Measurement of the associated production of a W boson and a charm quark in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ with the CMS detector at the LHC

Medida de la producción asociada de un bosón W y un quark charm en colisiones protón-protón a $\sqrt{s} = 13\,\text{TeV}$ con el detector CMS en el LHC

Realizado por

Sergio Sánchez Navas

Dirigido por

Dr. Juan Pablo Fernández Ramos

Dr. José María Hernández Calama

UNIVERSIDAD
COMPLUTENSE
MADRID

Facultad de Ciencias Físicas

Madrid, marzo 2023.

Measurement of the associated production of a W boson and a charm quark in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ with the CMS detector at the LHC

Medida de la producción asociada de un bosón W y un quark charm en colisiones protón-protón a $\sqrt{s} = 13\,\text{TeV}$ con el detector CMS en el LHC

Realizado por

Sergio Sánchez Navas

Para la obtención del título de Doctor en Física

Dirigido por

Dr. Juan Pablo Fernández Ramos

Dr. José María Hernández Calama

UNIVERSIDAD
COMPLUTENSE
MADRID

Facultad de Ciencias Físicas

Madrid, marzo 2023.

UNIVERSIDAD COMPLUTENSE MADRID

DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD DE LA TESIS PRESENTADA PARA OBTENER EL TÍTULO DE DOCTOR

D./Dña. Sergio Sánchez Navas,
estudiante en el Programa de Doctorado en Física,
de la Facultad de Ciencias Físicas de la Universidad Complutense de Madrid, como autor/a de la tesis presentada para la obtención del título de Doctor y titulada:

"Medida de la producción asociada de un bosón W y un quark charm en colisiones protón-protón a $\sqrt{s}$ = 13TeV con el detector CMS en el LHC"

"Measurement of the associated production of a W boson and a charm quark in proton-proton collisions at $\sqrt{s}=13$ TeV with the CMS detector at the LHC"

y dirigida por:
Dr. Juan Pablo Fernández Ramos
Dr. José María Hernández Calama

DECLARO QUE:

La tesis es una obra original que no infringe los derechos de propiedad intelectual ni los derechos de propiedad industrial u otros, de acuerdo con el ordenamiento jurídico vigente, en particular, la Ley de Propiedad Intelectual (R.D. legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, modificado por la Ley 2/2019, de 1 de marzo, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia), en particular, las disposiciones referidas al derecho de cita.

Del mismo modo, asumo frente a la Universidad cualquier responsabilidad que pudiera derivarse de la autoría o falta de originalidad del contenido de la tesis presentada de conformidad con el ordenamiento jurídico vigente.

En Madrid, a 15 de diciembre de 2022

Fdo.:

Esta DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD debe ser insertada en la primera página de la tesis presentada para la obtención del título de Doctor.

Agradecimientos

El desarrollo de este trabajo se ha realizado en las instalaciones del CIEMAT, situado en las inmediaciones del Campus de Ciudad Universitaria de la Universidad Complutense de Madrid. En la Facultad de Físicas de esta universidad inicié mi Licenciatura en Física en 2008 y aquí termino mi Doctorado en Física en 2023. Quiero aprovechar estas líneas para recordar a toda la gente que, de una forma u otra, ha contribuido a que mi trayectoria haya completado este círculo, o más bien una helicoidal, ya que no se vuelve al punto de partida sino mucho más adelante. Me gustaría emplear nombres propios pero probablemente sería una lista demasiado extensa, además del inasumible riesgo de olvidar a alguien por el camino y, por ello, ser condenado al más absoluto ostracismo.

First and foremost, como se dice en la lengua en que se ha redactado este trabajo, me gustaría expresar mi gratitud a mis directores de tesis, Juan Pablo y Chema, por su ayuda y consejo durante el desarrollo de este análisis, proporcionando no sólo códigos, twikis y referencias, sino también muy valiosas aportaciones en términos de la forma y el estilo a la hora de redactar la memoria; del contenido y del continente.

Me gustaría también agradecer el tiempo compartido con mis compañeros del CIEMAT, necesariamente segregados en la Becacueva, el Despachito y despacho de Neutrinos. Guardo muy buenos recuerdos de las excursiones al surf y al snowboard, de los *general meetings* y de los *after work*. También me guardo un poco de radón y una clavícula rota.

De las dos estancias en el CERN también me quedo con todo lo bueno que viví y lo bastante que aprendí, desde ponerme el casco para ir a por café hasta echar gasolina con un token electrónico. Curiosamente, mi primera visita al CERN fue en 2015 como turista y finalmente acabé trabajando allí durante siete meses. Cómo me acabó gustando Ginebra. Plus vite, La Citadelle!

Como causa directa de mi doctorado en Madrid debo mencionar mis tres años de estancia en Trondheim, Noruega. Allí cursé el máster en la Norges Teknisk-Naturvitenskapelige Universitet y quiero recordar a mis compañeros de la Leserom B3-136, con litros de café aguado noruego y no sabiendo muy bien si era por la mañana o era por la tarde, tanto por la eterna noche invernal como por la ausencia absoluta de oscuridad durante el verano. De allí también me traigo las auroras boreales, la barba congelada y un lamentable intento de acudir a clase en esquí de fondo tras una copiosa nevada. También saludar a los amigos que hice en Fru Lundgreens y durante mis curros en Good Omens y la Aktivitetshuset. Takk for meg! Skål!

Por supuesto, debo incluir en estos agradecimientos a todos los físicos de la UCM con los que pasé los primeros años de vida adulta. Un abrazo muy fuerte en especial a mis Zambranetis y Fish&Chips, con quienes hemos abierto y cerrado las

bibliotecas del campus, la piscina de verano y alguna que otra taberna. Últimamente nos vemos más a menudo, de boda en boda. ¡Muy braf por tod@s!

Necesariamente tiene que estar mencionado mi año de Erasmus en Gante, Bélgica. No fue el estilo de experiencia que se suele promocionar en los folletos universitarios, pero por supuesto aprendí lecciones muy valiosas en lo personal y tuve el honor de conocer a gente maravillosa. Una Rochefort 10 a tu salud y otra a tu memoria. Op ow muile!

También me gustaría remontarme un poco más atrás en el tiempo y recordar con cariño a mis amigos de la Secta Estudio, con los que crecí y aprendí a ser quien soy, dentro de unas razonables incertidumbres. A los que seguimos al pie del cañón, ¡salud!

A mi querido grupo de frikis de la Legión 501 - Spanish Garrison, por más eventos de Star Wars, más excursiones por toda la geografía peninsular, más trajes y más sonrisas.

A quienes me rodean hoy día, cada cual más de ahora o más de antes, muchas gracias por apoyarme en estos años de locura científica. Valoro cada cerveza fría y cada café caliente y sois parte de todo esto.

En otro orden de cosas, más etéreas, me gustaría también mostrar mi agradecimiento a la música, tanto la que he compuesto, como interpretado, como escuchado. Es fuente de inspiración, de control y de desahogo. También quiero reservar una línea al Muay Thai, porque *mens sana in corpore sano* con un ligero sazonado de violencia comedida.

No puede faltar un abrazo muy fuerte a mi familia por la confianza depositada y la paciencia invertida y a ellos sí quiero nombrarles. A mis padres Sonsoles y Juan Carlos por mostrarme el camino que yo mismo debía recorrer y por creer en mí en cualquier circunstancia. A mis tíos y tías Miguel y María Luisa, Mariángeles y Oli, a mi primo Miguel Ángel y mis primas Carmen, Lidia y Katia por las charlas sobre ciencia y demás asuntos del Universo. A mi primo Carlos, por enseñarme los primeros compases de batería y por resistir, ahora y siempre, contra todo y contra todos en Bernuy. A mis abuelas Deme y Teo, centenarias y centurionas, que me quieren y me cuidan desde que abrí los ojos.

Last but not least, quiero mostrar mi afecto y mi cariño a María, la *socia* que diría aquel señor de Pamplona, quien se ha convertido por méritos propios en un pilar fundamental del orden del caos. El último minuto se arrastra por la esfera cuando todo brilla con luz artificial.

...a mi abuelo Miguel...

Contents

List of Figures

List of Tables

Chapter 1

Introduction

"Die Wissenschaft sorgt dafür, dass Menschen Wahrheit und Objektivität erreichen; sie lehrt die Menschen, die Realität mit Staunen und Bewunderung zu akzeptieren, ganz zu schweigen von der tiefen Ehrfurcht und Freude vor der natürlichen Ordnung der Dinge."
— L. Meitner, Atomenergie und Frieden.

Forty years ago, the discovery of the W boson was announced following the analysis of the data from the Super Proton Synchrotron (SPS) at CERN [1]. Almost fifty years ago, the discovery of the J/Ψ meson at SLAC [2] and BNL [3] confirmed the existence of a fourth quark, the charm quark (c). Today, we aim to continue the collective effort to unveil the mysteries of the smallest corners of Nature, by studying the associated production of a W boson and a charm quark in order to extract information about the inner structure of the proton.

The proton is a quantum chromodynamics (QCD) bound state of quarks and gluons. The proton constituents, the partons, collectively give rise to its fundamental properties, such as its mass and spin. The complex proton structure has been a source of fundamental discoveries, like the experimental proof of the existence of quarks or the nature of the quark-gluon plasma state of matter. The proton energy shared by its constituents is described through the parton distribution functions (PDF). A precise knowledge of the content of protons is a necessary ingredient for making accurate predictions of both standard model (SM) and beyond-the-SM processes.

The associated production of a W boson and a single charm quark in proton-proton collisions is directly sensitive to the strange quark content of the colliding protons at an energy scale of the order of the W boson mass [4]. The sensitivity comes from the dominance of the sg→W+c contribution over the strongly Cabibbo suppressed process dg→W+c at tree level. Studies of this process therefore provide valuable information on the strange quark parton distribution function (PDF), which is one of the least constrained PDFs of the proton. Accurate measurements of the W+c production cross section and of the $R_c^{\pm} = \sigma(\mathrm{W}^+ + \bar{\mathrm{c}})/\sigma(\mathrm{W}^- + \mathrm{c})$ cross section ratio can be used to constrain the ratio between strange and non-strange sea quark PDFs, $R_s = (s + \bar{s})/(\bar{u} + \bar{d})$, and to probe the level of asymmetry between the s and $\bar{s}$ PDFs [5, 6].

If the strange quark contribution to W+c production were totally dominant,

then any deviation of $R_c^{\pm}$ from 1 would imply an asymmetry between s and $\bar{s}$ PDFs. However even if $s = \bar{s}$, the fact that $\bar{d} < d$ in the proton due to the d valence quark, results in $R_c^{\pm} < 1$ through the Cabibbo suppressed d-quark contribution. Therefore, a careful comparison of the $R_c^{\pm}$ measurement with theoretical predictions including all contributions is required to explore a potential strange quark asymmetry in the proton. The $R_c^{\pm}$ measurement and the comparison with theoretical predictions are presented in this work.

The production of W+c events is also a background for SM processes and searches beyond the SM with electroweak bosons and heavy quarks in the final state. The understanding of W+c production is therefore an important ingredient for several measurements at the LHC. Furthermore, precise measurements of W+c production can be used to verify the theoretical calculation of this process and its modeling in the currently available simulations.

The study of the W+c process is an interesting and contemporary field of work, with previous publications by the CMS [7–9], ATLAS [10] and LHCb [11] Collaborations using proton-proton collisions at center-of-mass energies of 7, 8 and 13 TeV. In those analyses, the charm quark is identified using either exclusive decay channels, e.g. $D^+(2010) \to D^0\pi^+$ with $D^0 \to K\pi^+$ (+c.c.), or inclusive final states where semileptonic or hadronic decays of charmed hadrons are recognized through the reconstruction of a muon or a secondary vertex inside a jet. In this study we select four separated data samples depending on the W boson leptonic decay (either to electron or muon, and a neutrino) and the charm tagging mode (muon or secondary vertex inside a jet). The combination of the measurements in the four channels, the use of the large data set collected at $\sqrt{s} = 13\,\text{TeV}$, and the reduction of systematic uncertainties lead to the most precise measurements to date.

For this analysis we are using data collected by the CMS detector between 2016 and 2018 of pp collisions at $\sqrt{s} = 13\,\text{TeV}$ corresponding to an integrated luminosity of $138\,\text{fb}^{-1}$. Inclusive and differential cross sections are measured as functions of the transverse momentum (p_{T}^{ℓ}) and pseudorapidity (η^{ℓ}) of the lepton from the W boson decay. The cross section measurements carried out in this work, together with other existing measurements in pp collisions at the Large Hadron Collider, combined with deep inelastic scattering data from electron-proton collisions at the HERA collider, will allow an improved determination of the strange quark PDF through a complex analysis using the predictions of the QCD theory.

The body of this thesis is divided into several chapters, covering the different theoretical and experimental aspects of the analysis. Chapter 2 is devoted to the theoretical background, with a brief description of the SM, the inner structure of the proton and the computational tools to simulate the physical processes involved in the process under study. Following that, the experimental setup is presented in Chapter 3, introducing the machine that produces the collisions (the Large Hadron Collider) and the detector that registers the particles produced in the collisions (the Compact Muon Solenoid). The reconstruction of the physics objects for the W+c final state is shown in Chapter 4. In Chapter 5 we present the analysis of the W+c production cross section, with emphasis on the event selection, the background determination and the study of the systematic effects affecting the measurements. Fi-

nally, the results of this investigation are presented in Chapter 6, the measurements of the cross sections unfolded to the particle (hadron) and parton levels, compared with predictions based on fixed-order perturbative QCD at the next-to-leading order (NLO) and next-to-NLO accuracies. The conclusions and overview are summarized in Chapter 7.

Chapter 2

Theoretical overview

"Ash nazg durbatulûk, ash nazg gimbatul, ash nazg thrakatulûk, agh burzum-ishi krimpatul."
— J.R.R. Tolkien, The Lord of the Rings: The Fellowship of the Ring.

High energy physics is a lively field of study nowadays. The search for the fundamental answers to the deepest mysteries of the Universe encourages scientists to develop complicated and beautiful theories that explain the observed phenomena and predict new features of Nature. In this chapter we will focus on the theoretical foundations for our topic of research, such as the fundamental particles and their interactions, and the structure of the proton. The mathematical and computational tools to describe them will be addressed in the next few sections.

2.1. The standard model of particle physics

The standard model of particle physics (SM) is the gauge quantum field theory that describes the fundamental constituents of matter and their interactions. It describes three out of the four fundamental forces. The SM was meant to be one theory to rule them all, one theory to find them, one theory to bring them all and in the Grand Unificaton bind them. But unfortunately, gravitation seems too stubborn to bend the knee and hand over the graviton.

As of today, it is the most precise scientific theory in terms of the accuracy between theoretical predictions and empiric measurements. It considers elementary particles as point-like, with no further substructure, arising from quantum fields [12, 13]. Visible matter is composed of fermions, with half-integer spin and obeying the Fermi–Dirac statistics, while bosons, with integer spin and obeying the Bose–Einstein statistics, are responsible for the weak (vector bosons, $W^{\pm}$ and Z), strong (gluons, g) and electromagnetic (photon, γ) interactions, as well as the process that provides matter with mass (Higgs boson, H). Fermions can also be further classified in two categories, quarks and leptons, each with three generations of two particles increasing in mass. The difference between each species is called flavour, and is defined by a set of quantum numbers. Lepton generations consist of a massive lepton (electron, muon and tau) with the electric charge of the electron, and its corresponding neutrino. Neutrinos were expected to be massless following the mathematical description of the theory, but instead it was found that they oscillate

between the three flavours mid-flight, a process that requires massive particles, although they may be small enough. Quark generations consist of an up-type quark with electric charge of +2/3 of that of the electron (up, charm and top), and a down-type quark with electric charge of -1/3 of that of the electron (down, strange and bottom). A corresponding antiparticle with the same mass, but opposite electric charge exists for each type of fermion, and is conventionally denoted by a bar on top of it. The elementary particles described by the SM are summarized in Table 2.1.

Table 2.1: Elementary particles of the SM. Quark and lepton antiparticles are not included.

Fermions		Bosons	
Quarks	Leptons	Gauge Bosons	Scalar Bosons
Up (u)	Electron (e)	Photon (γ)	
Down (d)	Ele. neutrino (ν_e)		
Charm (c)	Muon (μ)	$W^{\pm}$ and Z	Higgs (H)
Strange (s)	Muon neutrino (ν_μ)		
Top (t)	Tau (τ)	Gluons (g)	
Bottom (b)	Tau neutrino (ν_τ)		

As a gauge theory, the SM is described by a Lagrangian that is invariant under local transformations of the $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ symmetry group. The $SU(3)_C$ is the colour gauge group, the fundamental symmetry group of quantum chromodinamics (QCD). This theory describes the strong interaction between quarks as mediated by the exchange of eight massless gluons, also carriers of the colour charge. The $SU(2)_L \otimes U(1)_Y$ gauge group describes the electroweak interactions, a unified description of the electromagnetic and weak forces, where L denotes that only left-handed (negative helicity) particles participate in the interaction and the label Y refers to the weak hypercharge. Is is mediated by three massive vector bosons and one massless photon.

The SM as a whole theory requires 26 free parameters that can be measured experimentally: 3 gauge couplings, 6 mixing angles, 12 masses (6 of the quarks and 6 of the leptons), 2 boson masses, 2 CP–violating phases, and 1 CP–violating angle.

The Feynman diagram calculations are a graphical representation of the perturbation theory approximation that provides a visualization of particle interactions. The vertices represent the couplings of the interactions, open lines correspond to initial state or final state particles, and closed lines represent virtual particles that are created and reabsorbed. One axis represents time and the other one represents space. The boson mediating the reaction can be a virtual particle, meaning that its mass may be off-shell and therefore cannot be observed. We will often use these diagrams throughout this work.

2.1.1 Electromagnetic interaction

Electromagnetism is the interaction that affects all electrically charged particles. It can be seen as the unification and generalisation of electricity and magnetism

and is described by the theory of Quantum Electrodynamics (QED), being the first relativistic quantum field theory, where there is a full agreement between quantum mechanics and special relativity. This theory was developed in the decade 1950 by Tomonaga, Schwinger and Feynman [14–16]. Its mediator particle is the photon, a massless and electrically neutral gauge boson. Therefore, a self-coupling is not allowed, and an infinite reach is achieved, with its strength decreasing with the distance between the interacting particles. The coupling strength of the EM interaction is given by the so called fine structure constant:

$$\alpha_{\mathrm{EM}} = \frac{e^2}{4\pi\epsilon_0\hbar c} \approx \frac{1}{137}, \tag{2.1.1}$$

where e is the electric charge of the electron, ϵ_0 is the vacuum permittivity, $\hbar$ is the reduced Planck's constant and c is the speed of light in the vacuum. From now on we will use the convention of natural units: $\hbar = c = \epsilon_0 = 1$. The value of α_{EM} is found to increase at short distances or for interactions taking place at high energy scales. This effect is caused by vacuum fluctuations creating and annihilating virtual e^+e^- pairs surrounding the interacting particles. An effective charge-screening is produced, where the net-charge experienced in the interaction deviates from the charge carried by the real particles alone.

Mathematically, QED is an abelian gauge theory with a $\mathrm{U(1)_{EM}}$ symmetry in Minkowski space and its Lagrangian density can be written as follows:

$$\mathscr{L}_{\mathrm{U(1)_{EM}}} = -\frac{1}{4\mu_0} F^i_{\mu\nu} F_i^{\mu\nu} + \bar{\psi}(i\gamma^\mu D_\mu - m)\psi, \tag{2.1.2}$$

where

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{2.1.3}$$

is the electromagnetic tensor, γ^μ are the gamma-matrices, and

$$D_\mu = \partial_\mu - iA_\mu \tag{2.1.4}$$

is the covariant derivative. The four-potential A_μ is chosen such that the Lorentz condition $\partial_\mu A^\mu = 0$ can be applied as a gauge fixing condition. The quantum fields ψ and $\bar{\psi}$ of the interacting charged particles with mass m fulfill the Dirac equation:

$$(i\gamma^\mu\partial_\mu - m)\psi = 0. \tag{2.1.5}$$

2.1.2 Weak interaction

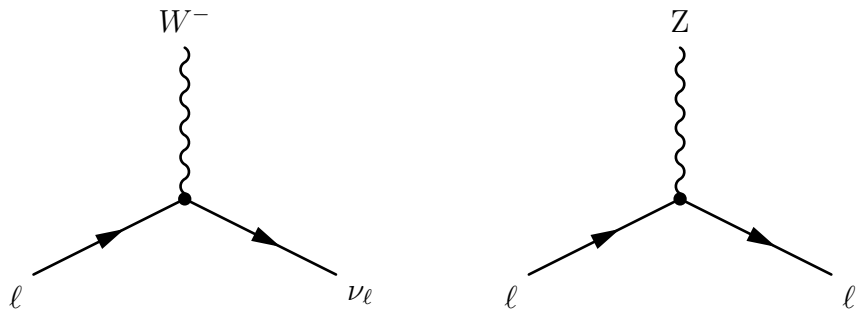
The weak force is the only one of the four fundamental interactions that may change the flavour of particles and violate charge-parity symmetry. It affects all the fermions of the SM as well as the Higgs boson, and is mediated by heavy intermediate vector bosons, resulting in very short lifetime and effective range. Their measured masses are [17]:

$$M_{W^\pm} = 80.377 \pm 0.012 \text{ GeV}, \qquad M_Z = 91.1876 \pm 0.0021 \text{ GeV}. \tag{2.1.6}$$

This interaction is also called quantum flavourdynamics, but it is rarely referred to as that since it is more commonly described through the electroweak theory, explained in Section 2.1.2.2.

