters of the likelihood fit.

Experimental uncertainties

Luminosity: the integrated luminosities of the 2016, 2017, and 2018 data-taking periods are individually measured with their corresponding uncertainty that amounts for 2.5%, 2.3% and 2.5%, respectively [150–152]. The luminosity for 2016 was re-evaluated [153] and the related uncertainty was reduced to 1.2%, the latest estimation of 2016 luminosity is in agreement with the previous one.

Pileup reweighting: variations on the number of additional pp interactions per event may modify the efficiency to pass the event selection criteria, hence, inducing a variation in the event yield. To account for this effect, simulation is reweighted using the information from the instantaneous luminosity profile in data. A systematic uncertainty is estimated by varying the assumed minimum-bias cross section of 69.2 mb by $\pm 4.6\%$ [154]. The uncertainty is propagated as shape uncertainty and is considered to be correlated between data-taking years.

Lepton selection: Muon and electron identification and isolation efficiencies are estimated as described in Ref. 3.3.1. Per-lepton corrections are derived for electrons and muons independently and as a function of the p_T and η of the leptons. Then, they are propagated depending on the number of leptons in the final state. Overall uncertainties in the muon and electron identification and isolation efficiencies are considered as shape uncertainties. The uncertainty is considered uncorrelated between the two flavours and correlated across the data-taking years. To account for differences in the

event topology between DY events and the events in the signal regions of the analysis, an additional cross-check is performed in a region enriched in tt̄+jets events. The difference between the efficiency measured in the tt̄+jets enriched sample and the one measured in DY events is taken as an additional systematic uncertainty, and found to be around to 1–2%. The uncertainty in the τ_h identification efficiencies and energy scale amount to 5% and 1.2%, respectively. They are estimated using $Z/\gamma \to \tau\tau$ events [155] and are dominated by statistical effects; for this reason it is treated as uncorrelated.

Trigger: efficiency on the trigger selection is estimated using the orthogonal trigger method described in 2.5.1. The efficiency is computed separately depending on the lepton multiplicity, η and p_T using a set of triggers based on p_T^{miss}. Trigger uncertainties amount to 1-2% and are treated as uncorrelated across years for all categories.

Jet energy scale and resolution: jet reconstruction is affected by resolution effects and by the energy of the jet. Eleven uncertainty sources are considered on the jet energy scale, while six components on the jet energy resolution are used to model the dependency on the detector region and p_T range [92]. Jet energy scale and resolution effects are propagated to the analysis by varying the energies of jets in simulated events within their uncertainties, recalculating all kinematic observables, and reapplying the event selection criteria. An extra uncertainty is used to account for the HCAL endcap issue described in Section 2.4.3. Besides, jet reconstruction directly affects p_T^{miss} estimation due to the unclustered energy. Variations on this quantity are propagated to the p_T^{miss} estimation as an additional uncertainty. The uncertainties of statistical origin are uncorrelated across years, while the other components are treated as correlated.

b-**tagging:** uncertainties on the b-tagging efficiency and mistag rate are evaluated as a function of the p_T and η [93]. Several sources of uncertainties are considered: the purity of the sample selected to derive the corrections, the statistics of that sample and the effect of jet energy scale uncertainties.

L1 prefiring issue: the trigger inefficiency due to the loss of transparency in the ECAL endcap cover in Section 2.4.3 is taken into account in the analyses as a correction factor with its associated uncertainty. Then it is propagated to the analysis as a shape uncertainty.

Nonprompt background estimation: the nonprompt estimation based on data-driven techniques has several sources of uncertainty. The first is the statistical uncertainty on the events selected in the MR and AR. Moreover, the multijet enriched sample is contaminated with events containing prompt leptons, which are subtracted, an additional systematic uncertainty takes into account inaccuracies in the subtraction procedure.

These uncertainties are accounted in the analyses by varying the nonprompt rates (f). Three nuisance parameters are introduced for each lepton flavour to model the uncertainty on the fake rate: one that modifies the overall scale, and one that changes the dependence on p_T and η respectively, but keeps the total number of estimated nonprompt events constant. The uncertainty is taken as the envelope of the three variations. Additionally, an extra uncertainty is applied to cover for the possible non-closure effects such as the ones shown in Fig. 4.14. The overall contribution of this uncertainty to the analyses is moderate and is higher for final states with τ_h.

Charge misidentification background estimation: the normalization uncertainty related to this background, which is estimated from data, is accounted by applying a 30 % in categories with electrons. The estimation is based on the data to simulation comparison in a region with two *ss* leptons with the m_{ee} within the Z mass peak.

Modelling uncertainties

Uncertainties due to missing higher-order diagrams in the theoretical calculations are evaluated by varying the scales μ_R and μ_F. The effect of such variations on the shape and normalization are taken as independent uncertainties and are applied to all background processes modelled with simulation. For the signal, only the effect on the shape is taken into account. The uncertainties related to the PDFs and α_s are also taken into account in the same manner, following the recommendations in Ref. [156].

For the tt̄W analysis, the effect of the initial and final state radiation (ISR and FSR) in the PS simulation of the signal process is also taken into account as a shape systematic uncertainty. The effect of the colour-reconnection model is additionally taken into account by comparing the tt̄W simulated events with alternative models [157].

A mismodelling of the ZZ+jets and WZ+jets backgrounds is observed at high jet multiplicity. An additional uncertainty of 30% is applied to WZ and ZZ events with at least three selected jets in order to account for mismodelling effects. Besides, a large number of b-tagged jets in WZ and ZZ process are either due to misidentified light-flavour jets, or to the inclusion of a phase space region that is not well described in simulation. An additional 40 (10)% uncertainty is applied to the predicted number of events with two or more (fewer than two) b-tagged jets following the studies in Ref. [133].

In the tt̄W analysis, the normalization of tt̄H, tH, tt̄γ, tZq are constrained to be within 20% [9], 50% [58], 8% [138] and 10% [158], following the latest CMS measurements and in order to cover for the extrapolation to the phase space of the analysis, which

contains events with high jet and b-tagging multiplicities. Other low cross section processes are assigned a 50% uncertainty given the limited knowledge of their cross section.

For the t$\bar{t}$H analysis, the uncertainties on the normalization, including missing higher-order corrections in the perturbative expansion, different choices of PDFs, and uncertainties in α_S for the signals and t$\bar{t}$W, t$\bar{t}$WW, and t$\bar{t}$Z backgrounds are taken from Ref. [117]. They amount to $^{+6.8}_{-9.9}$ and $^{+5.1}_{-7.3}$% for the t$\bar{t}$H and tH, and to $^{+13.5}_{-12.2}$, $^{+8.6}_{-11.3}$, and $^{+11.7}_{-10.2}$% for the t$\bar{t}$W, t$\bar{t}$WW, and t$\bar{t}$Z, respectively. Uncertainties on the cross section for other low rate processes estimated with Monte Carlo simulation (including tZq) are assigned a 50% uncertainty. The uncertainties on the branching ratio of the Higgs are also taken into account, and are extracted from Ref. [117].

Finally, conversions are assigned a 30-50% normalization uncertainty, following studies in Ref. [159].

Regarding the statistical uncertainty of the simulation samples in the signal and control regions used in the analysis, the Barlow-Beeston [160] method is used to account for it.

5

Measurement of $t\bar{t}W$ **production**

The $t\bar{t}W$ process is one of the main backgrounds when studding $t\bar{t}H$ (and tH) production in final states with multiple leptons. Furthermore, this process presents unique production features as the charge asymmetry and, unlike $t\bar{t}Z$, previous measurements showed a tension with the prediction. This motivates dedicated studies of $t\bar{t}W$ process. In the previous chapter the most important characteristics of $t\bar{t}W$ production and the experimental status prior to the studies reported in this thesis were given. In this chapter the measurement of the $t\bar{t}W$ inclusive cross section using the full Run-2 dataset at $\sqrt{s} = 13\,\text{TeV}$ is reported. Final states with two and three leptons are used.

The analysis presented in this chapter has been developed in collaboration with University of Ghent (UG), Université catholique de Louvain (UCL) and University of Zurich (UZH). I lead the analysis in the three lepton final state, performed studies on the systematic uncertainties, performed the signal extraction and contributing to all aspects of the analysis.

5.1   Analysis strategy for $t\bar{t}W$ **measurement**

In this thesis $t\bar{t}W$ process is studied in final states with leptons, where the W boson decays to a lepton (electron or muon) and a neutrino. Depending on the decay of the $t\bar{t}$ system, which can be semileptonic or dileptonic, final states with two or three leptons are targeted. Hence, the first classification of events is done according to lepton multiplicity in two non-overlapping categories with two leptons with equal sign of the charge ($2\ell ss$) or three leptons (3ℓ).

In each category a different strategy is followed: in the $2\ell ss$ category an MVA is trained in order to discriminate signal and background. In the 3ℓ signal region, events are further categorize to obtain several subcategories of events with different purity; then, the mass of the three lepton system is used as discriminating variable. In both

cases, events are divided depending on the lepton charge to be sensitive to the charge asymmetry expected in this process.

5.2 Event selection in the t̄tW signal regions

5.2.1 Dilepton category

At trigger level events are selected by a combination of single- and double-lepton triggers, with a p_T threshold that vary depending on the data-taking period. A summary of the trigger paths used with the corresponding thresholds is given in Table 5.1.

Trigger path	flavour	p_T thresholds [GeV]
Single-lepton	e	27-35
	μ	24
Double-lepton	ee	23 (leading)
	eμ	23 (leading)
	$\mu\mu$	17 (leading)
Triple-lepton	eee	16 (leading)
	eeμ	8 (leading)
	e$\mu\mu$	9 (leading)
	$\mu\mu\mu$	10-12 (leading)

TABLE 5.1: Summary of trigger paths used to record events. The threshold applied on the p_T of the leading lepton is shown.

Exactly two tight leptons are required with same electric charge, a p_T of at least 25 or 30 GeV for the leading lepton, depending on whether it is a muon or an electron, and a $p_T > 20$ GeV for the subleading lepton. The invariant mass of the *ss* dilepton pair should be larger than 30 GeV and the leptons should be separated by $\Delta R > 0.4$. In case both leptons are electrons, $m_{\ell\ell}$ should not be within 15 GeV of the Z mass. In addition, events must have a missing transverse energy of at least 30 GeV and contain at least two jets, out of which at least two pass the loose b-tagging requirement or at least one passes the medium requirement. The comparison between prediction and data is shown in Fig. 5.1 for various observables in the selection described above. In the figures the contribution from the charge misidentification background is labeled as *Charge misID* and low cross section processes with a marginal contribution are depicted under the label *Other*, this includes tW, tH, tZ, triboson, t̄tVV and t̄tt̄t production. As depicted, the agreement is reasonably good, with some discrepancies in the bins most populated by signal. The agreement improves after the fit to data.
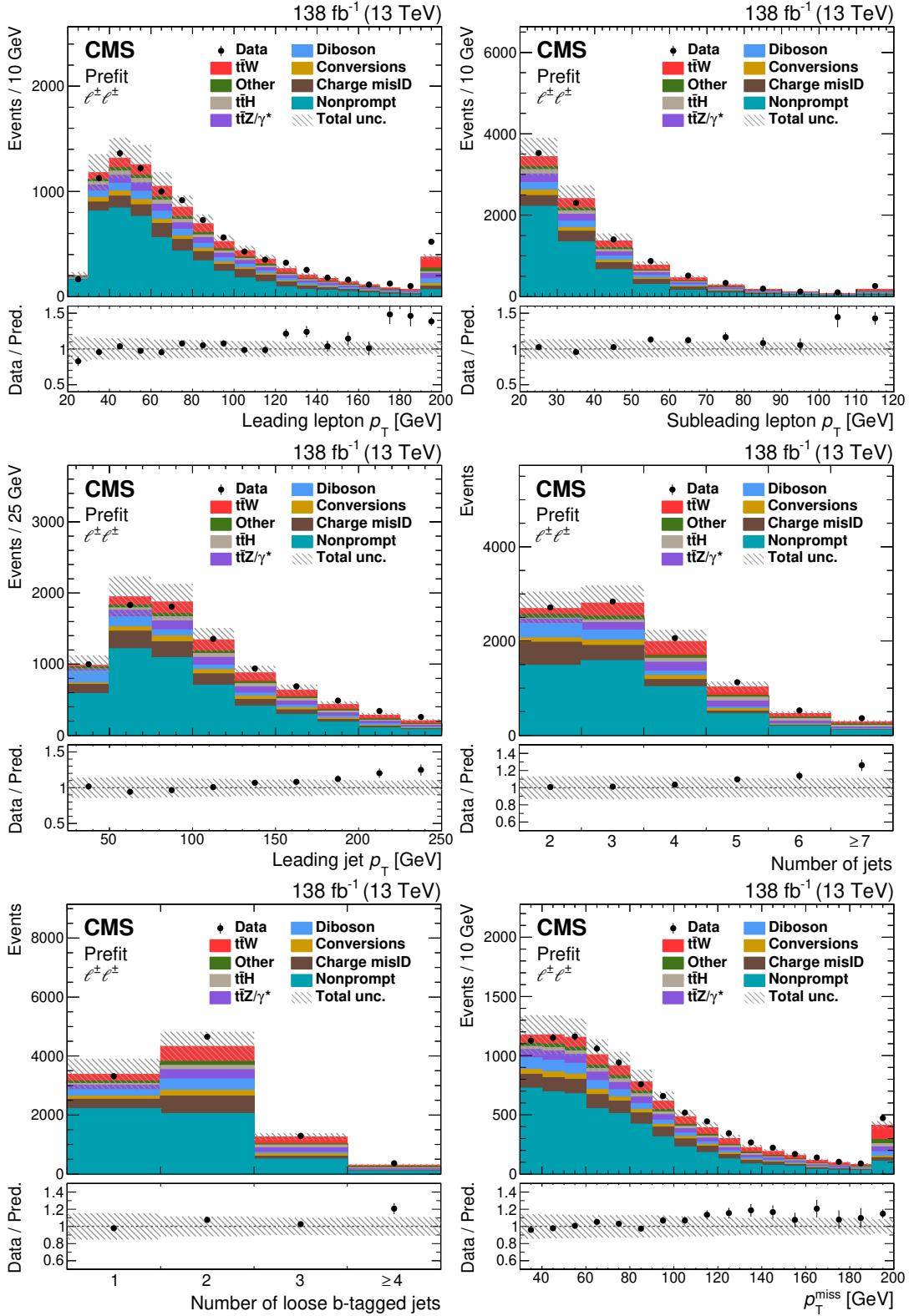
FIGURE 5.1: Comparison of the number of observed (points) and predicted (coloured histograms) events in the same-sign dilepton signal region. Several observables are depicted: the leading (top left) and subleading (top right) lepton p_T, the leading jet p_T (middle left), the number of jets (middle right), the number of b-tagged jets (bottom left) and the missing transverse momentum (bottom right). The predictions are shown before the fit to data. The vertical bars on the points represent the statistical uncertainties in the data, and the hatched bands the systematic uncertainty in the predictions. The last bins include the overflow contributions. In the lower panels, the ratio of the event yields in data to the overall sum of the predictions is presented.

5.2.2   Three-lepton category

In this category events are selected at trigger level by a combination of single-, double- and triple-lepton triggers. A summary of the trigger paths used with the corresponding threshold is given in Table 5.1.

Exactly three tight leptons, with $p_T > 25$ GeV for the leading lepton and $p_T > 15$ GeV for the other two leptons are required. To remove background contributions from low-mass resonances, the invariant mass of each pair of leptons in the event is required to be $m_{\ell\ell} > 12$ GeV. Besides, events with $|m_Z - 91.2| < 10$ GeV are excluded, where m_Z is the invariant mass for each pair of leptons of opposite sign of the charge and same flavour (OSSF), this criteria aims to remove DY contributions. Additionally, the sum of charges of the three leptons is required to be ± 1, and events must contain at least two jets, out of which at least one passes the medium criteria of the b-tagging discriminant.

5.3   Control regions for t̄tW analysis

Two control regions are constructed to improve the understanding of some of the leading irreducible backgrounds: the first one selects events with three leptons in the final state and is enriched in t̄tZ and WZ backgrounds. The second one targets events with four leptons and is used to constrain the ZZ and t̄tZ backgrounds.

Three lepton control region

The three lepton control region is defined following the selection of the 3ℓ signal category, but inverting the Z boson mass veto and not applying any requirements on the jet and b-tagged jet multiplicities. Then, events are classified in 12 bins as a function of the number of jets and b-tagged jets, as described in Table 5.2. The data to prediction comparison in the described categorization is shown in Fig. 5.2 (left). Events are later classified depending on the flavour of the leptons.

Four lepton control region

This control region is constructed using events where four leptons with $p_T > 25$, 15, 15, and 10 GeV, respectively, are present. Additionally, at least one OSSF lepton pair on the Z boson mass ($|m_Z - 91.2| < 10$ GeV) is required. Events are further classified in 4 bins depending on the jet (j) and b tag multiplicities (b), and on the presence of an

Category	Number of Jets	Number of b-tagged Jets
1	= 1	= 0
2	= 2	= 0
3	= 3	= 0
4	>3	= 0
5	= 2	= 1
6	= 3	= 1
7	= 4	= 1
8	> 4	= 1
9	= 2	> 1
10	= 3	> 1
11	= 4	> 1
12	> 4	> 1

TABLE 5.2: Event categorization in the three lepton control region.
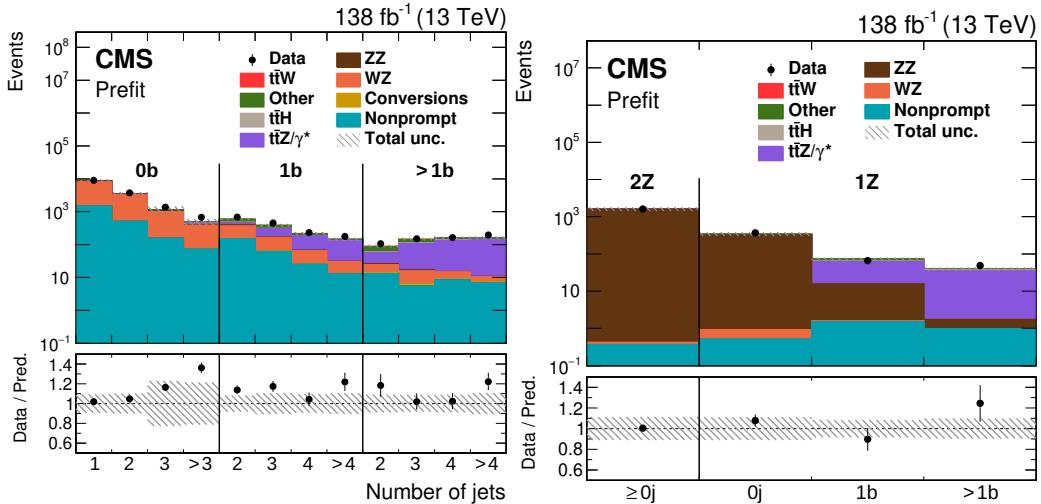
FIGURE 5.2: Comparison of the number of observed (points) and predicted (coloured histograms) events in the three lepton (left) and four lepton (right) control regions. The predictions are shown before the fit to data. The vertical bars on the points represent the statistical uncertainties in the data, and the hatched bands the systematic uncertainty in the predictions. In the lower panels, the ratio of the event yields in data to the overall sum of the predictions is presented.

extra OSSF lepton pair in the event (2Z). A summary of the event yields in data and prediction in this control region is shown in Fig. 5.2 (right).

5.4 Signal extraction for t̄tW measurement

Events in both the $2\ell ss$ and 3ℓ categories, as well as in the control regions, are combined to extract the signal by performing a maximum likelihood fit. After the selection applied in each of the signal regions, events are further categorize in order to improve

the signal against background discrimination, and certain discriminating distributions are provided to the fit to enhance the sensitivity.

5.4.1 Dilepton category

In the 2ℓss category a multiclass NN is trained to distinguish between events from the t̄tW signal process and those from three classes of background processes: t̄tZ/t̄tγ^* and t̄tH as a combined category, t̄tγ, and backgrounds with nonprompt leptons. The NN is trained using simulated t̄tW, t̄tZ/t̄tγ^*, t̄tH, and t̄tγ samples, and with simulated t̄t events as a source of nonprompt-lepton background events. A different set of events is used to train the discriminant and to perform the signal extraction. This NN has been developed by the UG and UCL groups, hence here a brief description of the most relevant items is given.

The NN exploits the reconstructed kinematics of the leptons and jets, the presence of additional jets, and the presence of b-tagged jets. The variables used as input are:

- lepton flavour category, *ee*, *μμ*, *eμ* or *μe*, depending on the flavour of the leading lepton

- lepton charge ($+1$ or -1)

- p_T and η of the leptons

- invariant mass of the lepton pair ($m_{\ell\ell}$)

- difference in azimuthal angle ($\Delta\phi_{\ell\ell}$) and pseudorapidity ($\Delta\eta_{\ell\ell}$) between the leptons

- transverse mass of each lepton and the p_T^{miss} in the event, $m_T(\ell_i, p_T^{\text{miss}})$

- p_T^{miss}, and the azimuthal angle difference between the $\vec{p}_T^{\text{miss}}$ and each of the leptons ($\Delta\phi(\vec{p}_T^{\text{miss}}, \ell_i)$)

- number jets passing the loose selection of the b-tagging algorithm, a minimum of one and a maximum of two b-tagged jets are considered

- number of jets that are not among the b-tagged ones

- p_T and η of the b-tagged jets and the leading non-b-tagged jets (up to four).

- azimuthal angle difference between the jets and the leading ($\Delta\phi(j_i, l_1)$) and sub-leading lepton ($\Delta\phi(j_i, l_2)$), respectively.

- invariant mass for the system constructed by each b-tagged jet and each lepton, $m(\ell_i, b_j)$

- ΔR between any pair of jets

- for b-tagged jets a label indicating whether they pass the medium selection of the b-tagging discriminant

- a label for each of the four leading jets, indicating whether they also pass the loose selection of the b-tagging discriminant (this can be the case for events with more than two jets that pass the loose b-tagging working point)

Six of the input variables are shown in Fig. 5.1.

The NN is constructed with two fully connected hidden layers, with 128 and 64 nodes respectively, and an output layer with four nodes, corresponding to the signal category and the three background categories defined above. The value that a certain event obtains in each output node can be interpreted as the probability for that event to belong to that category. Only the t̄tW output node is used for the signal extraction. Nevertheless, the described multiclassifier leads to a better discriminating power when compared to a binary classification.

5.4.2   Three-lepton category

In this category, after the event selection described above, events are further categorized using the the number of jets, number of b-tagged jets and sum of the lepton charges. Twelve subcategories are constructed, each of them with a different sensitivity and background composition. In particular, the events are classified into three categories based on the number of jets: exactly 2 jets, 3 jets, or more than three jets in the event. Within these categories, further divisions are made based on whether one or two of the jets are tagged as coming from a b quark. Finally, each of the six categories is divided according to the sum of the charge of the three leptons, that can be +1 or -1. The data to prediction comparison in the classification of the three lepton signal region is shown in Fig. 5.3.

In each of the subcategories a common single discriminant variable is chosen to be used in the maximum likelihood fit for the signal extraction. Several observables such as p_T^{miss}, the sum of p_T of the three leptons, H_T, $\Delta R(\ell\ell)$ or the invariant mass of the three lepton system have been studied as discriminant variables. The invariant mass of the three leptons, $m(3\ell)$, was selected, as it was the observable with the best discriminating power between signal and background contributions. In order to
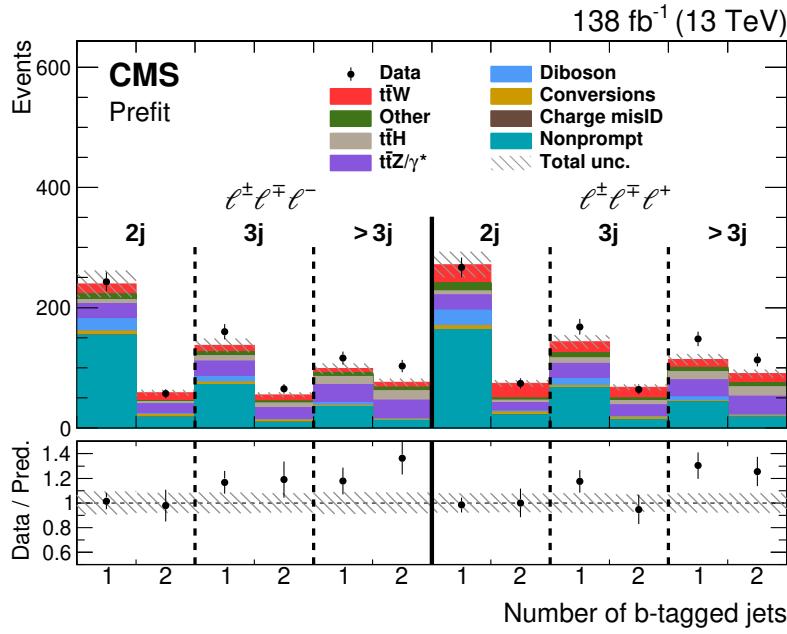
FIGURE 5.3: Comparison of the number of observed (points) and predicted (coloured histograms) events in the three lepton signal region. The predictions are shown before the fit to data. The vertical bars on the points represent the statistical uncertainties in the data, and the hatched bands the systematic uncertainty in the predictions. In the lower panel, the ratio of the event yields in data to the overall sum of the predictions is presented.

evaluate this the total total uncertainty on the signal strength was used as figure of merit.

5.5 Results

The number of events observed in each category is reported and compared to the expected value. A description of the signal extraction fit is also provided and the results for the t̄tW cross section is given. In addition, the cross section for t̄tW$^+$ and t̄tW$^-$ and the ratio between both production modes is obtained.

The fit is performed using both signal regions with $2\ell ss$ and 3ℓ, as well as, the three-lepton and four-lepton control regions. In the $2\ell ss$ category, the NN node targeting t̄tW is used as discriminating variable, dividing the events in 8 subcategories depending on the lepton flavour and charge. In Fig. 5.4 the output node is shown in events with positive (left) and negative (right) charge of the leptons; the plots are inclusive in flavour. As depicted in the figure, the NN provides good discrimination between signal and background, specially against the nonprompt background. The different shape between left and right figures is explained by the fact that the charge of the leptons is used as input during the training. Moreover, it is worth noting that in events

with positive lepton charge, the amount of signal is larger than in events with negative charge, while the amount of background remains constant.

In the 3ℓ signal region, the invariant mass of the three lepton system is used in the twelve categories described in the previous section. In Fig. 5.5 the $m(3\ell)$ is shown in four of the subcategories. The distributions shown have two b-tagged jets and two or three jets, respectively. Events are also divided depending on the charge of the leptons.

Process	$\ell+\ell+$	$\ell-\ell-$	$\ell\pm\ell\mp\ell+$	$\ell\pm\ell\mp\ell-$
t$\bar{\text{t}}$W	677 ± 21	355 ± 12	119.4 ± 9.2	65.3 ± 5.4
Nonprompt	2490 ± 600	2360 ± 570	325 ± 75	298 ± 71
Charge misidentification	520 ± 110	520 ± 111	—	—
t$\bar{\text{t}}$H	167 ± 34	169 ± 34	56 ± 12	57 ± 12
t$\bar{\text{t}}$Z/ t$\bar{\text{t}}\gamma$ *	335 ± 26	333 ± 26	145 ± 13	147 ± 13
Diboson	382 ± 88	285 ± 65	46.8 ± 9.1	38.0 ± 7.5
Other	178 ± 34	126 ± 27	43.4 ± 8.2	33.5 ± 7.4
Conversions	177 ± 54	192 ± 59	22.9 ± 7.1	24.0 ± 7.4
Total background	4250 ± 620	4000 ± 590	639 ± 80	600 ± 76
Total prediction	4920 ± 620	4350 ± 590	758 ± 81	663 ± 76
Data	5143	4486	834	744

TABLE 5.3: Number of predicted and observed events in the dilepton and trilepton signal regions before the fit to the data. The uncertainties in the predicted number of events include both the statistical and systematic components. The uncertainties in the total number of predicted background and background plus signal events in each channel are also given. The symbol "—" indicates that the corresponding background does not apply.

The total number of observed and predicted events obtained before the fit in the $2\ell ss$ and 3ℓ signal regions is given in Table 5.3. Good agreement between data and the total signal plus background prediction within the systematic uncertainties is observed for each channel.

Additionally, the fit includes the event yields in the three lepton control region, where events are classified in 48 bins grouped by the number of jets and b-tagged jets, and lepton flavours, and in the four lepton control region in four bins grouped by the number of jets, b-tagged jets, and Z candidates.

The systematic uncertainties are modelled as nuisance parameters, which are able to introduce modifications in the shape and normalization of the predicted events in the distributions included in the fit. The fit also takes as free parameters the signal strengths for t$\bar{\text{t}}$W, t$\bar{\text{t}}$Z, WZ and ZZ. The signal strengths can modify the the normalization of the related processes. In Figs. 5.4 and 5.5, the data to prediction comparison is
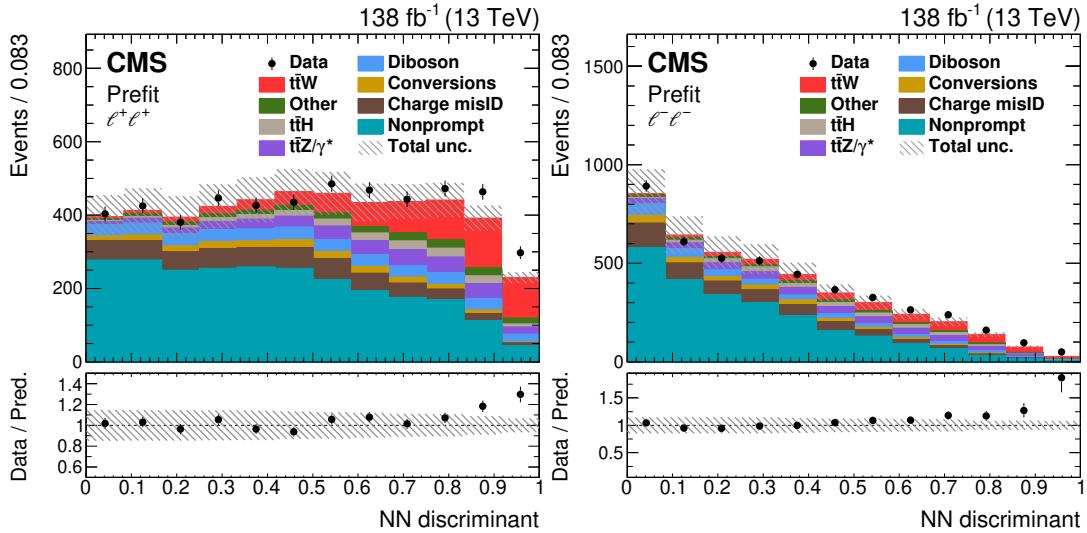
FIGURE 5.4: NN discriminant in the $2\ell ss$ signal region. Events are classified in the left (right) plot if they have positive (negative) lepton charge. The predictions are shown before the fit to data. The vertical bars on the points represent the statistical uncertainties in the data, and the hatched bands the systematic uncertainty in the predictions. In the lower panels, the ratio of the event yields in data to the overall sum of the predictions is presented.

shown before the fit to data: "prefit". A slight excess is shown in data for some of the bins enriched in t\=tW events. The distributions are also shown after the fit; "postfit" distributions for both signal regions are depicted in Figs. 5.6 and 5.7 showing good agreement. In appendix A, all distributions for both signal regions entering the fit are shown after the fit to data.

Measurement of the inclusive cross section

The signal strength modifier for t\=tW is found to be:

$$\mu_{t\bar{t}W} = 1.47 \pm 0.07 \text{ (stat.)}^{+0.09}_{-0.08} \text{ (syst.)}, \qquad (5.1)$$

and can be interpreted in terms of the cross section by taking into account the cross section used to normalize the expected prediction of t\=tW events. The measured cross section for the inclusive t\=tW production is:

$$\sigma_{t\bar{t}W} = 868 \pm 40 \text{(stat)} \pm 51 \text{(syst)} \text{fb}. \qquad (5.2)$$

The measurement is in agreement within 2 s.d. with respect to the prediction 722^{+71}_{-78} fb [18], computed at NLO and including corrections from an improved FxFx merging procedure. This result is in agreement with the corresponding ATLAS measurement [161], and with previous CMS and ATLAS measurements [12, 14, 106, 143].
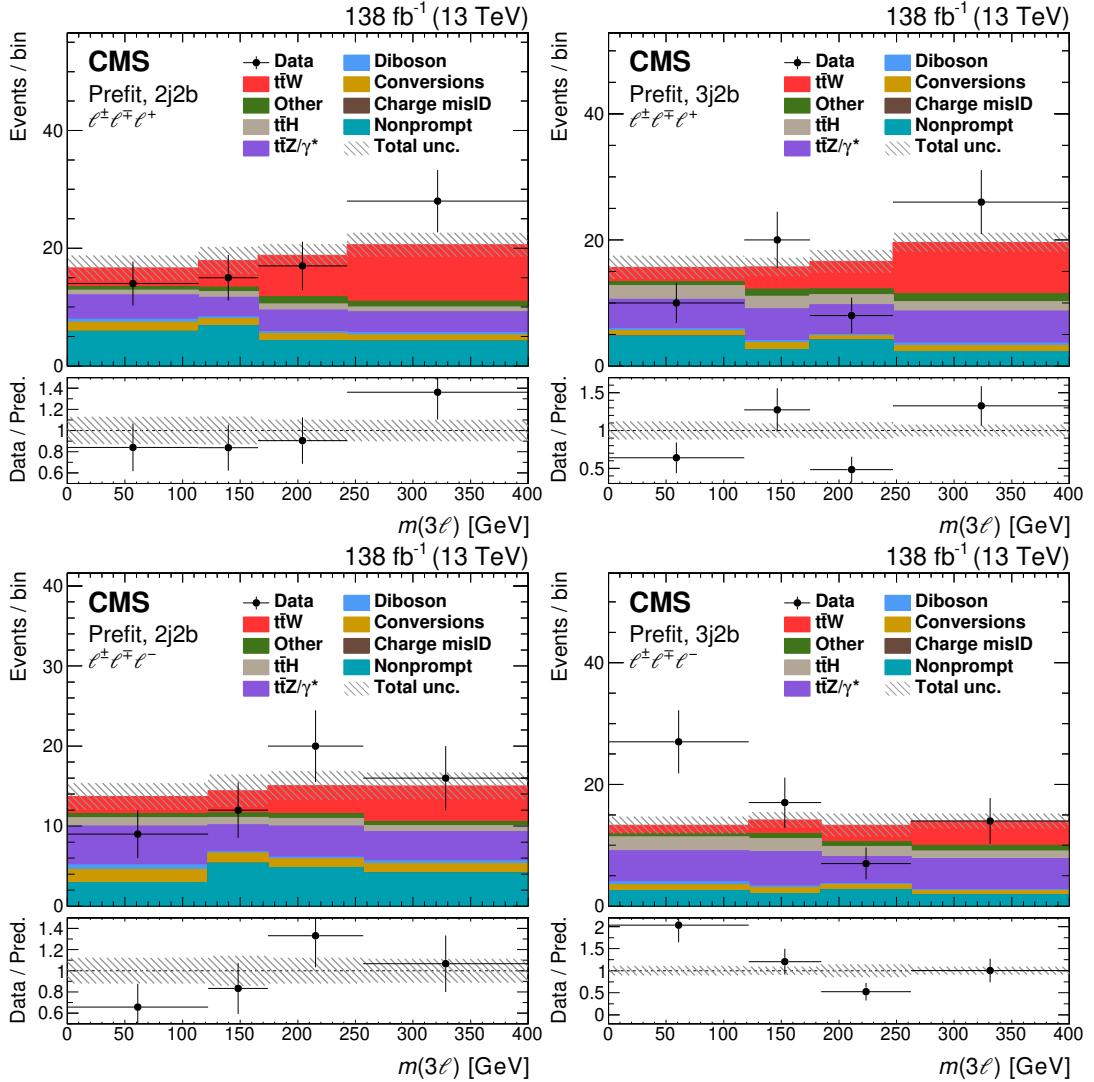
FIGURE 5.5: Comparison of the number of observed and predicted events in the 3ℓ signal region. The distribution of the $m(3\ell)$ variable is shown in four of the subcategories used in the signal extraction, in particular the ones with: two jets (left) and three jets (right) and in events with positive sum of the lepton charge (upper row) or negative charge (lower row). The predictions are shown before the fit to data. The vertical bars on the points represent the statistical uncertainties in the data, and the hatched bands the systematic uncertainty in the predictions. In the lower panels, the ratio of the event yields in data to the overall sum of the predictions is presented.

The discrepancies might be explained by a mismodelling on the signal simulation, which is challenging; regarding this, several developments have been presented in the recent years, as was detailed in the previous chapter. Those developments still need to be implemented in the Monte Carlo simulation used for the experimental data analysis. When compared to the previous CMS result [12], performed with 2016 data, the current measurement achieves a significant improvement in the precision, with a reduction on the statistical (systematic) uncertainty of a factor of three (more than two). The improvement on the statistical uncertainty can be attributed to the larger data sample used. Besides, the reduction on the systematic uncertainty is derived from the
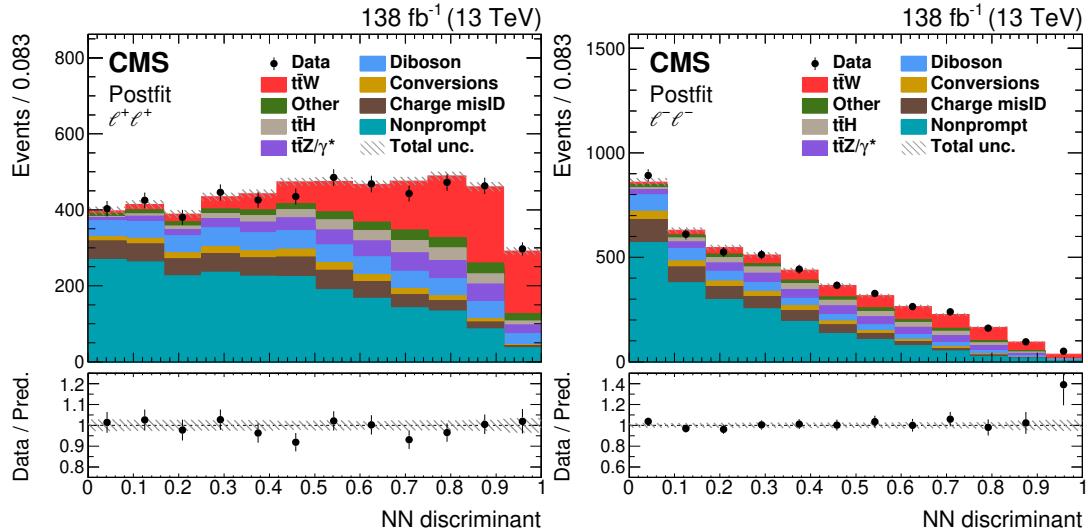
FIGURE 5.6: NN discriminant in the 2ℓss signal region. Events are classified in the left (right) plot if they have positive (negative) lepton charge. The predictions are shown after the fit to data. The vertical bars on the points represent the statistical uncertainties in the data, and the hatched bands the systematic uncertainty in the predictions. In the lower panels, the ratio of the event yields in data to the overall sum of the predictions is presented.

optimization of the analysis strategy and improved estimates of the dominant background contributions using data control samples. A more efficient selection of t̄tW events and a stronger background suppression of the nonprompt-lepton background contribution is achieved by the lepton selection with the dedicated lepton MVA discriminant. The enhanced b-tagging performance following the installation of the new pixel detector in 2017 and a more performant b-tagging algorithm, is also a source of improvement.

The cross section is also computed using only the 2ℓss and 3ℓ categories, respectively. In the case of 2ℓss category, the measurement is also provided by flavour of the leptons. The results are shown in Fig. 5.8. The 2ℓss category is the dominant one, as the signal to background proportion is greater. In the figure, the measured cross section is compared with the computation at NLO and including the improved merging procedure [18], and with the computation at NLO+NNLL [19].

Systematic and statistical Uncertainties

The contribution to the total systematic uncertainty from each of the sources considered and the total statistical uncertainty is detailed in Table 5.4. The leading systematic uncertainties of experimental origin are the ones related to the integrated luminosity measurement, the estimation of electron charge misidentification, and the b-tagged jet identification efficiency. Regarding the normalization uncertainties, t̄tH, VVV and t̄tVV give a non-negligible contribution to the total uncertainty in the measurement.
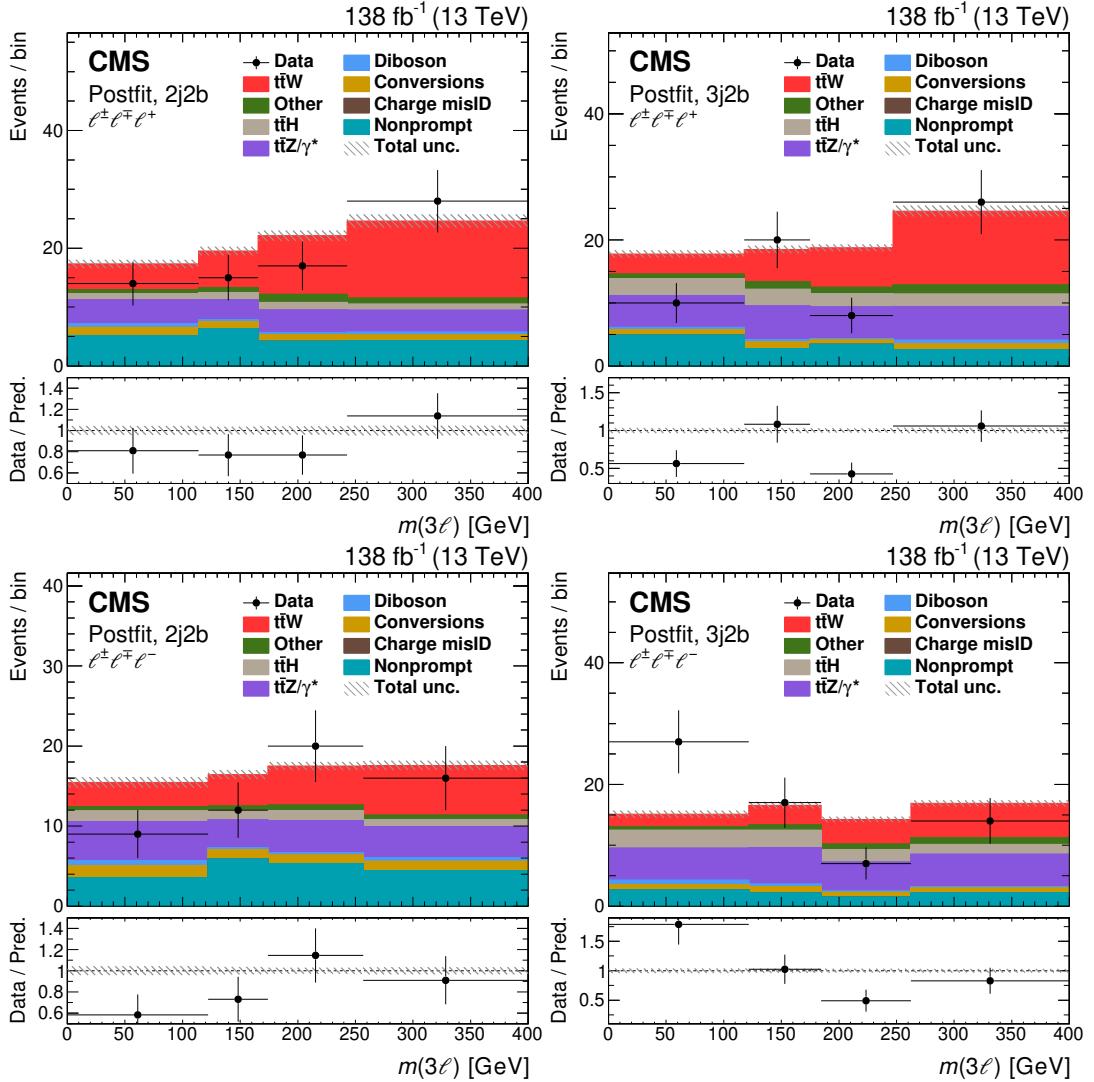
FIGURE 5.7: Comparison of the number of observed and predicted events in the 3ℓ signal region. The distribution of the $m(3\ell)$ variable is shown in four of the subcategories used in the signal extraction, in particular the ones with: two jets (left) and three jets (right) and in events with positive sum of the lepton charge (upper row) or negative charge (lower row). The predictions are shown after the fit to data. The vertical bars on the points represent the statistical uncertainties in the data, and the hatched bands the systematic uncertainty in the predictions. In the lower panels, the ratio of the event yields in data to the overall sum of the predictions is presented.

This is due to the similarities of the decay products of these rare processes with those of the t̄tW signal process. Modelling uncertainties associated with the simulation of t̄tW events and from the statistical uncertainty in the predicted number of signal and background events are also a sizable contribution.

Measurement of the t̄tW$^+$ and t̄tW$^-$ cross sections

As events are categorized depending on the charge of the leptons, this information can be exploit to measure the cross sections for t̄tW$^+$ and t̄tW$^-$ production. The fit described in the previous section is modified; in this case two parameters to modify the

Source	Uncertainty [%]
Experimental uncertainties	
Integrated luminosity	1.9
b-tagging efficiency	1.6
Trigger efficiency	1.2
Pileup reweighting	1.0
L1 inefficiency	0.7
Jet energy scale	0.6
Jet energy resolution	0.4
Lepton selection efficiency	0.4
Background uncertainties	
tt̄H normalization	2.6
Charge misidentification	1.6
Nonprompt leptons	1.3
VVV normalization	1.2
tt̄VV normalization	1.2
Conversions normalization	0.7
tt̄γ normalization	0.6
ZZ normalization	0.6
Other normalizations	0.5
tt̄Z normalization	0.3
WZ normalization	0.2
tZq normalization	0.2
tHq normalization	0.2
Modelling uncertainties	
tt̄W scale	1.8
tt̄W colour reconnection	1.0
ISR & FSR scale for tt̄W	0.8
tt̄γ scale	0.4
VVV scale	0.3
tt̄H scale	0.2
Conversions	0.2
Simulation statistical uncertainty	1.8
Statistical uncertainty	4.6
Total systematic uncertainty	5.8

TABLE 5.4: Sources of systematic and statistical uncertainty in the predicted signal and background event yields with their impact on the measured tt̄W production cross section, estimated after the fit to the data. Only systematic uncertainty sources with values greater than 0.1% are included in the table. The production cross sections of the tt̄W, WZ, ZZ, and tt̄Z processes are simultaneously constrained in the fit. The second-to-last row refers to the statistical uncertainty in the simulated event samples.

138 fb^{-1} (13 TeV)

	Nominal ±stat ±syst
ee	845 ±117 ±111
eμ	996 ±61 ±68
μμ	868 ±63 ±64
Dilepton	905 ±42 ±51
Trilepton	649 ±104 ±96
Combined	868 ±40 ±51
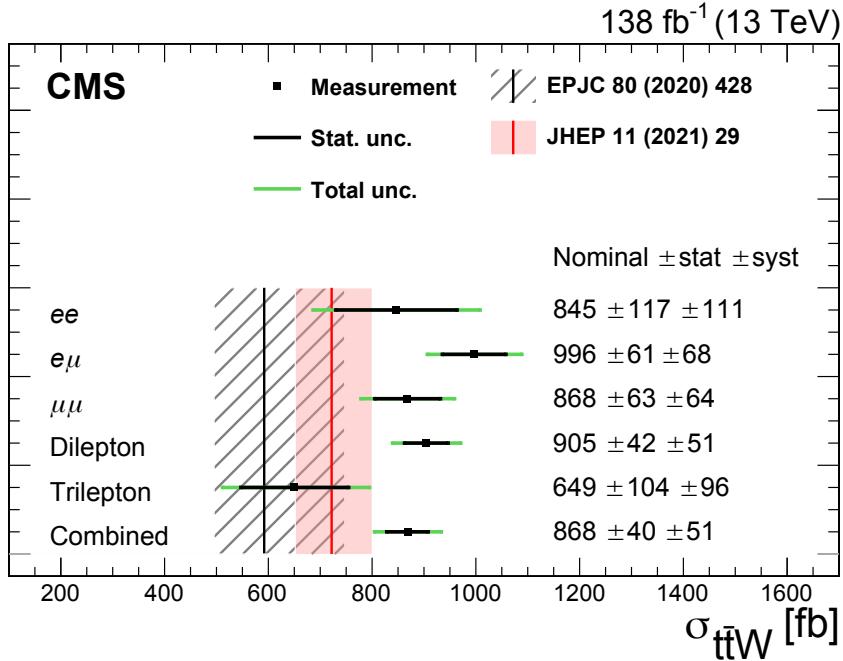
FIGURE 5.8: Cross section for t̄tW production, individual dilepton (*ee*, *eμ*, *μμ*, and combined) channels and the trilepton channel, as well as their combination. The inner black bars show the statistical uncertainty, and the outer green bars give the total uncertainty. The predictions from two SM calculations from Refs. [18, 19] are shown by the black and red vertical lines, with the associated bands corresponding to the total uncertainty.

signal strengths for t̄tW$^+$ and t̄tW$^-$ are used. Both parameters are extracted using the information from categories with positive (negative) sum of lepton charges simultaneously. Hence, the cross sections for t̄tW$^+$ and t̄tW$^-$ are obtained. This measurements are of special interest, as the production asymmetry gives information about the PDF of the proton. The results are shown in Table 5.5 with two theory computations serving as comparison. The sum of the two measured cross sections for t̄tW$^+$ and t̄tW$^-$ is in good agreement with the measured inclusive t̄tW cross section within the statistical uncertainties. As for the inclusive production, when comparing both cross sections to the theory computations in Refs. [18, 19], they are in agreement within 2 s.d. The values for the cross sections are, as well, depicted in Fig. 5.9 (left). Correlations between both cross sections are expected, arising from the use of common nuisance parameters and control regions in the fit.

Furthermore, the ratio between t̄tW$^+$ and t̄tW$^-$ production modes, $\sigma_{t\bar{t}W^+}/\sigma_{t\bar{t}W^-}$, is also computed, treating it as a single parameter and recomputing the maximum likelihood fit. In Fig. 5.9 (right) the negative log-likelihood scan is depicted, providing the best fit value for the $\sigma_{t\bar{t}W^+}/\sigma_{t\bar{t}W^-}$ with its 68 and 95% confidence level (CL) intervals. The best fit value with its corresponding uncertainties is quoted in Table 5.5, the systematic uncertainties are reduced when compared to the ones for the cross sections values,

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

TABLE 5.5: Measured and predicted production cross sections for $t\bar{t}W^+$ and $t\bar{t}W^-$ production, as well as of the $\sigma_{t\bar{t}W^+}/\sigma_{t\bar{t}W^-}$ ratio. The SM predictions quoted at NLO + NNLL accuracy are taken from Refs. [19, 162]. The SM predictions quoted at NLO accuracy and including corrections from an improved FxFx merging procedure (NLO + FxFx) have been provided by the authors of Ref. [18]. The theoretical uncertainties include scale variations and PDF uncertainties.
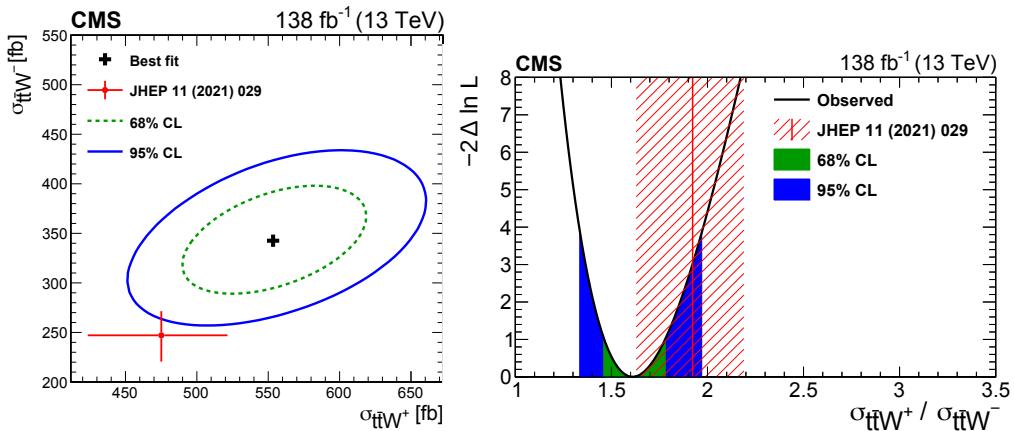
FIGURE 5.9: **Left:** Measured cross section for $t\bar{t}W^+$ vs. $t\bar{t}W^-$ production (black cross), along with the 68 (green dashed) and 95% (blue solid) CL intervals.
Right: Scan of the negative log-likelihood used in the measurement of the cross section ratio $\sigma_{t\bar{t}W^+}/\sigma_{t\bar{t}W^-}$. The best fit value is indicated, with the inner green and outer blue bands displaying the 68% and 95% CL intervals, respectively.
In both cases the SM prediction in Ref. [18] is shown in red.

this is expected as some uncertainty sources cancel out due to partial correlations. On Table 5.5 the theory value for the ratio is also provided. The uncertainty propagation from the $t\bar{t}W^-$ and $t\bar{t}W^+$ cross section theory predictions to the ratio assumes that there are no correlation between both, which results in an overestimation of the ratio uncertainty. The measured ratio is in agreement with the theory prediction in Ref. [18] within uncertainties.

5.6   Conclusions and prospects

The latest measurement for $t\bar{t}W$ inclusive cross section at $\sqrt{s} = 13\,\text{TeV}$ using the full Run-2 dataset has been presented. The analysis is performed in final states with $2\ell ss$ and 3ℓ. The obtained cross section is $\sigma_{t\bar{t}W} = 868 \pm 65$ fb, yielding the most precise

measurement up to date. The systematic uncertainty was significantly reduced with respect to the previous CMS result.

The leading uncertainties in the current analysis arise from the normalization of low cross section backgrounds, the luminosity, the nonprompt background estimation and b-tagging efficiencies. It is shown that, given the large dataset available, very low cross section production modes such as t̄tWW have a non-negligible impact on the measurement of production modes with similar final states. Hence, in order to provide high precision t̄tW measurements it is mandatory to improve the understanding of the low cross section backgrounds such as t̄tWW, tZq or triboson production. Besides, the nonprompt background plays a predominant role in this measurement, specially on the $2\ell ss$ region. The experience derived from this analysis showed the importance of reducing this background using the most efficient prompt lepton selection available. Several new selection techniques had been developed or are under development. In particular, the success achieved in jet identification by the so-called *ParticleNet* [163], is currently being extrapolated to lepton selection. This technique aims to make use of the maximal amount of information, and, in order to do so, it takes the information from unordered set of particles around the lepton and secondary vertices around the lepton. Using this information, a Graph Convolutional Neural Network architecture is used to discriminate prompt leptons from nonprompt ones. Finally, t̄tW modelling is another dominant source of uncertainty and has been the topic of an active discussing within the community.

The tension between the measurement of the t̄tW cross section presented in this thesis and the SM is confirmed by ATLAS. The t̄tW cross section at $\sqrt{s} = 13\,\text{TeV}$ using the full Run-2 dataset, reported by ATLAS is $\sigma_{t\bar{t}W} = 890 \pm 80$ fb [161].

In addition, the cross sections for $t\bar{t}W^+$ and $t\bar{t}W^-$ production and the ratio $\sigma_{t\bar{t}W^+}/\sigma_{t\bar{t}W^-}$ were reported in this thesis. A table summarizing the results for CMS, ATLAS and a comparison with the latest calculation, performed at NNLO(QCD)+NLO(EW) [21] is shown in Table 5.6.

	$\sigma_{t\bar{t}W^+}$ [fb]	$\sigma_{t\bar{t}W^-}$ [fb]	$\sigma_{t\bar{t}W^+}/\sigma_{t\bar{t}W^-}$
$\text{NNLO}_{\text{QCD}} + \text{NLO}_{\text{EW}}$	$497.5 \pm 6.6\% \pm 1.8\%$	$247.9 \pm 7.0\% \pm 1.8\%$	$2.007 \pm 2.1\%$
ATLAS	$585^{+6.0\%+8.0\%}_{-5.8\%-7.5\%}$	$301^{+9.3\%+11.6\%}_{-9.0\%-10.3\%}$	$1.95^{+10.8\%+8.2\%}_{-9.2\%-6.7\%}$
CMS	$553^{+5.4\%+5.4\%}_{-5.4\%-5.4\%}$	$343^{+7.6\%+7.3\%}_{-7.6\%-7.3\%}$	$1.61^{+9.3\%+4.3\%}_{-9.3\%-3.1\%}$

TABLE 5.6: Cross section for $t\bar{t}W^+$ and $t\bar{t}W^-$ measured by ATLAS and CMS, and latest theory prediction at $\text{NNLO}_{\text{QCD}} + \text{NLO}_{\text{EW}}$.

The measurements for $\sigma_{t\bar{t}W^+}$ and $\sigma_{t\bar{t}W^-}$ show also a tension with the latest, most precise, prediction; the agreement between the measurements and this calculation is at

the 2 s.d. level. The measurement of the ratio $\sigma_{t\bar{t}W^+}/\sigma_{t\bar{t}W^-}$ by the ATLAS collaboration is in good agreement with the prediction, whereas the CMS result shows a small tension.

Following these results, the need to further improve our knowledge of $t\bar{t}W$ production is clear, and is a key element for other measurements such as $t\bar{t}H$ production. In the extraction of the experimental results shown so far, $t\bar{t}W$ simulation at NLO (QCD) with the improved jet merging scheme [18] was not included, as this simulation was not available by the time this effort started. The use of such simulation and the performance of differential measurements of the cross section to test the modelling of $t\bar{t}W$ as a function of the relevant observable is mandatory. ATLAS already presented differential measurements [161] showing no clear preference for one of the simulation hypothesis used. As a natural continuation of the work presented in this thesis, I am performing measurements of the differential cross section in the $2\ell ss$ and 3ℓ final states. Nevertheless, the differential measurements are still dominated by statistics. Additionally, $t\bar{t}W$ production can be used to measure the leptonic charge asymmetry, as defined in Chapter 4. This measurement was already published by ATLAS [140] and is currently being performed within CMS, this will allow to improve our understanding of the top quark production.

6

Measurements of $t\bar{t}H$ and tH production in multilepton final states

The $t\bar{t}H$ and tH processes allow to study the coupling of the Higgs boson to the quark top at LO, providing a unique setup to measure the Yukawa coupling of the most massive fermion. Discrepancies in this coupling with respect to the SM expectation would be a clear indication of BSM physics. Given the low production cross sections of $t\bar{t}H$ and tH processes they are very challenging to measure; in particular, before the start of this thesis, $t\bar{t}H$ was observed only by combining all Higgs decay modes, while evidence for tH production was not claimed yet. Details about these production modes were provided in Chapter 4, including the experimental status previous to the studies presented in this thesis and in other final states.

In this chapter, the measurements related to $t\bar{t}H$ and tH production in final states with leptons, including hadronically decaying taus are reported. The selected final states target the Higgs decay to W^+W^- and $\tau^+\tau^-$ mainly, with a residual contribution from ZZ events, which are vetoed. These final states provide a better signal to background ratio when compared to $b\bar{b}$ final states and a larger branching ratio than $\gamma\gamma$ final states.

6.1 Analysis strategy for $t\bar{t}H$ and tH measurements

As already anticipated, in order to perform the analysis, events are classified in ten non-overlapping categories, depending on the lepton multiplicity in the final state. These categories are: $2\ell ss + 0\tau_h$, $3\ell + 0\tau_h$, $2\ell ss + 1\tau_h$, $1\ell + 1\tau_h$, $0\ell + 2\tau_h$, $2\ell os + 1\tau_h$, $1\ell + 2\tau_h$, $4\ell + 0\tau_h$, $3\ell + 1\tau_h$, and $2\ell + 2\tau_h$, where ss (os) denotes same-sign (opposite-sign), as both leptons are required to have equal (opposite) electric charge. The decay of the $t\bar{t}H$ system targeted in each of them is described in Table 6.1. A dedicated event
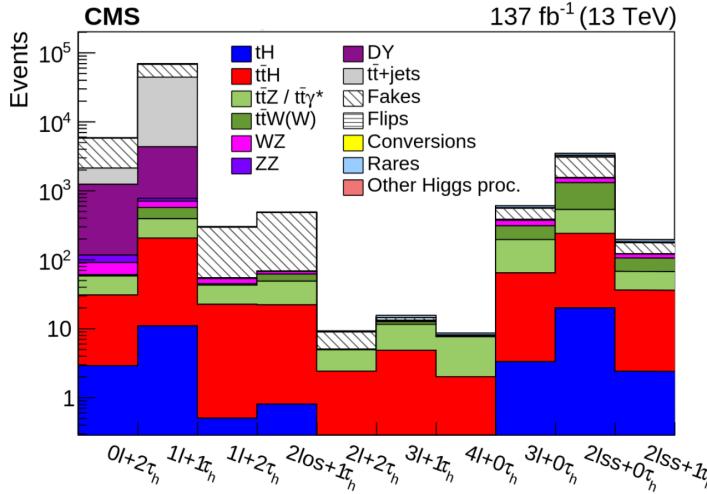
FIGURE 6.1: Expected event yields in each of the categories used in the analysis. Figure extracted from [164].

selection, detailed later in this chapter, is applied in each of the categories following closely the expected decays. In Fig. 6.1 the expected signal and background events in each of the categories are shown after the selection. As depicted, each category has a different background composition, as well as, different purity. The most sensitive categories in the analysis corresponds to the $2\ell ss + 0\tau_{\rm h}$, $3\ell + 0\tau_{\rm h}$ and $2\ell ss + 1\tau_{\rm h}$. In Fig. 6.1, the bins corresponding to those categories show a large signal purity, as well as, high total number of events. They are, also, sensitive to tH, as the number of tH events is sizable while the backgrounds are still under control. $1\ell + 1\tau_{\rm h}$ and $0\ell + 2\tau_{\rm h}$ categories present a signal yield —both for tt̄H and tH— similar to the one in the previous categories, nevertheless there is a large contribution from DY and tt̄ + jets backgrounds. The $2\ell os + 1\tau_{\rm h}$ and $1\ell + 2\tau_{\rm h}$ categories have an amount of signal slightly lower than the three most sensitive categories, while the nonprompt background is roughly a factor of 2-3 higher. Finally, the $4\ell + 0\tau_{\rm h}$, $3\ell + 1\tau_{\rm h}$, and $2\ell + 2\tau_{\rm h}$ categories present low background contributions but, due to the large lepton and τ multiplicities, these categories suffer from the low number of events and are not sensitive to tH production.