2.1.2.1 Neutral and charged weak interactions

There are two types of weak interactions, depending on the total electric charge of the current produced by the interacting fermions, also mediated by different intermediate vector bosons. The charged current interaction, mediated by the pair $W^{\pm}$, and the neutral current interaction, mediated by the Z boson.



(a) Charged current. (b) Neutral current.

Figure 2.1: Tree-level Feynman diagrams of the two fundamental weak interactions.

The fundamental leptonic vertex of the charged weak interaction is shown in Fig. 2.1a, where a charged lepton becomes its corresponding neutrino through the emission of a W^-. The reverse and crossed processes are also allowed, exchanging particles for antiparticles where needed. The same procedure can be found for quarks instead of leptons, where the incoming quark would become a different flavour with ± 1 electric charge. The weak charged coupling constant for this vertex is $g_w = \sqrt{4\pi\alpha_w}$. This process is responsible for the β-decay of hadrons, such as the neutron into a proton (Fig. 2.2) or the B meson into a D meson (Fig. 2.3), where we can observe both the quark and the lepton vertices. The charged current interaction is also responsible for other processes, like the kaon-antikaon oscillation, as shown in Fig. 2.4

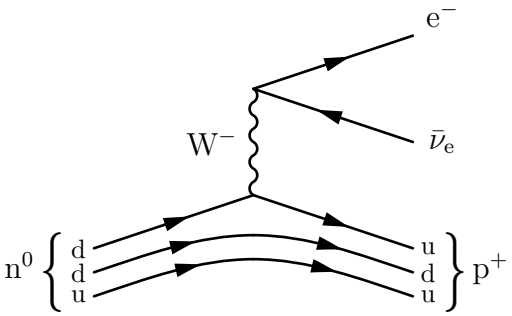


Figure 2.2: Decay of a neutron into a proton by emission of an electron and its antineutrino.

The fundamental vertex of a neutral current interaction is shown in Fig. 2.1b, where the same lepton enters and exits the reaction. The neutral coupling constant

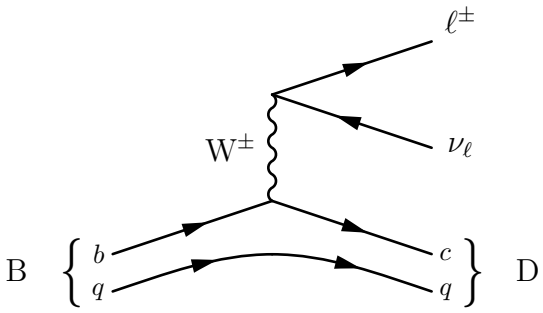


Figure 2.3: Decay of a generic B meson into a generic D meson through the emission of a charged lepton and its corresponding neutrino.

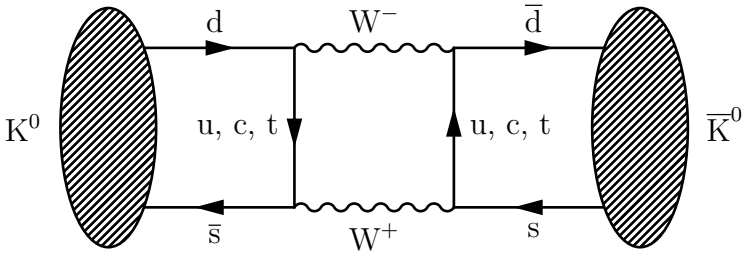


Figure 2.4: Kaon oscillation via exchange of W bosons.

of this vertex is g_z, and we will see later on that it is closely related to the charged one.

Most neutral weak processes are hidden by the equivalent electromagnetic one since the photon can couple to everything the Z does, and at low energies the photon mechanism clearly dominates. Therefore, neutrino scattering was used to confirm the existence of a neutral weak current, since neutrinos do not couple to the photon. In 1973, the GARGAMELLE experiment at CERN [19] was able to produce the first image of a neutral current interaction between a muon-antineutrino and an electron, shown in Fig. 2.5,

$$\bar{\nu}_\mu + e^- \rightarrow \bar{\nu}_\mu + e^-, \tag{2.1.7}$$

that suggested a neutral mediator as the theoretically predicted Z (Fig. 2.6).

There were also observed some neutrino-nucleon scattering events that revealed the neutral neutrino-quark interaction. These cross sections were much smaller than the corresponding charged current events, so that it needed to be a new interaction and not higher-order processes.

2.1.2.2 Electroweak unification and spontaneous symmetry breaking

Weak and electromagnetic interactions could be unified at high energy scales and took the form of the so-called Electroweak theory. Its formalism follows some historical development. Fermi tried to explain the nuclear β-decay through a current-current interaction between two fermions [20] in the decade of 1930. When parity violation was found in weak interactions, the leptonic current was then described as a vector-minus-axial ($V - A$) interaction. However, this theory is included in the class of non-renormalizable theories and further efforts were demanded. The

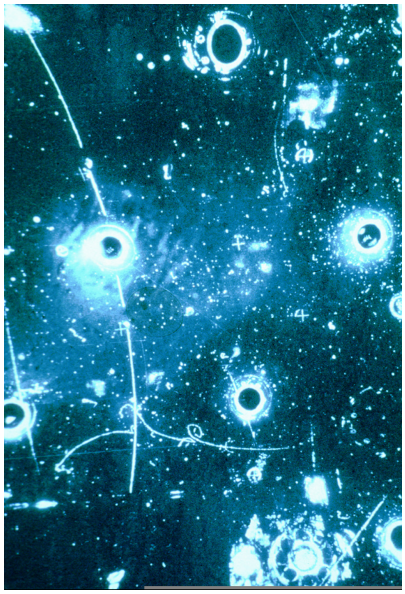


Figure 2.5: The first example of a single-electron neutral current at Gargamelle. An incoming antineutrino knocks an electron towards the left, creating a characteristic electronic shower with electron-positron pairs [18].

first attempt to construct a renormalizable gauge theory was the introduction of intermediate vector bosons $W^{\pm}$ with mass m_W. The charged current interaction can then be written using doublets of left-handed fermions,

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L \quad \text{and} \quad \begin{pmatrix} u \\ d \end{pmatrix}_L . \tag{2.1.8}$$

This doublet structure suggests the gauge group SU(2) for weak interactions, which is often called weak isospin. Only left-handed fermions interact this way and therefore the goup is usually denoted $\mathrm{SU(2)_L}$, while right-handed fermions transform as singlets. For an electroweak unification, In the decade of 1960, Glashow suggested that the gauge group of weak and electromagnetic interactions is the product of two groups [21–23], while Weinberg and Salam included the idea of spontaneous symmetry breaking using the Higgs mechanism [24–30] so that

$$SU(2)_L \otimes U(1)_Y \rightarrow U(1)_{EM}, \tag{2.1.9}$$

where Y stands for hypercharge.

A $SU(2) \otimes U(1)$ gauge theory contains four gauge bosons but only one should remain massless [31]. Using a complex SU(2) doublet to break the gauge symmetry would add four real degrees of freedom, of which three will become longitudinal degrees of freedom for the three massive gauge bosons and one physical Higgs field remains. This complex scalar SU(2) doublet is:

$$\mathbf{\Phi} = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} . \tag{2.1.10}$$

The Lagrangian

$$\mathscr{L} = \left(\partial_\mu \mathbf{\Phi}\right)^\dagger \left(\partial^\mu \mathbf{\Phi}\right) + \mu^2 \mathbf{\Phi}^\dagger \mathbf{\Phi} - \lambda \left(\mathbf{\Phi}^\dagger \mathbf{\Phi}\right)^2 \tag{2.1.11}$$

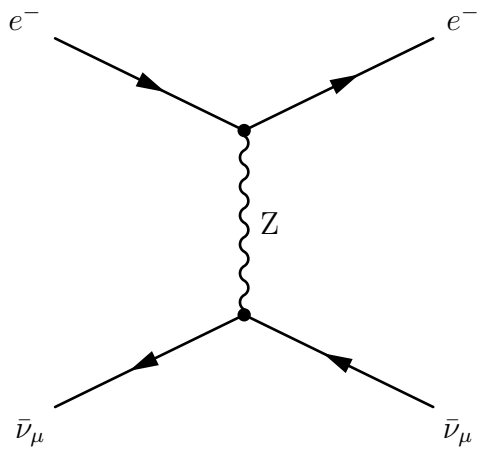


Figure 2.6: Neutral current interaction between an incoming muon-antineutrino and an electron.

is invariant under global SU(2) and U(1) transformations of $\mathbf{\Phi}$. The potential $V(\Phi) = \mu^2|\Phi|^2 + \lambda|\Phi|^4$ has minima at $|\langle\Phi^0\rangle|^2 = v^2 = |\mu^2|/\lambda$ if we demand $\mu^2 < 0$. This potential presents the famous shape of a "Mexican hat" or "the bottom of a wine bottle", as Higgs preferred to call it. Since the ground state of Φ is degenerated, we choose the vacuum expectation value in the ϕ^0 direction to avoid an electrically charged vacuum,

$$\langle 0|\mathbf{\Phi}|0\rangle = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix}, \tag{2.1.12}$$

where $v = \sqrt{\mu^2/\lambda}$ is the non-zero vacuum expectation value.

Electroweak symmetry breaking should leave $\mathrm{U_{EM}}(1)$ invariant and this is achieved using a combination of the Pauli matrices (τ^a) as $\mathbf{1} + \tau_3$. We expect the photon as a superposition of the $\mathrm{U(1)_Y}$ abelian gauge boson and the third component of the $\mathrm{SU(2)_L}$ non-abelian gauge boson. If we identify the electric charge as $Q \propto Y + \tau_3$ in the upper component ϕ^+ of the Higgs doublet we find the Gell-Mann−Nishijima relation $2Q = Y + \tau_3$. The weak isospin operator is related to the Pauli matrices as $I^a = \tau_a/2$, o that $Y = 2Q - I_3$.

Covariant derivatives can be used to gauge the model,

$$\partial^\mu \mathbf{\Phi} \to D^\mu \mathbf{\Phi} = \left(\partial^\mu + \frac{ig}{2}\,\boldsymbol{\tau}\cdot\boldsymbol{W}^\mu + \frac{ig'}{2}B^\mu\right)\mathbf{\Phi}, \tag{2.1.13}$$

introducing the couplings g and g' for the two groups. We can obtain the field strength tensors for the three $\mathrm{SU(2)_L}$ gauge fields $\boldsymbol{W}^\mu$ and the $\mathrm{U(1)_Y}$ field B^μ as

$$F_a^{\mu\nu} = \partial^\mu W_a^\nu - \partial^\mu W_a^\mu - g\epsilon_{abc}W_b^\mu W_c^\nu \tag{2.1.14}$$

$$G^{\mu\nu} = \partial^\mu B^\nu - \partial^\nu B^\mu, \tag{2.1.15}$$

where ϵ_{abc}, the Levi-Civita symbol, is a completely antisymmetric tensor that corresponds to the structure constants of SU(2). The Lagrangian describing the Higgs gauge sector is therefore

$$\mathscr{L} = -\frac{1}{4}F^2 - \frac{1}{4}G^2 + (D_\mu \mathbf{\Phi})^\dagger (D^\mu \mathbf{\Phi}) - V(\mathbf{\Phi}), \tag{2.1.16}$$

We require only one physical scalar field, so we use polar coordinates for the lower component, $\phi^0 = \rho e^{i\theta}$, and set $\phi^+ = \theta = 0$ by a SU(2) unitary transformation. We separate ρ into a vacuum expectation value v and fluctuations $h(x)$, so that the mass term in Eq. (2.1.16) becomes

$$\mathscr{L}_m = \frac{v^2}{2}\chi^\dagger \left(\frac{g}{2}\boldsymbol{\tau}\cdot\boldsymbol{W}^\mu + \frac{g'}{2}B^\mu\right)\left(\frac{g}{2}\boldsymbol{\tau}\cdot\boldsymbol{W}_\mu + \frac{g'}{2}B_\mu\right)\chi. \tag{2.1.17}$$

Knowing that $(\boldsymbol{\tau}\cdot\boldsymbol{W})^2 = \boldsymbol{W}^2$ and $\chi^\dagger \boldsymbol{\tau}\cdot\boldsymbol{W}^\mu \chi = -W_3^\mu$, we find

$$\mathscr{L}_m = \frac{g^2 v^2}{8}(W_1^2 + W_2^2) + \frac{v^2}{8}(gW_3^\mu - g'B^\mu)^2. \tag{2.1.18}$$

We can finally see the relation between the two fields in the first bracket with the two gauge bosons from the charged current interaction in the Fermi theory,

$$W^\pm = \frac{1}{\sqrt{2}}(W_1 \mp iW_2)^2. \tag{2.1.19}$$

Equivalently, we the second bracket can be identified as a new neutral massive gauge boson called Z,

$$Z^\mu = \frac{1}{\sqrt{g^2 + g'^2}}(gW_3^\mu - g'B^\mu) = \cos\theta_W W_3^\mu - \sin\theta_W B^\mu, \tag{2.1.20}$$

where we have introduced the Weinberg angle θ_W as a parametrization of the mixing of B and W_3. The mixing would disappear for $\theta_W = 0$ and the hypercharge would then equal the electric charge.

Last, but not least, the massless photon can be found at the combination of W_3 and B orthogonal to Z, which does not appear in the mass Lagrangian,

$$A^\mu = \frac{1}{\sqrt{g^2 + g'^2}}(g'W_3^\mu + gB^\mu) = \sin\theta_W W_3^\mu + \cos\theta_W B^\mu. \tag{2.1.21}$$

The mass Lagrangian in terms of physical fields then becomes

$$\mathscr{L}_m = \frac{1}{2}m_W^2 + W_\mu^+ W^{-\mu} + \frac{1}{2}m_Z^2 Z_\mu Z^\mu, \tag{2.1.22}$$

where the boson masses are

$$m_W = \frac{gv}{2}, \qquad m_Z = \frac{\sqrt{(g^2 + g'^2)}v}{2} = \frac{m_w}{\cos\theta_W} \qquad \text{and} \qquad m_A = 0, \tag{2.1.23}$$

as function of the coupling constants g and g', and the vacuum expectation value of the scalar Higgs bosons, which appears as the remaining degree of freedom of $\mathbf{\Phi}$. The mass ratio of the gauge bosons is apparently fixed by the Weinberg angle

$$\frac{m_W}{m_Z} = \cos\theta_W, \tag{2.1.24}$$

In the unified electroweak theory, g and g' are related to the electromagnetic coupling e and the Weinberg angle θ_W.

$$\tan\theta_W = \frac{g}{g'}, \qquad g_w = \frac{g_e}{\sin\theta_w}, \qquad g_z = \frac{g_e}{\sin\theta_w\cos\theta_w}, \tag{2.1.25}$$

where g_e is determined by the charge of the electron ($g_e = e\sqrt{4\pi}$).

Proof of this mechanism was obtained when the Higgs boson was finally discovered in 2012 by the CMS and ATLAS experiments at LHC [32–34], with the data collected during the years 2010 to 2012 at $\sqrt{s} = 7\,\text{TeV}$ and $8\,\text{TeV}$.

2.1.2.3 Fermion mass and flavour mixing

The electroweak symmetry breaking mechanism also gives mass to the elementary fermions as follows. The term in the Lagrangian associated to fermion mass

$$m(\bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R) \tag{2.1.26}$$

needs to be gauge invariant. Therefore we need the parameter m to become a SU(2) doublet. If we introduce a Yukawa coupling $y_f = m/v$ between the lepton doublet L, the scalar doublet Φ and the lepton singlet e_R,

$$\mathscr{L}_Y = -y_f\left(\bar{L}\Phi e_R + \bar{e}_R\Phi^\dagger L\right), \tag{2.1.27}$$

we obtain a SU(2) invariant term. The equivalent procedure can be followed if the fermion fields are quarks. Knowing that $Y(\Phi) = 1$ and $Y(L) = -1$, we can also obtain a $\mathrm{U}_Y(1)$ invariant mass term if the lepton singlet has hypercharge $Y(e_R) = -2$. This satisfies the Gell-Mann-Nishijima relation $2Q = Y + 2T_3$. This coupling generates masses and Yukawa interactions between the fermions and the Higgs. After inserting the vacuum expectation of the Higgs, we find a Dirac mass term in the form

$$\begin{aligned}\mathscr{L}_m &= -\frac{y_f v}{\sqrt{2}}\left[\begin{pmatrix}\bar{\nu}_e & \bar{e}\end{pmatrix}_L\begin{pmatrix}0\\1\end{pmatrix}e_R + \bar{e}_R\begin{pmatrix}0 & 1\end{pmatrix}_L\begin{pmatrix}\nu_e\\e\end{pmatrix}_L\right]\\ &= -m_f\left(\bar{e}_L e_R + \bar{e}_R e_L\right) = -m_f\bar{e}e,\end{aligned} \tag{2.1.28}$$

and we generate the masses for the down-like fermions with $\tau_3 = -1/2$ like the electron. For the up-like fermions we need the charge conjugated Higgs doublet $i\tau_2\Phi^*$.

The coupling y_f and therefore the mass m_f for the three generations of lepton and quarks are 3×3 arbitrary matrices in flavour space, so that the weak eigenstates $\boldsymbol{\psi} = \{\psi_e, \psi_\mu, \psi_\tau\}$ transform into mass eigenstates $\boldsymbol{\psi'} = \{\psi_1, \psi_2, \psi_3\}$ as

$$\bar{\boldsymbol{\psi}}_{\boldsymbol{L}}\boldsymbol{m}\boldsymbol{\psi}_{\boldsymbol{R}} = \bar{\boldsymbol{\psi}}_{\boldsymbol{L}}\boldsymbol{S}\boldsymbol{S}^\dagger\boldsymbol{m}\boldsymbol{T}\boldsymbol{T}^\dagger\boldsymbol{\psi}_{\boldsymbol{R}} = \bar{\boldsymbol{\psi}'}_{\boldsymbol{L}}\boldsymbol{m}_{\boldsymbol{D}}\boldsymbol{\psi'}_{\boldsymbol{R}}, \tag{2.1.29}$$

where $\boldsymbol{S}$ and $\boldsymbol{T}$ are unitary matrices and $\boldsymbol{m_D}$ is diagonal and positive. These new eigenstates are not diagonal, so there will be flavour mixing. Specifically, for the charged weak current J^μ with the mass eigenstates,

$$J^\mu = \bar{\boldsymbol{\nu}}_{\boldsymbol{L}} \gamma^\mu \boldsymbol{e_L} = \bar{\boldsymbol{\nu}}'_{\boldsymbol{L}} \gamma^\mu \boldsymbol{S}^\dagger_{\boldsymbol{\nu}} \boldsymbol{S_e} \boldsymbol{e}'_{\boldsymbol{L}}, \tag{2.1.30}$$

we find that there will be only one observable, $\boldsymbol{U} \equiv \boldsymbol{S}^\dagger_{\boldsymbol{\nu}} \boldsymbol{S_e}$. This gives us some freedom to choose mixing only for down-like quarks and neutrinos, meaning $\boldsymbol{S_e} = \boldsymbol{1}$ and $\boldsymbol{U} = \boldsymbol{S_\nu}$. The $\boldsymbol{U}$ matrices for quarks and neutrinos are called respectively Cabibbo-Kobayashi-Maskawa matrix (CKM matrix) [35, 36] and Pontecorvo-Maki-Nakagawa-Sakata matrix (PMNS matrix) [37, 38].