These analyses have been carried out in collaboration with the CERN, Institute of High Energy Physics (IHEP), Laboratoire Leprince-Ringuet (LLR), National Institute of Chemical Physics and Biophysic (NICPB), UCL, UZH and Cornell University. My contributions were the definition of control regions to improve the understanding of the irreducible backgrounds, the signal extraction, the coupling interpretation and the search for *CP* violation.

After the selection there is still a sizable background contribution in all the signal regions, therefore, further discrimination between signal and background is needed.

Signal category	tt̄H decay mode
$2\ell ss + 0\tau_{\mathrm{h}}$	$t \to b\ell^{\pm}\nu, t \to bqq', H \to WW \to \ell^{\pm}\nu qq'^{1}$
$3\ell + 0\tau_{\mathrm{h}}$	$t \to b\ell\nu, t \to bqq', H \to WW \to \ell\nu\ell\nu$
	$t \to b\ell\nu, t \to b\ell\nu, H \to WW \to \ell\nu qq'$
	$t \to b\ell\nu, t \to bqq', H \to ZZ \to \ell\ell qq'/\ell\ell\nu\nu$
$2\ell ss + 1\tau_{\mathrm{h}}$	$t \to b\ell\nu, t \to bqq', H \to \tau\tau \to \ell\nu\nu\tau_{\mathrm{h}}\nu^{2}$
$1\ell + 1\tau_{\mathrm{h}}$	$t \to bqq', t \to bqq', H \to \tau\tau \to \ell\nu\nu\tau_{\mathrm{h}}\nu$
$0\ell + 2\tau_{\mathrm{h}}$	$t \to bqq', t \to bqq', H \to \tau\tau \to \tau_{\mathrm{h}}\nu\tau_{\mathrm{h}}\nu$
$2\ell os + 1\tau_{\mathrm{h}}$	$t \to b\ell^{\pm}\nu, t \to bqq', H \to \tau\tau \to \ell^{\mp}\nu\nu\tau_{\mathrm{h}}\nu^{3}$
$1\ell + 2\tau_{\mathrm{h}}$	$t \to b\ell\nu, t \to bqq', H \to \tau\tau \to \tau_{\mathrm{h}}\nu\tau_{\mathrm{h}}\nu$
$4\ell + 0\tau_{\mathrm{h}}$	$t \to b\ell\nu, t \to b\ell\nu, H \to WW \to \ell\nu\ell\nu$
	$t \to b\ell\nu, t \to b\ell\nu, H \to ZZ \to \ell\ell qq'/\ell\ell\nu\nu$
$3\ell + 1\tau_{\mathrm{h}}$	$t \to b\ell\nu, t \to b\ell\nu, H \to \tau\tau \to \ell\nu\nu\tau_{\mathrm{h}}\nu$
$2\ell + 2\tau_{\mathrm{h}}$	$t \to b\ell\nu, t \to b\ell\nu, H \to \tau\tau \to \tau_{\mathrm{h}}\nu\tau_{\mathrm{h}}\nu$

TABLE 6.1: Summary of top quark and Higgs decays targeted in each signal region.
[1] The leptons should have same electric charge, which means if one is coming from the top (antitop) the other should come from a W^{+} (W^{-}).
[2] The leptons should have same electric charge, which means if one is coming from the top (antitop) the other should come from a τ^{+} (τ^{-}) decaying to light leptons.
[3] The leptons should have opposite sign of the charge, which means if one is coming from the top (antitop) the other should come from a τ^{-} (τ^{+}) decaying to light leptons.

Given the features of each signal region, different strategies are used to classify events and improve the sensitivity. In the $2\ell ss + 0\tau_{\mathrm{h}}$, $3\ell + 0\tau_{\mathrm{h}}$ and $2\ell ss + 1\tau_{\mathrm{h}}$ regions, where the number of events is large, a NN is trained to classify events as signal and background. Furthermore, as reducible backgrounds are sizable in these categories, and there is a significant number of tH events, this category would benefit from applying multiclassification, such as the one provided by a multiclass NN. In the rest of the categories, with lower number of signal events and large contribution from non-reducible backgrounds, a BDT classifier is used to discriminate the sum of signal processes from the total background.

6.2 Event selection in the tt̄H and tH signal regions

Events are selected at trigger level by a set of single-, double- and triple-lepton triggers; lepton+τ_{h} and double τ_{h} triggers are also used. The p_{T} thresholds applied vary slightly depending on the data-taking period. A summary of the trigger paths and

its corresponding thresholds for the leading lepton is given in Table 6.2. The trigger selection in each category is consistent with the lepton requirements in the specific category.

Trigger path	flavour	p_T thresholds [GeV]
Single-lepton	e	25-35
	μ	22-27
Double-lepton	ee	23 (leading)
	eμ	23 (leading)
	$\mu\mu$	17 (leading)
Triple-lepton	eee	16 (leading)
	eeμ	12 (leading)
	e$\mu\mu$	9 (leading)
	$\mu\mu\mu$	12 (leading)
Lepton+τ_h	e+τ_h	24 (e), 20 or 30 (τ_h)
	μ+τ_h	19-20 (μ), 20 or 27 (τ_h)
Double τ_h	τ_h τ_h	35-40

TABLE 6.2: Summary of trigger paths used to record events. The threshold applied on the p_T of the leading lepton is shown.

The p_T thresholds applied to leptons and τ_h in the offline event selection are above the trigger thresholds. The charge of the leptons and τ_h is required to match the selected decay mode, besides some categories have specific requirements on the lepton charge, such as the $2\ell ss + 0\tau_\text{h}$, where two ss leptons are required in order to reduce the background.

Moreover, several kinematic selections are applied. All signal regions veto events with invariant mass lower than 12 GeV, to reduce contributions from low mass resonances. Events from on-shell Z bosons are also rejected, by applying a selection on the invariant mass of an OSSF pair of leptons. Some of the regions also include requirements on the value of the missing transverse energy, as neutrinos are expected in the final states.

Additionally, requirements on the jet and b-tagged jet multiplicities are applied. These selections are designed to match the expected decay in each category, but also include events in which the jets are outside the acceptance. For categories sensitive to tH, the jet and b-tagged jet multiplicity requirements are relaxed: events that do not pass the nominal jet and b-tagging requirements, but have at least one light jet, that can be forward, and at least one jet that passes the medium b-tagging requirement, are also included in the selection. This is done to account for the fact that in tHq, the spectator quark is likely to be emitted in the forward direction, and only one b-tagged jet is expected.

Finally, to ensure there is no overlap between categories, events with a number of tight leptons larger than the strictly required are vetoed. Additionally, events with at least four leptons and $m_{4l} < 140$ GeV are rejected to be orthogonal to the H $\rightarrow 4\ell$ analysis.

A summary of the specific selection criteria applied in each category is provided in Tables 6.3 to 6.7.

Selection step	$2\ell ss + 0\tau_{\mathrm{h}}$	$2\ell ss + 1\tau_{\mathrm{h}}$
Trigger	Single- and double-lepton triggers	
Lepton p_{T}	$p_{\mathrm{T}} > 25 \: / \: 15$ GeV	$p_{\mathrm{T}} > 25 \: / \: 15$ GeV (e) or 10 GeV (μ)
Lepton η	$|\eta| < 2.5$ (e) or 2.4 (μ)	
$\tau_{\mathrm{h}} \: p_{\mathrm{T}}$	—	$p_{\mathrm{T}} > 20$ GeV
τ_{h}—η	—	$|\eta| < 2.3$
τ_{h} identification	—	very loose
Charge requirements	2 ss leptons	2 ss leptons
	and charge quality requirements	and charge quality requirements
		$\sum\limits_{\ell,\tau_{\mathrm{h}}} q = \pm 1$
Multiplicity of central jets[†]	$\geq$3 jets	$\geq$3 jets
b-tagging requirements[†]	$\geq$1 medium b-tagged jet or $\geq$2 loose b-tagged jets	
Missing transverse momentum	$L_{\mathrm{D}} > 30$ GeV	
Dilepton invariant mass	$|m_{\ell\ell} - m_{\mathrm{Z}}| > 10$ GeV[‡] and $m_{\ell\ell} > 12$ GeV	

TABLE 6.3: Event selections applied in the $2\ell ss + 0\tau_{\mathrm{h}}$ and $2\ell ss + 1\tau_{\mathrm{h}}$ categories. The p_{T} thresholds applied to the leading and subleading leptons are separated by slashes. The symbol "—" indicates that no requirement is applied.
[‡] Applied to OSSF lepton pairs and all electron pairs with same charge.
[†] Events not passing this requirements but meeting the relaxed jet and b-tagging selection, described in the text are included in the category.

6.3 Control regions in the t t̄H and tH analysis

Two control regions with three and four leptons in the final state, respectively are implemented. As in the t t̄W analysis, they are used to constrain the t t̄Z, WZ and ZZ backgrounds. The control regions are defined by inverting the Z veto in the $3\ell + 0\tau_{\mathrm{h}}$ and $4\ell + 0\tau_{\mathrm{h}}$ categories and removing the jet multiplicities requirements. Then, events are classified following the same approach as in the t t̄W analysis, as defined in Section 5.3. These regions enter the maximum Likelihood fit explained later, on Section 6.4. The classification of events in each of the control regions is depicted in Fig. 6.2, where postfit event yields and uncertainty are used.

Selection step	$3\ell + 0\tau_h$	$3\ell + 1\tau_h$
Trigger		Single-, double- and triple-lepton triggers
Lepton p_T		$p_T > 25 \ / \ 15 \ / \ 10\,\mathrm{GeV}$
Lepton η		$|\eta| < 2.5 \ (e)$ or $2.4 \ (\mu)$
$\tau_h \ p_T$	—	$p_T > 20\,\mathrm{GeV}$
τ_h—η	—	$|\eta| < 2.3$
τ_h identification	—	very loose
Charge requirements	$\sum\limits_{\ell} q = \pm 1$	$\sum\limits_{\ell,\tau_h} q = 0$
Multiplicity of central jets[†]		≥ 2 jets
b-tagging requirements[†]		≥ 1 medium b-tagged jet or ≥ 2 loose b-tagged jets
Missing transverse momentum		$L_D > 0$ if Nj > 4, $> 30\,\mathrm{GeV}$ if there is an SFOS lepton pair and Nj ≤ 3 or $> 45\,\mathrm{GeV}$ in any other case
Dilepton invariant mass		$m_{\ell\ell} > 12\,\mathrm{GeV}$ and $|m_{\ell\ell} - m_Z| > 10\,\mathrm{GeV}$[‡]
Four-lepton invariant mass	$m_{4\ell} > 140\,\mathrm{GeV}$[¶]	—

TABLE 6.4: Event selections applied in the $3\ell + 0\tau_h$ and $3\ell + 1\tau_h$ categories. The p_T thresholds applied to the leading, subleading and trailing leptons are separated by slashes. The symbol "—" indicates that no requirement is applied and Nj stands for the number of jets.
‡ Applied to OSSF lepton pairs.
¶ If the event contains two SFOS pairs of leptons that pass the loose lepton selection criteria.
† In the $3\ell + 0\tau_h$ events not passing this requirements but meeting the relaxed jet and b-tagging selection, described in the text are included in the category.

Selection step	$0\ell + 2\tau_h$	$1\ell + 1\tau_h$
Trigger	Double-τ_h trigger	Single-lepton and lepton+τ_h triggers
Lepton p_T	—	$p_T > 30 \ (e)$ or $25\,\mathrm{GeV} \ (\mu)$
Lepton η	—	$|\eta| < 2.1$
$\tau_h \ p_T$	$p_T > 40\,\mathrm{GeV}$	$p_T > 30\,\mathrm{GeV}$
$\tau_h \ \eta$		$|\eta| < 2.1$
τ_h identification	loose	medium
Charge requirements	$\sum\limits_{\tau_h} q = 0$	$\sum\limits_{\ell,\tau_h} q = 0$
Multiplicity of central jets		≥ 4 jets
b-tagging requirements		≥ 1 medium b-tagged jet or ≥ 2 loose b-tagged jets
Dilepton invariant mass		$m_{\ell\ell} > 12\,\mathrm{GeV}$

TABLE 6.5: Event selections applied in the $0\ell + 2\tau_h$ and $1\ell + 1\tau_h$ categories. The symbol "—" indicates that no requirement is applied.

6.4 Signal Extraction for tt̄H and tH processes

Events in all signal regions, as well as in the control regions, are combined to extract the rates for the signals by performing a maximum likelihood fit. As already stated, in each of the signal regions, events need to be further categorized to improve the

Selection step	$1\ell + 2\tau_h$	$2\ell + 2\tau_h$
Trigger	Single-lepton and lepton+τ_h triggers	Single- and double-lepton triggers
Lepton p_T	$p_T > 30$ (e) or 25 GeV (μ)	$p_T > 25$ / 10 (15) GeV (e)
Lepton η	$|\eta| < 2.1$	$|\eta| < 2.5$ (e) or 2.4 (μ)
τ_h p_T	$p_T > 30$ / 20 GeV	$p_T > 20$ GeV
τ_h—η	$|\eta| < 2.1$	$|\eta| < 2.3$
τ_h identification	medium	medium
Charge requirements	$\sum_{\ell,\tau_h} q = \pm 1$	$\sum_{\ell,\tau_h} q = 0$
Multiplicity of central jets	≥ 3 jets	≥ 2 jets
b-tagging requirements	≥ 1 tight b-tagged jet or ≥ 2 loose b-tagged jets	
Missing transverse momentum	—	$L_D > 0$ (if Nj > 4) / 30 GeV (if SFOS lepton pair and Nj $\leq$ 3)/ 45 GeV (else)
Dilepton invariant mass	$m_{\ell\ell} > 12$ GeV	

TABLE 6.6: Event selections applied in the $1\ell + 2\tau_h$ and $2\ell + 2\tau_h$ categories. The p_T thresholds applied to the leading and subleading leptons are separated by slashes. The symbol "—" indicates that no requirement is applied and Nj stands for number of jets.

Selection step	$2\ell os + 1\tau_h$	$4\ell + 0\tau_h$
Trigger	Single- and double-lepton triggers	Single-, double- and triple-lepton triggers
Lepton p_T	$p_T > 25$ / 15 GeV (e) or 10 GeV (μ)	$p_T > 25$ / 15 / 15 / 10 GeV
Lepton η	$|\eta| < 2.5$ (e) or 2.4 (μ)	
τ_h p_T	$p_T > 20$ GeV	—
τ_h—η	$|\eta| < 2.3$	—
τ_h identification	tight	—
Charge requirements	$\sum_\ell q = 0$ and $\sum_{\ell,\tau_h} q = \pm 1$	$\sum_\ell q = 0$
Multiplicity of central jets	≥ 3 jets	≥ 2 jets
b-tagging requirements	≥ 1 tight b-tagged jet or ≥ 2 loose b-tagged jets	
Missing transverse momentum	$L_D > 30$ GeV‡	$L_D > 0$ (if Nj > 4) / 30 GeV (if SFOS lepton pair and Nj $\leq$ 3)/ 45 GeV (else)
Dilepton invariant mass	$m_{\ell\ell} > 12$ GeV	$|m_{\ell\ell} - m_Z| > 10$ GeV§ and $m_{\ell\ell} > 12$ GeV
Four-lepton invariant mass	—	$m_{4\ell} > 140$ GeV¶

TABLE 6.7: Event selections applied in the $2\ell os + 1\tau_h$ and $4\ell + 0\tau_h$ categories. The p_T thresholds applied to the first, second, third and fourth leptons are separated by slashes. The symbol "—" indicates that no requirement is applied and Nj stands for number of jets.
$\ddagger$ Criteria only applied if the event contains two electrons.
$\S$ Applied to OSSF lepton pairs.
$\P$ The event contains two SFOS pairs of leptons passing the loose lepton selection criteria.

signal against background discrimination. Multiclass NN are used in the $2\ell ss + 0\tau_h$, $3\ell + 0\tau_h$ and $2\ell ss + 1\tau_h$ categories, while in the rest of signal regions BDTs are used. Besides, the control regions defined in Section 6.3 are also included in the fit. A
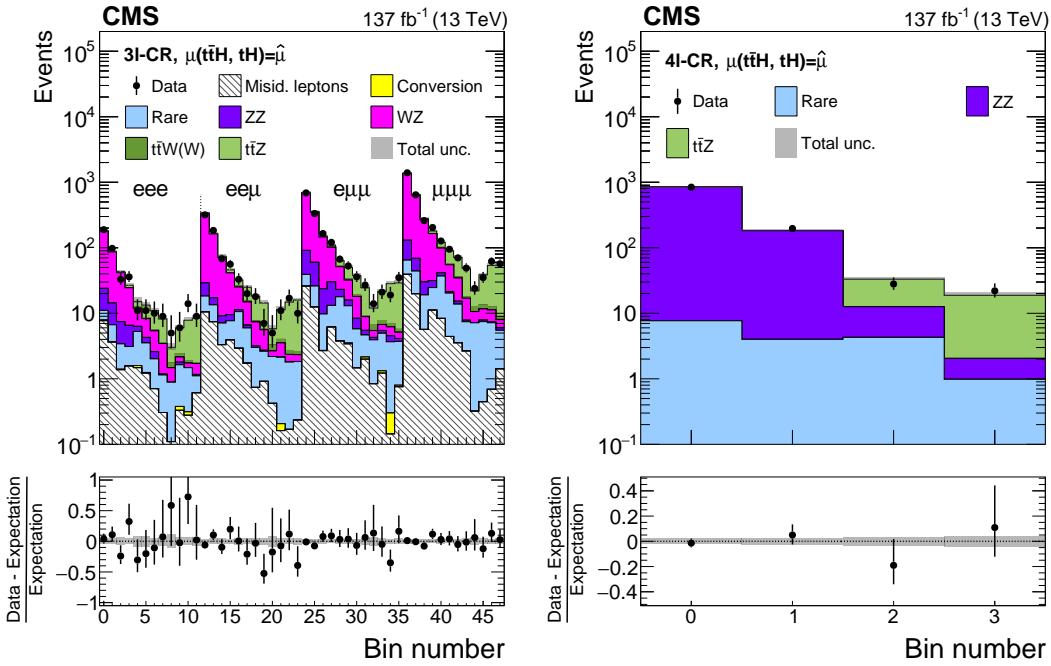
FIGURE 6.2: Comparison of the number of observed (points) and predicted (coloureded histograms) events in the three lepton (left) and four lepton (right) control regions. The predictions are shown after the fit to data. The vertical bars on the points represent the statistical uncertainties in the data, and the hatched bands the systematic uncertainty in the predictions. In the lower panels, the event yields in data are compared to the predictions.

graphic summary of the categorization of events entering the fit is given in Fig. 6.3.

The aforementioned MVA discriminants where developed by the IHEP and NICPB groups; for this reason in this section only a brief description of the most important items such as the input variables are given. Both types of discriminants take as input two composed variables, which are themselves based on MVA techniques: the hadronic top tagger and the Higgs jet tagger. The former is designed to find triplets of jets coming from the top quark decay while the latter allows to identify the jet produced in the Higgs boson decay in $2\ell ss + 0\tau_h$ category.

The hadronic top tagger is a BDT trained on $t\bar{t}$ simulated events, where combinations of three jets coming from the top quark decay are used as signal, and other combinations of three jets are considered as background. The BDT uses 16 input variables related to the jet properties, such as the b-tagging score, the likelihood for a jet to be quark- or gluon-induced, ΔR between the jets, and masses and p_T of different combinations of jets. In each event, all jets are considered and the triplets with highest and second-highest score in the discriminant are selected.
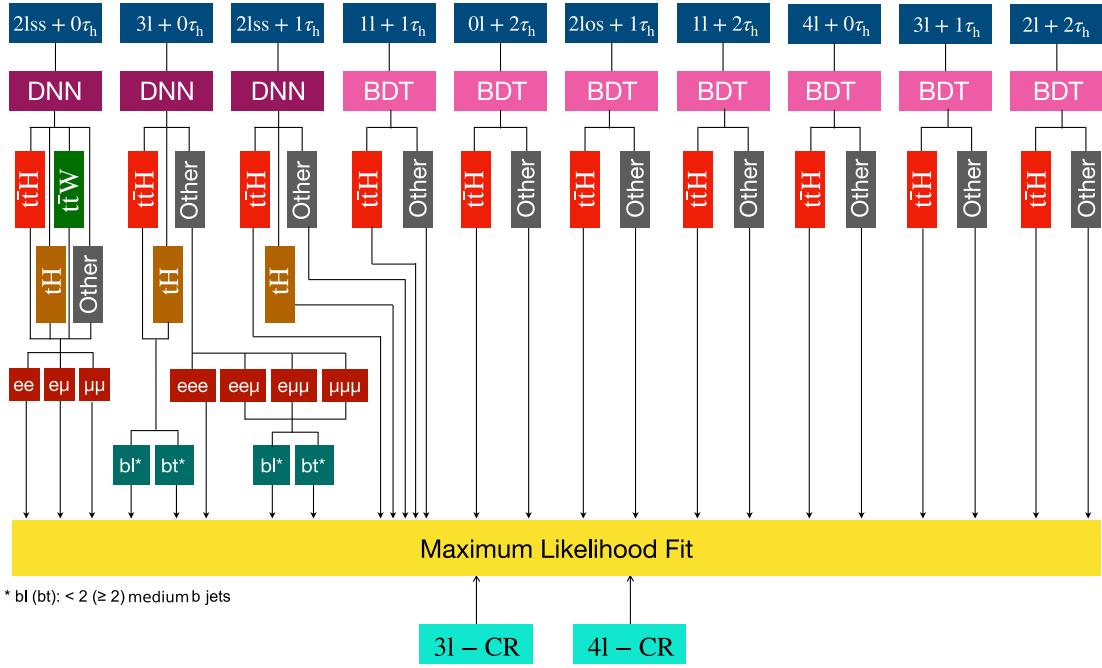
FIGURE 6.3: Categorization strategy used for the signal extraction, making use of MVA-based algorithms, jet and b tag multiplicities and flavour of the leptons. The ten signal regions with their corresponding subcategorization are used along with two control regions in the fit.

The Higgs jet tagger is a BDT-based discriminant trained in tt̄H and tt̄W simulated events which are required to have two leptons with same-sign of the charge and at least four jets. Events where the $H \to \ell\nu qq'$ are considered as signal, and the rest as background. Five input variables are used: the p_T, b-tagging score and quark-gluon likelihood of the jets, and the ΔR between the jets and any lepton passing the fakeable selection. All jets in the event not selected by the hadronic top tagger are considered, and the one with highest score is taken.

6.4.1   NN multiclassifiers

A NN discriminant is defined and optimize for each of the $2\ell ss + 0\tau_h$, $3\ell + 0\tau_h$ and $2\ell ss + 1\tau_h$ signal regions. NNs are trained on simulated samples of signal processes (tt̄H, tHq and tHW), the main irreducible backgrounds (tt̄Z, tt̄W an diboson) and tt̄ samples to model the nonprompt lepton background. A different set of events is used to train the discriminant and to perform the signal extraction.

In each of the categories, around 40 input variables are used; they include kinematic information of the leptons, τ_h, and jets, b-tagging information of the jets, angular distances between the objects in the event, missing transverse energy and the information

from the hadronic top tagger and Higgs boson jet tagger. The variables used on each category are detailed in Table 6.8.

The NN defined in the $2\ell ss + 0\tau_h$ has four output nodes, two targeting the t̄tH and tH signals, another one dedicated to t̄tW background and the fourth one dedicated to nonreducible backgrounds, labeled as *rest* node. For the $3\ell + 0\tau_h$ and $2\ell ss + 1\tau_h$ categories, three output nodes are used, two targeting each of the signals and a third one dedicated to all the backgrounds.

Events in each category are therefore classified according to the highest score in the different nodes of the NN. In the case of the $2\ell ss + 0\tau_h$ category, all nodes are further subcategorized depending on the flavour of the leptons. Similarly, in the $3\ell + 0\tau_h$ category, events in the t̄tH or tH node are further categorized depending on whether there is a medium b-tagged jet present in the event. Additionally, on the background node events are also classified depending on the flavour of the leptons.

6.4.2   BDT-based discriminants

Events in the signal regions not mentioned in the previous subsection are classified using BDT-based discriminants. This approach is more convenient, as these categories are not sensitive to tH and the number of signal events is limited. The BDT provides a binary classification that labels the event as signal (t̄tH and tH) or background.

The simulation samples use to model the background during the training are t̄tW, t̄tZ and t̄t samples; in the $1\ell + 1\tau_h$ and $0\ell + 2\tau_h$ a Drell-Yann sample is also used. Input variables used are similar to the one used to train the NNs. In categories with $0\ell + 2\tau_h$ and $1\ell + 1\tau_h$, the invariant mass of the leading lepton (or τ_h) and the (second) τ_h is reconstructed using the SVFit algorithm [165]. This algorithm provides the likelihood for the reconstructed visible p_T of the τ_h decays and p_T^{miss} on the event to be compatible with the Higgs boson mass. In addition, in those categories, a second hadronic top is constructed using the jets not tagged by the first hadronic tagger. On several categories, the decay angle of the two τ_h's (in the Higgs boson rest frame) is considered, the variable is denoted as $\cos\theta^*$. A summary of the input variables used in each category is given in Table 6.9.

6.5   Results: measurement of t̄tH and tH cross sections

In this section the observed number of events in each signal region used for the analysis are given with its comparison to predictions. Moreover, the values obtained for

Input variable	$2\ell ss + 0\tau_h$	$2\ell ss + 1\tau_h$	$3\ell + 0\tau_h$
Number of electrons in the category definition	✓	✓	✓
Cone-p_T of leptons	✓	✓	✓
η of leptons	✓	✓	✓
ϕ of leptons	✓	✓	✓
Transverse mass of leptons	✓	✓	✓
Leading lepton charge	✓	—	—
Presence of an OSSF lepton pair	—	—	✓
p_T of τ_h	—	✓	—
η of τ_h	—	✓	—
ϕ of τ_h	—	✓	—
Transverse mass of $\ell(+\tau_h)$ system	—	(✓)	✓
Invariant mass lepton+τ_h	—	✓	—
Charge sum of leptons(+τ_h)	—	(✓)	✓
Maximum lepton η	✓	—	—
Minimum ΔR leptons(+τ_h)	—	✓	✓
Minimum ΔR leptons+jets	✓	✓	✓
Minimum ΔR τ_h +jets	—	✓	—
Number of jets	✓	✓	✓
p_T of the jets	✓	✓	✓
η of the jets	✓	✓	✓
ϕ of the jets	✓	✓	✓
Jets average ΔR	✓	✓	✓
Number of forward jets	✓	✓	✓
p_T of leading forward jet	✓	✓	✓
$|\eta|$ of leading forward jet	✓	✓	✓
Min $\Delta\eta$ between leading forward jet and central jets	—	✓	✓
Number of loose b-tagged jets	✓	✓	✓
Number of medium b-tagged jets	✓	✓	✓
Invariant mass of loose b-tagged jets	—	✓	✓
Invariant mass of medium b-tagged jets	✓	—	—
L_D	✓	✓	✓
Hadronic top tagger	✓	✓	✓
Hadronic top p_T	✓	✓	✓
Higgs jet tagger	✓	—	—
Number of variables	36	41	37

TABLE 6.8: Summary of input variables to the NN discriminants defined for the $2\ell ss + 0\tau_h$, $2\ell ss + 1\tau_h$ and $3\ell + 0\tau_h$ regions. Variables related to leptons and jets such as lepton cone-p_T are take as many times as number of the corresponding object in the selection.

$t\bar{t}H$ and tH signal strengths are provided. In the maximum likelihood fit performed, the signal strengths for $t\bar{t}H$ and tH, the nuisance parameters that model the systematic uncertainties and the rates for $t\bar{t}W$, $t\bar{t}Z$, WZ and ZZ processes are left unconstrained. The rate of $t\bar{t}WW$ is considered to scale with the same factor as $t\bar{t}W$. Figures 6.4 to

Input variable	$0\ell+2\tau_h$	$1\ell+1\tau_h$	$1\ell+2\tau_h$	$2\ell os+1\tau_h$	$2\ell+2\tau_h$	$3\ell+1\tau_h$	$4\ell+0\tau_h$
Cone-p_T of lepton(s)	—	✓	✓	✓	—	✓	✓
p_T of τ_h('s)	✓	✓	✓	✓	✓	✓	—
Transverse mass of leptons	—	✓	✓	✓	—	—	—
Transverse mass of τ_h('s)	✓	✓	—	—	—	—	—
Transverse mass of lepton+τ_h's system	—	—	✓	—	—	—	—
$\tau_h\,\tau_h\,(\tau_h\,\ell)$ visible mass	✓	✓	✓	✓	✓	✓	—
$\tau_h\,\tau_h\,(\tau_h\,\ell)$ SVFit mass	✓	✓	—	—	—	—	—
Charge sum of lepton and τ_h	—	✓	—	—	—	—	—
Maximum η of leptons / τ_h('s)	✓	✓	✓	✓	—	—	—
ΔR between leptons	—	—	—	✓	—	—	—
ΔR between τ_h's	✓	—	✓	—	—	—	—
ΔR between leptons and τ_h	—	✓	—	—	—	—	—
ΔR between OS leptons and τ_h	—	—	✓	✓	—	—	—
ΔR between SS leptons and τ_h	—	—	✓	✓	—	—	—
Minimum ΔR between leptons and τ_h('s)	✓	—	—	✓	—	—	—
Minimum ΔR between leptons and jets	—	✓	✓	✓	✓	—	—
Minimum ΔR between τ_h('s) and jets	✓	—	—	✓	—	—	—
$\cos\theta^*$ of $\tau_h\,\tau_h\,(\tau_h\,\ell)$	✓	✓	✓	—	✓	—	—
Number of jets	—	✓	—	—	—	—	—
Average ΔR between jets	✓	✓	✓	✓	✓	—	—
Number of loose b-tagged jets	—	✓	—	—	—	—	—
Number of medium b-tagged jets	✓	—	—	—	—	—	—
Invariant mass of loose b-tagged jets	✓	✓	✓	✓	✓	—	—
L_D	✓	✓	✓	✓	✓	✓	✓
Presence of an OSSF lepton pair	—	—	—	—	—	✓	✓
Minimum invariant mass of loose lepton pairs	—	—	—	—	—	✓	✓
Hadronic top tagger	✓	✓	✓	✓	—	—	—
Second hadronic top tagger	—	✓	—	—	—	—	—
Hadronic top p_T	✓	—	✓	✓	—	—	—
Second hadronic top p_T	✓	—	—	—	—	—	—
Number of variables	18	18	17	18	9	9	7

TABLE 6.9: Summary of input variables to the BDT discriminant defined for the $0\ell+2\tau_h$, $1\ell+1\tau_h$, $1\ell+2\tau_h$, $2\ell os+1\tau_h$, $2\ell+2\tau_h$, $3\ell+1\tau_h$ and $4\ell+0\tau_h$ regions. Variables related to leptons and jets such as lepton cone-p_T are taken as many times as the number of the corresponding object in the selection.

6.7 show the number of observed events in all signal regions. The signal and background predictions are scaled by the values of the parameters of interest and of the nuisance parameters obtained from the ML fit, i.e. they are "postfit". In the figures the background contribution from nonprompt leptons is labeled as *misId. leptons*, the contribution from the charge misidentification background is labeled as *Flips*. Besides, low cross section processes with a marginal contribution are depicted under the label *Rare*, this includes tW and tZ production, the production of *ss* W boson pairs, triboson, and tt̄tt̄ production. Good agreement between observation and prediction is depicted in all signal regions.

The number of events measured in each of the categories and the signal and background prediction in each of them is given in Appendix B.

The production rate for tt̄H process is found to be $\mu_{tt̄H} = 0.92 \pm 0.19\,\text{(stat)}^{+0.17}_{-0.13}\,\text{(syst)}$ times the SM expectation. In terms of cross section, it is equivalent to a tt̄H production
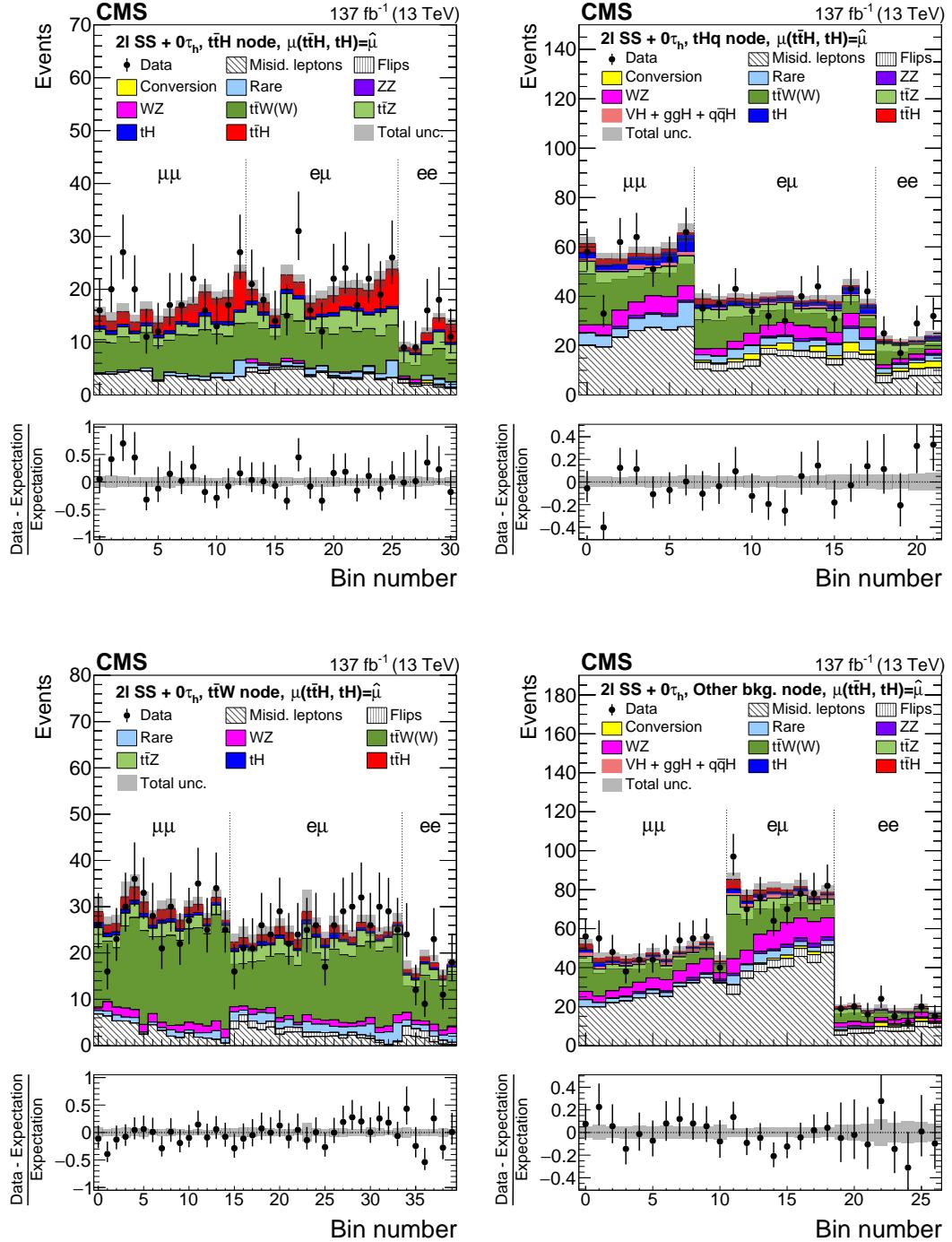
FIGURE 6.4: Comparison of the number of observed (points) and predicted (coloureded histograms) in the $2\ell ss + 0\tau_h$ signal region, for events with highest NN score in the tt̄H output node (upper left), tH output node (upper right), tt̄W output node (lower left) and other output node (lower right). The predictions are shown after the fit to data. The vertical bars on the points represent the statistical uncertainties in the data, and the hatched bands the systematic uncertainty in the predictions. In the lower panels, the event yields in data are compared to the predictions.
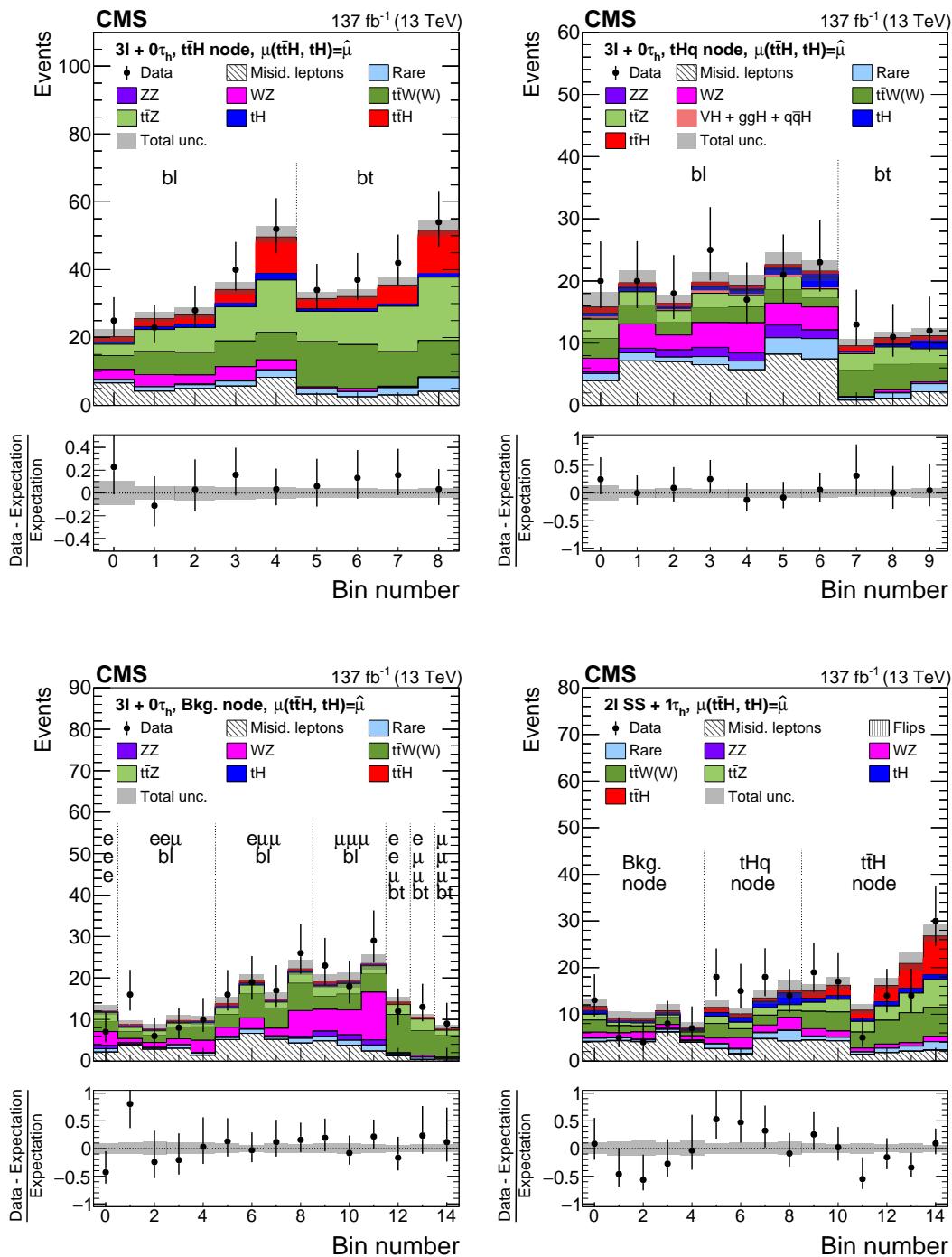
FIGURE 6.5: Comparison of the number of observed and predicted events in the $3\ell + 0\tau_\mathrm{h}$ signal region for events with highest NN score in the tt̄H output node (upper left), tH output node (upper right), and other output node (lower left), and events in the $2\ell ss + 1\tau_\mathrm{h}$ signal region categorized according to the NN discriminant (lower right). The predictions are shown after the fit to data. The vertical bars on the points represent the statistical uncertainties in the data, and the hatched bands the systematic uncertainty in the predictions. In the lower panels, the event yields in data are compared to the predictions.
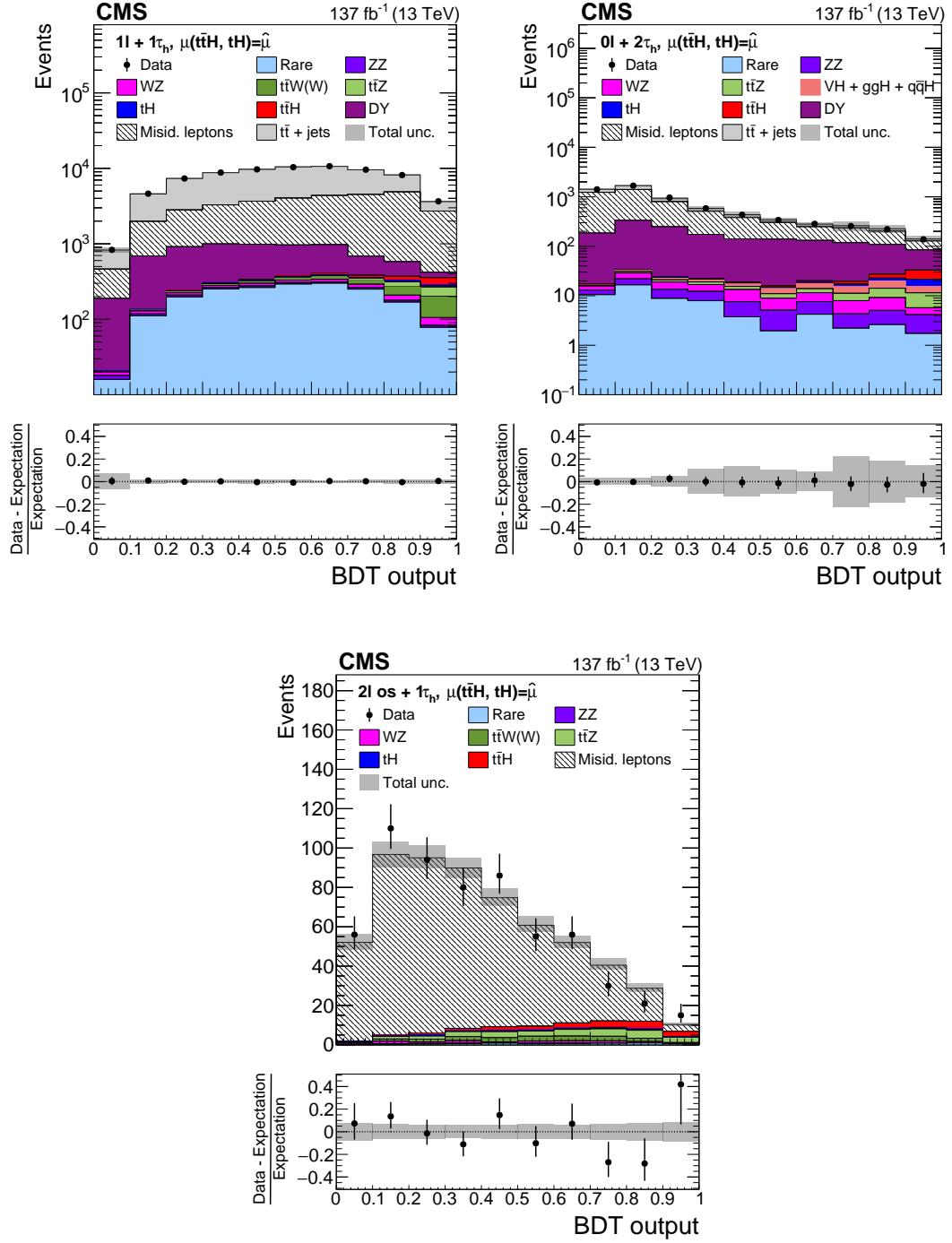
FIGURE 6.6: Comparison of the number of observed and predicted events in the signal regions: $1\ell + 1\tau_h$ (upper left), $0\ell + 2\tau_h$ (upper right), $2\ell os + 1\tau_h$ (lower). The predictions are shown after the fit to data. The vertical bars on the points represent the statistical uncertainties in the data, and the hatched bands the systematic uncertainty in the predictions. In the lower panels, the event yields in data are compared to the predictions.
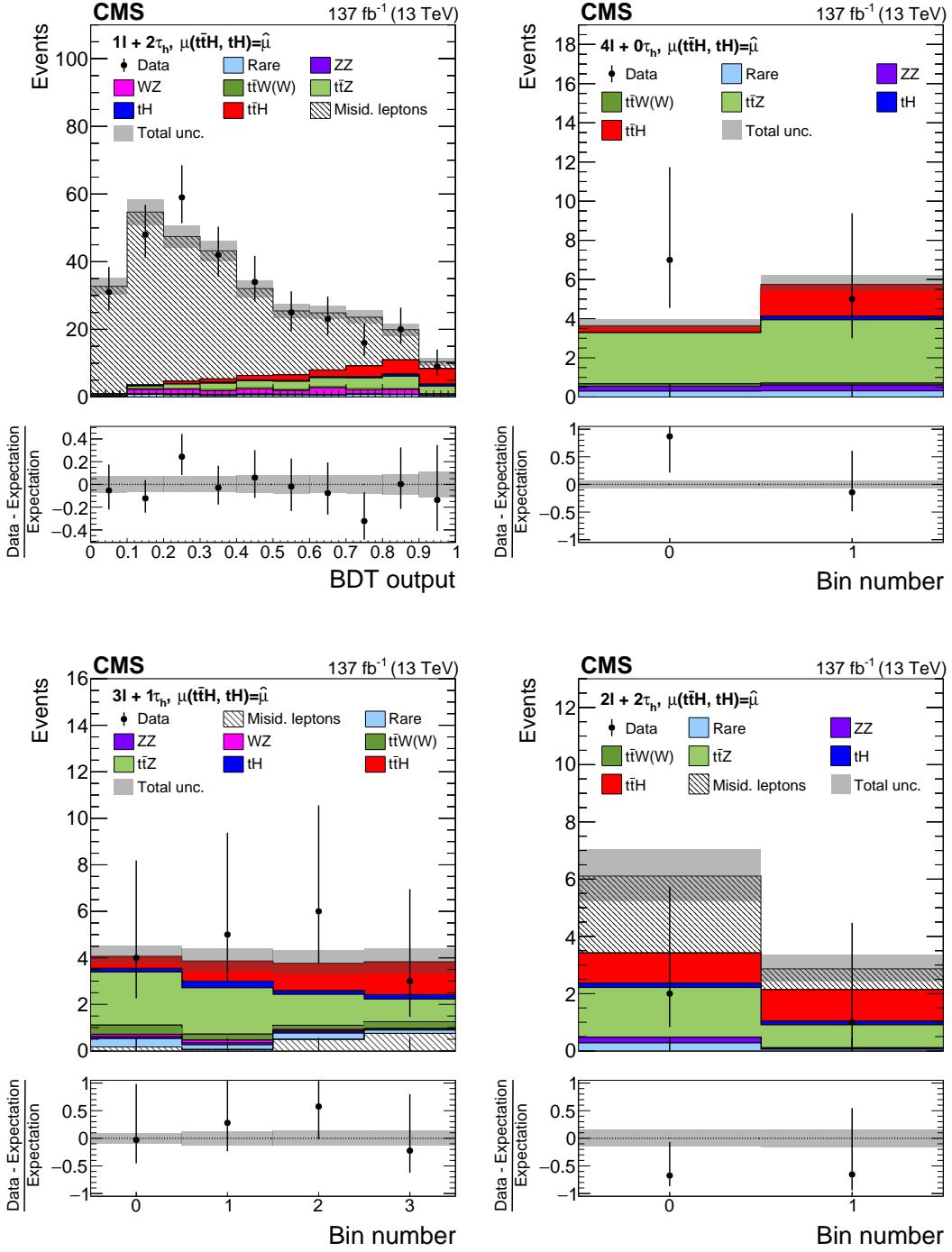
FIGURE 6.7: Comparison of the number of observed and predicted events in the signal regions: $1\ell + 2\tau_h$ (upper left), $4\ell + 0\tau_h$ (upper Right), $3\ell + 1\tau_h$ (lower left) and $2\ell + 2\tau_h$ (lower right). The predictions are shown after the fit to data. The vertical bars on the points represent the statistical uncertainties in the data, and the hatched bands the systematic uncertainty in the predictions. In the lower panels, the event yields in data are compared to the predictions.
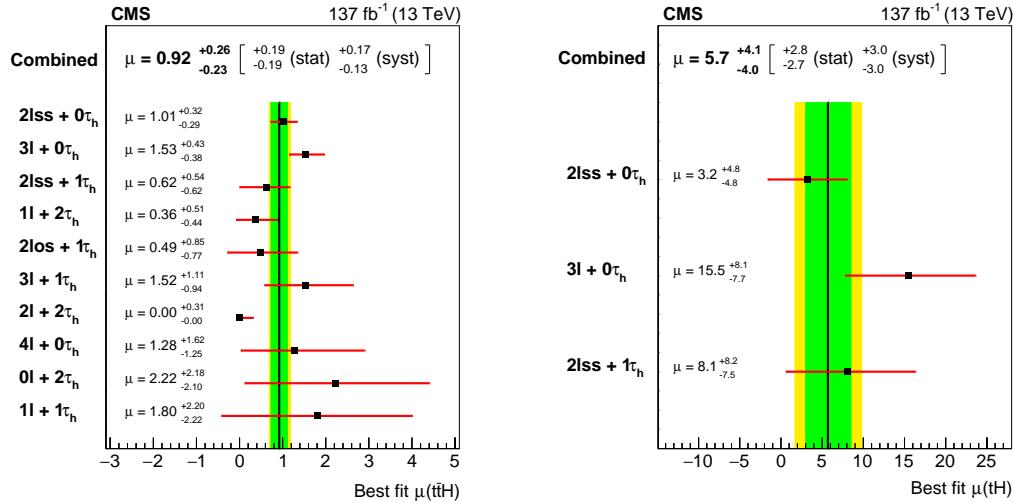
FIGURE 6.8: Signal production rates (μ), for tt̄H (left) and tH (right), in units of their rate of production expected in the SM. The result is shown in each of the ten categories individually and for the combination. The central value of the signal strength in the $2\ell + 2\tau_h$ is constrained to be greater than zero.

cross section of:

$$\sigma_{\text{tt̄H}} = 466 \pm 96 \text{ (stat)}^{+70}_{-56} \text{ (syst) fb.} \tag{6.1}$$

The signal rate for tH is measured to be $\mu_{\text{tH}} = 5.7 \pm 2.7$ (stat) ± 3.0 (syst) times the SM expectation. The corresponding cross section for tH production is:

$$\sigma_{\text{tH}} = 510 \pm 200(\text{stat}) \pm 220 \text{ (syst) fb.} \tag{6.2}$$

Assuming that the tH process has the same production rate as expected in the SM, the observed (expected) significance of tt̄H signal amounts to 4.7 (5.2) s.d. Analogously, the significance of tH signal is 1.4 (0.3) s.d. when assuming the tt̄H process has the SM production rate.

Figure 6.8 summarizes the values obtained for both signal strengths in each of the signal regions. For tt̄H, the result is shown in the ten signal regions, while for tH the result is given in the three most sensitive signal regions. Good agreement is shown among all the measurements and the combined one. The error band in each channel indicates how sensitive it is to the measurement.

The rate for tt̄Z and tt̄W processes is also extracted from the fit. For tt̄Z, the production rate is $\mu_{\text{tt̄Z}} = 1.03 \pm 0.14$ (stat + syst) times the SM expectation, hence it is in good agreement with the SM prediction. The tt̄W production rate is found to be $\mu_{\text{tt̄W}} = 1.43 \pm 0.21$ (stat + syst) times their SM expectation. Hence, $\mu_{\text{tt̄W}}$ is above the SM

prediction, the sensitivity to this background arise from the node dedicated to tt̄W in the $2\ell ss + 1\tau_{\mathrm{h}}$, which provides a region very pure in tt̄W events. The discrepancy is in agreement with previous CMS and ATLAS results [12, 14], and motivates the need for a dedicated measurement of the tt̄W inclusive cross section.

Figure 6.9 illustrates the two-dimensional confidence regions showing the simultaneous measurement of the signal rates ($\mu_{\mathrm{tt̄H}}$ and μ_{tH}); the simultaneous measurement of the tt̄H signal rate alongside tt̄W and tt̄Z production rates, as well as the simultaneous extraction of tt̄W and tt̄Z rates. The 68% and 95% confidence level contours are evaluated. In each of the fits the parameters not shown in the plot are profiled. From the confidence regions one can infer the correlation between the parameters of interest of the fit. Such correlations are found to be moderate, showing the discriminating power achieved by the NNs. Most of the parameters are in good agreement with the SM within the 68% CL. The tt̄W rate, shows a small tension but remains in agreement with the SM at the 95% CL

Systematic and statistical uncertainties

The impact of each source of systematic uncertainty and the statistical uncertainty in the measurements of tt̄H and tH production rates is given in Table 6.10. The uncertainties with a larger contribution are the statistical uncertainty of the observed data; the uncertainty related to the τ_{h} reconstruction efficiency; the uncertainties related to the estimation of the misidentified leptons and flips backgrounds; and the theoretical uncertainties, which affect the normalization and the shape of the discriminating observables for the tt̄H and tH signals as well as for the main irreducible backgrounds.

6.6   Coupling interpretation

So far in this chapter it has been assumed that the Yukawa coupling of the top quark to the Higgs boson, and the coupling of vector bosons to the Higgs boson, were consistent with the SM. In this section this assumption is dropped, and the production rates, $\mu_{\mathrm{tt̄H}}$ and μ_{tH}, are parameterized as a function of the coupling modifiers. These modifiers are defined as the ratio of the coupling with respect to its SM expected value: $\kappa_t = y_t/y_t^{SM}$ and $\kappa_V = \lambda_V/\lambda_V^{SM}$.

The tt̄H production is affected only by κ_t; and, at leading order, no kinematic dependence is present, hence only the scaling of tt̄H production rate with κ_t^2 is considered. Instead, tH production is sensitive to both κ_t and κ_V, and an effect in the kinematic distributions is considered besides the scaling on the production rate. The changes in the kinematic properties of tH events will affect the acceptance of such events and the
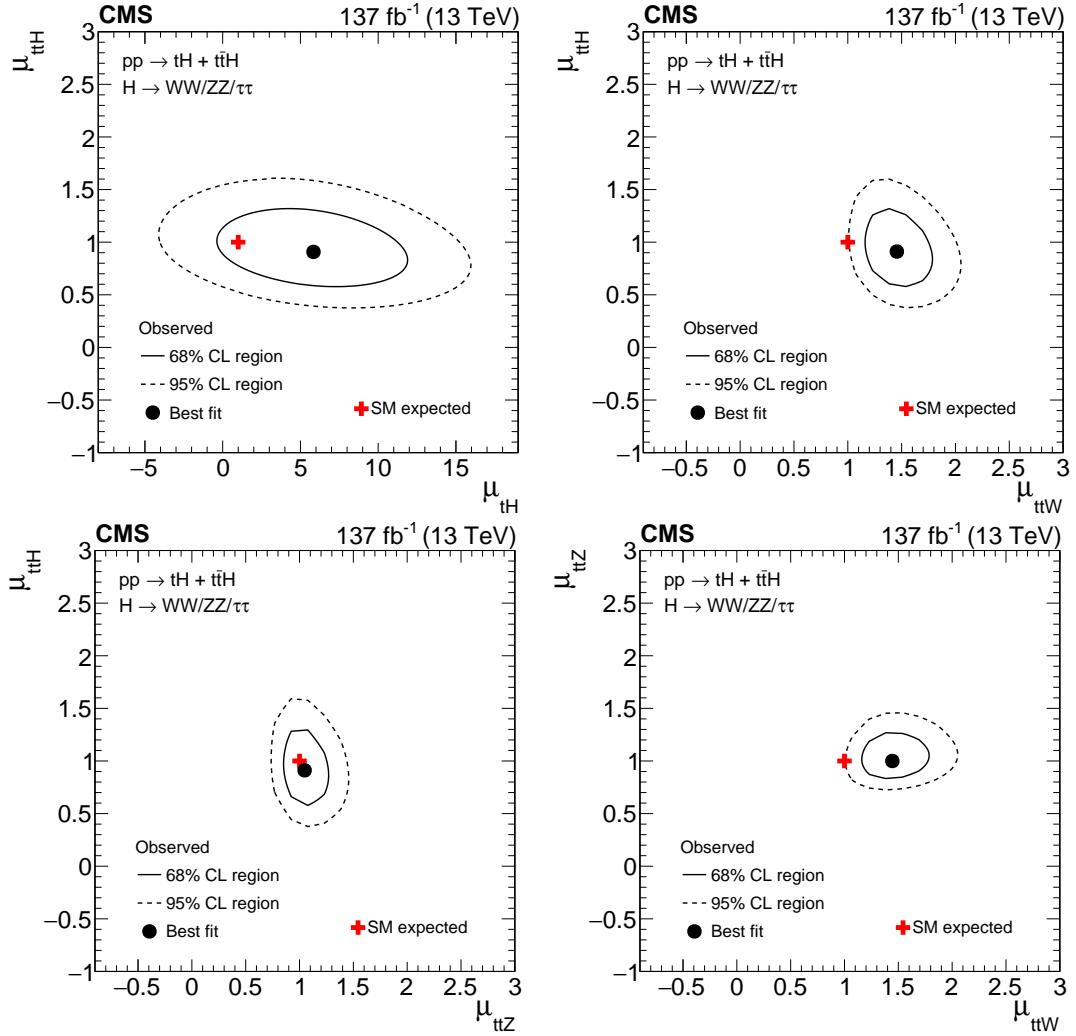
FIGURE 6.9: Two-dimensional contours of the likelihood function, as a function of the production rates of tt̄H and tH signals and tt̄Z and tt̄W backgrounds. The production rates not shown in the axes of the plots are profiled.

efficiency of the analysis selection, this effect is depicted in Fig. 6.10. Where $\kappa_t / \kappa_V = 1$ represents the SM scenario and is shown to have the lowest value of the acceptance $\times$ efficiency.