The measured values of the CKM matrix are [17]

$$\begin{pmatrix} |V_{ud}| & |V_{us}| & |V_{ub}| \\ |V_{cd}| & |V_{cs}| & |V_{cb}| \\ |V_{td}| & |V_{ts}| & |V_{tb}| \end{pmatrix} = \begin{pmatrix} 0.97373 & 0.2243 & 0.00382 \\ 0.221 & 0.975 & 0.0408 \\ 0.0086 & 0.0415 & 1.014 \end{pmatrix}, \tag{2.1.31}$$

where we observe that the matrix elements with value closest to one enhance the interaction vertices between those quarks and the W boson, whereas the lowest values suppress the corresponding processes.

2.1.3 Strong interaction

The strong interaction is the fundamental force that binds hadrons and nucleons together. Hadrons are the bound state of quarks, with mesons consisting of one quark and one antiquark, and baryons containing three quarks or antiquarks. Nucleons are protons and neutrons, the two types of baryons that populate the atom nucleus. When doubly charged baryons like Δ^{++} were experimentally found, an additional degree of freedom was required in addition to flavour, so that three identical quarks would not violate the Pauli principle. This new quantum number is called colour charge and can present three different values, conventionally called red, green and blue.

From the mathematical point of view, the strong interaction is described by the theory of Quantum Chromodynamics (QCD), a non-abelian gauge theory represented by the $\mathrm{SU(3)_C}$ group, mediated by an octet of massless gluons that also carry colour charge. This theory was developed in the decade of 1970 by Fritzsch, Leutwyler and Gell-Mann [39]. The Lagrangian density can be written as follows:

$$\mathscr{L}_{\mathrm{SU(3)_C}} = -\frac{1}{4} F^i_{\mu\nu} F_i^{\mu\nu} + \sum_f \bar{\psi}_f (i\gamma^\mu D_\mu - m)\psi_f, \tag{2.1.32}$$

where

$$F^i_{\mu\nu} = \partial_\mu G^i_\nu - \partial_\nu G^i_\mu - g_S f_{ijk} G^j_\mu G^k_\mu \tag{2.1.33}$$

$$D_\mu = \partial_\mu - \mathrm{i} g_S \frac{\lambda_i}{2} G^i_\mu, \tag{2.1.34}$$

with $F^i_{\mu\nu}$ being the field strength tensor of the gluon fields G^i_μ (i = 1, 2, ..., 8), and g_s as the gauge coupling constant of the strong interaction. The last term in the field tensor is the gluon self-interaction, a consequence of the non-commutative character of the Gell-Mann 3×3 matrices

$$[\lambda^i, \lambda^j] = 2\mathrm{i} f_{ijk} \lambda^k, \tag{2.1.35}$$

where f_{ijk} are the structure constants of the group, completely antisymmetric in the three indices, generalizing the antisymmetry of the Levi-Civita symbol ϵ_{ijk} of SU(2). The quark fields ψ_f form a colour triplet for each flavour f, as a consequence of the SU(3) symmetry:

$$\psi_f = \begin{pmatrix} \psi^r_f \\ \psi^g_f \\ \psi^b_f \end{pmatrix}, \tag{2.1.36}$$

2.1.3.1 Perturbative QCD and renormalization

The gauge coupling constant can also be expressed in terms of the strong coupling constant α_S as

$$\alpha_S = \frac{g_S^2}{4\pi}, \tag{2.1.37}$$

in the line of the QED fine structure constant. However, the value of this element depends on the energy scale of the interaction, so that the calculations of QCD observables can be expressed as a perturbative expansion in the strong coupling constant, therefore labeling the theory as perturbative QCD (pQCD). Keeping α_S at small values, any observable can be approximated by a truncation of the power series, where the coefficients are derived from the Feynman rules. When calculating higher order corrections in pQCD, ultraviolet divergences appear. Different schemes to avoid these infinities can be derived by introducing a cut-off scale μ, so that the relevant parameters like the strong coupling constant or the quark masses are renormalized to their effective value. In order to keep physical observables independent from any arbitrary scale, the renormalization group equation (RGE), or equivalently the β-function, must hold:

$$\begin{aligned} \beta(\alpha_S)\frac{\partial}{\partial \alpha_S} &= \mu^2 \frac{\partial \alpha_S}{\partial \mu^2} \\ &= -\alpha_S(b_0 + b_1\alpha_S + b_2\alpha_S^2 + \mathcal{O}(\alpha_S^3)), \end{aligned} \tag{2.1.38}$$

where the factors b_n depend on the employed renormalization scheme. At leading order, the solution of this equation yields

$$\alpha_S(\mu^2) = \frac{12\pi}{(11n - 2f)} \ln\left(\frac{\mu^2}{\Lambda^2_{\mathrm{QCD}}}\right), \tag{2.1.39}$$

where n is the number of colours (3 in the SM) and f is the number of quark flavours below a given energy scale Q^2 (6 in the SM), while $Q^2 = \Lambda^2_{\mathrm{QCD}}$ is the threshold that

determines the scale at which the equation blows up and α_S diverges. The determination scale of α_S is usually chosen as the mass of the Z^0 boson. This shows that the strong coupling constant is hardly a constant, but *runs* with Q^2. As b_0 is negative, the strong coupling constant α_S decreases at short distances or high energy scales, while increasing exponentially at large distances or low energy scales, thus leading to an effect called asymptotic freedom. Therefore, quarks interact weakly at high energies, permitting perturbative calculations as if they were free particles, while the dominant interaction is strong at low energies, forcing the confinement of quarks and gluons within composite colour-neutral hadrons. If we tried to separate two quarks from each other, the gluon self-coupling terms in the Lagrangian would spontaneously create a cloud of virtual gluons around each quark, increasing the potential between both of them with the distance. Eventually, at the range of the typical size of a hadron, the energy becomes high enough to create a new quark-antiquark pair and two new hadrons are formed. As opposed to the screening effect observed for the electromagnetic interaction, this effect is referred to as anti-screening, and explains why quarks cannot be observed as free particles.

In a similar way, quark masses also run with Q^2 and their values depend on the renormalization scheme. As free quarks are not observable, their masses cannot be determined via direct measurements but may be inferred from other mass-dependent observables in accelerators and colliders. Finally, it is interesting to mention that the gauge symmetry remains unbroken, so $m_g = 0$, although confinement means the gluon has an effective, dynamical mass.

2.2. The proton structure

The proton is the most stable hadron with an estimated minimum lifetime almost larger than the known age of the Universe [17]. Protons are baryons, composed of three valence quarks (two up and one down), that provide the effective electric charge, and a sea of quarks and gluons in between. This sea experiences continuous fluctuations through creation and annihilation of pairs of particles and antiparticles, which can be probed through high-energy collisions at accelerators such as the LHC [40].

The proton structure was extensively investigated in Deep Inelastic Scattering (DIS) experiments at fixed target and at lepton-proton colliders, that confirmed the composite nature of the proton. Feynman proposed that the constituents of hadrons could be point-like elements that he called partons [41], while Bjorken applied this concept to electron-proton DIS [42]. Currently, they are identified with quarks and gluons.

2.2.1 Cross section and parton distribution functions

In any given collision event, the probability density for having a parton with a certain longitudinal momentum fraction x at a momentum scale Q^2 is described by the so-called parton distribution function (PDF) $f_i(x, Q^2)$. The PDFs are considered as intrinsic properties of the nucleon and process independent. However, they cannot

be derived from first principles and need to be determined in QCD fits to measured cross sections. Taking advantage of the factorization theorem [43–45], the interaction can be separated into a hard part, that carries the largest momentum transfer, can be described with perturbative QCD, and depends on the renormalization scale μ_R, and a soft interaction, that includes non-perturbative contributions and is separated by a factorization scale μ_F, needed to handle infrared divergences arising from this approach. A graphic representation of this factorization is shown in Fig. 2.7.

We can express the cross section for any scattering process between two protons with four-momenta P1 and P2, at a center-of-mass energy of $\sqrt{s} = (P1 + P2)$, as:

$$\sigma(\mathrm{pp} \to X) = \sum_{a,b} \int dx_1 dx_2 f_a(x_1, \mu_F^2) f_b(x_2, \mu_F^2) \sigma_{ab\to X}(x_1 P_1, x_2 P_2, \sqrt{s}, \mu_F^2, \mu_R^2), \tag{2.2.1}$$

where the sum runs over all partons a and b of each proton, since the initial state momentum of the partons is not known. The partonic cross section $\sigma_{(1,2\to X)}$ is computed as the phase–space integral of the corresponding matrix element squared and can be calculated perturbatively, whereas the nonperturbative contributions are described by the DGLAP (Dokshitzer-Gribov-Lipatov-Altarelli-Paris) equation [46]:

$$\frac{d}{dQ^2} f_i(x, Q^2) = \sum_j \frac{\alpha_S(Q^2)}{2\pi} \int_x^1 \frac{dz}{z} P_{i\to j}(z) f_j(x/z, Q^2), \tag{2.2.2}$$

where $P_{i\to j}(z)$ is the splitting kernel, and represents the probability of a parton i becoming a parton j carrying a fraction z of its original momentum.

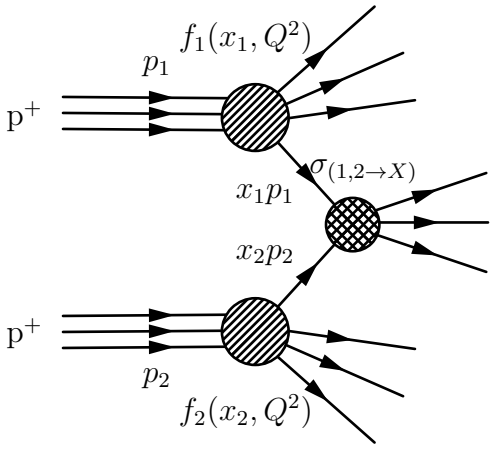


Figure 2.7: Generic pp collision where one parton from each proton carries away a fraction of the momentum $x_i p_i$. The cross section of the hard process (hatched) is denoted by $\sigma_{(1,2\to X)}$, and the soft processes (shaded) are included in the PDFs f_1 and f_2.

The first order Feynman diagrams in α_S are called Leading Order (LO) terms and superior ones are called Next-to-Leading Order (NLO), Next-to-Next-to-Leading Order (NNLO), etc. The calculations for each order have their own benefits and

setbacks, aiming for a balance between accuracy in the prediction, calculation effort and computational time. An example of the difference in precision between NLO and NNLO can be seen in Fig. 2.8. Calculations of PDFs using experimental data are shown in Fig. 2.9. The representation of $x \cdot f$ provides information about the relative contribution of each flavour with respect to the proton momentum. Thus, the integral of each curve represents the fraction of the average proton momentum carried away by each flavour, with the sum of all contributions being 1. It is worth noting that the area of u_v is twice the area of d_v, due to the fact that there are two up valence quarks and only one down valence quark. These valence quarks are more prevalent at higher parton momenta, whereas gluons greatly dominate at lower fractions. In fact, the integral of the gluon curve is approximately 0.5, meaning that gluons carry away almost half of the proton momentum.

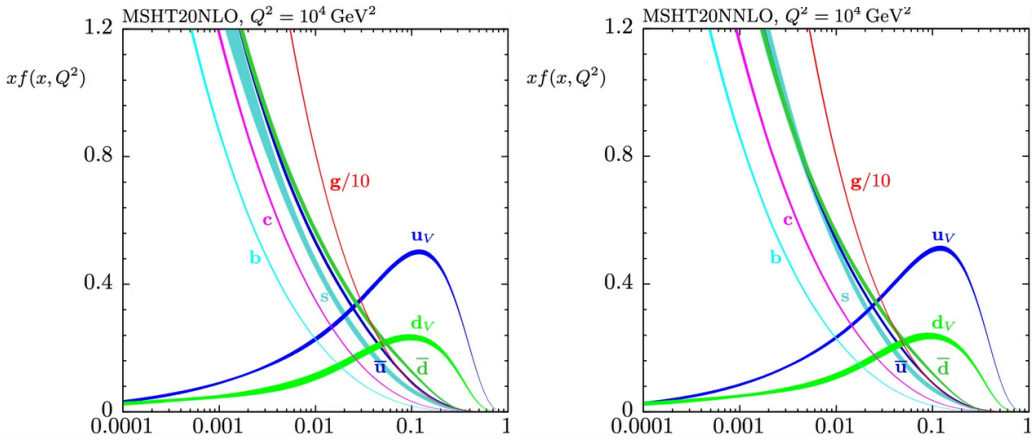


Figure 2.8: PDF predictions at NLO (left) and NNLO (right) for different partons at the energy scale of $Q^2 = 10^4\,\text{GeV}^2$. The bands reflect the uncertainties at the 68% confidence level [47].

2.2.2 The strange quark content of the proton

The strange quark distribution is one of the least constrained PDFs of the proton [5]. It is quite relevant to study since it is one of the dominant sources of uncertainty in measurements like the $W^{\pm}$ boson mass at hadron colliders [49, 50], usually dominated by the uncertainty from PDFs.

The main constraints of the s-quark distribution arise from the analysis of charm production in charged current DIS of a passing (anti)neutrino and a nucleus, with a final state of two muons [51–55]. The muon carrying the higher p_T is considered to originate from the neutrino scattering, whereas the oppositely charged muon is assumed to originate from the decay of a charmed particle. These measurements are highly sensitive to the strange quark content of the nucleon sea, since the contribution from down quarks to charm production is Cabibbo suppressed, as seen in Eq. 2.1.31.

The inclusive production of electroweak bosons at hadron colliders, such as Tevatron or the LHC, provides constrains for all light quark PDFs and is therefore indirectly sensitive to the strange quark distribution in the proton. The production

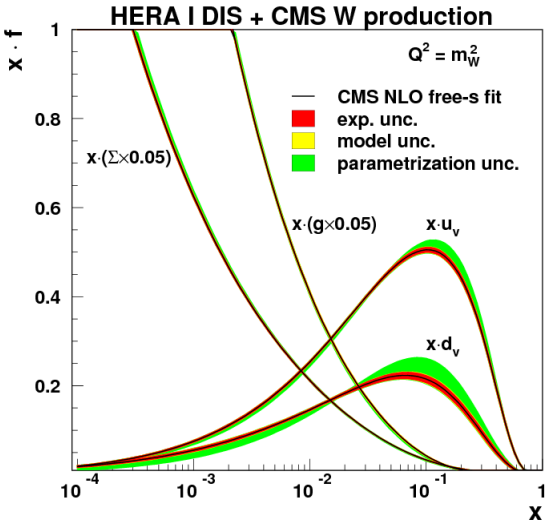


Figure 2.9: PDFs shown as functions of x, obtained by using HERA DIS data and CMS measurements of W-boson production in the free-s NLO QCD analysis, at the scale of m_W^2. The sea distribution is defined as $\Sigma = 2(\bar{u} + \bar{d} + \bar{s})$. The full band represents the total uncertainty. The individual contributions from the experimental, model, and parametrization uncertainties are represented by the bands of different shades. The gluon and sea distributions are scaled down by a factor of 20 [48].

of a W boson associated with a single charm quark probes more directly the strange quark content of the proton, as W+c is dominantly produced by the hard scattering of a strange quark and a gluon at leading order. We will describe this in detail in Section 5.1.

The strange quark distribution is often presented relative to the PDFs of the other light sea quarks via the so-called strangeness suppression R_s:

$$R_s(x, \mu^2) = \frac{s(x, \mu^2) + \bar{s}(x, \mu^2)}{\bar{u}(x, \mu^2) + \bar{d}(x, \mu^2)}. \tag{2.2.3}$$

Calculations of the strange PDF and R_s using experimental data are shown in Fig 2.10, performed in the CMS W+c analysis at 8 Tev [8], which is a precursor to this work.

One way to study if the strange quark and antiquarks behave differently in the nucleon sea is to measure the ratio:

$$R_c^{\pm} = \frac{W^+ + \bar{c}}{W^- + c}, \tag{2.2.4}$$

where deviations from 1 would indicate that one process is favoured with respect to the other. A priori, we expect a value slightly lower than 1 since the process dg→W$^-$+c is enhanced with respect to its charge conjugate due to the presence of the d valence quark in the proton, despite the fact that the overall d-to-c process is Cabibbo suppressed. Different contributions in W+c production from strange

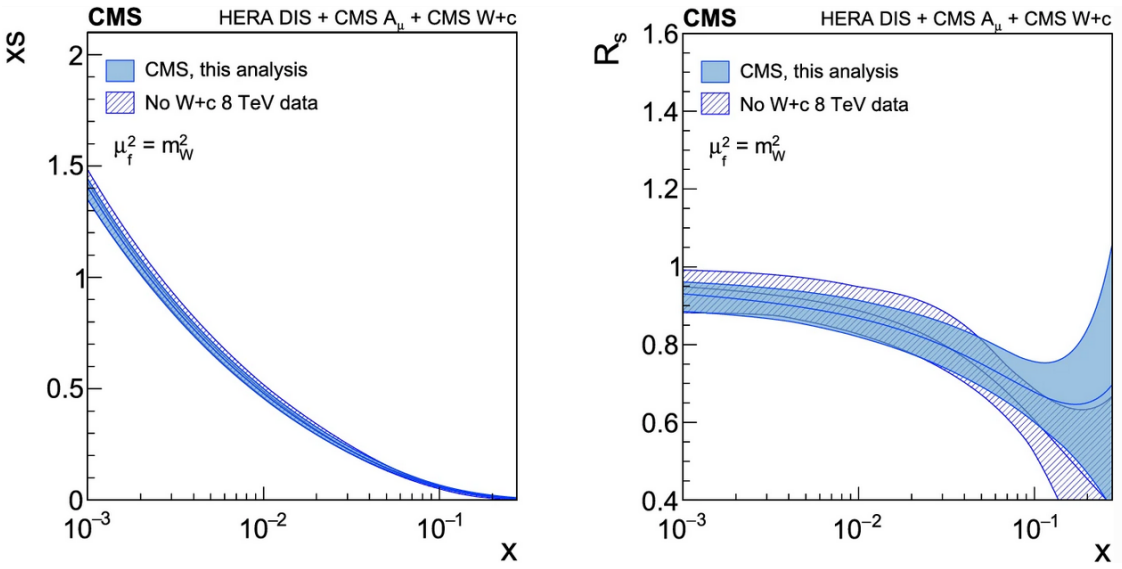


Figure 2.10: The strange quark distribution (left) and the strangeness suppression factor (right) as a function of x at the factorization scale of m_W^2. Both plots show a comparison between calculations using (shaded) or excluding (hatched) the data from the W+c at 8 TeV analysis [8].

quarks and antiquarks in the proton would modify the expected value for $R_c^{\pm}$. The $R_c^{\pm}$ measurement performed in the CMS W+c analysis at 8 Tev is shown in Fig. 2.11. The measurement of $R_c^{\pm}$ in proton-proton collisions at 13 TeV with the highest precision up to date is one of the main results of this thesis work.

2.3. Theoretical calculations and Monte Carlo simulations

The way to test these theories and compare them with experimental data is necessarily through computer simulations of the probabilistic distributions of the physical processes at the desired order of perturbations, as well as a thorough representation of a virtual detector. Several dedicated state-of-the-art programs may be interfaced in order to recreate each step for a full event simulation, namely, matrix elements calculations, parton shower, matching and merging, fragmentation and hadronization, decay branching ratios, interaction of the particles with the detector and signal readout. Figure 2.12 illustrates these steps in the simulation of proton-proton collisions.

The calculation of leading order matrix elements of parton level cross sections has already been implemented in Monte Carlo event generators, such as MadGraph [57] or pythia [58]. MadGraph takes as input the initial and final state particles, as well as the specified order in QCD and QED, and returns the desired amplitude. MadGraph5 [59, 60] is the latest version of this matrix element generator, written in Python. pythia is another standard tool for event generation in high-energy collisions, modeling the evolution of the hard interaction into a complex

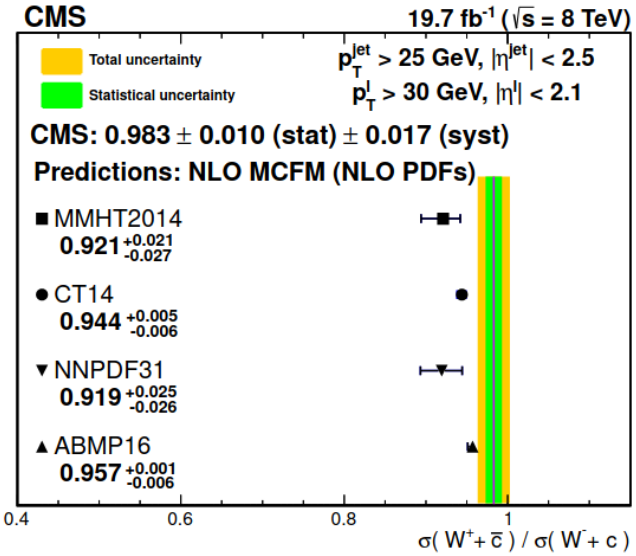


Figure 2.11: Experimental measurement of $R_c^{\pm}$ from the CMS W+c at 8 TeV analysis compared with theoretical predictions at NLO in QCD using several PDF sets [8].

multiparticle final state. PYTHIA8 is completely rewritten in C++, and may offer a replacement for most external applications, especially for analyses of LHC physics.