The definition of the likelihood function given in the previous section is modified. In this case, the signal strengths are fixed to their SM value, and tt̄H and tH yields are scaled according to their κ_t and κ_V dependence. Then, the likelihood function is evaluated using several hypothesis for κ_t and κ_V, while the rate modifiers for tt̄W, tt̄Z, WZ and ZZ, as well as the nuisance parameters, are profiled. In this model, the coupling of the Z and W to the Higgs boson is assumed to be identical. Additionally, other Higgs boson couplings are assumed to take its SM value. The branching fractions of the Higgs boson decay modes are dependent on the couplings, and they are

Source	$\Delta\mu_{\text{tt̄H}}/\mu_{\text{tt̄H}}$ [%]	$\Delta\mu_{\text{tH}}/\mu_{\text{tH}}$ [%]
Trigger efficiency	2.3	8.1
e, μ reconstruction and identification efficiency	2.9	7.1
τ_{h} identification efficiency	4.6	9.1
b-tagging efficiency and mistag rate	3.6	13.6
Misidentified leptons and flips	6.0	36.8
Jet energy scale and resolution	3.4	8.3
MC sample and sideband statistical uncertainty	7.1	27.2
Theory-related sources affecting acceptance and shape of distributions	4.6	18.2
Normalization of MC-estimated processes	13.3	12.3
Integrated luminosity	2.2	4.6
Total systematic uncertainty	18.5	52.6
Statistical uncertainty	20.9	48.0

TABLE 6.10: Sources of systematic and statistical uncertainties in the predicted signal and background event yields with their impact on the measured tt̄H and tH signal strengths, estimated after the fit to the data.
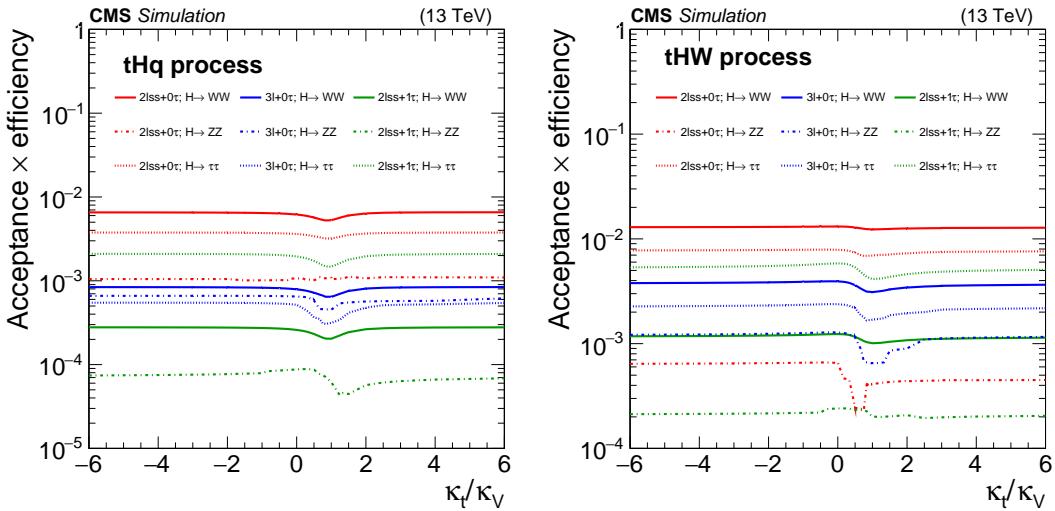
FIGURE 6.10: Probability for tH events produced by the tHq (left) and tHW (right) production modes to pass the event selection criteria for each of the signal regions sensitive to tH as a function of the ratio κ_t/κ_V.

modified accordingly in the fit. The relevant κ values are κ_τ, κ_V, κ_μ and $\kappa_{Z\gamma}$. The contribution from H $\to \mu\mu$ and H $\to Z\gamma$ is expected to be negligible, hence we consider $\kappa_{Z\gamma} = \kappa_\mu = 1$. Moreover, light fermions are treated universally with the parameter $\kappa_f = \kappa_\mu = \kappa_\tau$.

The likelihood function for different values of κ_t, profiling κ_V is shown in Fig. 6.11 (left). Limits on κ_t are derived from the crossing of the $-2\Delta\mathcal{L}$ with the χ^2 distribution quantiles. The parameter κ_t is found to be within the intervals: $-0.9 < \kappa_t < -0.7$ and
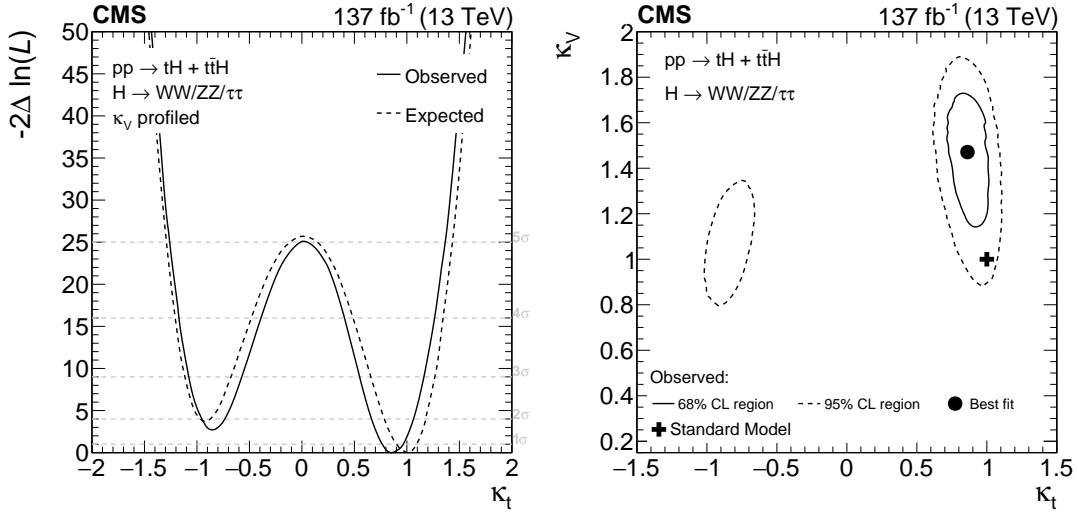
FIGURE 6.11: Likelihood function as a function of κ_t, profiling over κ_V (left), and as a function of both κ_t and κ_V (right).

$0.7 < \kappa_t < 1.1$ at 95% CL Both the SM and the ITC scenario are in agreement with the observation at 95% CL Besides, the two-dimensional contours at 68% and 95% CL as a function of κ_t and κ_V are depicted in Fig. 6.11 (right). Good agreement with the SM is shown and the ITC is excluded at 68% CL

6.7 CP violation in the top-Higgs coupling

The interpretation of the measurement in terms of the top-Higgs coupling can be extended to search for CP violation. To do so, the most sensitive categories of the analysis ($2\ell ss + 0\tau_h$, $2\ell ss + 1\tau_h$ and $3\ell + 0\tau_h$) are used.

As explained in Section 4.1.2, the Lagrangian for the Yukawa interaction can be written in terms of a CP-even and a CP-odd part. Different proportions of both CP scenarios would alter the normalizations of tt̄H and tH processes, as well as, introduce modifications on the distributions of kinematic observables. To exploit these kinematic differences on the observables between the CP-even and CP-odd scenarios, a dedicated MVA discriminant is trained in events selected by the tt̄H signal node of each of the NNs introduced in Section 6.4.1. These discriminants were implemented by the UCL, UZH and IHEP groups. I performed all other steps of the analysis.

6.7.1 CP-discrimination

Three BDTs are trained, one on each of the signal regions, using tt̄H simulated events that are selected as tt̄H-like by the NNs. The simulation used contains events with both

the SM and the *CP*-odd hypotheses for the coupling. Events used for the training and testing of the model are not used for the signal extraction.

Several input observables are considered, including basic kinematic properties of the objects in the final state, angular distances between the objects, missing transverse energy information, the score provided by the hadronic top tagger and the Higgs jet tagger, the transverse mass of each of the leptons, and the invariant mass of the reconstructed t$\bar{\text{t}}$H system, which is defined as:

$$M_{\text{t}\bar{\text{t}}\text{H}} = \| \sum_i p^{\ell_i} + \vec{p}_{\text{T}}^{\text{miss}} + \sum_{i \leq k} p^{\text{jet}_i} \| \tag{6.3}$$

where $\| \cdot \|$ denotes the norm in the Minkowski space and the second sum runs over the expected number of jets given by the t$\bar{\text{t}}$H subsequent decays in the final state corresponding to the category, i.e., $k = 6$ (4) in the final state with $2\ell ss + 0\tau_{\text{h}}$ ($2\ell ss + 1\tau_{\text{h}}$ and $3\ell + 0\tau_{\text{h}}$). If more than the expected number of jets are present the appropriate subset of jets is chosen ordered by highest p_{T}. The complete list of observables used in the training of each of the BDTs is given in Table 6.11.

The shape of two of the most discriminating input variables are shown, for the $2\ell ss + 0\tau_{\text{h}}$ and $3\ell + 0\tau_{\text{h}}$ categories, in Fig. 6.12. In addition, the shape of the output of each of the classifiers is shown in Fig. 6.13.

6.7.2   Signal extraction

A maximum likelihood fit is performed, using the three signal regions and including events from the two control regions as well. The categorization is very similar to the one applied in the previous sections, but events in the t$\bar{\text{t}}$H node are further classified depending on the score of the BDT-based algorithms used to distinguish the *CP* scenario. A graphic summary of the event categorization before the fit is provided in Fig. 6.14.

In this interpretation the likelihood fit is, again, redefined. In this case, two new parameters of interest are included: κ_t and $\tilde{\kappa}_t$, which are respectively the *CP*-even and *CP*-odd top-Higgs Yukawa coupling modifiers. From equation 4.3 it can be derived that κ_t is proportional to $\cos(\alpha)\kappa_{Htt}$ while $\tilde{\kappa}_t$ is proportional to $\sin(\alpha)\kappa_{Att}$.

The yields for t$\bar{\text{t}}$H can be parameterized as:

$$A \cdot \kappa_t^2 + B \cdot \tilde{\kappa}_t^2 \tag{6.4}$$

Variable description	$2\ell ss + 0\tau_h$	$2\ell ss + 1\tau_h$	$3\ell + 0\tau_h$
p_T of leading and subleading jet	—	—	✓
Cone-p_T of leptons	✓	✓	✓
p_T of τ_h	—	✓	—
η of leptons	✓	✓	—
η of τ_h	—	✓	—
ϕ of leptons	✓	✓	—
ϕ of τ_h	—	✓	—
Transverse mass of leptons	✓	—	—
ΔR of lepton i (i=1,2) to its closest jet	✓	✓	✓
Invariant mass of the reconstructed tt̄H system ($M_{tt̄H}$)	✓	✓	✓
$\Delta\eta$ of two jets with highest b score in the laboratory frame ($\Delta\eta_{B-B}$)	✓	✓	✓
$\Delta\eta$ of the two leptons in frame of two most-likely b jets	✓	✓	—
$\Delta\eta$ of two jets with highest b score in the dilepton system frame	✓	✓	—
$\Delta\eta$ of two jets with highest b score in the ℓ_1-ℓ_2 system frame	—	—	✓
$\Delta\eta$ of two jets with highest b score in the ℓ_1-ℓ_3 system frame	—	—	✓
$\Delta\phi$ of the two leptons in frame of two most-likely b jets	—	✓	—
$\Delta\phi$ of two jets with highest b score in the dilepton system frame	—	✓	—
Average ΔR among all jets	✓	✓	—
Jet multiplicity	✓	✓	—
p_T^{miss}	✓	✓	—
Azimuthal angle of $\vec{p}_T^{\,miss}$	✓	✓	—
Highest score of the hadronic top tagger	✓	✓	—
Higgs jet tagger	—	✓	—
Angle of tt̄ and H boson in tt̄H-system	—	✓	—
Angle between two t in tt̄-frame	—	✓	—
$\Delta R_{l_3-l_1} = \sqrt{(\eta_{\ell_3} - \eta_{\ell_1})^2 + (\phi_{\ell_3} - \phi_{\ell_1})^2}$	—	—	✓
$\Delta R_{l_1-l_2} = \sqrt{(\eta_{\ell_1} - \eta_{\ell_2})^2 + (\phi_{\ell_1} - \phi_{\ell_2})^2}$	—	—	✓
$\Delta R_{l_2-l_3} = \sqrt{(\eta_{\ell_2} - \eta_{\ell_3})^2 + (\phi_{\ell_2} - \phi_{\ell_3})^2}$	—	—	✓
$\eta_{jet1} - \eta_{jet2}$	—	—	✓
$p_T^{jet1} + p_T^{jet2} + p_T^{jet3} + p_T^{miss}$	—	—	✓
Total number of variables	19	25	16

TABLE 6.11: Input observables for each of the BDT-based discriminant aiming to distinguish tt̄H *CP*-even vs. tt̄H *CP*-odd like events. A check mark (✓) indicates the variable is used in a given final state, whereas a long dash (—) indicates the variable is not used in that final state.

While the yields for tH depend also on κ_V, and can be parameterized as:

$$A \cdot \kappa_t^2 + B \cdot \tilde{\kappa}_t^2 + C \cdot \kappa_V + D \cdot \kappa_t \kappa_V \tag{6.5}$$

This parametrization shows the need of classifying the signal events in two scenarios for tt̄H (pure *CP*-even, or SM; and pure *CP*-odd). For tH four scenarios are needed:

- Pure *CP*-even (SM): $\kappa_t = 1$, $\tilde{\kappa}_t = 0$, $\kappa_V = 1$.

- Pure *CP*-odd: $\kappa_t = 0$, $\tilde{\kappa}_t = 1$, $\kappa_V = 0$.

- $\kappa_t = 0$, $\tilde{\kappa}_t = 0$, $\kappa_V = 1$

- $\kappa_t = 1$, $\tilde{\kappa}_t = 0$, $\kappa_V = 0$
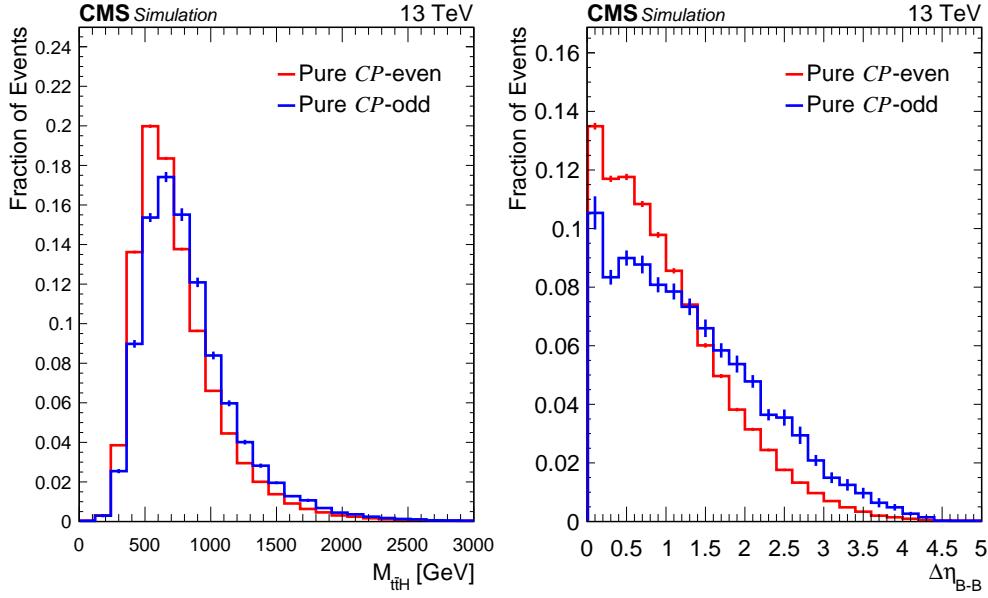
FIGURE 6.12: Two of the most important input variables to the BDT used for *CP* discrimination in the $2\ell ss + 0\tau_\text{h}$ (left) and $3\ell + 0\tau_\text{h}$ (right) signal regions. The reconstructed mass of the t$\bar{\text{t}}$H system is shown for the $2\ell ss + 0\tau_\text{h}$ region, and the $\Delta\eta$ of the two jets with highest b score in the laboratory frame is depicted for the $3\ell + 0\tau_\text{h}$. The vertical bars represent the statistical uncertainty originating from the limited amount of simulated events. When not visible, the bars are smaller than the line width.

Each signal component can be obtained with the dedicated signal simulations described in Section 4.5, and is scaled in the fit with the corresponding relevant κ modifier.

In this signal extraction process, other coupling modifiers different than κ_t and $\tilde{\kappa}_t$ are considered to have its SM value. This includes κ_V and, as a result, the relevant branching fractions are also fixed to the SM expectation value. Moreover, κ_t and $\tilde{\kappa}_t$, the nuisance parameters and the normalization of the leading nonreducible backgrounds are treated as free parameters in the fit.

An alternative parametrization is constructed using t$\bar{\text{t}}$H signal strength and the parameter $|f_{CP}^{\text{Htt}}|$:

$$|f_{CP}^{\text{Htt}}| = \frac{\tilde{\kappa}_t{}^2}{\kappa_t^2 + \tilde{\kappa}_t{}^2} \tag{6.6}$$

This parameterization allows to probe fractional *CP*-odd contributions.

Systematic uncertainties

The systematic uncertainties affecting the measurement are the ones summarised in Section 4.7. An extra uncertainty is added to take into account the fact that t$\bar{\text{t}}$H simulation used in these studies are LO. Hence, to cover for discrepancies between NLO
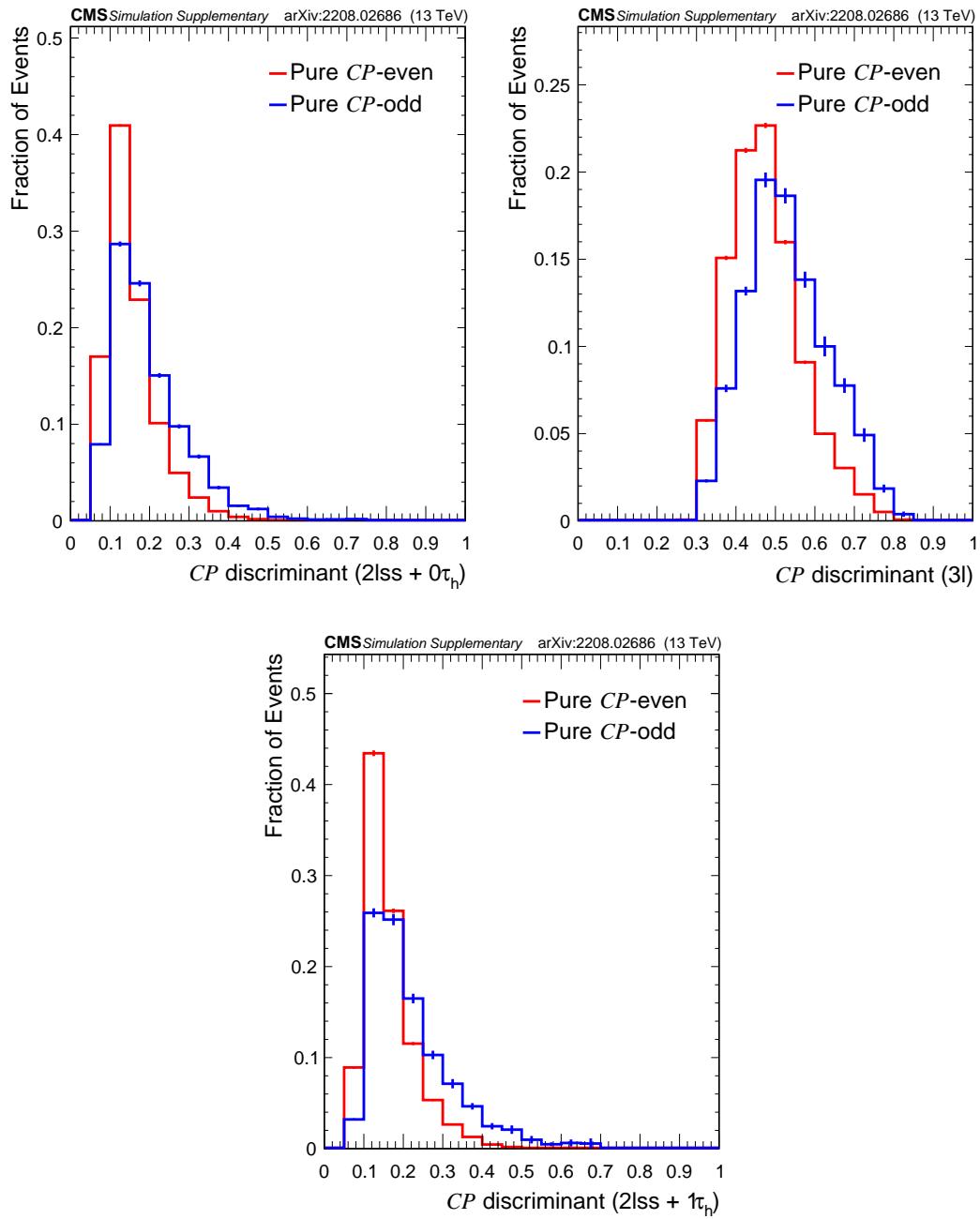
FIGURE 6.13: Output score of the BDT-based algorithm used for CP discrimination in $2\ell ss + 0\tau_{\text{h}}$ (upper left), $3\ell + 0\tau_{\text{h}}$ (upper right) and $2\ell ss + 1\tau_{\text{h}}$ (lower) categories. The vertical bars represent the statistical uncertainty originating from the limited amount of simulated events. When not visible, the bars are smaller than the line width.
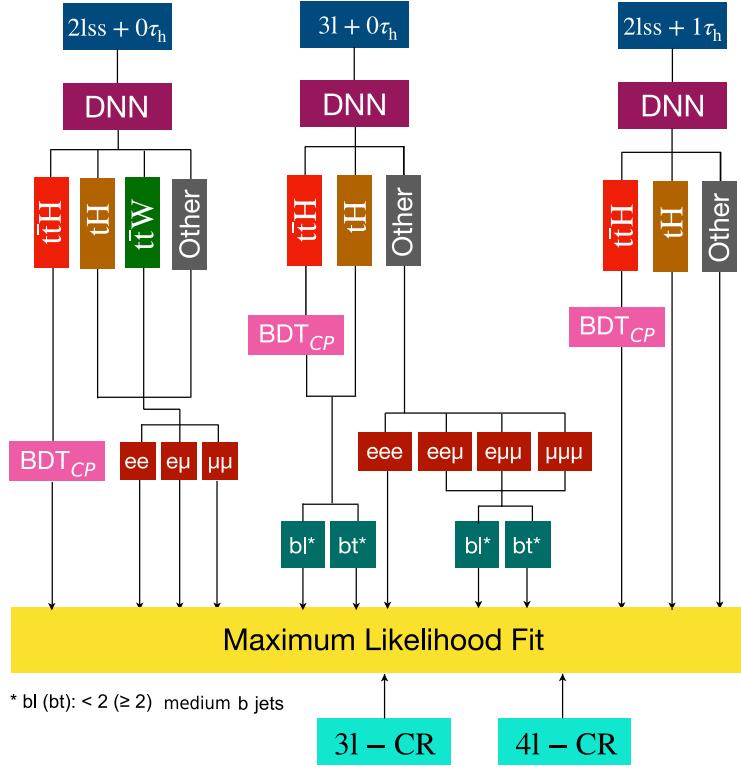
FIGURE 6.14: Categorization strategy used for the signal extraction in the *CP* interpretation context. It makes use of MVA-based algorithms, jet and b tag multiplicities and flavour of the leptons and the BDT-based *CP* discriminants.

and LO simulation, an uncertainty is constructed as the ratio of yields with LO and NLO samples.

6.7.3   Results

Figure 6.15 shows the observed and predicted events in each of the three tt̄H signal regions. Events are classified using the BDTs depending on its likelihood to be a *CP*-even or a *CP*-odd event. In the figures the contribution from the charge misidentification background is labeled as *Charge mism*, *Conv.* stands for conversions and low cross section processes with a marginal contribution are depicted under the label *Rare*, this includes tW and tZ production, the production of *ss* W boson pairs, triboson, and tt̄tt̄ production. Good agreement is shown after the fit to data is performed.

The two-dimensional likelihood function as a function of κ_t and $\tilde{\kappa}_t$ is depicted in Fig. 6.16. The corresponding confidence regions at 68% and 95% CL are established using the crossing of the $-2\Delta\mathcal{L}$ with the χ^2 distribution quantiles. The value on the κ_t and $\tilde{\kappa}_t$ phase space preferred by the data is in good agreement with the SM within the 68% CL The non-zero postfit value of $\tilde{\kappa}_t$ is driven by a small excess in the most *CP*-odd enriched regions of the $2\ell ss + 0\tau_h$ and $3\ell + 0\tau_h$ categories.

Parameter	68% CL	95% CL
	Expected	
κ_t	(0.87, 1.14)	(0.74, 1.27)
$\tilde{\kappa}_t$	(-0.71, 0.71)	(-1.01, 1.01)
	Observed	
κ_t	(0.89, 1.17)	(-1.09, -0.74) or (0.77, 1.3)
$\tilde{\kappa}_t$	(0.37, 1.16) or (-1.16, -0.37)	(-1.4, 1.4)

TABLE 6.12: One-dimensional confidence intervals at 68% and 95% CL for κ_t (fixing $\tilde{\kappa}_t$ to the SM) and $\tilde{\kappa}_t$ (fixing κ_t to the SM). The upper part of the table shows the expected limits while the lower part shows the observed limits.

One-dimensional confidence intervals can be established on κ_t, fixing $\tilde{\kappa}_t$ to the SM; analogously limits can be set on $\tilde{\kappa}_t$ when fixing κ_t to its SM value. Such limits are quoted on Table 6.12.

The results are also shown in terms of $|f_{CP}^{\mathrm{Htt}}|$: the expected and observed likelihood function with respect to $|f_{CP}^{\mathrm{Htt}}|$ is shown in Fig. 6.17, where $\mu_{t\bar{t}H}$ is profiled. The best fit value is found to be $|f_{CP}^{\mathrm{Htt}}| = 0.59$ with an interval of (0.24, 0.81) at 68% CL and the pure CP-odd hypothesis is excluded at 95 % CL

The precision in these results is limited by the size of the dataset. Additionally, the leading systematic uncertainties arise from theoretical uncertainties on t̄tH simulation and the uncertainty applied to cover NLO vs. LO differences in t̄tH predictions.

6.7.4   Combination with other Higgs decay modes

The results presented in the previous section are combined with other CMS t̄tH measurements using final states not considered in this analysis. In particular, the measurements of the CP structure of the Higgs boson coupling using t̄tH production mode is extended by considering $H \to \gamma\gamma$ [126] and $H \to ZZ \to 4\ell$ [166] final states.

Correlations between uncertainties are carefully considered. Nuisance parameters modelling the same source of uncertainty are taken as correlated, this includes uncertainties on: the luminosity, PU effects, p_T^{miss}, b-tagging efficiency and theoretical calculations on the signal rates and the backgrounds modelled with simulation. Uncertainties on the jet energy scale and resolution are treated as uncorrelated, as different components are taken in each of the analysis. Uncertainties on the lepton identification are taken as uncorrelated, following the fact that the dominant component of this uncertainty in the multilepton channel is due to the extrapolation of the efficiency measured in DY to a t̄t phase space.
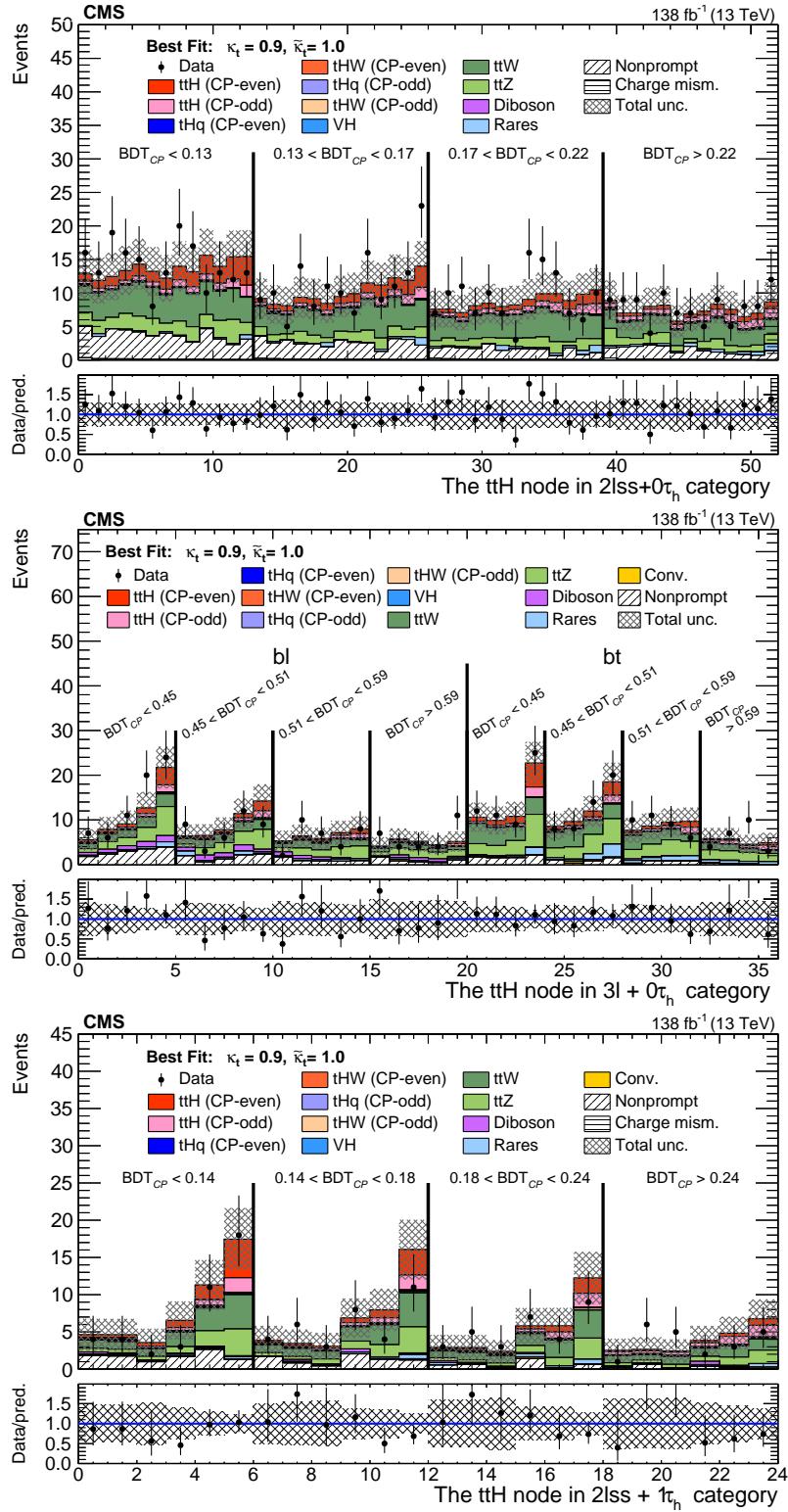
FIGURE 6.15: Comparison of the number of observed (points) and predicted (coloureded histograms) events in the t̄tH node, categorized depending on the score of the *CP* discriminant for the $2\ell ss + 0\tau_h$ (top), $3\ell + 0\tau_h$ (center) and $2\ell ss + 1\tau_h$ (bottom). For the $2\ell ss + 1\tau_h$ bl (bt) denotes events with less than (at least) two b-tagged jets. The t̄tH *CP*-even (red) and *CP*-odd (pink) predictions and the background predictions are shown after the fit to data. The vertical bars on the points represent the statistical uncertainties in the data, and the hatched bands the systematic uncertainty in the predictions. In the lower panels, the event yields in data are compared to the predictions.
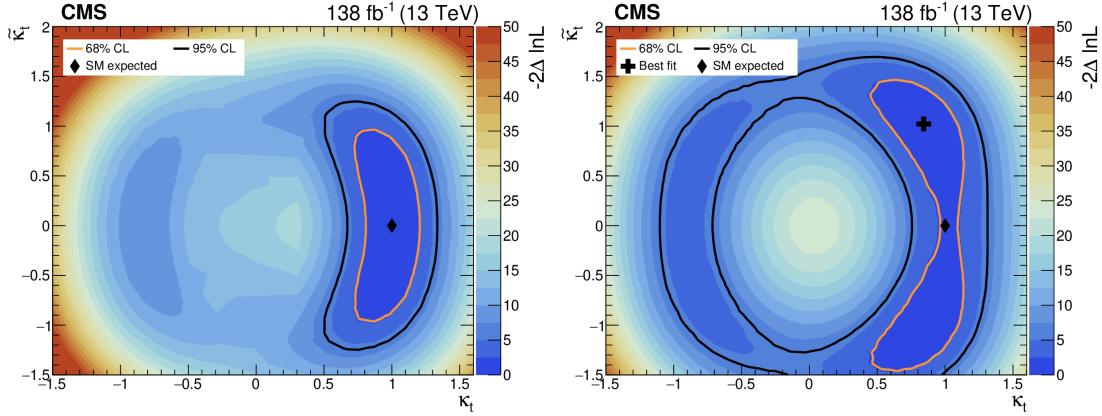
FIGURE 6.16: Likelihood scan as a function of κ_t and $\tilde{\kappa}_t$: expected limits (left) and observed limits (right). The black cross shows the best value for κ_t and $\tilde{\kappa}_t$ given by the fit. The black diamond shows the expected SM values for κ_t and $\tilde{\kappa}_t$. Both 68 and 95% CL limits are shown. κ_V and H branching fractions are kept to their SM values.
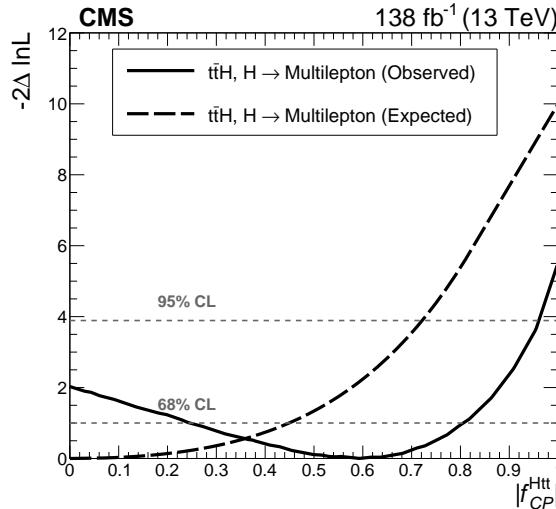
FIGURE 6.17: Likelihood scan as a function of $|f_{CP}^{Htt}|$ for multilepton final estates. The solid (dashed) line shows the observed (expected) scan.