Fixed order calculations in pQCD can be used to compute inclusive quantities in strong processes [61]. The computation of complex Feynman diagrams limits the precision of these calculations. Furthermore, we would not be able to produce predictions for exclusive quantities even if we could perform computations at arbitrary order. This is a direct consequence of the presence of collinear and soft divergences in fixed order calculations with a definite final state. Thus, the computation of certain inclusive quantities can only be achieved through the summation of different final states, where these divergences cancel.

Matrix element calculations at higher orders become more computationally demanding, so that corrections to LO or NLO calculations have to be approximated. Parton shower algorithms simulate additional radiations by evaluating the probability of an emission between two evolution scales μ_1 and μ_2, with $\mu_1 > \mu_2$. Parton shower evolution is also considered universal and independent of the production process of the partons. Soft and collinear divergences are avoided by excluding emissions below a chosen cutoff scale μ_0, so that a finite emission probability is maintained.

Parton shower simulations must be matched and merged with the corresponding matrix element calculations of a given QCD process [62]. This is needed to avoid double counting of events produced by either program. Hard emissions are treated as in matrix element computations while soft and collinear emissions are handled by the parton shower simulation, and matching between both regions is smooth. For LO events, one can introduce a scale that bans contributions over the cutoff for the parton shower and below it for the matrix element. For NLO events, there are some virtual corrections that produce additional double counting. These can be handled by subtracting those parton shower contributions from the matrix element

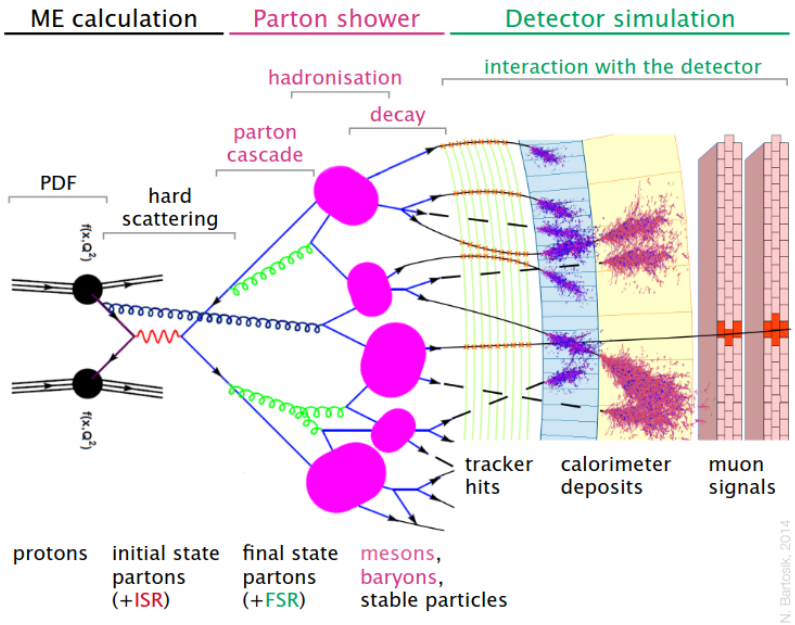


Figure 2.12: Schematic view of the typical event-generation process for a proton-proton collision [56].

calculation, although a fraction of events with negative weight might be generated.

Despite the divergences arising from real and virtual corrections, NLO or even NNLO generators are preferred. MadGraph5_aMC@NLO [63] is capable of computing tree-level and one-loop amplitudes for arbitrary processes at LO and NLO accuracy. Furthermore, additional real emissions are considered in the calculations so that the generated events can include higher order corrections. Another relevant event generator is POWHEG [64–66], which provides calculations of heavy quark production processes at NLO accuracy. It is a method for interfacing NLO calculations with parton shower generators that overcomes the problem of events with negative weight. The hardest radiation is generated first, with a method that yields only events with positive weight using the exact NLO matrix elements. Both generators can be interfaced with other programs for parton shower and hadronization and provide suitable matching and merging procedures for such cases.

We mentioned previously that quarks are not free particles and must form bound states called hadrons. This hadronization process cannot be calculated in pQCD and must be modelled following experimental data. The fragmentation function describes the fraction of the longitudinal momentum of the quark carried away by the new hadron. The string model [67] considers that short range $q\bar{q}$ fluctuations of the vacuum may fall into the potential of the strong field between two drifting quarks. The colour charges of the new quarks may break the string connecting the former and create new hadrons, carrying away a fraction of the original energy and momentum with their own new string. This fragmentation proceeds by iteratively breaking one hadron apart from either endpoint.

Alternatively, Monte Carlo for FeMtobarn processes (MCFM) [68, 69] is a program that gives predictions at parton level for a wide range of processes at

hadron colliders, especially those containing W, Z and H bosons and heavy quarks. Most processes are calculated at NLO, with some others also available at NNLO or even N^{3}LO. Some NLO electroweak effects are also considered and spin correlations are included in the decay. MCFM implements transverse momentum and jet veto resummation for the production of colour singlet final states. Vetoing energetic jet activity suppresses backgrounds and enables new physics searches at the LHC, but the large logarithms introduced by a veto scale need to be resummed.

The underlying event (UE) is the collection of any other interactions, not associated with the hard scattering process of interest. The UE consists of particles from the hadronization of beam-beam remnants, of multiple-parton interactions, and their associated initial and final state radiation. These include hadrons from the fragmentation of spectator partons, that do not exchange relevant p_{T} in the collision, and additional parton-parton scatterings within the same pp collision, with lower p_{T} than the hard scattering. Those particles in the low-p_{T} spectrum cannot be described by pQCD, so the UE models implemented in MC generators require corrections to experimental data using specific sets of parameters called tune [70–72].

A realistic comparison between experimental data and simulated events requires an estimation of the reconstruction efficiencies and the misidentification rates of particles. Therefore, the response of the detector to the traversing particles must be simulated as well. GEANT4 [73] is a toolkit developed at CERN to simulate the passage of particles through matter, including tracking, geometry, physics models and hits from 250 eV up to the TeV energy range. It was implemented in C++ exploiting software engineering and object-oriented technology. It has been widely used in different scientific environments, such as particle physics, nuclear physics, accelerator design, space engineering and medical physics.

Chapter 3

The CMS experiment at the LHC

"Don't be too proud of this technological terror you've constructed. The ability to destroy a planet is insignificant next to the power of the Force."
— Darth Vader, Star Wars Episode IV: A New Hope.

In this chapter we describe the most powerful tool we have so far to explore the smallest constituents of matter. The Large Hadron Collider accelerates protons or heavy ions to *almost* the speed of light and makes them collide in four detection points, while respectfully not destroying our planet nor transporting us to other dimensions or universes. The Compact Muon Solenoid is capable of detecting the resulting particles and measuring their properties very precisely, which allows us to infer the subjacent processes at the collision point, test the currently accepted theories and look for exotic phenomena.

3.1. The Large Hadron Collider

The Large Hadron Collider (LHC) [74] at CERN is the largest particle accelerator ever built. It consists of a ring of 26.7 km of superconducting electromagnets and radio-frequency cavities located 100 m underground in the Swiss-French border. The LHC began operation in 2009 and provided proton collisions at a centre-of-mass energy, $\sqrt{s}$, of 7 TeV in 2010 and $\sqrt{s} = 8$ TeV in 2011 and 2012, during what is called Run 1. The Long Shutdown 1 period took place in order to update and upgrade the machine and then new data-taking activities started in 2015, Run 2, which increased $\sqrt{s}$ up to 13 TeV, providing collisions until 2018. After the Long Shutdown 2, a new era has just begun in July 2022, Run 3, with collisions at $\sqrt{s} =$ 13.6 TeV and expected to reach 14 TeV.

The LHC takes advantage of the tunnel built for the Large Electron-Positron collider (LEP) [75], that was dismantled in 2000, in order to avoid the cost of a new excavation while maintaining the Earth's crust as shielding for radiation. The tunnel was built at a mean depth of 100 m, due to the geological properties of the area between Jura and Lake Geneva, and with a smooth slope of 1.4% in order to minimize the depth of the shafts with the constraint of a connection point to the already existing Super Proton Synchrotron (SPS) [76].

In this work we will focus on proton-proton (pp) collisions, so let us describe

briefly how to obtain a large and pure sample of high-energy protons. Particles cannot be accelerated directly to energies of the order of TeV, such as those of the LHC design. Therefore, it is required to follow a series of coordinated boosts in a collider chain [77], as shown in Fig. 3.1. Starting with a single bottle of hydrogen gas, that may be replaced just twice per year, an electric field removes the electrons from the hydrogen atoms. These bare protons are then accelerated up to 50 MeV with a linear accelerator (LINAC 2) and injected into a circular accelerator (Booster) where they reach an energy of 1.4 GeV. After that, the protons are consecutively sent through the Proton Synchrotron (PS) and the SPS, increasing their energies up to 25 and 450 GeV, respectively. The final step is the injection of protons into the LHC, where a total of 2808 bunches, separated by 25 ns and containing $\sim 1.1 \times 10^{11}$ protons each, are boosted to final energies of 6.5 TeV, circling the perimeter 11 245 times per second.

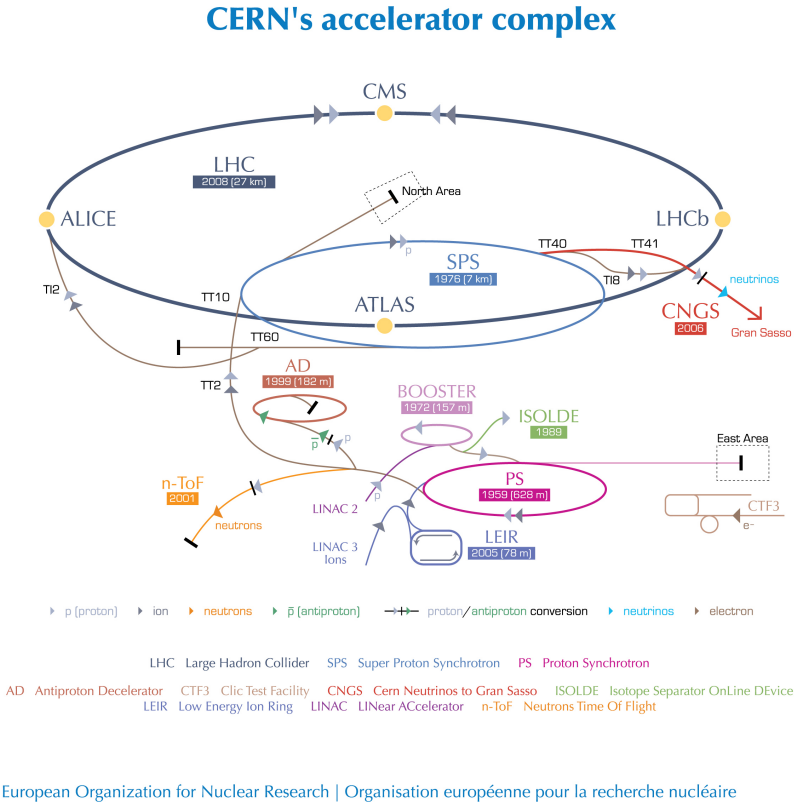


Figure 3.1: The LHC is the last ring (dark grey line) in a complex chain of particle accelerators. The smaller machines are used in a chain to help boost the particles to their final energies and provide beams to a whole set of smaller experiments, which also aim to uncover the mysteries of the Universe [78]. This configuration is the one used during the data-taking period of this work.

One of the main advantages of a circular accelerator versus a linear one is that the energy of the particles can be increased gradually each loop inside the same pipe. However, this also poses one of the major challenges, such as containing the beams in a closed trajectory without significant loss of energy. Thousands of magnets of different varieties and sizes are used to direct the beams around the accelerator, keeping them stable and precisely aligned. These include 1232 dipole magnets, 15

metres in length, which bend the beams, and 392 quadrupole magnets, each 5–7 metres long, which focus the beams. Just prior to collision, insertion magnets are used to "squeeze" the particles closer together to increase the chances of collisions. The main dipoles generate powerful 8.3 T magnetic fields, more than 100 000 times more powerful than the Earth's magnetic field, which is only possible when a current of 11 080 amperes flows through a superconducting coil.

The state of superconductivity, where electricity is conducted without any resistance, is achieved at very low temperatures through cryogenic techniques. The LHC is the largest cryogenic system in the world and one of the coldest places on Earth, operating at a temperature of 1.9 K (-271.3°C), colder than outer space. The superconducting coils of niobium-titanium (NbTi) are kept at those temperatures by a closed circuit of superfluid liquid helium.

Finally, proton beams must travel along the accelerator without interacting with any other material, which would cause scattering of the energy. For this matter, the pipes are kept at ultra-high vacuum. The pressure in these pipes is in the order of 10^{-10} to 10^{-11} mbar, a vacuum similar to interstellar space.

Acceleration to top energies at the LHC is achieved using superconducting radio frequency (RF) cavity systems. Each beam travels through eight cavities made of niobium sputtered on copper performing at maximum accelerating voltage of 2 MV at 400 MHz. This oscillating fields also tighten the proton beams into discrete bunches.

Such is the level of precision required in these facilities that even the effect of the Moon has to be taken into consideration. Tidal forces due to gravitational attraction cause the level of water to rise and fall periodically every day. The Earth's crust is made of elastic rocks that also feel that pull and rises by some 25 cm in the Geneva area. This stretching causes a variation of 1 mm in the circumference of the LHC, which is then translated into changes in beam energy that must be taken into consideration [79].

One proton beam is injected at the LHC in a clockwise trajectory, while another is sent anticlockwise. This way, collisions can be forced at the four huge detectors: ATLAS [80], CMS [81], LHCb [82] and ALICE [83]. ATLAS and CMS are general purpose experiments that study a broad spectrum of physics topics, whereas LHCb focuses on b quark physics and CP volation. ALICE mainly studies Quark Gluon Plasma physics using heavy ion collisions.

This kind of experiments is so vast and ambitious and require such a large deploy of resources, time and personnel that it can only be accomplished by a cooperating international community, which once again reminds us of the importance of public investment in the scientific and technological development.

3.1.1 Luminosity and pileup

We define the instantaneous luminosity ($\mathscr{L}$) as the number of protons crossing a unit surface (cm^2) per unit time (s). The number of events produced per unit time

is:

$$\frac{dN}{dt} = \mathscr{L} \cdot \sigma, \tag{3.1.1}$$

where $N = \mathcal{L} \cdot \sigma$ is the total number of events, $\mathcal{L} = \int \mathscr{L} \cdot dt$ is the total luminosity integrated during the data-taking period and σ is the cross section of the process. Therefore, in order to maximize the probability of finding rare processes, with low cross section, it is imperative to have as high luminosity as possible. The design value of the LHC of $\mathscr{L} = 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ provides inelastic pp collision rates of 10^9 events per second or 10^{15} events per year, which in turn increases the chance of finding interesting events of high energy processes with very low cross sections and event rates.

The integrated luminosity is measured in the scale unit of inverse barns and cross sections in barns, typically in the range of pico- ($1\,\mathrm{pb} = 10^{-12}\,\mathrm{b}$) and femtobarns ($1\,\mathrm{fb} = 10^{-15}\,\mathrm{b}$).

The instantaneous luminosity is measured in the hadronic forward calorimeters of CMS and calibrated using Van der Meer scans [84], varying the separation of the proton beams. Luminosity is then calculated from the beam overlap width, obtained from the fit of the profile as a function of the separation. Figure 3.2 shows the total luminosity delivered by the LHC and the one recorded by the CMS experiment for the full Run 2. The less than 10% inefficiency of the CMS recorded luminosity is due to the idle time between each data acquisition run.

In the previous lines we have seen that higher luminosities contribute to higher chances of finding processes with low cross section. However, nothing comes for free and this also presents a major inconvenience to take into account. High luminosity means large amount of bunch crossings and large density of protons per bunch, which in turn means that multiple pp interactions may happen from the same (in-time) or adjacent (out-of-time) bunch crossing of the collision with the hardest transferred momentum. This effect is called pileup (PU) and its correct determination is important in high energy analyses. Figure 3.3 shows the PU profile and the average number of PU interactions for each data-taking year and the full Run 2 average. These plots use only data that passed the so-called golden certification of CMS and LHC standard values for the minimum bias inelastic cross section. This means that all sub-detectors were flagged to be ok for any kind of usage in physics analysis.

3.2. The CMS Detector

The Compact Muon Solenoid is the second largest detector of the LHC, located in the northernmost part of the accelerator ring, near Cessy, at what is called Point 5. The central feature of the CMS apparatus is a superconducting solenoid of 6 m internal diameter, providing a magnetic field of 3.8 T, embedded in a rather compact detector of an overall length of 22 m, diameter of 15 m, and weight of 14 000 tonnes. Within the solenoid volume are a silicon pixel and strip tracker, a lead-tungstate crystal electromagnetic calorimeter, and a brass and scintillator hadron calorimeter, each composed of a barrel and two endcap sections. Forward calorimeters extend

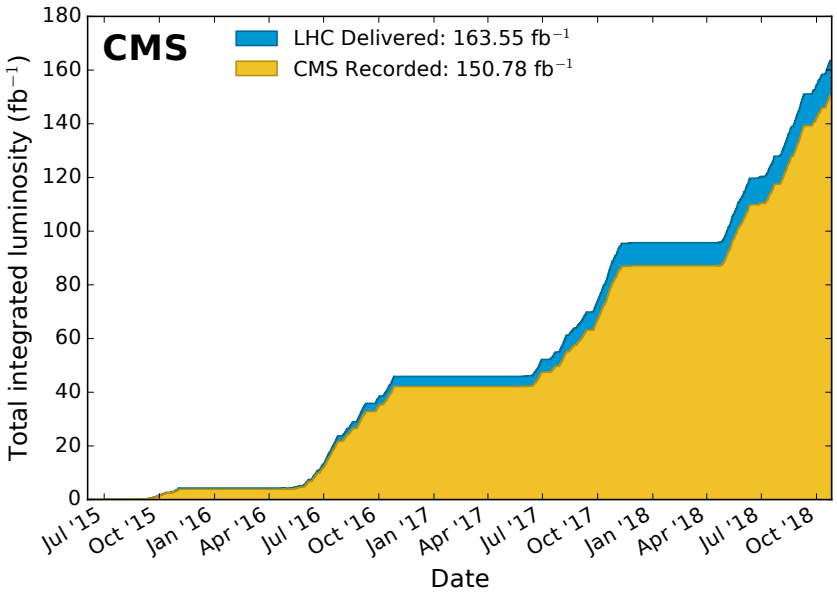


Figure 3.2: Cumulative delivered and recorded luminosity versus time for 2015-2018 during stable beams for pp collisions at nominal center-of-mass energy [85].

the coverage provided by the barrel and endcap detectors. Muons are measured in gas-ionization detectors embedded in the steel flux-return yoke outside the solenoid.

The detector requirements for CMS to meet the goals of the LHC physics programme are described in Ref. [81] as follows:

- Good muon identification and momentum resolution over a wide range of momenta and angles, good dimuon mass resolution, and the ability to determine unambiguously the charge of muons with momentum lower than 1 TeV.

- Good charged-particle momentum resolution and reconstruction efficiency in the inner tracker. Efficient triggering and offline tagging of τ's and b jets, requiring pixel detectors close to the interaction region.

- Good electromagnetic energy resolution, good diphoton and dielectron mass resolution, wide geometric coverage, π^0 rejection, and efficient photon and lepton isolation at high luminosities.

- Good missing-transverse-energy ($p_{\mathrm{T}}^{\mathrm{miss}}$) and dijet-mass resolution, which requires hadron calorimeters with a large hermetic geometric coverage and with fine lateral segmentation.

The design of CMS meets these requirements and we will briefly describe each subdetector following the extensive characterization found in Ref. [81], as depicted in Figs. 3.4 and 3.5. The procedure followed for aligning the detector is described in Ref. [87].