The addition of the three final states increases the sensitivity significantly, as can be inferred from the left plot of Fig 6.18, that shows the expected likelihood scan as a function of $|f_{CP}^{Htt}|$. The observed result is also depicted in Fig. 6.18 (right), showing a best fit value of $|f_{CP}^{Htt}| = 0.28$ and allowing for a fractional CP-odd contribution with an interval of $|f_{CP}^{Htt}| < 0.55$ at 68% CL The pure CP-odd scenario is excluded with 3.7 s.d and the result is compatible with the SM within the 95% CL

The two-dimensional confidence intervals as a function of κ_t and $\tilde{\kappa}_t$ are shown in Fig. 6.19. The plot is symmetric with respect to the line $\tilde{\kappa}_t = 0$, hence there are two points corresponding to the best fit, here we only show one for simplicity. The black
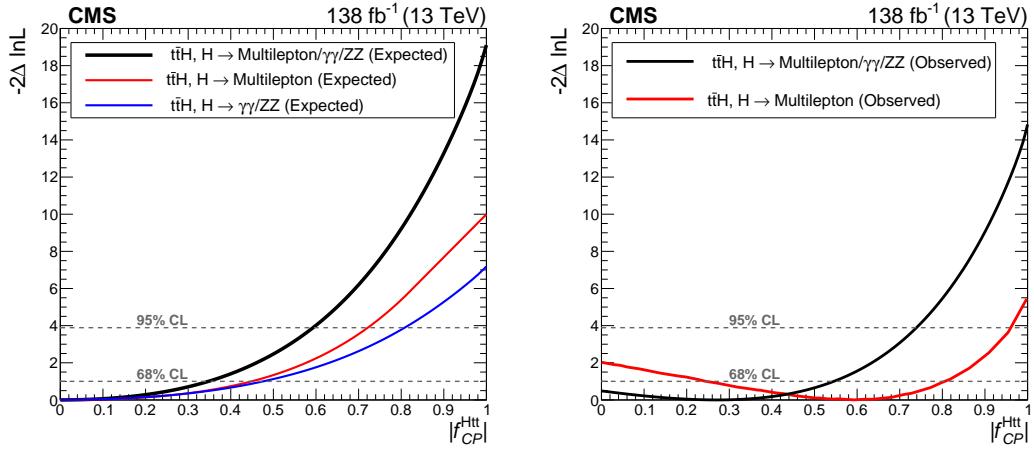
FIGURE 6.18: Likelihood scan as a function of $|f_{CP}^{\text{Htt}}|$. The left plot shows the expected likelihood scan for multilepton final states, $H \to \gamma\gamma$, and $H \to ZZ$ final states, and the combination of the three. The right plot shows the observed likelihood scan for multilepton final states and the combination of multilepton, $H \to \gamma\gamma$ and $H \to ZZ$ final states.
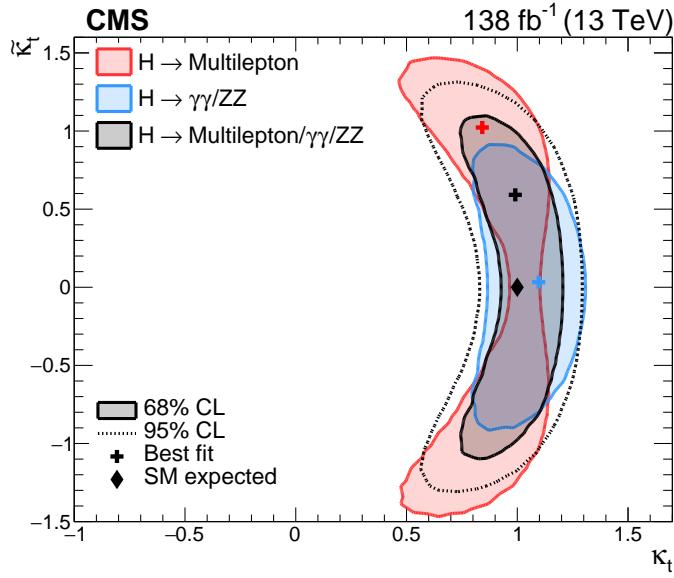
FIGURE 6.19: Likelihood scan as a function of κ_t and $\tilde{\kappa}_t$. Two-dimensional confidence intervals at 68% CL are depicted as shaded areas, for the multilepton final states (red), the combination of $H \to \gamma\gamma$ and $H \to ZZ$ channels (blue), and the combination of the three channels (black). The 95% CL for the combination is show as a dashed line. The best fit for each is shown as a cross of the corresponding colour.

diamond shows the SM expected value. The nontrivial correlation between the measurements are the source of the change in the best fit value and shape of the confidence regions. One-dimensional limits set κ_t to be within (0.86, 1.26) at 95% C.L, when $\tilde{\kappa}_t$ is fixed to the SM; and $\tilde{\kappa}_t$ to be within (-1.07, 1.07) at 95% CL when κ_t is set to its SM value. The limits are also computed at 68% CL, setting κ_t to be within (0.96, 1.16) and $\tilde{\kappa}_t$ to be within (-0.86, 0.85). Results are in good agreement with the SM.

6.8 Conclusions and prospects

A measurement of t$\bar{\text{t}}$H and tH processes in final states with multiple leptons is performed. The selected final states target the Higgs boson decaying to W^+W^- and $\tau^+\tau^-$ mainly. Several categories are defined depending on the lepton multiplicity, where each category presents different purity, leading background, and total number of expected events. In each of them dedicated MVA-based algorithms are used to separate t$\bar{\text{t}}$H and tH signals from the main backgrounds.

The signal strength for t$\bar{\text{t}}$H is measured to be: $\mu_{\text{t}\bar{\text{t}}\text{H}} = 0.92 \pm 0.19 \text{ (stat)}^{+0.17}_{-0.13} \text{ (syst)}$, yielding an observed (expected) significance of 4.7 (5.2) standard deviations with respect to the background only hypothesis, when fixing tH rate to its SM value. The tH production rate is also measured to be $\mu_{\text{tH}} = 5.7 \pm 2.7 \text{ (stat)} \pm 3.0 \text{ (syst)}$, corresponding to a significance of 1.4 (0.3) s.d. for tH production when fixing the t$\bar{\text{t}}$H production rate to its SM predicted value. This result allows to claim the observation of t$\bar{\text{t}}$H production mode in multilepton final states. Regarding tH production, the sensitivity is still very limited given the low cross section of this production mode, as detailed in the previous chapter.

The main uncertainties are related to the estimation of the nonprompt lepton background, and the modelling uncertainties for the signals and the backgrounds estimated with simulation. As already stated in Section 5.6, an efficient prompt lepton selection is crucial in this kind of analyses, and new techniques are being developed and will allow to reduce the related systematic uncertainties during Run-3. Regarding nonreducible backgrounds, t$\bar{\text{t}}$W is the predominant one in t$\bar{\text{t}}$H multilepton measurements. The measurement of the inclusive cross section presented in this thesis represents an important step to understand this background. Nevertheless, its modelling is challenging and more effort is still needed, for example the performance of dedicated differential measurements are crucial to improve our knowledge of this process.

Besides, the measurement is used to study the top-Higgs coupling. Firstly, the top-Higgs and W-Higgs couplings are tested. The κ_t modifier is constrained to be within the two intervals: $-0.9 < \kappa_t < -0.7$ and $0.7 < \kappa_t < 1.1$ at 95% CL

Secondly, the three most sensitive categories are used to search for CP violation in the top-Higgs coupling. The measurements are found to be in good agreement with the SM and exclude the pure CP-odd hypothesis with more than 2 s.d. Furthermore, the result using final states with multiple leptons is combined with previously published CMS measurements using events where $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ \rightarrow 4\ell$. This further constrains the limits on κ_t and $\tilde{\kappa}_t$, which are set to be within (0.86, 1.26) and (-1.07,
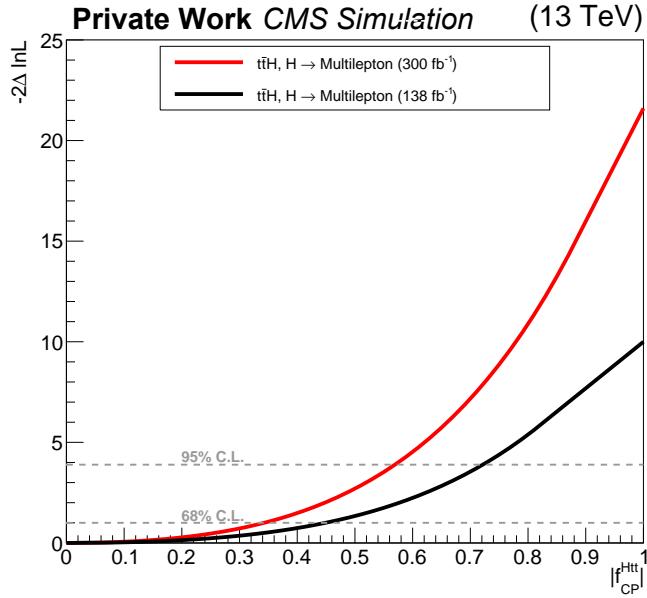
FIGURE 6.20: Prospect on the expected likelihood scan as a function of $|f_{CP}^{Htt}|$ using the expected total luminosity accumulated at the end of Run-3.

1.07) intervals at 95% CL, respectively. The pure *CP*-odd scenario is excluded with 3.7 s.d. This measurement is still statistically limited and would benefit significantly from Run-3 data. In Fig. 6.20 a projection of the expected sensitivity on $|f_{CP}^{Htt}|$ using 300 fb^{-1} is given. In this projection, the systematic uncertainties are kept at the same level as the ones achieved during Run-2.

Furthermore, the current amount of data allows to perform differential measurements of the t̄tH cross section. In particular, it is interesting to study the cross section as a function of the kinematic properties of the t̄tH system, such as the transverse momentum of the t̄tH system and the visible mass of the system. These measurements would allow to disentangle the effects of modified Higgs boson self-coupling values from other effects such as the presence of anomalous top-Higgs couplings [124]. Currently I am involved in this measurement with CMS.

Summary and conclusions

The work presented in this thesis makes use of proton-proton collisions recorded by the CMS experiment during the Run-2 of the LHC at $\sqrt{s} = 13\,\text{TeV}$. The unprecedented amount of analyzed data —138 fb^{-1}— and the improvements in the experimental techniques —some of which are presented in this thesis— has allowed to measure low cross section processes and study the properties of the Higgs boson and the top quark with high precision. The studies reported in this thesis use final states with multiple leptons and production modes where the top quark is generated in association with a Higgs or a W boson. Standard Model measurements have been performed and, in some cases, interpretations of such measurements in the context of Beyond the Standard Model (BSM) physics have been presented as well.

The first study presented in this thesis is the measurement of the cross section for the associated production of a W boson with a top quark-antiquark pair ($t\bar{t}W$). This process allows to study the interaction between the top quark and the W boson. Additionally, its unique production features yield a sizable asymmetry between $t\bar{t}W^+$ and $t\bar{t}W^-$, which can be used to extract information about the parton distribution functions of the proton. A good understanding of the modelling of $t\bar{t}W$ is mandatory to perform measurements of the associated production of the Higgs boson with a top quark-antiquark pair ($t\bar{t}H$) and the four top production ($t\bar{t}t\bar{t}$) in final states with leptons, as $t\bar{t}W$ represents a sizable background to those processes. In this thesis, the measurement of the $t\bar{t}W$ cross section has been reported, achieving the most precise measurement up to date. These studies have been performed in final states with two leptons with electric charges of the same-sign ($2\ell ss$) and in events with three leptons (3ℓ), using different strategies. In the $2\ell ss$ category, which is the one with a higher number of signal events, multivariate analysis techniques have been used to discriminate the signal from the background. In the 3ℓ category a categorization depending on the jet, b-tagged jet multiplicities and lepton charge has been performed; besides, the invariant mass of the three lepton system have been used as discriminating variable.

The cross section is measured to be

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

which is the most precise measurement of the $\sigma_{t\bar{t}W}$ up to date and improves the previous CMS result [12], reducing the systematic uncertainty by a factor larger than two.
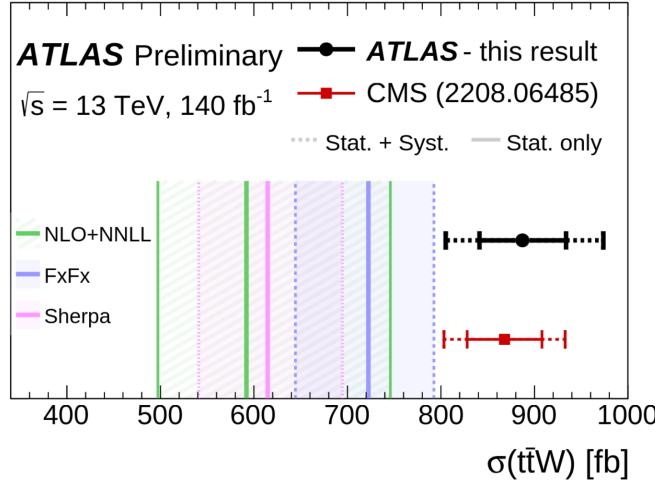
FIGURE 6.21: Comparison of the t$\bar{\text{t}}$W cross section measurement reported by AT-
LAS [161] and CMS (this thesis) with the theoretical predictions from SHERPA, the
MADGRAPH_AMC@NLO +PYTHIA 8 FxFx prescription including EWK corrections [18],
and the NLO+NNLL prediction [19]. Figure extracted from Ref. [161]

The measurement shows some tension with the SM prediction computed at NNLO
(QCD) + NLO (EWK) of $497.5 \pm 6.6\% \pm 1.8\%$ [21] and is in agreement with the one
reported by ATLAS [161]. The t$\bar{\text{t}}$W cross section measured by ATLAS and CMS are
shown in figure 6.21. Measuring this quantity is mandatory to understand t$\bar{\text{t}}$W mod-
elling, which is an open question under debate within the community. Additionally,
the cross section for t$\bar{\text{t}}$W$^+$ and t$\bar{\text{t}}$W$^-$ production modes has been measured separately,
and the ratio between both cross sections has also been reported.

In this thesis the measurements of the t$\bar{\text{t}}$H and tH production in final states with
leptons, including hadronically decaying taus, have been presented. These challenging
measurements allow to study the Yukawa coupling between the two most massive
particles in the SM. The selected final states target the Higgs decay modes to W^+W^-
and $\tau^+\tau^-$ mainly, with a residual contribution from ZZ events. Events were classified
in ten categories depending on the number of leptons present in the final state. A
dedicated selection was then applied in each category based on the number of jets
and b-tagged jets in the event. The usage of multivariate analysis techniques was
needed to improve the discrimination between signal and background. The measured
production rate is

$$\mu_{\text{t}\bar{\text{t}}\text{H}} = 0.92 \pm 0.19(\text{stat})^{+0.17}_{-0.13}(\text{syst})$$

times the SM expectation, which corresponds to an observed (expected) sensitivity of
4.7 (5.2) standard deviations. This measurement allows to claim the first observation
of t$\bar{\text{t}}$H process using only final states with multiple leptons. The tH production rate
has also been measured, with an observed (expected) sensitivity of 1.4 (0.3) s.d. .

Furthermore, the measurement has been interpreted considering possible anomalous couplings, allowing the coupling of the Higgs boson to the top quark and the Higgs boson to the W boson to acquire values different to its SM prediction. Modifications on such couplings would induce a change in the cross section for t$\bar{\text{t}}$H and tH, as well as, changes in the kinematic observables for tH production. Limits have been set on the coupling modifiers of the Higgs to the top (κ_t), which is found to be within the two intervals:

$$-0.9 < \kappa_t < -0.7 \text{ and } 0.7 < \kappa_t < 1.1 \text{ at 95\% CL,}$$

assuming that the coupling of the Higgs boson to the W boson has the value predicted by the SM.

The SM Higgs boson is a pseudo-scalar invariant under Charge-Parity (*CP*) transformations. In this thesis, t$\bar{\text{t}}$H production mode has been used to search for *CP* violation in the top-Higgs coupling. Results are in good agreement with the SM and allow to set limits on the κ_t and $\tilde{\kappa}_t$, which are the *CP*-even and *CP*-odd coupling modifiers, respectively. The pure *CP*-odd hypothesis has been excluded with more than 2 s.d.. The result has been combined with previous measurements to include events in which the Higgs boson decays to $\gamma\gamma$ and $ZZ \to 4\ell$. This allows to set limits on κ_t and $\tilde{\kappa}_t$, which are constrained to be within (0.86, 1.26) and (-1.07, 1.07) at 95% CL, respectively. Such limits are shown in Fig. 6.22. With the combined analysis, the pure *CP*-odd scenario is excluded with more than 3 s.d..

In this thesis a new muon identification technique has been developed. The constant revision and improvement of lepton reconstruction, identification and isolation is crucial for any CMS analysis using this kind of particle in the final state. The new MVA-based algorithm had been designed to improve the muon identification efficiency achieved during Run-2. The algorithm was shown to have a higher efficiency and to perform better under scenarios with high number of simultaneous collisions than the cut-based selection used as standard during Run-2. This second feature is of particular interest given higher luminosity expected to be recorded during Run-3 of the LHC.

In summary, this thesis exploited the possibilities of final states with multiple leptons to measure the associated production of a boson with top quarks. In particular, t$\bar{\text{t}}$H process was measured with enough sensitivity to claim an observation. The measurement of this process has allowed to study the top Yukawa coupling and search for *CP* violation. Besides, the t$\bar{\text{t}}$W process has been measured providing the most precise
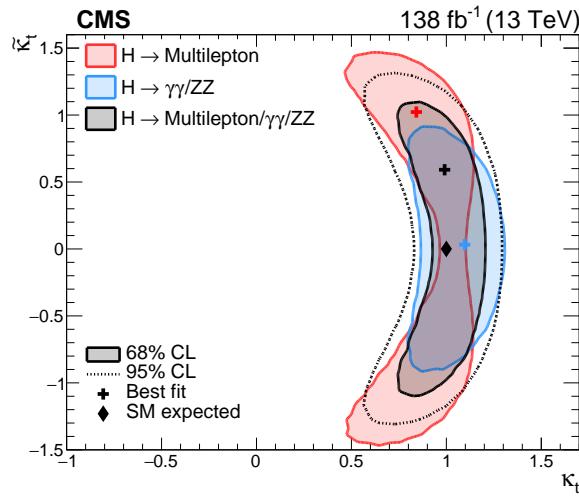
FIGURE 6.22: Likelihood scan as a function of κ_t and $\tilde{\kappa}_t$. Two-dimensional confidence intervals at 68% CL are depicted as shaded areas, for the multilepton final states (red), the combination of H $\rightarrow \gamma\gamma$ and H $\rightarrow$ ZZ channels (blue), and the combination of the three channels (black). The 95% CL for the combination is show as a dashed line. The best fit for each is shown as a cross of the corresponding colour.

measurement of the production cross section of this process at the CERN LHC. Finally, a new MVA-based algorithm for muon identification has been developed, with features that ensure its successful usage during Run-3 of the LHC.

Resumen y conclusiones

El trabajo presentado en esta tesis hace uso de colisiones protón-protón recopiladas por el experimento CMS durante el Run-2 del LHC, a una energía en centro de masas de $\sqrt{s} = 13$ TeV. La cantidad sin precedente de datos analizados —138 fb^{-1}— y las mejoras en las técnicas experimentales —algunas de las cuales se presentan en esta tesis— han permitido medir procesos con secciones eficaces bajas y estudiar con alta precisión las propiedades del bosón de Higgs y el quark top. Los estudios presentados se centran en estados finales con múltiples leptones y modos de producción en los que el quark top se genera en asociación con un bosón de Higgs o un bosón W. Se han realizado medidas en el contexto del Modelo Estándar (ME) y, en algunos casos, dichas medidas se han interpretadas en el contexto de la física más allá del ME.

El primer estudio presentado en esta tesis es la medida de la sección eficaz inclusiva de producción del bosón W en asociación con un par de quarks top-antitop (t$\bar{\text{t}}$W). Este proceso permite estudiar la interacción entre el quark top y el bosón W. Además, sus características únicas a nivel de producción conducen a una asimetría considerable entre los procesos t$\bar{\text{t}}$W$^+$ y t$\bar{\text{t}}$W$^-$. El estudio de dicha asimetría se pueden usar para extraer información acerca de las funciones de distribución de los partones en el protón. Una buena compresión del modelado del proceso t$\bar{\text{t}}$W es crucial a la hora de estudiar tanto la producción asociada del bosón de Higgs con un par de quarks top-antitop (t$\bar{\text{t}}$H), como la producción asociada de cuatro quarks top (t$\bar{\text{t}}$t$\bar{\text{t}}$), ya que t$\bar{\text{t}}$W es un fondo importante al estudiar dichos procesos en estados finales con múltiples leptones. En esta tesis se ha presentado la medida de la sección eficaz inclusiva del proceso t$\bar{\text{t}}$W más precisa hasta la fecha. Dicha medida se ha llevado a cabo en estados finales con dos leptones con el mismo valor de carga eléctrica (2ℓss) y estados finales con tres leptones (3ℓ), utilizando diferentes estrategias de análisis en cada una. En la categoría 2ℓss, que es la que tiene un mayor número de sucesos de señal, se han utilizado técnicas de análisis multivariante para discriminar la señal del fondo. En la categoría 3ℓ se ha realizado una clasificación en función del número de jets, del número de jets etiquetados como provenientes de un b, y de la carga de los leptones; usando como variable discriminadora la masa invariante del sistema de tres leptones.

La sección eficaz medida es

$$\sigma_{\text{t}\bar{\text{t}}\text{W}} = 868 \pm 40(\text{stat}) \pm 51(\text{syst})\text{fb,}$$
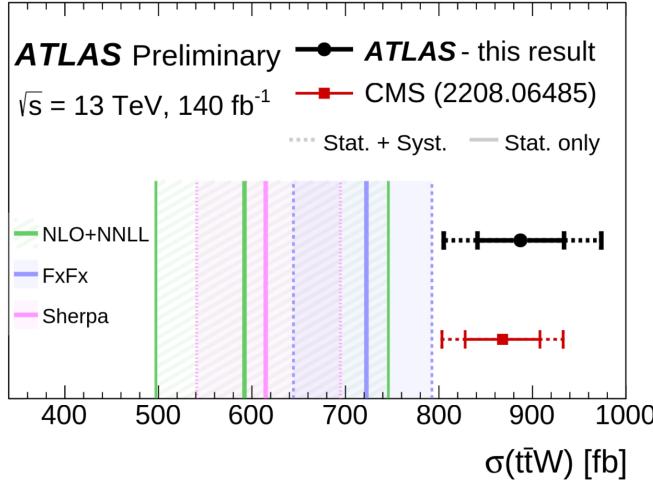
FIGURE 6.23: Comparación de la sección eficaz de producción de t̄tW medida por ATLAS [161] y CMS (esta tesis) junto con las perdiciones teóricas proporcionadas por SHERPA, MADGRAPH_AMC@NLO +PYTHIA 8 FxFx incluyendo correcciones electrodébiles [18], y la predicción aNLO+NNLL. [19]. La figura ha sido extraída de la Ref. [161]

que mejora el anterior resultado publicado por CMS [12], reduciendo la incertidumbre sistemática casi a la mitad. La medida muestra cierta tensión con la predicción del ME calculada a NNLO (QCD) + NLO (EWK) de $497.5 \pm 6.6\% \pm 1.8\%$ [21]. El resultado obtenido concuerda con el presentado por ATLAS [161], como se puede observar en la figura 6.23. Esta medida es necesaria para entender el modelado del proceso t̄tW, que es un tema en debate dentro de la comunidad de física de altas energías. Además, la sección eficaz también se mide para los modos de producción t̄tW$^+$ y t̄tW$^-$ por separado, y se ha proporcionado también la medida del cociente entre ambas secciones eficaces.

En esta tesis han presentado las medidas de los modos de producción del bosón de Higgs t̄tH y tH, en estados finales con leptones, incluyendo leptones τ que se desintegran hadrónicamente. Esta compleja medida permite estudiar el acoplamiento de Yukawa entre las dos partículas mas masivas de ME. Los estados finales seleccionados en este estudio se producen cuando el bosón de Higgs se desintegra en W$^+$W$^-$ y $\tau^+\tau^-$, principalmente, con una contribución residual de sucesos ZZ. Los sucesos fuero clasificados en 10 categorías dependiendo del número de leptones presentes en el estado final. En cada una de ellas se aplicó una selección específica, basada en el número de jets y en el número de jets etiquetados como provenientes de un quark b. Posteriormente, se utilizaron diferentes técnicas de análisis multivariante (MVA) para mejorar la discriminación entre la señal y el fondo. El cociente entre la sección eficaz de producción medida y predicha es:

$$\mu_{t\bar{t}H} = 0.92 \pm 0.19(\text{estad.})^{+0.17}_{-0.13}(\text{sist.})$$

lo que corresponde a una sensibilidad observada (esperada) de 4.7 (5.2) desviaciones estándar (d.e.). Esta medida permite afirmar que se ha observado por primera vez el proceso t̄tH usando solamente estados finales con leptones. También se ha medido la tasa de producción del proceso tH, con una sensibilidad observada (esperada) de 1.4 (0.3) d.e.

Además, la medida anterior ha sido interpretada considerando posibles acoplamientos anómalos, permitiendo que el acoplamiento entre el bosón de Higgs y el quark top y entre el bosón de Higgs y el bosón W adquieran valores diferentes a su predicción en el ME. La modificación de tales acoplamientos induciría un cambio en la sección eficaz de los procesos t̄tH y tH, así como cambios en los observables cinemáticos para el proceso tH. Se han establecido límites en el modificador del acoplamiento del Higgs al top (κ_t), que se encuentra dentro de los dos intervalos:

$$-0.9 < \kappa_t < -0.7 \text{ y } 0.7 < \kappa_t < 1.1 \text{ al 95\% de nivel de confianza,}$$

asumiendo que el acoplamiento entre el bosón de Higgs y el W toma el valor predicho por el ME.

El bosón de Higgs propuesto en el ME es un invariante pseudo-escalar bajo transformaciones de Carga-Paridad (CP). En esta tesis, se ha utilizado el modo de producción t̄tH para buscar violación de la simetría CP en el acoplamiento top-Higgs. Los resultados obtenidos están en concordancia con el ME y permiten establecer límites en κ_t y $\tilde{\kappa}_t$, que son los modificadores de acoplamiento CP-par y CP-impar, excluyendo la hipótesis puramente CP-impar con más de 2 d.e. Se ha combinado este resultado con medidas dónde el bosón de Higgs se desintegra a $\gamma\gamma$ y ZZ. Esto ha permitido establecer límites en κ_t y $\tilde{\kappa}_t$, que deben estar dentro de los intervalos (0.86, 1.26) y (-1.07,1.07) al 95% de nivel de confianza, respectivamente. Estos límites se muestran en la figura 6.24 El escenario puramente CP-impar se excluye con más de 3 d.e.

En esta tesis se ha desarrollado una nueva técnica de identificación de muones. La revisión constante y la mejora en la reconstrucción, identificación y aislamiento de leptones son cruciales para cualquier análisis de CMS que utilice este tipo de partícula en el estado final. El nuevo algoritmo, basado en técnicas de análisis MVA, ha sido diseñado para mejorar la eficiencia de identificación de muones con respecto a la lograda durante el Run-2. Se ha mostrado que el algoritmo tiene una eficiencia superior y un mejor rendimiento en escenarios con un alto número de colisiones simultáneas en comparación con la selección utilizada como estándar durante Run-2. Esta segunda característica es de particular interés dada la mayor luminosidad registrada durante Run-3.
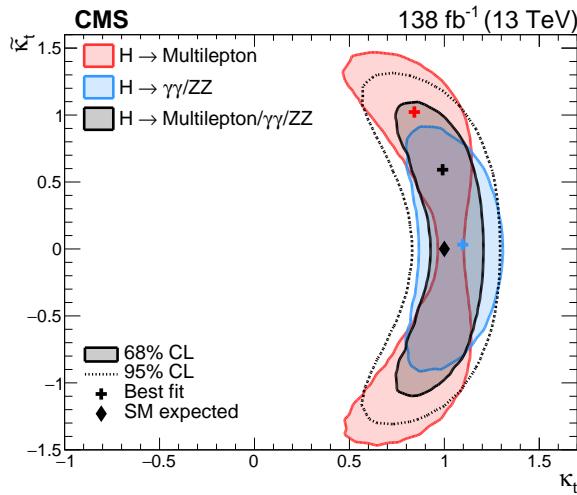
FIGURE 6.24: Función de máxima verosimilitud en función de κ_t y $\tilde{\kappa}_t$. Las áreas sombreadas muestran los intervalos de confianza bidimensionales al 68% de nivel de confianza, para estados finales con múltiples leptones (rojo), la combinación de los canales dónde H $\rightarrow \gamma\gamma$ y H $\rightarrow ZZ$ (azul) y la combinación de los tres canales (negro). Los intervalos al 95% de confianza se muestran con una línea discontinua para la combinación de los tres canales. Las cruces representan el punto mínimo del ajuste.