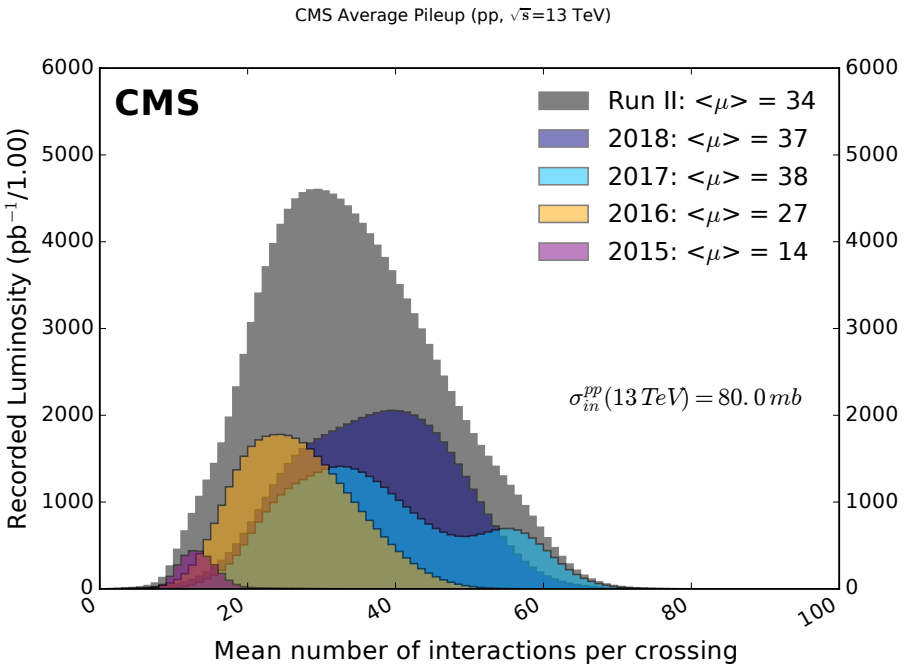


Figure 3.3: Distribution of the average number of interactions per crossing (pileup) for pp collisions 2015-2018 and full Run 2. The overall mean values and the minimum bias inelastic cross section are also shown [85].

3.2.1 Coordinate System

First of all we should define the coordinate system we are going to use throughout the description of the detector as well as in the analysis. The CMS detector is designed over a Cartesian right-handed system, with its origin centered at the nominal collision point. The x axis points radially inward toward the center of the LHC and the y axis points vertically upward, defining the transverse x-y plane perpendicular to the beam direction. Therefore, the z axis points along the beam direction toward the Jura mountains from LHC Point 5 or anticlockwise in the geographical map. Moreover, the geometry of CMS encourages to define a set of cylindrical coordinates maintaining the z axis. We can then measure the azimuthal angle ϕ from the x axis in the x-y plane and the polar angle θ from the z axis. A more convenient notation for the polar angle is the pseudorapidity, defined as $\eta = -\ln[\tan(\theta/2)]$, where η is zero in the x-y plane and $\pm\infty$ in the beam axis, $\eta = 1$ corresponds to $\theta \simeq 40°$ and $\eta = 2.5$ to $10°$. If we define the transverse momentum of the particle as $p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2}$, we can express the four vector of any particle using $(E, p_{\mathrm{T}}, \eta, \phi)$ as well as (E, p_x, p_y, p_z) through the transformation:

$$
\begin{aligned}
p_x &= p_{\mathrm{T}} \cos\phi \\
p_y &= p_{\mathrm{T}} \sin\phi \\
p_z &= p_{\mathrm{T}} \sinh\eta.
\end{aligned}
$$

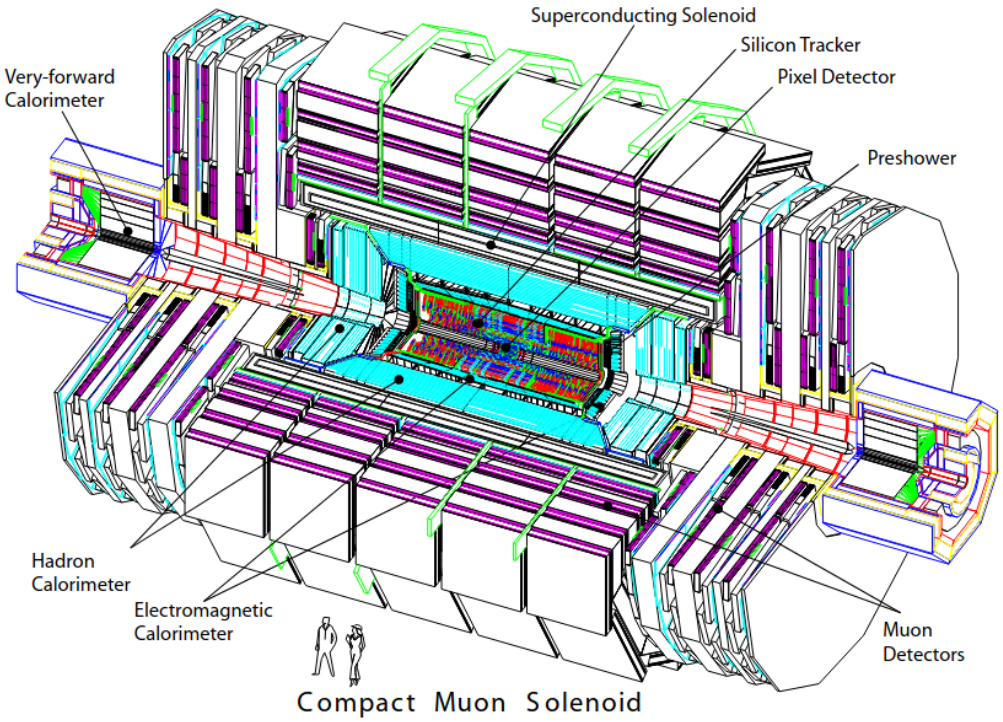


Figure 3.4: Schematic view of the CMS detector and its subdetectors [81].

3.2.2 Superconducting Magnet

The magnet for the CMS detector is the major element of the whole experiment in terms of size, weight and structural rigidity and it is used as the principal support structure for all the other detectors [88].

The main parameters of the CMS magnet which are related to the physics requirement are the magnetic field of 3.8 T, the yoke diameter of 14 m, the axial yoke length of 21.6 m and the total weight of approximately 10 800 tonnes. The yoke is composed of 5 wheels and 2 endcaps, divided into three disks each, serving as an additional hadron absorber and supporting the muon detectors. The superconducting solenoid is made of four layers of NbTi that allows a current flow of 19.14 kA at cryogenic temperature of 4.7 K. The intense magnetic field bends the trajectories of charged particles allowing the measurement of their momentum and electric charge.

3.2.3 Inner Tracking System

The inner tracking system of CMS provides the measurement of the trajectories of charged particles from the collision point, as well as a precise reconstruction of primary and secondary vertices. It covers a cylindrical volume of 5.8 m in length and 2.5 m in diameter [89].

Due to requirements on granularity, response time and radiation endurance, the tracker design is entirely based on silicon detector technology. It consists of a pixel detector of 1440 pixels, with three barrel layers between 4.4 cm and 10.2 cm, and a silicon strip tracker of 15 148 strips, with ten barrel detection layers reaching

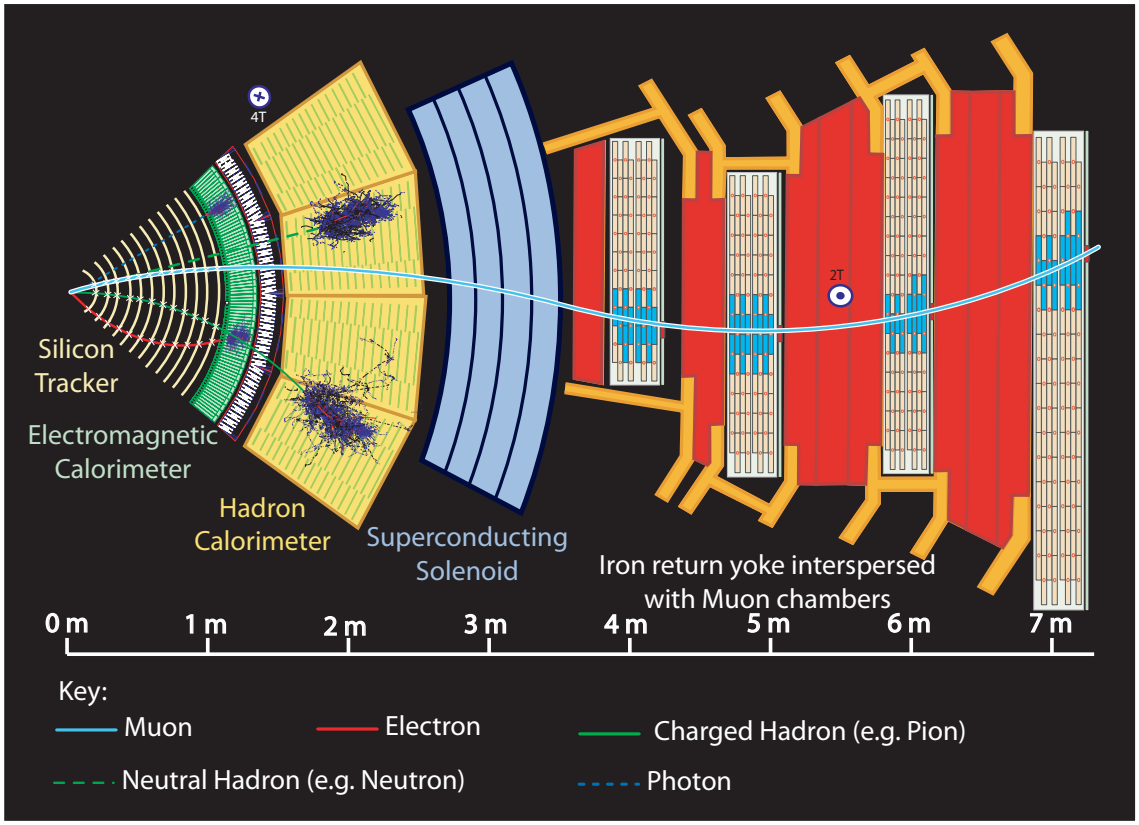


Figure 3.5: Slice view of the CMS detector and the particle trajectories [86].

a radius of $\rho = 110\,\text{cm}$. At the endcaps, there are two disks in the pixel detector and twelve disks in the strip tracker on each side of the barrel, covering a pseudorapidity of $|\eta| < 2.5$. The $200\,\text{m}^2$ of active silicon area made the CMS tracker the largest silicon tracker at the time [90].

The silicon tracker used in 2016 measured nonisolated charged particles of $1 < p_\text{T} < 10\,\text{GeV}$ and $|\eta| < 1.4$ with track resolutions of 1.5% in p_T and 25–90 (45–150) μm in the transverse (longitudinal) impact parameter [89], defined as the minimum transverse (longitudinal) distance of the track to the interaction point. At the start of 2017, a new pixel detector was installed [91], adding a fourth barrel layer closer to the beam pipe, at 2.9 cm of the beam axis, and several new stations in the forward region. A comparison between both configurations is shown in Fig. 3.6. The upgrade pixel detector was aimed to improve the precision in measurements of secondary vertices. The upgraded tracker measures average resolutions of 1.5% in p_T and 20–75 μm in the transverse impact parameter [92] for nonisolated particles of $1 < p_\text{T} < 10\,\text{GeV}$ up to $|\eta| < 3.0$. The impact parameter is defined as the distance of closest approach of the track to the collision point.

3.2.4 Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL) is in charge of measuring the energy of electrons and photons [93]. It consists of 75 848 lead tungstate ($PbWO_4$) crystals, which cover pseudorapidities of $|\eta| < 1.48$ in the barrel region (EB) and $1.48 <$

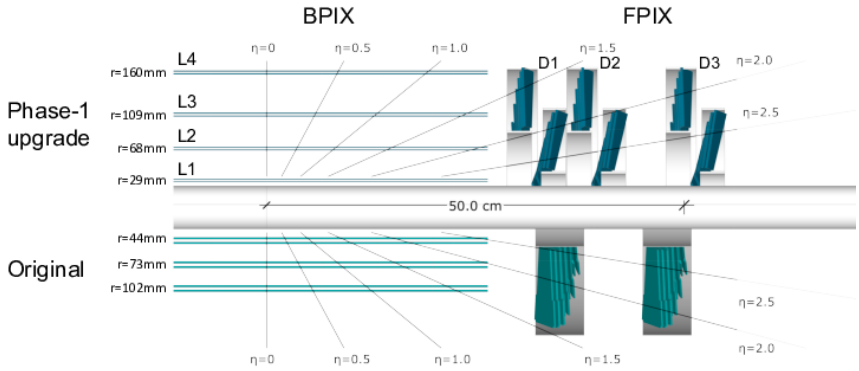


Figure 3.6: Layout of the CMS Phase-1 pixel detector compared to the original detector layout, in longitudinal view [91].

$|\eta| < 3.0$ in the two endcap regions (EE). There are two preshower detectors, one in fron of each EE, consisting of alternating layers of lead radiators and silicon strip sensors.

The ECAL is a homogeneous calorimeter, where the crystals function as both absorber and scintillator at the same time. This means that the electromagnetic cascades generated by the electrons and photons are absorbed inside the crystals and measured using photodetectors. The barrel photodetectors are avalanche photodiodes (APDs) and the ones in the endcaps are vacuum phototriodes (VPTs).

The lead tungstate crystals are 25.8 radiation lengths thick in the barrel and 24.7 in the endcaps. The high density ($8.28\,\mathrm{g/cm^3}$), short radiation length (0.89 cm) and small Molière radius (2.2 cm) result in a fine granularity and a compact calorimeter that is also radiation resistant. For uniform illumination of a crystal with 120 GeV electrons a resolution of 0.5% was achieved [94].

3.2.5 Hadron Calorimeter

The Hadron Calorimeter (HCAL) is located surrounding the tracker and the ECAL and will measure the energy and direction of hadron jets in order to calculate the missing transverse energy, a key to estimate neutrino directions and energies [95].

The HCAL is designed as a sampling calorimeter, meaning that it is made of active material inserted between copper absorber plates, and is conformed of four elements, the barrel (HB), the outer barrel (HO), the endcap disks (HE), and the forward calorimeters (HF). The barrel is limited by the outer extent of the electromagnetic calorimeter (R = 1.77 m) and the inner extent of the solenoid (R = 2.95 m). This constrains the amount of material installed to absorb the hadronic

shower, so an outer hadron calorimeter or tail catcher is also placed outside the magnet.

Both the HB and HE suffer the 3.8 T field of the magnet, so they are necessarily built of non-magnetic materials like copper alloy and stainless steel. The absorber plates are 5 cm thick in the barrel and 8 cm thick in the endcap. The active elements of the entire central hadron calorimeter are 4 mm thick plastic scintillator tiles read out using wavelength-shifting (WLS) plastic fibers.

In the region $|\eta| < 1.74$, the HCAL cells have widths of 0.087 in pseudorapidity and 0.087 in azimuth (ϕ). In the η-ϕ plane, and for $|\eta| < 1.48$, the HCAL cells map on to 5×5 arrays of ECAL crystals to form calorimeter towers projecting radially outwards from close to the nominal interaction point. For $|\eta| > 1.74$, the coverage of the towers increases progressively to a maximum of 0.174 in $\Delta\eta$ and $\Delta\phi$. Within each tower, the energy deposits in ECAL and HCAL cells are summed to define the calorimeter tower energies, which are subsequently used to provide the energies and directions of hadronic jets.

To extend hermeticity beyond $|\eta| = 3$, the HF calorimeters placed at 11.2 m from the interaction point cover up to $|\eta| = 5.2$ using a Cherenkov-based, radiation-hard technology. Since they use steel as absorber and quartz fibers as the sensitive material, they are most receptive to the electromagnetic component of hadronic showers and also serve as luminosity monitors.

When combining measurements from the tracker and the calorimeters, the jet energy resolution amounts to 15% at 10 GeV, 8% at 100 GeV, and 4% at 1 TeV, as opposed to about 40%, 12%, and 5% obtained when only the ECAL and the HCAL calorimeters are used [96]. The calibration of the hadronic calorimeters is described in [97] and leads to a typical precision of less than 3%.

3.2.6 Forward Detectors

The Zero Degree and CASTOR calorimeters are made of quartz fibers and plates embedded in tungsten absorbers. They are located far away from the main structure of CMS and measure particles that emerge very parallel to the beam line from the collision point.

3.2.6.1 CASTOR

The CASTOR (Centauro And STrange Object Research) [98] calorimeter was proposed, built, and installed in the CMS experiment with the purpose of studying very forward particle production in heavy ion (HI) and pp collisions. CASTOR is located at 1 cm surrounding the LHC beam line, with an outer radius of 40 cm and a length of 160 cm, at 14.4 m in the negative z direction from the CMS interaction point. It extends the CMS acceptance to the very forward pseudorapidity range, $-6.6 < \eta < -5.2$. The bulk of the mass is a tungsten absorber, and Cherenkov photons are produced in quartz, both nonmagnetic materials. The location and design of CASTOR are optimized for the study of electromagnetic and hadronic showers from particles emitted very parallel to the beam direction. This detector

is focused on searches for deeply penetrating particles, as well as measurements of generic properties of particle production at forward rapidities in inelastic proton and nuclear collisions.

3.2.6.2 ZDC

A set of two zero degree calorimeters (ZDC) [99] are designed to complement the CMS very forward region. The ZDCs measure neutral particles, such as neutrons and photons, for the heavy-ion and low-luminosity pp collisions at $|\eta| \geq 8.3$. The ZDC calorimeter sections are sampling calorimeters with the core of each structure consisting of a tungsten-plate and quartz-fiber-ribbon stack, with a design similar to HF and CASTOR. A significant advantage of this technology is that the calorimeter is compact, extremely fast, and radiation hard. Quartz fibers were chosen as the active media of the ZDC calorimeters because of their unique radiation hardness features and the intrinsic speed of the Cherenkov effect.

In order to measure neutrons and very forward photons the calorimeters are located at roughly 140 m on each side of the interaction vertex. This neutral particle absorbers were built with a detector slot of 1 m length each, 96 mm width and 607 mm height. It contains copper absorber bars of 90 cm length and a transversally segmented ionizing chamber serving as a pp luminosity monitor. The design of each individual ZDC includes two independent calorimeter sections, the electromagnetic (EM) section and the hadronic (HAD) section.

3.2.7 The Muon System

The muon system is the main component of the CMS detector, as suggested by its middle name, and has been designed for the accurate identification and measurement of muons. These particles rarely interact with matter and can easily escape the inner components of the detector. Therefore, the muon system is placed outside the magnet and consists of three types of gaseous detectors that cover the external volume; drift tubes (DT) in the barrel region, cathode strip chambers (CSC) in the endcap region, and resistive plate chambers (RPC) in both the barrel and endcap [100].

The DT system is designed to provide muon track reconstruction and Level-1 trigger selection including muon charge identification and an accurate p_{T} measurement, as well as single bunch–crossing identification with good time resolution. It covers the pseudorapidity region $|\eta| < 1.2$, where the neutron-induced background is small, the muon rate is low and the 3.8 T magnetic field is uniform and mostly contained in the steel yoke. Each DT working unit is called chamber and consists on standard rectangular drift cells, with a central wire at a voltage of 3600 V, two electrodes (cathodes) on the sides and two more above and below the wires, each of them at -1800 V. The volume is filled with a gas mixture of Ar(85%)/CO2(15%), corresponding to a drift time of 380 ns at the maximum path of 21 mm, thus keeping the number of active channels to an affordable value while maintaining a negligible occupancy to avoid the need for multi-hit electronics. Passing muons ionize the gas and the electric field causes an electron drift that is then collected by the wire.

The DT system consists of 240 muon chambers alternating with the layers of the return yoke of the magnet and composed of three groups, called Super–Layers (SL), of four staggered layers of independent drift cells. Two of the SLs have the wires parallel to the beam direction and are devoted to trigger and position measurements in the $r-\phi$ coordinate, while the other SL has wires perpendicular to the beam direction and measures the z coordinate, providing a 3D measurement of the muon track. The outermost station (MB4) is equipped with chambers containing only the two $r-\phi$ SLs.

A group of chambers around the same value of r is called "station", and across a fixed ϕ is called "sector". There are three stations of 12 sectors with four station types and one station of 14 sectors in each of the five wheels of the CMS barrel reaching a total of 172 000 channels. The four station types are called MB1, MB2, MB3, and MB4 from inside to outside, where MB stands for Muon Barrel. Although the chambers of station type MB4 in sectors 4 and 10 are physically divided in two parts, they are considered as a single unit. This layout is shown in Fig. 3.7 for one wheel.

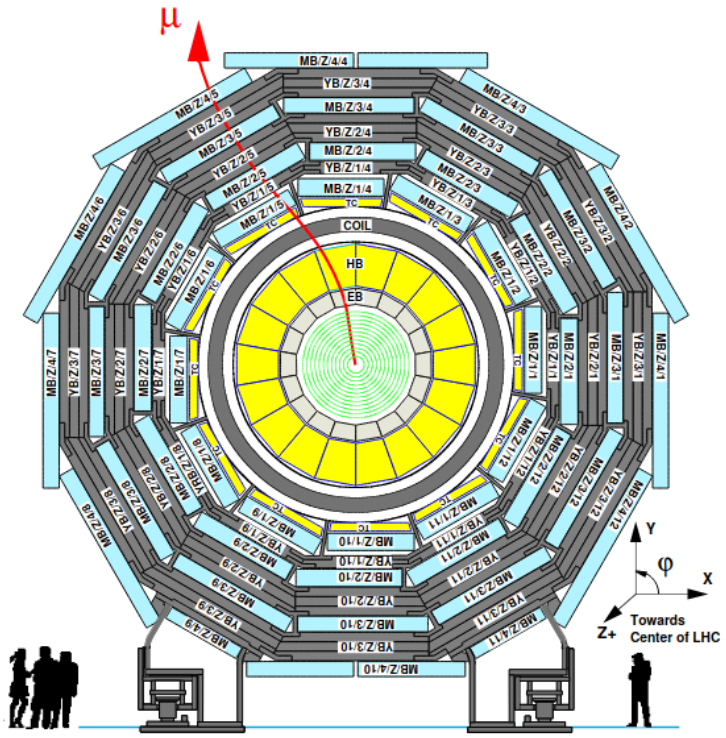


Figure 3.7: Layout of the CMS barrel muon DT chambers in one of the 5 wheels. The chambers in each wheel are identical with the exception of wheels -1 and +1 where the presence of cryogenic chimneys for the magnet shortens the chambers in 2 sectors [81].

In order to measure the drift time in each cell, chambers contain Time to Digital Converter (TDC) units as well as dedicated electronics to perform the Level-1 trigger algorithm, that we will describe later. For tracking purposes, the signals recorded by the TDC units are converted to position coordinates, called "hits", calculated using the drift velocity. In each chamber, a linear fit to the hits recon-

structs track segments, which are then used for the offline reconstruction of the muon tracks. For triggering purposes, the DT Local Trigger algorithm searches for trigger segments of aligned hits using dedicated electronics and algorithms.

The drift cells of each chamber are offset by a half-cell width with respect to their neighbour to eliminate dead spots in the efficiency, as depicted in the top part of Fig. 3.10. The two sets of four chambers in each station are separated as much as possible to achieve the best angular resolution. The fine segmented chambers in cells, and the good precision in measuring the drift time, allow the track segments to be reconstructed with a spatial resolution better than 100 μm along $r\phi$, and 150 μm along $r - z$.

In the 2 endcap regions of CMS the muon system uses cathode strip chambers (CSC), due to high muon rates and background levels, as well as a large and non-uniform magnetic field. Their main qualities are their fast response time, fine segmentation, and radiation resistance. These multiwire proportional chambers, consisting of alternating layers of anode wire arrays and cathode panels inside a gas volume, cover the range between $|\eta|$ values of 0.9 and 2.4. There are four stations of CSCs in each endcap, with chambers positioned perpendicular to the beam line and alternating with the flux return plates. The cathode strips of each chamber are aligned along the r axis and provide a precise measurement in the r-ϕ bending plane. The anode wires are installed almost perpendicular to the strips and provide measurements of η and the beam-crossing time of a muon. The gas mixture for these cells is 50% CO_2, 40% Ar, and 10% CF_4. Offline reconstruction efficiency of simulated single-muon samples is typically 95-99% except in the regions around $|\eta| = 0.25$ and 0.8 (the regions between 2 DT wheels) and $|\eta| = 1.2$(the transition region between the DT and CSC systems), where the efficiency drops. Negligible punch-through, meaning high-energy hadron shower remnants that escape the HCAL, reaches the system due to the amount of material in front of the muon system, which exceeds 16 interaction lengths.

Both DT and CSC chambers are capable of determining the p_T and electric charge of passing muons by precisely measuring their bending curvature in cooperation with the inner tracker.

A complementary dedicated trigger system consisting of resistive plate chambers (RPC) was added in both the barrel and endcap regions in order to reduce the uncertainty in the background rates and to increase the ability of the muon system to measure the correct beam-crossing time at full LHC luminosity. The chambers are filled with a gas mixture consisting of 95.2% $C_2H_2F_4$, 4.5% i-C_4H_{10}, 0.3% SF_6. Water vapour is added to the mixture until it reaches a relative humidity of 40–50%. The RPCs provide a fast, independent, and highly-segmented trigger with a sharp p_T threshold over the range $|\eta| < 1.6$. The RPCs are double-gap chambers, operated in avalanche mode to ensure good operation at high rates.

There are six layers of RPCs embedded in the barrel muon system, two in each of the first two stations, and one in each of the last two stations. The redundancy in the first two stations allows the trigger algorithm to also reconstruct low-p_T tracks that may not reach the outer two. In the endcap region, there is a plane of

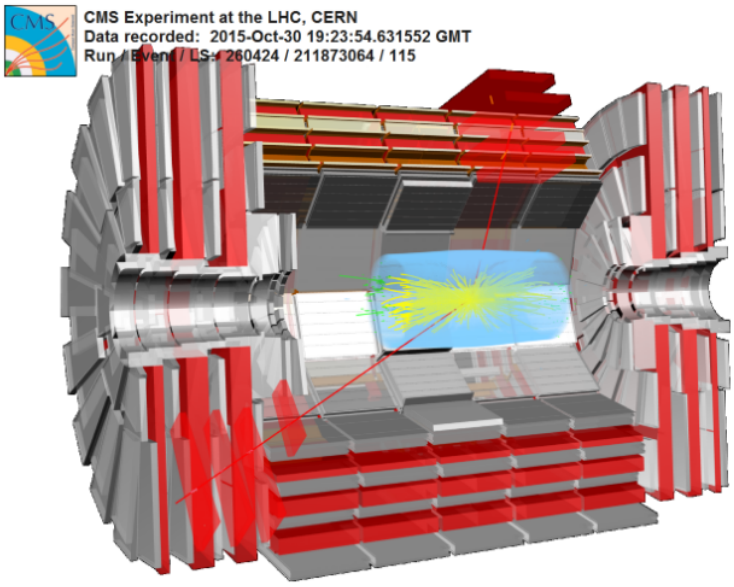


Figure 3.8: A pp collision event with two reconstructed muon tracks superimposed on a cutaway image of the CMS detector, rotated around the y axis so that the inner tracker appears offset relative to its true position in the center of the detector. The four layers of muon chambers are alternated with three layers of the steel flux-return yoke. The reconstructed invariant mass of the muon pair is 2.4 TeV. One muon is reconstructed in the barrel with a p_{T} of 0.7 TeV, while the second muon is reconstructed in the endcap with p_{T} of 1.0 TeV [101].

RPCs in each of the first three stations so that the coincidences may help reduce the background, improve the time resolution for bunch crossing identification, and achieve a good p_{T} resolution. A final alignment system measures the positions of the muon detectors with respect to each other and to the inner tracker in order to optimize the muon momentum resolution.

All the muon chambers are aligned roughly perpendicular to the muon trajectories and distributed to provide hermetic coverage for $|\eta| < 2.4$. The barrel DTs cover $|\eta| < 1.2$ while the endcap CSCs cover $0.9 < |\eta| < 2.4$. The RPCs cover $|\eta| < 1.9$ [101]. The muon detection system has nearly 1 million electronic channels.

The DT chambers in the second station were built by the CIEMAT-CMS group, as well as their readout electronics and part of the alignment system between the tracker and the muon system. CIEMAT members take essential part in the daily activity of the the trigger, alignment, online and offline monitoring, and in the muon reconstruction software.

3.2.8 Trigger and processing

The main function of the CMS trigger system is to reduce the vast amount of data produced at the LHC collision point to an affordable level. Given that the majority of collisions are uninteresting and can be discarded, events of interest are selected using a two-tiered trigger system. The first level (L1) is made of custom

hardware and uses information from the calorimeters and muon detectors to select events at a rate of around 100 kHz within a fixed latency of about 4 μs [102]. The second level is called the high-level trigger (HLT) and consists of a farm of commercially available processors running a version of the full event reconstruction software optimized for fast processing, which further reduces the event rate to around 1 kHz before data storage [103]. In total, there are $\approx 10^8$ data channels checked in each bunch crossing.

3.2.8.1 L1 trigger

The L1 trigger is a hardware system that decides if an event is accepted or rejected using information from the calorimeter and muon detectors for every bunch crossing. The basic arrangement of this system is shown in Fig. 3.9. The trigger primitives (TP) from the calorimeters and the muon detectors are processed in several steps before the combined event information is evaluated in the Global Trigger (GT) and a decision is made whether to accept the event or not.

The L1 calorimeter trigger works on two stages, a regional calorimeter trigger (RCT) and a global calorimeter trigger (GCT). The RCT receives the transverse energies and quality flags from the ECAL and HCAL towers. This is then processed in parallel so that e/γ candidates and regional E_T sums are sent as outputs. The GCT then analyzes those candidates further, finds jets using the E_T sums, and calculates global quantities such as E_T^{miss} . Its output is four isolated and nonisolated e/γ candidates, twelve jets, and several other global quantities.

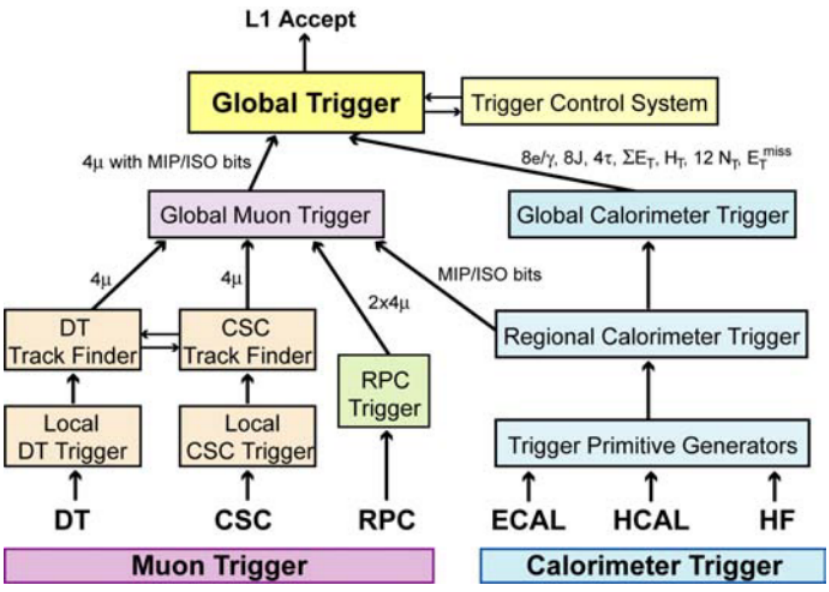


Figure 3.9: Architecture of the Level-1 Trigger, where each subdetector provides candidates to the following step [103].

The muon trigger is built using all three muon detector systems to ensure good coverage and redundancy. Specifically, for the DT and CSC systems the front-end trigger electronics identifies track segments or hit patterns. These segments

are transmitted via optical fibers to regional track finders in the electronics service cavern approximately 90 m from the experimental cavern. Pattern recognition algorithms are applied to identify muon candidates and measure their momenta from their bending in the magnetic field. The DT track finder (DTTF) and CSC track finder (CSCTF) share information for efficient coverage in the η region of overlap. The RPCs send the hits directly from the front-end electronics to pattern comparator trigger (PACT) logic boards that select muon candidates. A maximum of 4 (CSCTF, DTTF) or 8 (RPC) muon candidates are then sent to the global muon trigger (GMT) every bunch crossing. Each candidate carries a p_{T}, charge, (η, ϕ)-position and a quality code. The purpose of the GMT is to improve trigger efficiency, reduce trigger rates and suppress background by merging muon candidates found by more than one system. An example of the DT Local trigger can be found in Fig. 3.10.

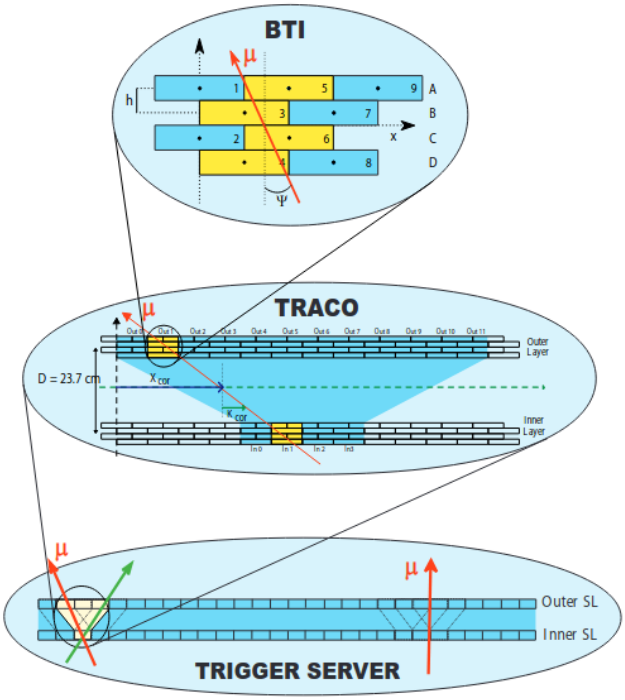


Figure 3.10: The main components of the DT Local Trigger in the ϕ view of a muon chamber. The Bunch and Track Identifiers (BTIs) detect hit alignments within each superlayer, the Track Correlators (TRACOs) search for a proper matching between superlayers, and the Trigger Server selects the best two candidates in the chamber [104].

The GT is the final system that accepts or rejects an event at L1 using the candidates coming from the GCT and GMT. The set of values of each candidate goes through a menu of 128 algorithms based on simple combinatorial logic (AND-OR-NOT), where any condition may be used as a trigger or as a veto. This menu has a wide variety of options, from simple single-object selections with p_{T} above a threshold to jet multiplicity or complex combinations of objects with topological conditions among them.

3.2.8.2 HLT

The last step for the event selection is the HLT, consisting of a single processor farm composed of regular computers running Scientific Linux, called the event filter farm (EVF). The validation of any event follows a similar procedure as the offline processing. Objects accepted by the L1 such as electrons, muons, and jets are reconstructed and go through the scrutiny of whether they are relevant for data analysis or not. The HLT path is a set of algorithmic processing steps of increasing complexity and refinement that both reconstructs physics objects and makes selections on these objects. In order to reduce the CPU required as well as the latency, a key feature of the algorithms is to only reconstruct the necessary regions of the the CMS detector, such as the muon chambers indicated by the L1 and the corresponding road in the tracker for the validation of a given muon [105]. The reduction rate is $\mathcal{O}(10^2)$, with an output rate of roughly 1 kHz with a decision time of a few hundred milliseconds.

The output rate of the L1 and HLT can be further adjusted by prescaling the number of events that pass the selection criteria of specific algorithms. Events accepted by the HLT are then stored by a software process called storage manager, written on disk and sent to offline processing for physics analysis.

3.2.8.3 Offline computing infrastructure

The data produced by the HLT is processed and analyzed by a worldwide computing infrastructure developed specifically for LHC, the Worldwide LHC Computing Grid (WLCG) [106]. The goal of the LHC Computing Project is to build and maintain a data storage and analysis infrastructure for all LHC users. This is achieved with a globally distributed grid that shares the funding and responsibilities amongst member organizations, while benefiting from robustness and data security through redundancy amongst multiple centres.

The CMS offline computing system is arranged in four tiers [107]. A single Tier-0 centre at CERN accepts data from the CMS Online Data Acquisition System, archives the data and performs prompt first pass reconstruction. Several Tier-1 centres in CMS collaborating countries provide services for data archiving, reconstruction, calibration, skimming and other data-intensive analysis tasks. Smaller but more numerous Tier-2 centres provide capacity for analysis, calibration activities and Monte Carlo simulation (see Fig. 3.11). CIEMAT hosts one of these centers. Tier-3 centres provide interactive resources for local groups and additional best-effort computing capacity for the collaboration. A CMS-CERN Analysis Facility centre at CERN provides fast turnaround computing services local to the experiment.

The raw data coming from the Tier-0 is processed into the more manageable Analysis Object Data (AOD), and its reduced versions such as miniAOD and nanoAOD. Events are further stored and analysed in ROOT files [109], that contain all the relevant information and kinematics, providing a tree–like structure of branches, leaves and histogram entries. Physics analyses such as the one presented in this work are performed using an object–oriented dedicated framework, available to the whole collaboration, called CMS Software (CMSSW) [110], with a local version

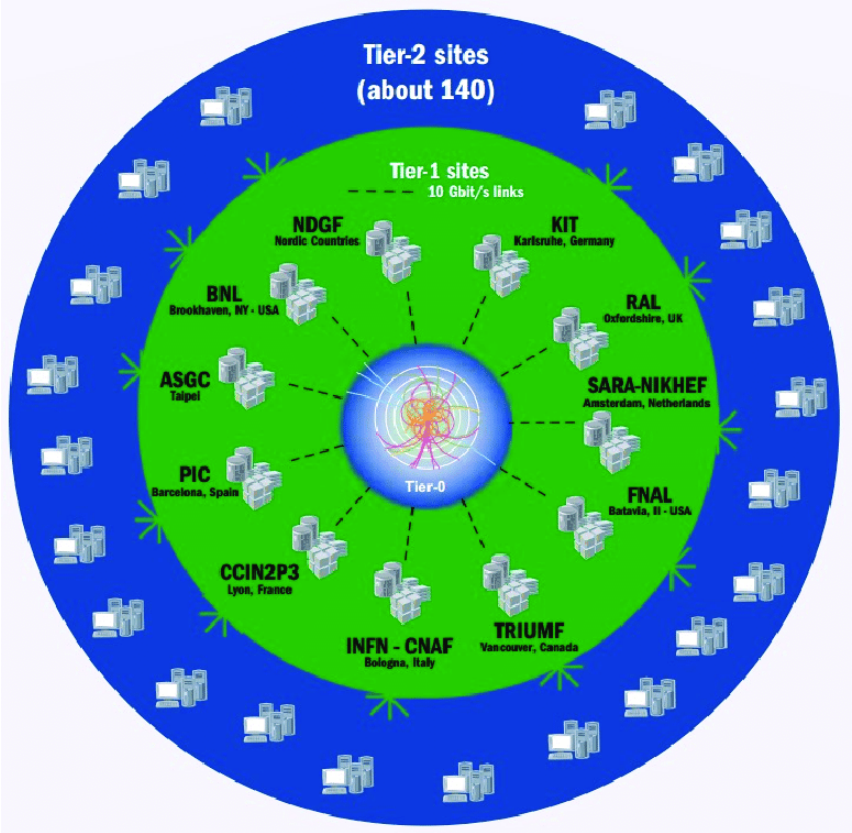


Figure 3.11: WLCG Tier-1 and Tier-2 connections [108].

at CIEMAT, accessible with custom C++ and Python code.

Chapter 4

Reconstruction of the W+c final state

"Ph'nglui mglw'nafh Cthulhu R'lyeh wgah'nagl fhtagn."
— H.P. Lovecraft, The Call of Cthulhu .

Proton collisions in the LHC produce a large number of particles that interact with the different parts of the CMS detector, losing energy which is converted to electronic signals by the detector sensors, as depicted in Fig. 3.5. If we are to roll back the film and investigate what kind of process took part in the original event of interest, we must use these signals to reconstruct trajectories, momenta and energies of the particles, and to build more complex objects, such as particle jets and interaction vertices. This is the way to study high-energy physics such as our topic of interest, the W+c production. In this chapter we will focus on the reconstruction techniques of the building blocks of our analysis.

4.1. Physics objects and particle flow

The particle-flow (PF) reconstruction and global event description [96, 111] is designed to identify and reconstruct every particle in any event using information from all subdetectors. First of all, we must identify the flavour of the particle in order to determine the particle direction and energy.

- Photons are identified as ECAL energy clusters not linked to the extrapolation of any charged particle trajectory to the ECAL. The energy of photons is obtained from the ECAL measurement.
- Electrons are identified as a primary charged particle track and potentially many ECAL energy clusters corresponding to this track extrapolation to the ECAL and to possible bremsstrahlung photons emitted along the way through the tracker material. The energy of electrons is determined from a combination of the track momentum at the main interaction vertex, the corresponding ECAL cluster energy, and the energy sum of all bremsstrahlung photons attached to the track.
- Muons are identified as tracks in the central tracker consistent with either a track or several hits in the muon system, and associated with calorimeter

deposits compatible with the muon hypothesis of minimum ionizing particle. The energy of muons is obtained from the corresponding track momentum.