En conclusión, esta tesis he realizado una medida de los procesos $t\bar{t}H$ y tH, que son dos de los modos de producción del bosón de Higgs más complejos de estudiar debido a su baja sección eficaz. Dicha medida proporciona la primera observación de $t\bar{t}H$ utilizando estados finales con múltiples leptones y permite estudiar el acoplamiento entre el quark top y el bosón de Higgs con una precisión sin precedentes. La medida del proceso $t\bar{t}H$ ha sido utiliza para interpretar el acoplamiento en el contexto de la física más allá del modelo estándar, buscando violación de la simetría *CP*. Los resultados están de acuerdo con el Modelo Estándar y establecen los límites más restrictivos dados por CMS para el acoplamiento *CP*-impar entre el Higgs y el top. El proceso $t\bar{t}W$ también se ha estudiado en detalle, proporcionando la medida más precisa de la sección eficaz de producción de este proceso. Esta medida es fundamental para comprender la modelización de $t\bar{t}W$, que es un tema de discusión dentro de la comunidad. En esta tesis he presentado un nuevo algoritmo basado en MVA para seleccionar muones; este nuevo algoritmo mejora el utilizado previamente en CMS y muestra un comportamiento estable y robusto en escenarios con un alto número de colisiones simultáneas. Este hecho garantiza el uso exitoso de este algoritmo durante Run-3

En resumen, esta tesis se aprovecha las posibilidades que brindan los estados finales con múltiples leptones para medir la producción de un bosón en asociación con quarks

top. En particular el proceso t$\bar{\text{t}}$H se ha medido con precisión suficiente como para afirmar su observación. La medida de este proceso ha permitido estudiar el acoplamiento de Yukawa del quark top y buscar por violación de la simetría CP. Además, se ha medido la sección eficaz del proceso t$\bar{\text{t}}$W con la máxima precisión hasta la fecha en el LHC del CERN. Finalmente, un nuevo algoritmo basado en técnicas de análisis multivariante se ha desarrollado, dicho algoritmo tiene características que garantizan su uso exitoso durante el Run-3 del LHC.

A

Additional distributions used in the $t\bar{t}W$ measurement

In this appendix all distributions used for the signal extraction described in Chapter 5 are provided. Figure A.2 shows the NN discriminant in the $2\ell ss$ signal region, events are classified depending on the lepton charge and flavour. Plots are shown after the fit showing good agreement between data and simulation.

Figure A.4 shows the $m(3\ell)$ variable in the 3ℓ signal region for the subcategories not shown in Fig. 5.7. All distributions shown are post-fit and show a good agreement between data and simulation.
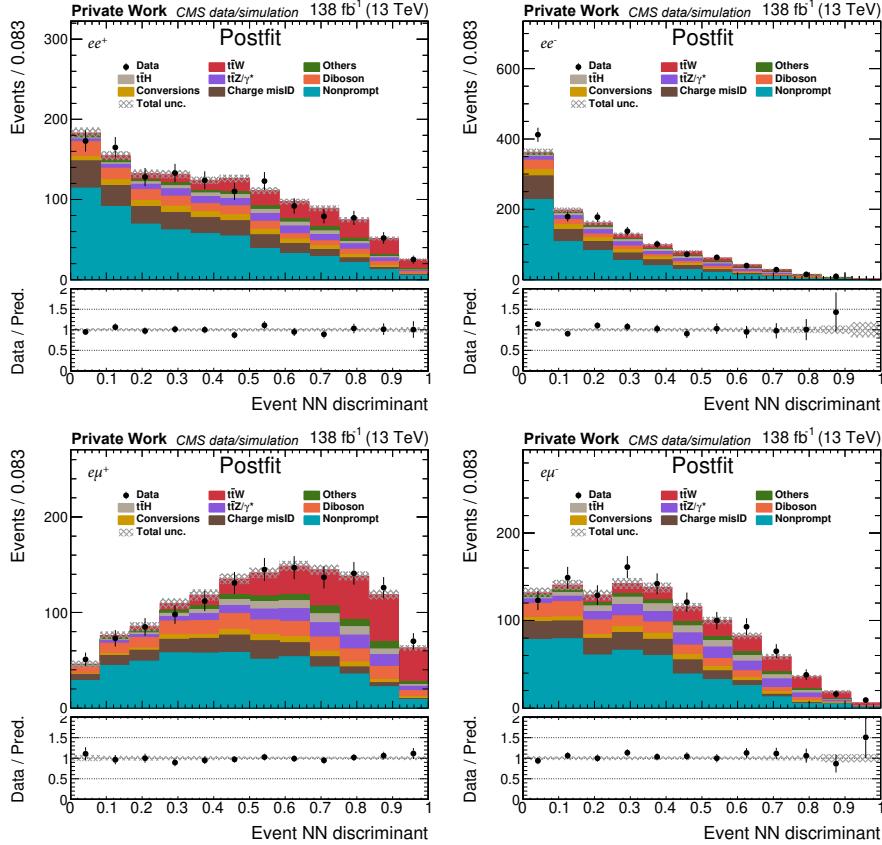
FIGURE A.1: NN discriminant in the 2ℓss signal region. Events are classified according to the lepton flavour, *ee* (upper row), *eμ* (lower row); and sign of the leptons charges, positive (left) and negative (right) . Plots are postfit, The vertical bars on the points represent the statistical uncertainties in the data, and the hatched bands the systematic uncertainty in the predictions. In the lower panels, the ratio of the event yields in data to the overall sum of the predictions is presented.
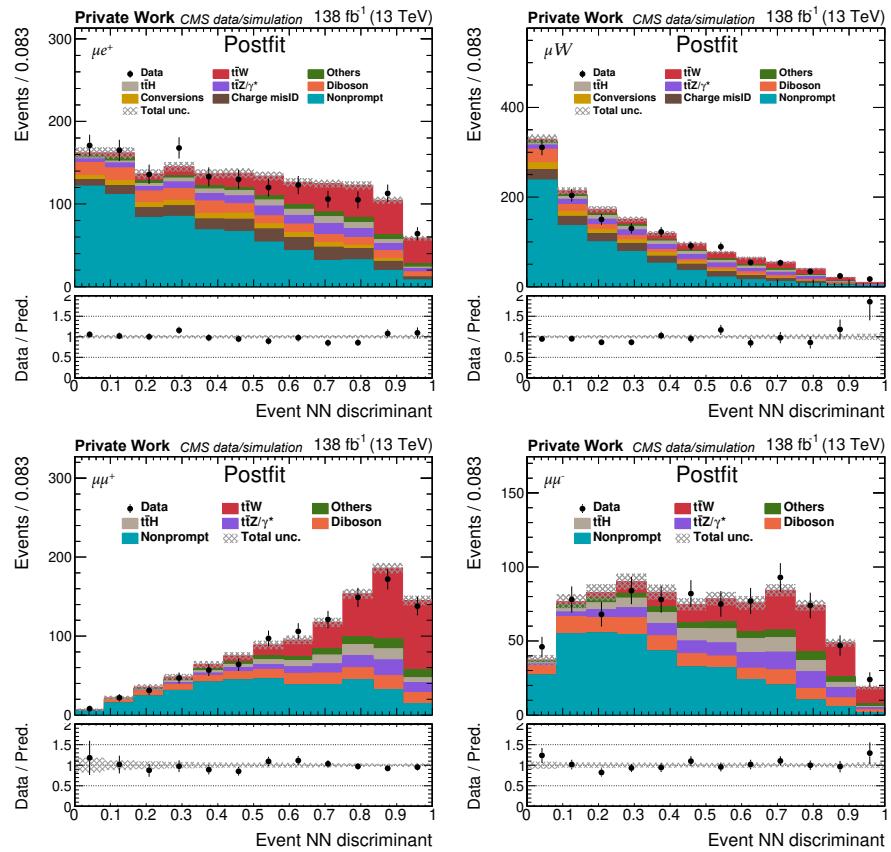
FIGURE A.2: NN discriminant in the $2\ell ss$ signal region. Events are classified according to the lepton flavour, μe (upper row) and $\mu\mu$ (lower row); and sign of the leptons charges, positive (left) and negative (right) . Plots are postfit, The vertical bars on the points represent the statistical uncertainties in the data, and the hatched bands the systematic uncertainty in the predictions. In the lower panels, the ratio of the event yields in data to the overall sum of the predictions is presented.
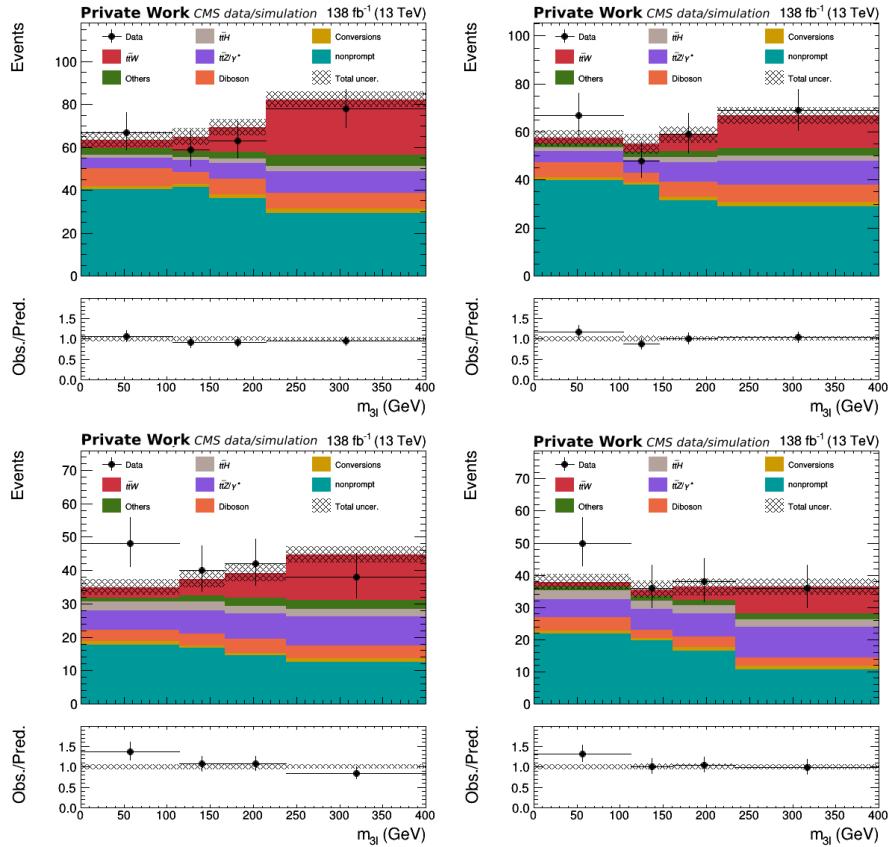
FIGURE A.3: $m(3\ell)$ variable in the 3ℓ signal region. Events are subcategorized as follows: two jets and one b-tagged jet (upper row) and exactly three jets and one b-tagged (lower row). Plots on the left (right) contain events with a positive (negative) sum of the lepton charges. Plots are postfit, The vertical bars on the points represent the statistical uncertainties in the data, and the hatched bands the systematic uncertainty in the predictions. In the lower panels, the ratio of the event yields in data to the overall sum of the predictions is presented.
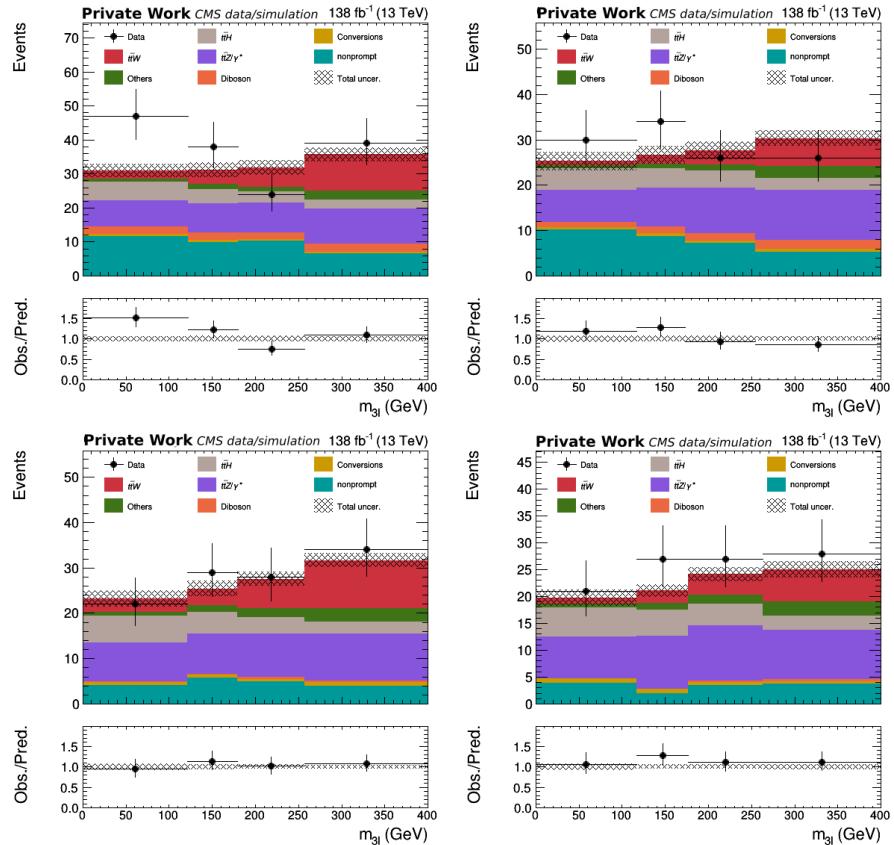
FIGURE A.4: $m(3\ell)$ variable in the 3ℓ signal region. Events are subcategorized as follows: at least three jets and one b-tagged (upper row), at least three jets and two b-tagged jets (lower row). Plots on the left (right) contain events with a positive (negative) sum of the lepton charges. Plots are postfit, The vertical bars on the points represent the statistical uncertainties in the data, and the hatched bands the systematic uncertainty in the predictions. In the lower panels, the ratio of the event yields in data to the overall sum of the predictions is presented.

B

Yields for the t$\bar{\text{t}}$H and tH measurements

In this appendix the number of events selected in each of the ten analysis channels defined to study t$\bar{\text{t}}$H and tH processes in Chapter 6 is reported. The information is given in Tables B.1 and B.2 .

Process	$2\ell ss + 0\tau_{\text{h}}$	$3\ell + 0\tau_{\text{h}}$	$2\ell ss + 1\tau_{\text{h}}$
t$\bar{\text{t}}$H	222 ± 51	61 ± 15	28.9 ± 6.4
tH	119 ± 85	20 ± 14	12.7 ± 9.0
t$\bar{\text{t}}$Z + t$\bar{\text{t}}$X	322 ± 25	145 ± 11	29.6 ± 3.3
t$\bar{\text{t}}$W + t$\bar{\text{t}}$WW	1153 ± 64	171.1 ± 9.5	47.4 ± 6.5
WZ	296 ± 31	89.7 ± 9.7	19.4 ± 2.9
ZZ	31.2 ± 3.3	16.2 ± 1.6	1.6 ± 0.3
Misidentified leptons	1217 ± 91	140 ± 11	52.0 ± 9.6
Flips	121 ± 19	—	—
Rare backgrounds	222 ± 48	41.0 ± 8.9	13.3 ± 3.1
Conversion	42 ± 12	5.6 ± 1.6	—
ggH + qqH + VH + t$\bar{\text{t}}$VH	35.3 ± 4.0	3.4 ± 0.3	1.8 ± 0.3
Total expected background	3517 ± 85	627 ± 20	179 ± 13
Data	3738	744	201

TABLE B.1: Number of events selected in the $2\ell ss + 0\tau_{\text{h}}$, $3\ell + 0\tau_{\text{h}}$ and $2\ell ss + 1\tau_{\text{h}}$ analysis categories, compared to the event yields expected from the t$\bar{\text{t}}$H and tH signals and from background processes. The expected event yields are computed for the values of nuisance parameters and of the POI obtained from the ML fit. The best fit values of the POI amount to $\mu_{\text{t}\bar{\text{t}}\text{H}} = 0.92$ and $\mu_{\text{tH}} = 5.7$. Quoted uncertainties represent the sum of statistical and systematic components. The symbol "—" indicates that the corresponding expected contribution is smaller than 0.1 events.

Process	$1\ell + 1\tau_h$	$0\ell + 2\tau_h$	$2\ell os + 1\tau_h$	$1\ell + 2\tau_h$
$t\bar{t}H$	183 ± 41	24.4 ± 6.0	19.1 ± 4.3	19.3 ± 4.2
tH	65 ± 46	16 ± 12	4.8 ± 3.4	2.6 ± 1.9
$t\bar{t}Z + t\bar{t}X$	203 ± 24	27.1 ± 3.8	25.5 ± 2.9	20.3 ± 2.1
$t\bar{t}W + t\bar{t}WW$	254 ± 34	3.8 ± 0.5	17.4 ± 2.4	2.6 ± 0.4
WZ	198 ± 37	42.5 ± 8.7	8.4 ± 1.6	11.8 ± 2.2
ZZ	98 ± 13	34.2 ± 4.8	1.9 ± 0.3	1.8 ± 0.3
DY	4480 ± 460	1430.0 ± 220	519 ± 28	250 ± 16
$t\bar{t}$+jets	41900 ± 1900	861 ± 98	—	—
Misidentified leptons	25300 ± 1900	3790 ± 220	—	—
Rare backgrounds	1930 ± 420	60 ± 14	5.9 ± 1.3	5.6 ± 1.3
Conversion	—	—	0.5 ± 0.2	—
ggH + qqH + VH + $t\bar{t}VH$	38.5 ± 3.6	26.7 ± 3.6	0.8 ± 0.1	—
Total expected background	73550 ± 610	6290 ± 130	584 ± 27	295 ± 16
Data	73736	6310	603	307

Process	$4\ell + 0\tau_h$	$3\ell + 1\tau_h$	$2\ell + 2\tau_h$
$t\bar{t}H$	2.0 ± 0.5	4.0 ± 0.9	2.2 ± 0.5
tH	0.2 ± 0.2	0.8 ± 0.6	0.3 ± 0.2
$t\bar{t}Z + t\bar{t}X$	5.9 ± 0.4	6.6 ± 0.7	2.5 ± 0.3
$t\bar{t}W + t\bar{t}WW$	0.2 ± 0.0	1.1 ± 0.2	—
ZZ	0.6 ± 0.2	0.3 ± 0.1	0.2 ± 0.0
Misidentified leptons	—	1.5 ± 0.9	3.4 ± 0.9
Rare backgrounds	0.6 ± 0.1	1.0 ± 0.3	0.3 ± 0.1
Conversion	—	—	—
Total expected background	7.4 ± 0.5	11.5 ± 1.3	6.8 ± 1.0
Data	12	18	3

TABLE B.2: Number of events selected in six of the analysis categories compared to the event yields expected from the $t\bar{t}H$ and tH signals and from background processes. The expected event yields are computed for the values of nuisance parameters and of the POI obtained from the ML fit. The best fit values of the POI amount to $\mu_{t\bar{t}H} = 0.92$ and $\mu_{tH} = 5.7$. Quoted uncertainties represent the sum of statistical and systematic components. The symbol "—" indicates that the corresponding expected contribution is smaller than 0.1 events.

List of Figures

List of Tables

Bibliography

[1]   P.W. Higgs. "Spontaneous Symmetry Breakdown without Massless Bosons". *Phys. Rev.* 145 (4 1966), pp. 1156–1163. DOI: `10.1103/PhysRev.145.1156`. URL: `https://link.aps.org/doi/10.1103/PhysRev.145.1156`.

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

[3]   Peter W. Higgs. "Broken symmetries, massless particles and gauge fields". *Phys. Lett.* 12 (1964), pp. 132–133. DOI: `10.1016/0031-9163(64)91136-9`.

[4]   Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". *Phys. Rev. Lett.* 13 (1964). Ed. by J. C. Taylor, pp. 508–509. DOI: `10.1103/PhysRevLett.13.508`.

[5]   G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. "Global Conservation Laws and Massless Particles". *Phys. Rev. Lett.* 13 (1964). Ed. by J. C. Taylor, pp. 585–587. DOI: `10.1103/PhysRevLett.13.585`.

[6]   T. W. B. Kibble. "Symmetry breaking in nonAbelian gauge theories". *Phys. Rev.* 155 (1967). Ed. by J. C. Taylor, pp. 1554–1561. DOI: `10.1103/PhysRev.155.1554`.

[7]   Serguei Chatrchyan et al. "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC". *Phys. Lett. B* 716 (2012), pp. 30–61. DOI: `10.1016/j.physletb.2012.08.021`. arXiv: `1207.7235 [hep-ex]`.

[8]   Georges Aad et al. "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), pp. 1–29. DOI: `10.1016/j.physletb.2012.08.020`. arXiv: `1207.7214 [hep-ex]`.

[9]   Albert M Sirunyan et al. "Measurement of the Higgs boson production rate in association with top quarks in final states with electrons, muons, and hadronically decaying tau leptons at $\sqrt{s} = 13$ TeV". *Eur. Phys. J. C* 81.4 (2021), p. 378. DOI: `10.1140/epjc/s10052-021-09014-x`. arXiv: `2011.03652 [hep-ex]`.

[10] Armen Tumasyan et al. "Search for *CP* violation in ttH and tH production in multilepton channels in proton-proton collisions at $\sqrt{s} = 13$ TeV". *JHEP* 07 (2023), p. 092. DOI: `10.1007/JHEP07(2023)092`. arXiv: `2208.02686 [hep-ex]`.

[11] M. Aaboud et al. "Measurement of the $t\bar{t}Z$ and $t\bar{t}W$ cross sections in proton-proton collisions at $\sqrt{s}$ with the ATLAS detector". *Phys. Rev. D* 99 (2019), p. 072009. DOI: `10.1103/PhysRevD.99.072009`. arXiv: `1901.03584 [hep-ex]`.

[12] Albert M Sirunyan et al. "Measurement of the cross section for top quark pair production in association with a W or Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV". *JHEP* 08 (2018), p. 011. DOI: `10.1007/JHEP08(2018)011`. arXiv: `1711.02547 [hep-ex]`.

[13] The CMS Collaboration. *Measurement of the associated production of a Higgs boson with a top quark pair in final states with electrons, muons and hadronically decaying τ leptons in data recorded in 2017 at $\sqrt{s} = 13$ TeV*. Tech. rep. CMS-PAS-HIG-18-019. 2018.

[14] The ATLAS Collaboration. *Analysis of $t\bar{t}H$ and $t\bar{t}W$ production in multilepton final states with the ATLAS detector*. Tech. rep. ATLAS-CONF-2019-045. Oct. 2019. URL: `http://cds.cern.ch/record/2693930`.

[15] Georges Aad et al. "Search for $t\bar{t}H/A \rightarrow t\bar{t}t\bar{t}$ production in the multilepton final state in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". *JHEP* 07 (2023), p. 203. DOI: `10.1007/JHEP07(2023)203`. arXiv: `2211.01136 [hep-ex]`.

[16] Albert M Sirunyan et al. "Search for production of four top quarks in final states with same-sign or multiple leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV". *Eur. Phys. J. C* 80.2 (2020), p. 75. DOI: `10.1140/epjc/s10052-019-7593-7`. arXiv: `1908.06463 [hep-ex]`.

[17] Rikkert Frederix, Davide Pagani, and Marco Zaro. "Large NLO corrections in $t\bar{t}W^{\pm}$ and $t\bar{t}t\bar{t}$ hadroproduction from supposedly subleading EW contributions". *JHEP* 02 (2018), p. 031. DOI: `10.1007/JHEP02(2018)031`. arXiv: `1711.02116 [hep-ph]`.

[18] Rikkert Frederix and Ioannis Tsinikos. "On improving NLO merging for $t\bar{t}W$ production". *JHEP* 11 (2021), p. 029. DOI: `10.1007/JHEP11(2021)029`. arXiv: `2108.07826 [hep-ph]`.

[19] Anna Kulesza et al. "Associated top quark pair production with a heavy boson: differential cross sections at NLO+NNLL accuracy". *Eur. Phys. J. C* 80.5 (2020), p. 428. DOI: `10.1140/epjc/s10052-020-7987-6`. arXiv: `2001.03031 [hep-ph]`.

[20] Armen Tumasyan et al. "Measurement of the cross section of top quark-antiquark pair production in association with a W boson in proton-proton collisions at $\sqrt{s} = 13$ TeV". *JHEP* 07 (2023), p. 219. DOI: `10.1007/JHEP07(2023)219`. arXiv: `2208.06485 [hep-ex]`.

[21]   Luca Buonocore et al. "Precise Predictions for the Associated Production of a
       W Boson with a Top-Antitop Quark Pair at the LHC". *Phys. Rev. Lett.* 131.23
       (2023), p. 231901. DOI: 10.1103/PhysRevLett.131.231901. arXiv: 2306.16311
       [hep-ph].

[22]   Aram Hayrapetyan et al. "Muon identification using multivariate techniques
       in the CMS experiment in proton-proton collisions at $\sqrt{s}$ = 13 TeV" (Oct. 2023).
       Accepted by JINST. arXiv: 2310.03844 (hep-ex).

[23]   Albert M Sirunyan et al. "Search for supersymmetry in final states with two
       oppositely charged same-flavor leptons and missing transverse momentum in
       proton-proton collisions at $\sqrt{s}$ = 13 TeV". *JHEP* 04 (2021), p. 123. DOI: 10.1007
       /JHEP04(2021)123. arXiv: 2012.08600 [hep-ex].

[24]   Armen Tumasyan et al. "First measurement of the top quark pair production
       cross section in proton-proton collisions at $\sqrt{s}$ = 13.6 TeV". *JHEP* 08 (2023),
       p. 204. DOI: 10.1007/JHEP08(2023)204. arXiv: 2303.10680 [hep-ex].

[25]   Steven Weinberg. *The quantum theory of fields. Vol. 2: Modern applications*. Cam-
       bridge University Press, Aug. 2013. ISBN: 978-1-139-63247-8, 978-0-521-67054-8,
       978-0-521-55002-4. DOI: 10.1017/CBO9781139644174.

[26]   R. L. Workman and Others. "Review of Particle Physics". *PTEP* 2022 (2022),
       p. 083C01. DOI: 10.1093/ptep/ptac097.

[27]   A. Salam and J.C. Ward. "Electromagnetic and weak interactions". *Physics Let-
       ters* 13.2 (1964), pp. 168–171. ISSN: 0031-9163. DOI: https://doi.org/10.1016
       /0031-9163(64)90711-5. URL: https://www.sciencedirect.com/science/ar
       ticle/pii/0031916364907115.

[28]   S. Weinberg. "A Model of Leptons". *Phys. Rev. Lett.* 19 (21 1967), pp. 1264–1266.
       DOI: 10.1103/PhysRevLett.19.1264. URL: https://link.aps.org/doi/10.11
       03/PhysRevLett.19.1264.

[29]   S. L. Glashow. "Partial-symmetries of weak interactions". *Nuclear Physics* 22.4
       (1961), pp. 579–588. ISSN: 0029-5582. DOI: https://doi.org/10.1016/0029-55
       82(61)90469-2. URL: https://www.sciencedirect.com/science/article/pi
       i/0029558261904692.

[30]   H. Fritzsch, M. Gell-Mann, and H. Leutwyler. "Advantages of the color octet
       gluon picture". *Physics Letters B* 47.4 (1973), pp. 365–368. ISSN: 0370-2693. DOI:
       https://doi.org/10.1016/0370-2693(73)90625-4. URL: https://www.scien
       cedirect.com/science/article/pii/0370269373906254.

[31]   Chen-Ning Yang and Robert L. Mills. "Conservation of Isotopic Spin and Iso-
       topic Gauge Invariance". *Phys. Rev.* 96 (1954). Ed. by Jong-Ping Hsu and D.
       Fine, pp. 191–195. DOI: 10.1103/PhysRev.96.191.

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

[33] S.Sakata Z. Maki M. Nakagawa. "Remarks on the unified model of elementary particles". *Prog. Theor. Phys.* 28 (1962), pp. 870–880. DOI: 10.1143/PTP.28.870.

[34] Q. R. Ahmad et al. "Measurement of the Rate of $\nu_e + d \rightarrow p + p + e^-$ Interactions Produced by ^8B Solar Neutrinos at the Sudbury Neutrino Observatory". *Phys. Rev. Lett.* 87 (7 2001), p. 071301. DOI: 10.1103/PhysRevLett.87.071301. URL: https://link.aps.org/doi/10.1103/PhysRevLett.87.071301.

[35] T. Yanagida. "Horizontal Symmetry and Masses of Neutrinos". *Prog. Theor. Phys.* 64 (1980), p. 1103. DOI: 10.1143/PTP.64.1103.

[36] A. D. Dolgov. "Baryogenesis, 30 years after". *25th ITEP Winter School of Physics*. July 1997. DOI: 10.1080/01422419808240874. arXiv: hep-ph/9707419.

[37] A. D. Sakharov. "Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe". *Pisma Zh. Eksp. Teor. Fiz.* 5 (1967), pp. 32–35. DOI: 10.1070/PU1991v034n05ABEH002497.

[38] V. L. Fitch J.H. Christenson J. W. Cronin and R. Turlay. "Evidence for the 2π Decay of the K_2^0 Meson". *Phys. Rev. Lett.* 13 (1964), pp. 138–140. DOI: 10.1103/PhysRevLett.13.138.

[39] R Aaij et al. "First observation of *CP* violation in the decays of B_s^0 mesons". *Phys. Rev. Lett.* 110.22 (2013), p. 221601. DOI: 10.1103/PhysRevLett.110.221601. arXiv: 1304.6173 [hep-ex].

[40] Roel Aaij et al. "Observation of CP Violation in Charm Decays". *Phys. Rev. Lett.* 122.21 (2019), p. 211803. DOI: 10.1103/PhysRevLett.122.211803. arXiv: 1903.08726 [hep-ex].

[41] K. Abe et al. "Constraint on the matter–antimatter symmetry-violating phase in neutrino oscillations". *Nature* 580.7803 (2020). [Erratum: Nature 583, E16 (2020)], pp. 339–344. DOI: 10.1038/s41586-020-2177-0. arXiv: 1910.03887 [hep-ex].