- Charged hadrons are identified as charged particle tracks neither identified as electrons, nor as muons. The energy of charged hadrons is determined from a combination of the track momentum and the corresponding ECAL and HCAL energies, corrected for the response function of the calorimeters to hadronic showers.
- Neutral hadrons are identified as HCAL energy clusters not linked to any charged hadron trajectory, or as a combined ECAL and HCAL energy excess with respect to the expected charged hadron energy deposit. The energy of neutral hadrons is obtained from the corresponding corrected ECAL and HCAL energies.

After correctly identifying all the particles in an event, we sum their corresponding energies and infer the presence of undetected neutrinos by missing transverse energy $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$.

4.2. Track reconstruction

Charged particles leave a track when travelling across the CMS Tracker, as described in Section 3.2.3. The first step of the reconstruction process is a local reconstruction in this subdetector [89, 112]. It consists of the collection of signals above adjustable thresholds in close-by pixel and strip channels into hits, and then estimating the cluster positions and their uncertainties defined in a local coordinate system. In the pixel detector, this is performed in the readout chips of the sensors. Offline, pixel clusters are formed from adjacent pixels, including both side-by-side and corner-by-corner adjacent cells. Hits from the local reconstruction in the local coordinate system are then used to obtain estimates for the tracks of the traversing charged particles in the global coordinate system. This translation takes into account discrepancies between the assumed and actual location and surface deformation of detector elements as found through the alignment process [113], as well as adding the uncertainty in the detector element location to the intrinsic uncertainty in the local hit position.

The tracking software used at CMS is the Combinatorial Track Finder (CTF), an adaptation of the combinatorial Kalman filter [114, 115], which is also an extension of the Kalman filter [116] to allow pattern recognition and track fitting in the same framework. Track reconstruction is produced by a process called iterative tracking using multiple passes of the CTF, where each iteration excludes tracks found in the previous one. The first iteration looks for the easiest tracks to find, called prompt tracks, that originate near the pp interaction point. These tracks must have $p_{\mathrm{T}} > 0.8\,\mathrm{GeV}$ and three pixel hits. The next iteration recovers prompt tracks that have only two pixel hits. The following one searches for low-p_{T} prompt tracks. Iterations 3–5 try to recover tracks not found in the previous iterations and find tracks that originate outside the beam spot, defined as the 3-D profile of the luminous region where the LHC beams collide in the CMS detector. Furthermore,

each iteration proceeds in four steps. First, initial track candidates with two or three hits are provided by a seed with the initial estimate of the trajectory and its uncertainties. Second, a Kalman filter extrapolates the seed trajectories along the expected flight path, searching for new hits that could match the track candidate. Third, the tracks are fitted again in order to resolve ambiguity and provide the best possible estimate of the parameters of each trajectory. Finally, quality flags are set, from loose to tight, and tracks that fail certain specified criteria are rejected. Seed generation and final track selection are the main differences between all iterations of the process.

Track reconstruction using Kalman Filter techniques achieve efficiencies above 99% on single muons with $p_{\mathrm{T}} > 1\,\mathrm{GeV/c}$. However, LHC events with a high density of charged particles produce high rate of fake combinatorial tracks for low p_{T} ranges, and nuclear interactions in the tracker material reduce the tracking efficiency for charged hadrons. Despite that, these techniques are able to reconstruct charged tracks with $p_{\mathrm{T}} > 300\,\mathrm{MeV/c}$ using the tight selection with an average efficiency $\sim$95%, keeping the fake rate at the per mill level [117].

4.3. Vertex reconstruction

In the context of high luminosity, there will be several collisions at the same time in the beam spot region. First of all, we are interested in finding the primary vertex (PV), the event vertex of the hard scattering. Furthermore, in this particular analysis we will make use of secondary vertices (SV) inside of jets in order to identify the decay of charmed hadrons. The specifics of each case are described as follows.

4.3.1 Primary vertex reconstruction

The identification of the PV is a key element for the study of high-energy processes, and consists of three steps using solely information from the tracker: selection of the tracks, clustering of the tracks that appear to originate from the same interaction vertex, and fitting for the position of each vertex using its associated tracks [89].

Track selection aims to find those produced promptly in the primary interaction region, by imposing requirements on the maximum value of significance, defined as the value of a variable over its uncertainty, of the transverse impact parameter (< 5) relative to the centre of the beam spot, the number of strip and pixel hits associated with a track (≥ 2 pixel layers, pixel+strip ≥ 5), and the normalized χ^2 from a fit to the trajectory (< 20). There is no requirement on the p_{T} of the tracks. The selected tracks are then clustered on the basis of their z-coordinates at their point of closest approach to the centre of the beam spot. This is achieved using a deterministic annealing (DA) algorithm [118], balancing the efficiency for resolving nearby vertices in cases of high pileup against the possibility of splitting a single interaction vertex into more than one cluster of tracks.

Finally, the vertex candidates containing at least two tracks are then fitted using an adaptive vertex fitter [119] to compute the best estimate of vertex param-

eters. It provides the position (x, y, z), the covariance matrix, and indicators for the success of the fit, such as the number of degrees of freedom for the vertex and weights of the tracks used in the reconstruction, reflecting the likelihood that they genuinely belong to the vertex.

Furthermore, we will additionally require for the good vertex selection that the number of degree of freedom in the vertex fit is larger than 4, and the position of the vertex must satisfy $|z| < 24\,\text{cm}$ along the beam line from the nominal centre of the detector and a radius of $|\rho| < 2\,\text{cm}$ in the transverse plane.

For all the vertices reconstructed using this method per bunch crossing, the event PV is identified as the reconstructed vertex with the largest value of summed physics-object p_{T}^2 [9]. The other vertices are considered as pileup vertices.

4.3.2 Secondary vertex reconstruction

A key to the selection of events for this analysis is the reconstruction of SVs inside of jets coming from the decay of charmed hadrons. Given the relatively long lifetime of those hadrons, of the order of $10^{-12}\,\text{s}$, these SVs must be displaced from the PV, meaning that a minimum flight distance must be identified. Displaced SVs are reconstructed with either the Simple Secondary Vertex (SSV) [120] or the Inclusive Vertex Finder (IVF) [121, 122] algorithms, which differ in the tracks used.

The SSV algorithm uses the tracks constituting the jet that are not shared with the PV to follow the adaptive vertex fitter technique. It also includes the flight distance significance, being the flight distance divided by its uncertainty, as a discriminating variable.

The IVF also follows the adaptive vertex fitter, but it starts from a displaced track with respect to the PV, completely independent of jet reconstruction, as a seed track. It then tries to build a vertex from nearby tracks in terms of their separation distance in three dimensions and their angular separation around the seed track. IVF vertices are then associated to the closest jet in a cone of $\Delta\text{R} = 0.3$.

4.4. Lepton reconstruction

The correct reconstruction of leptons is another key point of high-energy analyses. Leptonic and semileptonic decays of massive particles are the most abundant ones and therefore a special treatment is required. We will focus on electrons and muons, since taus have a very short lifetime [17].

4.4.1 Electrons

Electrons leave a distinctive signal in the ECAL as an isolated energy deposit that is also associated with a trace in the silicon tracker [123, 124]. They may interact with the detector material and emit bremsstrahlung photons that in turn split into an electron-positron pair. Thus, the final detection signature may no longer be a single particle but a shower of multiple electrons and photons. Clustering from

the collection particles into a single object is required to calculate the energy of the original electron. Moreover, the loss of momentum caused by the bremsstrahlung radiation changes the curvature of the flying electron in the tracker. Another dedicated tracking algorithm, based on the Gaussian Sum Filter (GSF), is used to estimate the track parameters for electrons [125]. Electron reconstruction in CMS is fully integrated into the PF framework and follows the same principles as other particles.

The initial steps of the energy reconstruction algorithm lead to the clustering of ECAL crystals with energies over a predefined threshold. The cluster containing most of the energy deposited in any specific region of is defined as the seed cluster, with a minimum transverse energy above 1 GeV. Superclusters (SC) are then arranged in a small window in η and an extended window in ϕ around the electron direction to include photon conversions and bremsstrahlung losses around the seed cluster. ECAL clusters, SCs, GSF tracks and generic tracks associated with electrons are then linked into blocks of particles using the PF algorithm. Objects meeting the loose selection requirements with an associated GSF track are labeled as electrons; otherwise as photons. Electrons in jets are more difficult to resolve because the energy and position of the associated supercluster are contaminated by other particle deposits. Therefore, we will only consider isolated electrons in our analysis and exclude electrons in jets, as will be described in Section 5.3.

Background sources for prompt electrons can originate from photon conversions, misidentified hadrons, and semileptonic decays of b or c quarks. Identification of electrons can be achieved using two different techniques. The first one is cut-based and works on sequential requirements using seven identification variables, while the other is based on a multivariate discriminant. In this analysis we will only use the cut-based method [123]. The first variable to reject electron and photon backgrounds involves isolation energy sums. The combined PF isolation applies information from momentum of charged hadrons and energy from photons and neutral hadrons in a cone $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.3$ around electrons. It is defined as:

$$I_{comb} = I_{ch.had} + \max(0, I_{nt.had} + I_{\gamma} - I_{PU}), \tag{4.4.1}$$

where $I_{ch.had}$ is the isolation of charged hadrons, $I_{nt.had}$ is the isolation of neutral hadrons, I_{γ} is the isolation of photons, and I_{PU} is the PU contribution of neutral particles from pileup vertices. This last one is estimated and subtracted with the jet area method described in Ref. [126]. A threshold is then set for the electron's relative combined PF isolation I_{comb}/E_{T}. The next variable exploits the shape of the electromagnetic shower in the ECAL, knowing that a wider shower profile is expected from two photons coming from the decay of neutral hadrons inside a jet than the one from a single incident electron. The hadronic over electromagnetic energy ratio (H/E) is then defined as the energy deposited in the HCAL in a cone of radius $\Delta R = 0.15$ around the SC direction over the energy of the electron candidate. The variable $\sigma_{i\eta i\eta}$ is defined as the second moment of the log-weighted distribution of crystal energies in η, calculated in the 5×5 matrix around the most energetic crystal in the SC and rescaled to units of crystal size. This quantity is tuned to reject ECAL noise and essentially depends on the distance between two crystals in η. The distribution of $\sigma_{i\eta i\eta}$ is expected to be narrow for electron showers, and broad

for two-photon showers that arise from neutral meson decays. The other relevant variables are tracker-related, such as $|1/E - 1/p|$, combining the SC energy E and the track momentum p at the point of closest approach to the vertex. The $|\Delta\eta_{\text{in}}^{\text{seed}}|$ is defined as $|\eta_{\text{seed}} - \eta_{\text{track}}|$, where ηseed is the position of the seed cluster in η, and ηtrack is the track η extrapolated from the innermost track position. In the same way, $|\Delta\phi_{\text{in}}|$ uses the SC energy-weighted position in ϕ instead of the seed cluster ϕ as $|\phi_{\text{SC}} - \phi_{\text{track}}|$. Finally, if photon conversion takes place inside the tracker volume, it is very probable that the first hit of the new electron tracks is not located in the innermost tracker layer, whereas no missing hits are expected in those layers for prompt electrons.

There are four working points generally used in CMS, depending on the values required for the set of parameters described above. The veto working point, corresponding to an average signal efficiency of about 95%, aims to reject events with more reconstructed electrons than expected from the signal topology. The loose working point refers to a signal efficiency of around 90%, and is chosen for analyses with low backgrounds to electrons. The medium working point results from an average signal efficiency of around 80%, and may be used for generic measurements involving heavy bosons. Finally, the tight working point corresponds to an efficiency of around 70% for genuine electrons, and is used for analyses with larger backgrounds.

The electron momentum is estimated by combining the energy measurement in the ECAL with the momentum measurement in the tracker. The momentum resolution for electrons with $p_{\text{T}} \approx 45\,\text{GeV}$ from $\text{Z} \to \text{e}^+\text{e}^-$ decays ranges from 1.6 to 5%. This variation depends on the electron η, being generally better in the barrel region than in the endcaps, and also on the bremsstrahlung energy emitted by the electron as it traverses the material in front of the ECAL [123, 124].

4.4.2 Muons

Muons are reconstructed using the muon detectors, as well as the inner tracker to measure their momentum. In fact, we can make a distinction between standalone, global, and tracker muons, depending on the detector where they are reconstructed. Standalone muons are built clustering hits from DTs or CSCs as seeds and gathering CSC, DT, and RPC hits along the muon trajectory using a Kalman-filter technique. Tracker muons are built matching tracker tracks with $p_{\text{T}} > 0.5\,\text{GeV}$ and a total momentum $\text{p} > 2.5\,\text{GeV}$ to a muon segment of at least one layer of the DTs or CSCs. Finally, global muons are built by matching standalone-muon tracks with tracker tracks with the Kalman filter and checking if the parameters of the two tracks propagated onto a common surface are compatible. We will only use global muons in this analysis taking advantage of the higher efficiency in reconstruction due to the activation of more than one muon detector plane and the inner track information.

A kink-finding algorithm splits the muon track in the tracker detector into two separate tracks at several places along the trajectory. For each split the algorithm makes a comparison between the two separate tracks, requiring that the two tracks

are compatible with being a single track. This kinks would indicate the decay of charged pions or kaons into a muon and a neutrino. The neutrino escapes the detector and the track of the new muon suffers a sudden change of direction.

Once we have selected our global muons, we can further classify them depending on how pure we want them to be, using some variables such as track fit χ^2, the number of hits per track or the degree of matching between tracker tracks and standalone-muon tracks.

Loose muon identification (ID) requires the selected candidate only to be either a tracker or a global muon. This aims to identify prompt muons and those from light and heavy flavour decays with a low rate of the misidentification of punch-through.

Medium muon ID is applied to a loose muon with a tracker track that uses hits from more than 80% of the inner tracker layers it traverses. This is optimized for prompt and heavy flavour decay muons.

Tight muon ID requires a muon reconstructed as both a tracker and a global, with a tracker track that uses hits from at least six layers of the inner tracker including at least one pixel hit and a segment matching in at least two of the muon stations. The global muon fit must have goodness-of-fit per degree of freedom $\chi^2/\text{dof} < 10$ and include at least one hit from the muon system. It must be compatible with the primary vertex, considering impact transverse and longitudinal parameters $|dxy| < 0.2\,\text{cm}$ and $|dz| < 0.5\,\text{cm}$. This aims to suppress muons from decay in flight and from hadronic punch-through.

Soft muon ID is optimized for low-p_T muons in physics analyses involving heavy flavour quarks. It is a tracker muon with a tracker track that uses hits from at least six layers of the inner tracker including at least one pixel hit. This muon may be loosely compatible with the primary vertex, with $|dxy| < 0.3\,\text{cm}$ and $|dz| < 20\,\text{cm}$.

High momentum muon ID refers to muons with $p_\text{T} > 200\,\text{GeV}$. It is reconstructed as both a tracker and a global muon and has the same initial requirements as the tight muon, but the global muon fit χ^2/dof is removed. This removes inefficiencies at high p_T when muons radiate as they pass through the steel flux-return yoke and produce additional hits in the muon chambers.

In order to resolve prompt muons from those inside of a jet coming from hadronic decays we must evaluate some isolation criteria. This variable is defined as the sum of the transverse momenta of the charged hadrons and the transverse energy of the neutral hadrons and photons, around a cone centred on the lepton direction with radius $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.4$:

$$I_{comb} = \sum_{ch.had} p_\text{T} + \sum_{nt.had} E_T + \sum_{\gamma} E_T - \frac{1}{2}\sum_{PU} p_\text{T}, \tag{4.4.2}$$

where the sums run over the charged hadrons, neutral hadrons and photons, with a distance ΔR to the lepton, and the pile-up term accounts for the additional energy from adjacent collisions. Selection criteria will be relative to the transverse momentum of the muon candidate. The factor one-half accounts for the expected ratio of

neutral to charged particle production in hadronic interactions.

Matching muons to tracks measured in the silicon tracker results in a relative transverse momentum resolution, for muons with p_T up to 100 GeV, of 1% in the barrel and 3% in the endcaps. The p_T resolution in the barrel is better than 7% for muons with p_T up to 1 TeV [101].

4.5. Jet reconstruction

Having identified and removed from the PF blocks muons, electrons, and isolated photons, we focus on hadrons from jet fragmentation and hadronization. Quarks and gluons emerging from pp collisions cannot live alone due to the asymptotic freedom of colour-carrying particles, as seen in Section 2.1.3.1. This hadronization process also involves the creation of quark-antiquark pairs from the vacuum due to the high energy density. Each member of the pair would find suitable partners from nearby interactions, leading to the formation of a colour string, which in turn may produce high multiplicity hadronic jets in the form of a collimated spray.

For each event, hadronic jets are clustered from these reconstructed particles using the infrared and collinear safe anti-k_T algorithm [127, 128] with a distance parameter of $\Delta R = 0.4$. There are three types of jets, depending on the the combination of individual contributions from each subdetector: Calorimeter jets (CALO), Jet-Plus-Track (JPT) jets and Particle-Flow (PF) jets. CALO jets are reconstructed from energy deposits in the calorimeter towers, consisting of at least one HCAL cells and the corresponding ECAL crystals. JPT jets are reconstructed CALO jets improved by adding tracking information, according to the Jet-Plus-Track algorithm [129]. The PF jets are reconstructed by clustering the four-momentum vectors of PF candidates, described earlier. We will only use PF jets in this analysis.

Jet momentum is determined as the vectorial sum of all particle momenta in the jet, and is found from simulation to be within 5 to 10% of the true momentum over the whole p_T spectrum and detector acceptance. On average, 65% of the jet energy is carried by charged hadrons, 25% by photons, and 10% by neutral hadrons. Pileup can contribute with additional tracks and calorimetric energy depositions to the jet momentum. To mitigate this effect, charged particles identified to be originating from pileup vertices are discarded and an offset correction is applied to correct for remaining contributions. Jet energy corrections (JEC) are derived from simulation to bring the measured response of jets to that of particle level jets on average. In situ measurements of the momentum balance in dijet, photon + jet, Z + jet, and multijet events are used to account for any residual differences in the jet energy scale (JES) between data and simulation [130]. The (JER) amounts typically to 15–20% at 30 GeV, 10% at 100 GeV, and 5% at 1 TeV [130]. Additional selection criteria are applied to each jet to remove jets potentially dominated by anomalous contributions from various subdetector components or reconstruction failures.

The PF jet momentum and spatial resolutions are greatly improved with respect to calorimeter jets, as the use of the tracking detectors and of the high granularity of ECAL allows resolution and measurement of charged hadrons and photons

inside a jet, which together constitute ∼85% of the jet energy [131].

In order to mitigate the impact of PU on the jet reconstruction, charged hadrons can be identified as coming from pileup by associating their track with a pileup vertex, thus removing them from the list of reconstructed particles used to form physics objects in the event. This algorithm is called pileup charged-hadron subtraction (CHS) [96].

4.5.1 Jet flavour tagging

The nature of the particles coming from the original interaction may be better identified if we are able to correctly classify jets as originating from light-flavour or heavy-flavour quarks. Current techniques make use of machine learning and neural networks such as DeepJet [132, 133] and DeepCSV [134, 135] to identify light-flavour jets from u, d, s quarks or gluons, charm jets from c quarks, and heavy-flavour jets from b quarks. These *taggers* evaluate the probability or likelihood of a jet of being associated with one flavour or another, making use of information from the tracks and secondary vertices to reconstruct the flight path and disintegration of the hadron. Further comparison of these algorithms using simulation and real data is used to better understand the performance of the detector. In general, the reconstruction of simulated data tends to identify heavy-flavour jets more efficiently than in observed data. Therefore, multiplicative scale factors are used to correct the differences in efficiency.

The distributions of the tagging variables for c jets lie in between the distributions for b and light-flavour jets [136]. The reason for this is the shorter lifetime of the charmed hadrons, as well as the lower secondary vertex multiplicity and the smaller c quark mass. Since efficiently identifying charm jets has proven to be particularly challenging, in this analysis we will not use this kind of techniques, but rather a simplified and more inclusive selection of jets that provide both the charm tagging and the electric charge of the original charm quark. We will use two strategies, requiring that the jet that potentially contains the charmed hadron either includes a muon from the semileptonic decay of the c quark, or a displaced secondary vertex from the hadronic decay of the c quark. We will discuss this selection in detail in Section 5.3.2.