[42] M. Fukugita and T. Yanagida. "Baryogenesis Without Grand Unification". *Phys. Lett. B* 174 (1986), pp. 45–47. DOI: 10.1016/0370-2693(86)91126-3.

[43] David Atwood et al. "CP violation in top physics". *Phys. Rept.* 347 (2001), pp. 1–222. DOI: 10.1016/S0370-1573(00)00112-5. arXiv: hep-ph/0006032.

[44] V. Rubin and W. K. Ford. "Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions". *Astrophys. J.* 159 (1970), pp. 379–403. DOI: 10.1086/150317.

[45] Leszek Roszkowski, Enrico Maria Sessolo, and Sebastian Trojanowski. "WIMP dark matter candidates and searches—current status and future prospects". *Rept. Prog. Phys.* 81.6 (2018), p. 066201. DOI: 10.1088/1361-6633/aab913. arXiv: 1707.06277 [hep-ph].

[46] J. Wess and B. Zumino. "A lagrangian model invariant under supergauge transformations". *Physics Letters B* 49.1 (1974), pp. 52–54. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(74)90578-4. URL: https://www.sciencedirect.com/science/article/pii/0370269374905784.

[47] A. Pich. "Effective field theory: Course". *Les Houches Summer School in Theoretical Physics, Session 68: Probing the Standard Model of Particle Interactions* (1998). arXiv: hep-ph/9806303. URL: https://arxiv.org/pdf/hep-ph/9806303.pdf.

[48] Christoph Englert et al. "Approaching robust EFT limits for CP-violation in the Higgs sector". *Phys. Rev. D* 99.9 (2019), p. 095007. DOI: 10.1103/PhysRevD.99.095007. arXiv: 1901.05982 [hep-ph].

[49] G. Altarelli and G. Parisi. "Asymptotic Freedom in Parton Language". *Nucl. Phys. B* 126 (1977), pp. 298–318. DOI: 10.1016/0550-3213(77)90384-4.

[50] Richard D. Ball et al. "Parton distributions for the LHC Run II". *JHEP* 04 (2015), p. 040. DOI: 10.1007/JHEP04(2015)040. arXiv: 1410.8849 [hep-ph].

[51] Stefano Frixione, Paolo Nason, and Carlo Oleari. "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method". *JHEP* 11 (2007), p. 070. DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph].

[52] 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), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph].

[53] Torbjörn Sjöstrand et al. "An introduction to PYTHIA 8.2". *Comput. Phys. Commun.* 191 (2015), p. 159. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].

[54] Johannes Bellm et al. "Herwig 7.0/Herwig++ 3.0 release note". *Eur. Phys. J. C* 76.4 (2016), p. 196. DOI: 10.1140/epjc/s10052-016-4018-8. arXiv: 1512.01178 [hep-ph].

[55] Christian Bierlich et al. "A comprehensive guide to the physics and usage of PYTHIA 8.3" (Mar. 2022). DOI: 10.21468/SciPostPhysCodeb.8. arXiv: 2203.11601 [hep-ph].

[56] Albert M Sirunyan et al. "A measurement of the Higgs boson mass in the diphoton decay channel". *Phys. Lett. B* 805 (2020), p. 135425. DOI: `10.1016/j.physletb.2020.135425`. arXiv: `2002.06398 [hep-ex]`.

[57] LHC Higgs Cross Section WG. *LHC Higgs Cross Section WG Picture Gallery.* `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHWGCrossSectionsFigures`. Accessed: 23/08/2023.

[58] The CMS Collaboration. "A portrait of the Higgs boson by the CMS experiment ten years after the discovery". *Nature* 607.7917 (2022), pp. 60–68. DOI: `10.1038/s41586-022-04892-x`. arXiv: `2207.00043 [hep-ex]`.

[59] The ATLAS Collaboration. "A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery". *Nature* 607.7917 (2022). [Erratum: Nature 612, E24 (2022)], pp. 52–59. DOI: `10.1038/s41586-022-04893-w`. arXiv: `2207.00092 [hep-ex]`.

[60] Vardan Khachatryan et al. "Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV". *Eur. Phys. J. C* 75.5 (2015), p. 212. DOI: `10.1140/epjc/s10052-015-3351-7`. arXiv: `1412.8662 [hep-ex]`.

[61] Georges Aad et al. "Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at $\sqrt{s} = 7$ and 8 TeV in the ATLAS experiment". *Eur. Phys. J. C* 76.1 (2016), p. 6. DOI: `10.1140/epjc/s10052-015-3769-y`. arXiv: `1507.04548 [hep-ex]`.

[62] Albert M Sirunyan et al. "Inclusive search for highly boosted Higgs bosons decaying to bottom quark-antiquark pairs in proton-proton collisions at $\sqrt{s} = $ 13 TeV". *JHEP* 12 (2020), p. 085. DOI: `10.1007/JHEP12(2020)085`. arXiv: `2006.13251 [hep-ex]`.

[63] Georges Aad et al. "Measurements of Higgs bosons decaying to bottom quarks from vector boson fusion production with the ATLAS experiment at $\sqrt{s} = $ 13 TeV". *Eur. Phys. J. C* 81.6 (2021), p. 537. DOI: `10.1140/epjc/s10052-021-09192-8`. arXiv: `2011.08280 [hep-ex]`.

[64] Albert M Sirunyan et al. "Evidence for Higgs boson decay to a pair of muons". *JHEP* 01 (2021), p. 148. DOI: `10.1007/JHEP01(2021)148`. arXiv: `2009.04363 [hep-ex]`.

[65] Makoto Kobayashi and Toshihide Maskawa. "CP Violation in the Renormalizable Theory of Weak Interaction". *Prog. Theor. Phys.* 49 (1973), pp. 652–657. DOI: `10.1143/PTP.49.652`.

[66] S. Abachi et al. "Observation of the top quark". *Phys. Rev. Lett.* 74 (1995), pp. 2632–2637. DOI: `10.1103/PhysRevLett.74.2632`. arXiv: `hep-ex/9503003`.

[67] F. Abe et al. "Observation of top quark production in $\bar{p}p$ collisions". *Phys. Rev. Lett.* 74 (1995), pp. 2626–2631. DOI: `10.1103/PhysRevLett.74.2626`. arXiv: `hep-ex/9503002`.

[68] Michal Czakon and Alexander Mitov. "Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders". *Comput. Phys. Commun.* 185 (2014), p. 2930. DOI: `10.1016/j.cpc.2014.06.021`. arXiv: `1112.5675 [hep-ph]`.

[69] LHC Top Working Group. *LHCTopWG Summary Plots.* `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCTopWGSummaryPlots`. Accessed: 20/09/2023. 2023.

[70] Georges Aad et al. "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), p. 124. DOI: `10.1007/JHEP07(2020)124`. arXiv: `2002.07546 [hep-ex]`.

[71] Albert M Sirunyan et al. "Observation of Single Top Quark Production in Association with a Z Boson in Proton-Proton Collisions at $\sqrt{s}$ =13 TeV". *Phys. Rev. Lett.* 122.13 (2019), p. 132003. DOI: `10.1103/PhysRevLett.122.132003`. arXiv: `1812.05900 [hep-ex]`.

[72] Georges Aad et al. "Observation of four-top-quark production in the multilepton final state with the ATLAS detector". *Eur. Phys. J. C* 83.6 (2023), p. 496. DOI: `10.1140/epjc/s10052-023-11573-0`. arXiv: `2303.15061 [hep-ex]`.

[73] Armen Tumasyan et al. "Evidence for four-top quark production in proton-proton collisions at $\sqrt{s} = 13$ TeV". *Phys. Lett. B* 844 (2023), p. 138076. DOI: `10.1016/j.physletb.2023.138076`. arXiv: `2303.03864 [hep-ex]`.

[74] Qing-Hong Cao, Shao-Long Chen, and Yandong Liu. "Probing Higgs Width and Top Quark Yukawa Coupling from $t\bar{t}H$ and $t\bar{t}t\bar{t}$ Productions". *Phys. Rev. D* 95.5 (2017), p. 053004. DOI: `10.1103/PhysRevD.95.053004`. arXiv: `1602.01934 [hep-ph]`.

[75] Stephen Myers. "The LEP collider, from design to approval and commissioning" (Oct. 1991). DOI: `10.5170/CERN-1991-008`.

[76] Lyndon Evans and Philip Bryant. "LHC Machine". *Journal of Instrumentation* 3.08 (2008), S08001. DOI: `10.1088/1748-0221/3/08/S08001`.

[77] K. Schindl. "The injector chain for the LHC". *9th LEP Performance Workshop.* CERN-PS-99-018-DI, CERN-PS-99-18-DI. Mar. 1999, pp. 47–52.

[78] The CMS Collaboration. *CMS Luminosity - Public results.* `https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults`. Accessed: 01/06/2023.

[79] G. L. Bayatian et al. "CMS Physics: Technical Design Report Volume 1: Detector Performance and Software". CERN-LHCC-2006-001, CMS-TDR-8-1, CERN-LHCC-2006-001, CMS-TDR-8-1 (2006). URL: https://cds.cern.ch/record/922757.

[80] Aram Hayrapetyan et al. "Development of the CMS detector for the CERN LHC Run 3" (Sept. 2023). arXiv: 2309.05466 [physics.ins-det].

[81] Tai Sakuma. "Cutaway diagrams of CMS detector" (2019). URL: https://cds.cern.ch/record/2665537.

[82] M. A. Akl et al. *CMS Technical Design Report for the Muon Endcap GEM Upgrade*. Tech. rep. CERN-LHCC-2015-012. 2015. URL: https://cds.cern.ch/record/2021453.

[83] CERN. *LHC Computing Grid*. 2023. URL: http://www.cern.ch/lcg/.

[84] A. M. Sirunyan et al. "Particle-flow reconstruction and global event description with the CMS detector". *JINST* 12.10 (2017), P10003. DOI: 10.1088/1748-0221/12/10/P10003. arXiv: 1706.04965 [physics.ins-det].

[85] W. Adam et al. *Track reconstruction in the CMS tracker*. Tech. rep. CERN-CMS-NOTE-2006-041. 2005. URL: https://cds.cern.ch/record/934067.

[86] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The anti-k_t jet clustering algorithm". *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].

[87] Emil Bols et al. "Jet Flavour Classification Using DeepJet". *JINST* 15.12 (2020), P12012. DOI: 10.1088/1748-0221/15/12/P12012. arXiv: 2008.10519 [hep-ex].

[88] The CMS Collaboration. "Performance of the DeepJet b tagging algorithm using 41.9/fb of data from proton-proton collisions at 13TeV with Phase 1 CMS detector". CMS-DP-2018-058 (2018). URL: http://cds.cern.ch/record/2646773.

[89] Albert M Sirunyan et al. "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector". *JINST* 14.07 (2019), P07004. DOI: 10.1088/1748-0221/14/07/P07004. arXiv: 1903.06078 [hep-ex].

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

[91]   Vardan Khachatryan et al. "Measurements of inclusive W and Z cross sections in pp collisions at $\sqrt{s} = 7$ TeV". *JHEP* 01 (2011), p. 080. DOI: `10.1007/JHEP01(2011)080`. arXiv: `1012.2466 [hep-ex]`.

[92]   Vardan Khachatryan et al. "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV". *JINST* 12.02 (2017), P02014. DOI: `10.1088/1748-0221/12/02/P02014`. arXiv: `1607.03663 [hep-ex]`.

[93]   A. M. Sirunyan et al. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV". *JINST* 13.05 (2018), P05011. DOI: `10.1088/1748-0221/13/05/P05011`. arXiv: `1712.07158 [physics.ins-det]`.

[94]   Anita Prinzie and Dirk Van den Poel. "Random multiclass classification: Generalizing random forests to random MNL and random NB". *Lecture Notes in Computer Science* 4653 (2007), p. 349. DOI: `10.1007/978-3-540-74469-6_35`.

[95]   Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning.* Springer Series in Statistics. New York, NY, USA: Springer New York Inc., 2001.

[96]   Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning.* `http://www.deeplearningbook.org`. MIT Press, 2016.

[97]   A. M. Sirunyan et al. "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV". *JINST* 13.06 (2018), P06015. DOI: `10.1088/1748-0221/13/06/P06015`. arXiv: `1804.04528 [physics.ins-det]`.

[98]   R. Fruhwirth. "Application of Kalman filtering to track and vertex fitting". *Nucl. Instrum. Meth. A* 262 (1987), pp. 444–450. DOI: `10.1016/0168-9002(87)90887-4`.

[99]   Albert M Sirunyan et al. "Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV". *JINST* 15.10 (2020), P10017. DOI: `10.1088/1748-0221/15/10/P10017`. arXiv: `2006.10165 [hep-ex]`.

[100] Vardan Khachatryan et al. "Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV". *JINST* 10.06 (2015), P06005. DOI: `10.1088/1748-0221/10/06/P06005`. arXiv: `1502.02701 [physics.ins-det]`.

[101] Wolfgang Adam et al. "Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC". CMS-NOTE-2005-001, CERN-CMS-NOTE-2005-001 (Jan. 2005). URL: `https://cds.cern.ch/record/815410`.

[102] The CMS Collaboration. "Muon identification and isolation efficiencies with 2017 and 2018 data". CMS-DP-2018-042 (2018). URL: https://cds.cern.ch/record/2629364.

[103] The CMS Collaboration. "Muon reconstruction performance during Run II". CMS-DP-2019-022 (2019). URL: http://cds.cern.ch/record/2682902.

[104] Albert M Sirunyan et al. "Measurement of properties of $B_s^0 \to \mu^+\mu^-$ decays and search for $B^0 \to \mu^+\mu^-$ with the CMS experiment". *JHEP* 04 (2020), p. 188. DOI: 10.1007/JHEP04(2020)188. arXiv: 1910.12127 [hep-ex].

[105] Albert M Sirunyan et al. "Observation of t$\bar{\text{t}}$H production". *Phys. Rev. Lett.* 120.23 (2018), p. 231801. DOI: 10.1103/PhysRevLett.120.231801. arXiv: 1804.02610 [hep-ex].

[106] Aram Hayrapetyan et al. "Observation of four top quark production in proton-proton collisions at s=13TeV". *Phys. Lett. B* 847 (2023), p. 138290. DOI: 10.1016/j.physletb.2023.138290. arXiv: 2305.13439 [hep-ex].

[107] Albert M Sirunyan et al. "Measurement of the associated production of a single top quark and a Z boson in pp collisions at $\sqrt{s} = 13$ TeV". *Phys. Lett. B* 779 (2018), pp. 358–384. DOI: 10.1016/j.physletb.2018.02.025. arXiv: 1712.02825 [hep-ex].

[108] Armen Tumasyan et al. "Measurement of the inclusive and differential WZ production cross sections, polarization angles, and triple gauge couplings in pp collisions at $\sqrt{s} = 13$ TeV". *JHEP* 07 (2022), p. 032. DOI: 10.1007/JHEP07(2022)032. arXiv: 2110.11231 [hep-ex].

[109] Armen Tumasyan et al. "Measurements of the electroweak diboson production cross sections in proton-proton collisions at $\sqrt{s} = 5.02$ TeV using leptonic decays". *Phys. Rev. Lett.* 127 (2021), p. 191801. DOI: 10.1103/PhysRevLett.127.191801. arXiv: 2107.01137 [hep-ex].

[110] Albert M Sirunyan et al. "Measurements of the pp $\to$ WZ inclusive and differential production cross section and constraints on charged anomalous triple gauge couplings at $\sqrt{s} = 13$ TeV". *JHEP* 04 (2019), p. 122. DOI: 10.1007/JHEP04(2019)122. arXiv: 1901.03428 [hep-ex].

[111] Armen Tumasyan et al. "Search for electroweak production of charginos and neutralinos in proton-proton collisions at $\sqrt{s} = 13$ TeV". *JHEP* 04 (2022), p. 147. DOI: 10.1007/JHEP04(2022)147. arXiv: 2106.14246 [hep-ex].

[112] Aram Hayrapetyan et al. "Muon identification using multivariate techniques in the CMS experiment in proton-proton collisions at $\sqrt{s} = 13$ TeV". Accepted by JINST (Oct. 2023). arXiv: 2310.03844 [hep-ex].

[113] Andreas Hocker et al. "TMVA - Toolkit for Multivariate Data Analysis" (Mar. 2007). arXiv: physics/0703039.

[114] The CMS Collaboration. *Performance of reconstruction and identification of tau leptons in their decays to hadrons and tau neutrino in LHC Run-2*. Tech. rep. Geneva: CERN, 2016. URL: https://cds.cern.ch/record/2196972.

[115] The CMS Collaboration. "Performance of the DeepTau algorithm for the discrimination of taus against jets, electron, and muons" (2019). URL: https://cds.cern.ch/record/2694158.

[116] A. M. Sirunyan et al. "Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV". *JINST* 13.10 (2018), P10005. DOI: 10.1088/1748-0221/13/10/P10005. arXiv: 1809.02816 [hep-ex].

[117] D. de Florian et al. "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector". 2/2017 (Oct. 2016). DOI: 10.23731/CYRM-2017-002. arXiv: 1610.07922 [hep-ph].

[118] Stefano Catani et al. "Higgs Boson Production in Association with a Top-Antitop Quark Pair in Next-to-Next-to-Leading Order QCD". *Phys. Rev. Lett.* 130.11 (2023), p. 111902. DOI: 10.1103/PhysRevLett.130.111902. arXiv: 2210.07846 [hep-ph].

[119] Federico Demartin et al. "tWH associated production at the LHC". *Eur. Phys. J. C* 77.1 (2017), p. 34. DOI: 10.1140/epjc/s10052-017-4601-7. arXiv: 1607.05862 [hep-ph].

[120] J R Andersen et al. "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties" (July 2013). Ed. by S Heinemeyer et al. DOI: 10.5170/CERN-2013-004. arXiv: 1307.1347 [hep-ph].

[121] P. Artoisenet et al. "A framework for Higgs characterisation". *JHEP* 11 (2013), p. 043. DOI: 10.1007/JHEP11(2013)043. arXiv: 1306.6464 [hep-ph].

[122] Vardan Khachatryan et al. "Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV". *Phys. Rev. D* 92.1 (2015), p. 012004. DOI: 10.1103/PhysRevD.92.012004. arXiv: 1411.3441 [hep-ex].

[123] Federico Demartin et al. "Higgs production in association with a single top quark at the LHC". *Eur. Phys. J. C* 75.6 (2015), p. 267. DOI: 10.1140/epjc/s10052-015-3475-9. arXiv: 1504.00611 [hep-ph].

[124] Fabio Maltoni et al. "Trilinear Higgs coupling determination via single-Higgs differential measurements at the LHC". *Eur. Phys. J. C* 77.12 (2017), p. 887. DOI: 10.1140/epjc/s10052-017-5410-8. arXiv: 1709.08649 [hep-ph].

[125] M. Aaboud et al. "Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector". *Phys. Lett. B* 784 (2018), pp. 173–191. DOI: 10.1016/j.physletb.2018.07.035. arXiv: 1806.00425 [hep-ex].

[126] Albert M Sirunyan et al. "Measurements of $t\bar{t}H$ Production and the CP Structure of the Yukawa Interaction between the Higgs Boson and Top Quark in the Diphoton Decay Channel". *Phys. Rev. Lett.* 125.6 (2020), p. 061801. DOI: 10.1103/PhysRevLett.125.061801. arXiv: 2003.10866 [hep-ex].

[127] Georges Aad et al. "*CP* Properties of Higgs Boson Interactions with Top Quarks in the $t\bar{t}H$ and tH Processes Using $H \rightarrow \gamma\gamma$ with the ATLAS Detector". *Phys. Rev. Lett.* 125.6 (2020), p. 061802. DOI: 10.1103/PhysRevLett.125.061802. arXiv: 2004.04545 [hep-ex].

[128] The CMS Collaboration. "Measurement of the $t\bar{t}H$ and tH production rates in the $H \rightarrow b\bar{b}$ decay channel with $138\,\mathrm{fb}^{-1}$ of proton-proton collision data at $\sqrt{s} = 13\,\mathrm{TeV}$". CMS-PAS-HIG-19-011 (2023). URL: http://cds.cern.ch/record/2868175.

[129] Morad Aaboud et al. "Search for the standard model Higgs boson produced in association with top quarks and decaying into a $b\bar{b}$ pair in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". *Phys. Rev. D* 97.7 (2018), p. 072016. DOI: 10.1103/PhysRevD.97.072016. arXiv: 1712.08895 [hep-ex].

[130] The ATLAS Collaboration. "Probing the *CP* nature of the top-Higgs Yukawa coupling in $t\bar{t}H$ and tH events with $H \rightarrow b\bar{b}$ decays using the ATLAS detector at the LHC" (Mar. 2023). arXiv: 2303.05974 [hep-ex].

[131] Albert M Sirunyan et al. "Measurements of production cross sections of the Higgs boson in the four-lepton final state in proton–proton collisions at $\sqrt{s} = 13\,\mathrm{TeV}$". *Eur. Phys. J. C* 81.6 (2021), p. 488. DOI: 10.1140/epjc/s10052-021-09200-x. arXiv: 2103.04956 [hep-ex].

[132] Georges Aad et al. "Higgs boson production cross-section measurements and their EFT interpretation in the 4ℓ decay channel at $\sqrt{s} =13$ TeV with the ATLAS detector". *Eur. Phys. J. C* 80.10 (2020). [Erratum: Eur.Phys.J.C 81, 29 (2021), Erratum: Eur.Phys.J.C 81, 398 (2021)], p. 957. DOI: 10.1140/epjc/s10052-020-8227-9. arXiv: 2004.03447 [hep-ex].

[133] Albert M Sirunyan et al. "Evidence for associated production of a Higgs boson with a top quark pair in final states with electrons, muons, and hadronically decaying τ leptons at $\sqrt{s} = 13$ TeV". *JHEP* 08 (2018), p. 066. DOI: 10.1007/JHEP08(2018)066. arXiv: 1803.05485 [hep-ex].

[134] Albert M Sirunyan et al. "Measurement of top quark pair production in association with a Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV". *JHEP* 03 (2020), p. 056. DOI: 10.1007/JHEP03(2020)056. arXiv: 1907.11270 [hep-ex].

[135] Georges Aad et al. "Measurements of the inclusive and differential production cross sections of a top-quark–antiquark pair in association with a Z boson at $\sqrt{s} = 13$ TeV with the ATLAS detector". *Eur. Phys. J. C* 81.8 (2021), p. 737. DOI: 10.1140/epjc/s10052-021-09439-4. arXiv: 2103.12603 [hep-ex].

[136] The ATLAS Collaboration. *Inclusive and differential cross section measurements of $t\bar{t}Z$ production in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, including EFT and spin correlations interpretations*. Tech. rep. ATLAS-CONF-2023-065. 2023. URL: http://cds.cern.ch/record/2873519.

[137] Armen Tumasyan et al. "Measurement of the inclusive and differential $t\bar{t}\gamma$ cross sections in the dilepton channel and effective field theory interpretation in proton-proton collisions at $\sqrt{s}$ =13 TeV". *JHEP* 05 (2022), p. 091. DOI: 10.1007/JHEP05(2022)091. arXiv: 2201.07301 [hep-ex].

[138] Armen Tumasyan et al. "Measurement of the inclusive and differential $t\bar{t}\gamma$ cross sections in the single-lepton channel and EFT interpretation at $\sqrt{s} = 13$ TeV". *JHEP* 12 (2021), p. 180. DOI: 10.1007/JHEP12(2021)180. arXiv: 2107.01508 [hep-ex].

[139] Georges Aad et al. "Measurements of inclusive and differential cross-sections of combined $t\bar{t}\gamma$ and $tW\gamma$ production in the eμ channel at 13 TeV with the ATLAS detector". *JHEP* 09 (2020), p. 049. DOI: 10.1007/JHEP09(2020)049. arXiv: 2007.06946 [hep-ex].

[140] Georges Aad et al. "Search for leptonic charge asymmetry in $t\bar{t}W$ production in final states with three leptons at $\sqrt{s} = 13$ TeV". *JHEP* 07 (2023), p. 033. DOI: 10.1007/JHEP07(2023)033. arXiv: 2301.04245 [hep-ex].

[141] G. Aad et al. "Measurement of the $t\bar{t}W$ and $t\bar{t}Z$ production cross sections in pp collisions at $\sqrt{s}$ [8] with the ATLAS detector". *JHEP* 11 (2015), p. 172. DOI: 10.1007/JHEP11(2015)172. arXiv: 1509.05276 [hep-ex].

[142] Vardan Khachatryan et al. "Observation of top quark pairs produced in association with a vector boson in pp collisions at $\sqrt{s} = 8$ TeV". *JHEP* 01 (2016), p. 096. DOI: 10.1007/JHEP01(2016)096. arXiv: 1510.01131 [hep-ex].

[143] Georges Aad et al. "Evidence for $t\bar{t}t\bar{t}$ production in the multilepton final state in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". *Eur. Phys. J. C* 80.11 (2020), p. 1085. DOI: 10.1140/epjc/s10052-020-08509-3. arXiv: 2007.14858 [hep-ex].

[144] D. de Florian et al. "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector". 2/2017 (Oct. 2016). DOI: 10.23731/CYRM-2017-002. arXiv: 1610.07922 [hep-ph].

[145] Rikkert Frederix and Ioannis Tsinikos. "Subleading EW corrections and spin-correlation effects in $t\bar{t}W$ multi-lepton signatures". *Eur. Phys. J. C* 80.9 (2020), p. 803. DOI: 10.1140/epjc/s10052-020-8388-6. arXiv: 2004.09552 [hep-ph].

[146] John M. Campbell, R. Keith Ellis, and Ciaran Williams. "Vector boson pair production at the LHC". *JHEP* 07 (2011), p. 018. DOI: 10.1007/JHEP07(2011)018. arXiv: 1105.0020 [hep-ph].

[147] LHC Higgs Combination Group ATLAS CMS. *Procedure for the LHC Higgs boson search combination in Summer 2011*. Tech. rep. CMS-NOTE-2011-005, ATL-PHYS-PUB-2011-011. Aug. 2011. URL: https://cds.cern.ch/record/1379837.

[148] J. S. Conway. "Incorporating Nuisance Parameters in Likelihoods for Multi-source Spectra". *PHYSTAT 2011*. 2011, pp. 115–120. DOI: 10.5170/CERN-2011-006.115. arXiv: 1103.0354 [physics.data-an].

[149] S. S. Wilks. "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses". *Annals Math. Statist.* 9.1 (1938), pp. 60–62. DOI: 10.1214/aoms/1177732360.

[150] The CMS Collaboration. *CMS luminosity measurements for the 2016 data-taking period*. CMS Physics Analysis Summary CMS-PAS-LUM-17-001. 2017. URL: http://cds.cern.ch/record/2257069.

[151] The 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. 2017. URL: http://cds.cern.ch/record/2621960.

[152] The CMS Collaboration. *CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13\,TeV$*. CMS Physics Analysis Summary CMS-PAS-LUM-18-002. 2018. URL: http://cds.cern.ch/record/2676164.

[153] Albert M Sirunyan et al. "Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS". *Eur. Phys. J. C* 81.9 (2021), p. 800. DOI: 10.1140/epjc/s10052-021-09538-2. arXiv: 2104.01927 [hep-ex].

[154] Albert M Sirunyan et al. "Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV". *JHEP* 07 (2018), p. 161. DOI: `10.1007/JHEP07(2018)1 61`. arXiv: `1802.02613 [hep-ex]`.

[155] A. M. Sirunyan et al. "Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV". *JINST* 13.10 (2018), P10005. DOI: `10.1088/1748-0221/13/10/P10005`. arXiv: `1809.028 16 [hep-ex]`.

[156] Jon Butterworth et al. "PDF4LHC recommendations for LHC Run II". *J. Phys. G* 43 (2016), p. 023001. DOI: `10.1088/0954-3899/43/2/023001`. arXiv: `1510.03865 [hep-ph]`.

[157] Armen Tumasyan et al. "CMS pythia 8 colour reconnection tunes based on underlying-event data". *Eur. Phys. J. C* 83.7 (2023), p. 587. DOI: `10.1140/epjc/s 10052-023-11630-8`. arXiv: `2205.02905 [hep-ex]`.

[158] Armen Tumasyan et al. "Inclusive and differential cross section measurements of single top quark production in association with a Z boson in proton-proton collisions at $\sqrt{s} = 13$ TeV". *JHEP* 02 (2022), p. 107. DOI: `10.1007/JHEP02(2022 )107`. arXiv: `2111.02860 [hep-ex]`.

[159] A. M. Sirunyan et al. "Search for electroweak production of charginos and neutralinos in multilepton final states in proton-proton collisions at $\sqrt{s} = 13$ TeV". *JHEP* 03 (2018), p. 166. DOI: `10.1007/JHEP03(2018)166`. arXiv: `1709.054 06 [hep-ex]`.

[160] Roger J. Barlow and Christine Beeston. "Fitting using finite Monte Carlo samples". *Comput. Phys. Commun.* 77 (1993), pp. 219–228. DOI: `10.1016/0010-4655 (93)90005-W`.

[161] The ATLAS Collaboration. *Measurement of the total and differential cross-sections of $t\bar{t}W$ production in pp collisions at 13 TeV with the ATLAS detector*. Tech. rep. ATLAS-CONF-2023-019. 2023. URL: `http://cds.cern.ch/record/2855337`.

[162] Alessandro Broggio et al. "Top-quark pair hadroproduction in association with a heavy boson at NLO+NNLL including EW corrections". *JHEP* 08 (2019), p. 039. DOI: `10.1007/JHEP08(2019)039`. arXiv: `1907.04343 [hep-ph]`.

[163] Huilin Qu and Loukas Gouskos. "ParticleNet: Jet Tagging via Particle Clouds". *Phys. Rev. D* 101.5 (2020), p. 056019. DOI: `10.1103/PhysRevD.101.056019`. arXiv: `1902.08570 [hep-ph]`.

[164] Sergio Sánchez Cruz. "Measurements of $t\bar{t}H$ and tH associated production at CMS". *PoS* ICHEP2020 (2021), p. 080. DOI: `10.22323/1.390.0080`.

[165] Lorenzo Bianchini et al. "Reconstruction of the Higgs mass in events with Higgs bosons decaying into a pair of τ leptons using matrix element techniques". *Nucl. Instrum. Meth. A* 862 (2017), pp. 54–84. DOI: `10.1016/j.nima.2017.05.001`. arXiv: `1603.05910` `[hep-ex]`.

[166] Albert M Sirunyan et al. "Constraints on anomalous Higgs boson couplings to vector bosons and fermions in its production and decay using the four-lepton final state". *Phys. Rev. D* 104.5 (2021), p. 052004. DOI: `10.1103/PhysRevD.104.052004`. arXiv: `2104.12152` `[hep-ex]`.