4.6. Missing transverse momentum

The decay of a W boson into a charged lepton and its corresponding neutrino cannot be precisely measured because of the amount of energy carried away by the neutrino that escapes the detector. Nevertheless, we can estimate the unmeasured (missing) momentum using the PF event reconstruction described earlier and bearing in mind a key feature of hadron colliders. The net momentum in the plane transverse to the beam is known to be nearly zero in pp collisions. We can therefore compute the sum of the transverse momenta of all the PF candidates in an event and define the negative resulting vector as the the missing transverse momentum vector, with its magnitude denoted as $p_{\mathrm{T}}^{\mathrm{miss}}$ [137]:

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} = -\sum \vec{p}_{\mathrm{T}} \tag{4.6.1}$$

The $p_{\mathrm{T}}^{\mathrm{miss}}$ relies on the accurate measurement of the reconstructed physics objects, namely muons, electrons, photons, hadronically decaying taus, jets, and unclustered energy (E_U). The E_U is the contribution from the PF candidates not associated with any of the previous physics objects.

The $\vec{p}_{\mathrm{T}}$ is modified to account for corrections to the energy scale (JES) of the reconstructed jets in the event. Specifically, the particle-flow objects clustered in a jet are replaced by the corrected jet. The JES uncertainties are less than 3% for jets within the tracker acceptance and 1–12% for those outside. The jet energy resolution (JER) uncertainties typically range between 5–20%. The muon energy scale uncertainty is 0.2%, and the electron and photon energy scale uncertainties are 0.6% in the barrel and 1.5% in the endcap [138]. The uncertainties related to the leptons are significantly smaller than those of JES and JER and are therefore not considered.

There are some processes, like $\mathrm{Z} \to \ell^+\ell^-$ or γ decays, that do not produce neutrinos, but the recorded events also present some amount of $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$. This may be due to the finite resolution of the calorimeters and other instrumental effects, such as insensitive areas at the location of cables and pipes. The analysis of these events is a useful tool to calibrate the detector and provide a reliable measurement of $\mathrm{E}_{\mathrm{T}}^{\mathrm{miss}}$ for those processes that do present neutrinos (or any other exotic particle for that matter).

Chapter 5

Analysis of W+c production

"En fait, nous sommes une liberté qui choisit, mais nous ne choisissons pas d'être libres: nous sommes condamnés à la liberté."
— J.P. Sartre, L'être et le néant.

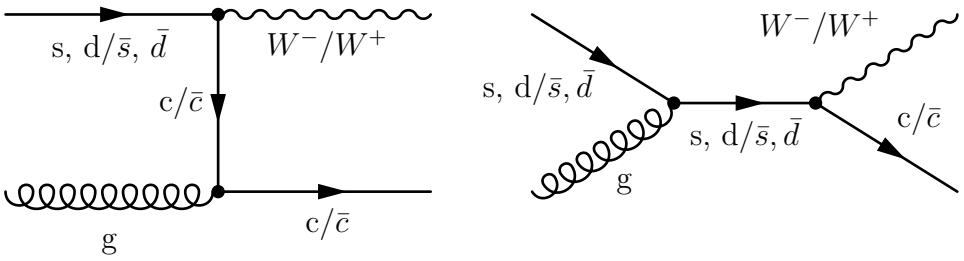
After describing the mathematical and experimental tools required for a high energy physics investigation, we can now proceed with the detailed analysis of the associated production of a W boson and a charm quark (W+c production). Following a general overview of the analysis, we introduce the data and simulated samples used for the measurements and the selection strategy followed to isolate the W+c signal events and evaluate the background contributions. Finally, we evaluate the systematic effects that affect our measurements and quantify their impact.

5.1. Overview of the analysis

As mentioned earlier, the associated production of a W boson and a single charm quark in pp collisions is directly sensitive to the strange quark content of the colliding protons at an energy scale of the order of the W boson mass. This is due to the dominance of the sg→W+c contribution, following the CKM-matrix elements in Eq. 2.1.31, over the strongly Cabibbo suppressed process dg→W+c at tree level, as depicted in Fig. 5.1. Furthermore, the production of W +c events provides a useful calibration sample for the measurements and searches at the LHC involving electroweak bosons and c quarks in the final state [139, 140]. Precise measurements of W + c production can be used to verify the theoretical calculation of this process and its modeling in the currently available MC event generators.

Previous publications, like the ones by the CMS [7–9], ATLAS [10] and LHCb [11] Collaborations at center-of-mass energies of 7, 8 and 13 TeV, identify the charm quark via exclusive decay channels of charmed hadrons, or inclusive final states reconstructed through a muon or a secondary vertex inside a jet. In this study we follow the strategy developed in the CMS 8 TeV W+c analysis, selecting four separated data samples depending on the W boson leptonic decay (either to electron or muon, and a neutrino) and the charm tagging mode (muon or SV inside a jet).

We are using data collected by the CMS detector between 2016 and 2018 of pp collisions at $\sqrt{s} = 13\,\text{TeV}$ corresponding to an integrated luminosity of $138\,\text{fb}^{-1}$. The



(a) (b)

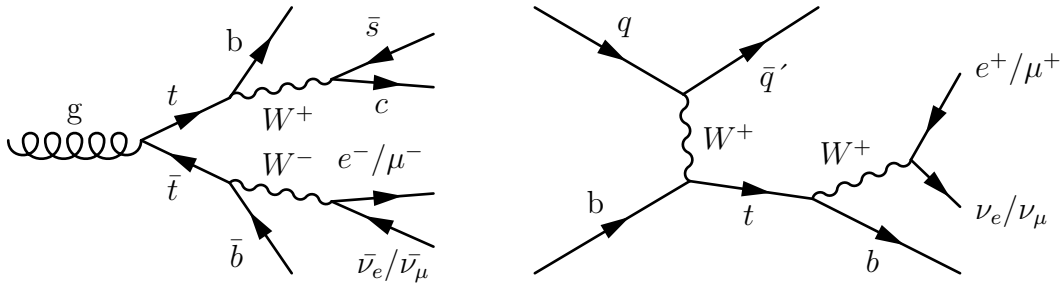
Figure 5.1: Leading order diagrams for the associated production of a W boson and a charm quark. The electric charges of the W boson and c quark have opposite sign.

cross sections $\sigma(\mathrm{W}+\mathrm{c}) \equiv \sigma(\mathrm{pp} \to \mathrm{W}+\mathrm{c}+\mathrm{X})\mathcal{B}(\mathrm{W} \to \ell\nu)$ and the cross section ratio $R_c^{\pm} \equiv \sigma(\mathrm{W}^+ + \bar{\mathrm{c}})/\sigma(\mathrm{W}^- + \mathrm{c})$ are measured. Inclusive and differential cross sections are measured as functions of the transverse momentum (p_{T}^{ℓ}) and pseudorapidity (η^{ℓ}) of the lepton from the W boson decay.

The first step of the analysis consists on the selection of W+c events. We will identify the W boson through its leptonic decay into an electron or a muon, and a neutrino. We do not consider the channel where the W boson decays to a tau lepton. The electrons or muons with enough transverse momentum and within the acceptance of the detector will fire the online trigger selection of the event. Events with an electron or muon coming from the decay of a tau lepton, which in turn comes from the decay of a W boson, are considered as a background. This contribution is small, around 1%, given the relatively small leptonic decay fraction of the tau lepton (∼17%) and the lower transverse momentum of the resulting electron or muon that in most cases is below the trigger threshold.

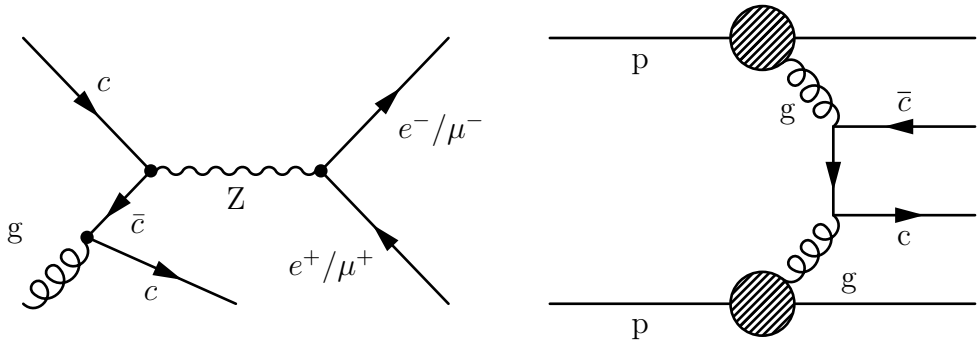
Electrons are detected at the ECAL, muons are reconstructed using the muon chambers and the escaping neutrinos will be estimated as missing transverse momentum. Since the W boson is quite massive, the resulting electron or muon will carry high momentum. In addition, the lepton from the W decay will be isolated from surrounding particles, contrary to leptons produced inside jets. The c quark cannot be measured directly but can be inferred through the reconstruction of the jet of particles coming from its fragmentation and hadronization.

The signal signature, an isolated, high-momentum lepton and a heavy flavour jet, can also be mimicked by other processes, the so-called backgrounds, that will be suppressed through the selection process described below, and their remaining contributions will be evaluated. Backgrounds include top quark-antiquark pair production ($t\bar{t}$, Fig. 5.2a), single top production (Fig. 5.2b), diboson processes (WW, WZ, and ZZ, Figures 5.4), the production of a Z boson (or a virtual photon) in association with jets (Z+jets, Fig. 5.3a), W+$c\bar{c}$ or W+$b\bar{b}$ events, and events composed uniquely of jets produced through the strong interaction, referred to as QCD multijet events (Fig. 5.3b).



(a) $t\bar{t}$ production. (b) Single top production.

Figure 5.2: Feynman diagrams of top quark background processes.



(a) Z+jets production. (b) Multijet QCD production.

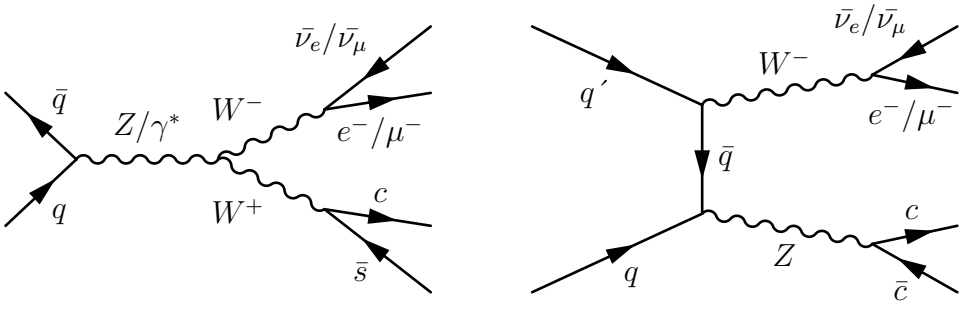
Figure 5.3: Feynman diagrams of Z+jets and QCD background processes.

One of the fundamental aspects of this analysis is the technique applied to reduce or eliminate background processes and achieve a c jet enriched signal sample. In the W+c process, the electric charges of the lepton from the W decay and the c quark always have opposite sign (OS), whereas most backgrounds present an even number of final states with same sign (SS) and OS electric charges with equal probability and identical kinematics. We can then obtain an effectively enriched sample of c jets by statistical OS-SS substraction, applying negative weights to the SS events in our analysis.

We use MC simulations to estimate the background contributions after OS-SS substraction. We will also use MC W+c samples to calculate the fraction of signal events that fulfill the reconstruction and selection criteria of the analysis. This efficiency, defined as the number of selected signal events divided by the number of produced events in the kinematical region of the analysis, is an important ingredient in the cross section measurements.

5.2. Data and simulated samples

This analysis has been performed using data from pp collisions at a center-of-mass of 13 TeV collected by the CMS experiment corresponding to an integrated lu-



(a) WW production. (b) WZ production

Figure 5.4: Feynman diagrams of diboson background processes.

minosity of 138 fb^{-1} during the 2016 (36.3 fb^{-1}), 2017 (41.5 fb^{-1}) and 2018 (59.7 fb^{-1}) data-taking periods [141–143].

Events with a high-p_{T} lepton from the decay of a W boson are selected online by a single-lepton trigger path that requires the presence of an electron (muon) candidate with minimum p_{T} that varies for each data-taking period. Each trigger path is the lowest non-prescaled one available for each year, all of them still looser than the minimum p_{T} required for the analysis. Therefore, we require that the event fires the trigger paths corresponding to an electron (muon) candidate with minimum p_{T} of 27, 32, and 32 GeV (24, 27, and 24 GeV) during the 2016, 2017, and 2018 data-taking periods, respectively.

Samples of signal and background events are simulated using MC event generators based on a fixed-order perturbative QCD calculation, supplemented with parton showering and multiparton interactions. Signal samples of W boson plus jets are generated with MadGraph5_aMC@NLO v2.6.3. Two sets of samples were produced, a set with up to four partons in the final state calculated with LO accuracy, and a second set with up to two partons computed at NLO accuracy. The scheme for merging the events produced on the basis of the matrix elements calculations for the hard, partonic process, and those produced with parton showers, in the samples generated at LO is MLM [63] whereas at NLO is FxFx [144]. The parton shower, hadronization and the underlying events are modelled by pythia v8.212. The pythia8 parameters for the underlying event modeling are set to the CP5 tune (2017, and 2018 samples) and CUETP8M1 tune (2016 sample), respectively. For 2017 and 2018 samples, the NNLO NNPDF 3.1 [145] set is used. The MC samples for 2016 conditions include the NLO NNPDF 3.0 [146] as PDF set. The matching scale $m^2 = (10\,\text{GeV})^2$ and the factorization and normalization scales are set at $q^2 = M_{\text{W}}^2 + (p_{\text{T}}^{\text{W}})^2$.

Regarding the main backgrounds, a sample of $\text{t}\bar{\text{t}}$ events is generated with powheg v2.0, interfaced with pythia8. The Z+jets background is generated with MadGraph5_aMC@NLO under the same conditions as the W+jets samples. The cross sections of the Z/W + jets simulations are normalized to the NNLO cross section predictions from FEWZ [147]. We also consider small contributions

from WW, WZ and ZZ diboson events generated with PYTHIA8. The single top quark t–channel, tW, and s–channel processes are simulated with POWHEG and MADGRAPH5_AMC@NLO, respectively, and interfaced with PYTHIA8 for parton showering and hadronization.

The generated events are processed through the full GEANT4 detector simulation and trigger emulation. These events are then reconstructed using the same algorithms as those used to reconstruct collision data.

The simulated signal sample is composed of W bosons accompanied by jets originating from quarks of all flavours (b, c, and light) and gluons. Simulated W+jets events are classified according to the flavour of the outgoing generated partons. A W+jets event is categorized as W+b if at least one bottom quark was generated in the hard process. It is classified as W+c signal event if it contains an odd number of charm quarks in the final state, as expected from the presence of a weak charged current exchange. Only outgoing quarks from the hard interaction are considered. Events containing an even (non-zero) number of charm quarks are assigned to the W+$c\bar{c}$ category. Remaining events, where no bottom or charm quark was produced, are defined as W+usdg. Only generator level partons with angular separation with respect to the generator lepton from the W-boson decay larger than $\Delta R > 0.4$ are considered.

The simulated samples generated using Monte Carlo methods, with their cross sections, for the signal and main background processes are listed in Table 5.1.

Table 5.1: Simulated signal and background datasets and their corresponding generators with parton shower and merging schemes.

Process	Generator
W+jets (LO) W+1jet W+2jets W+3jets W+4jets	MADGRAPH5_AMC@NLO MLM
W+jets (NLO)	MADGRAPH5_AMC@NLO FxFx
Z+jets	MADGRAPH5_AMC@NLO MLM
$t\bar{t}$ semileptonic $t\bar{t}$ dileptonic $t\bar{t}$ hadronic	POWHEG + PYTHIA8
t (s channel)	MADGRAPH5_AMC@NLO + PYTHIA8
t (t channel) t ($\bar{t}$ channel) t (tW channel) t ($\bar{t}$W channel)	POWHEG + PYTHIA8
WW WZ ZZ	PYTHIA8

5.2.1 Simulation corrections

These simulated samples are corrected by a series of data/simulation SFs that take into account the discrepancies between data and MC due to inefficiencies in identification and reconstruction of the various physical objects. Most of these SFs are provided by dedicated groups of CMS and we will describe in detail those that we have calculated specifically for this analysis.

5.2.1.1 Pileup

In many cases, it is useful to know the total number of pp interactions per bunch crossing. The most straightforward way of doing this is to use the instantaneous luminosity, given by a method involving the total pp inelastic cross section and the LHC orbit frequency, necessary to convert from the instantaneous luminosity, which is a per-time quantity, to a per-collision quantity. This quantity can be computed on a per-lumi section basis, where a lumi section is the fundamental unit of CMS luminosity calculation. The pileup estimation obtained from the luminosity can be compared with the number of reconstructed primary vertices per event. As the latter is affected by the vertex reconstruction efficiency, these two quantities will not necessarily agree. However, these can be used to measure the vertex reconstruction efficiency or validate the results. The value obtained from the instantaneous luminosity is a measure of the average pileup during a single lumi section. The distribution of pileup for individual events will therefore be a Poisson distribution around this average.

The Monte Carlo simulated samples, in addition to the hard scattering collision, incorporate a number of pileup collisions which are registered together when building the event data. To reach an optimal agreement with the experimental data, the simulated distributions are reweighted assuming a total inelastic pp cross section of 69.2 mb.

5.2.1.2 Charm hadron fragmentation fractions and decay branching ratios

To account for discrepancies between PYTHIA8 and data in the charm fragmentation fractions ($c \to D_h$), defined as the probabilities for c quarks to hadronize as particular charm hadrons, and branching ratios ($D_h \to \ell$ or $D_h \to h$), we have updated the simulation to match the charm fractions and decay branching ratios to more modern values. We have reweighted the charm fractions to match the values in [148] and the decay branching ratios with [149]. Given the large number of hadronic decay modes, we have reweighted only those with 3 or more tracks in the final state and a branching ratio >5%. The weights used in the analysis are listed in Tables 5.2 and 5.3. For each value of the weight, its error was propagated from the original relative error.

5.2.1.3 Lepton efficiencies

Simulated samples are corrected for differences between data and MC description in lepton trigger, reconstruction and identification efficiencies (ϵ_ℓ). Lepton

Table 5.2: The fragmentation fractions have been reevaluated with the latest set of PDG values applying the weights shown here. For each value of w the error was propagated from the original relative error.

Fragmentation fractions
$w_{c\to D^0} = 1.08 \pm 0.01$
$w_{c\to D^\pm} = 0.83 \pm 0.01$
$w_{c\to D_s} = 0.83 \pm 0.05$
$w_{c\to \Lambda_c} = 1.74 \pm 0.06$

Table 5.3: The decay branching ratios have been updated to more modern values with the weights shown here. For each value of w the error was propagated from the original relative error.

Branching ratios								
$w_{D^0\to\ell}$	=	0.99 ± 0.01	$w_{D^0\to K\pi\pi\pi}$	=	1.11 ± 0.02	$w_{D^\pm\to K\pi e\nu}$	=	14.90 ± 0.60
$w_{D^\pm\to\ell}$	=	1.03 ± 0.02	$w_{D^0\to K\pi\pi\pi\pi_0}$	=	1.12 ± 0.02	$w_{D_s\to KK\pi}$	=	4.90 ± 0.13
$w_{D_s\to\ell}$	=	0.82 ± 0.25	$w_{D^\pm\to K\pi\pi}$	=	1.02 ± 0.01	$w_{\Lambda_c\to pK\pi}$	=	2.94 ± 0.15
$w_{\Lambda_c\to\ell}$	=	0.76 ± 0.10	$w_{D^\pm\to K\pi\pi\pi_0}$	=	5.20 ± 0.15			

efficiencies are evaluated by dedicated teams within the Collaboration devoted to understand the reconstructed objects in data and simulation. With that purpose, clean samples of dilepton events in the Z mass peak are evaluated with the tag-and-probe method [150] and a correction factor $\epsilon_\ell^{\text{data}}/\epsilon_\ell^{\text{MC}}$, binned in p_{T} and η of the leptons is computed. The difference in performance in data and MC is corrected by the centrally provided ratio. These scale factors (SF) are applied to the simulation as weights on an event-by-event basis. In addition, for 2016 data-taking period, the SF was calculated separately for the beginning and the end of the period, so that the final SF for 2016 is estimated as the weighted sum of the two SFs using the integrated luminosities corresponding to each subperiod.

The muon momentum scale is also corrected separately to compensate for biases caused by detector misalignment or magnetic field uncertainties. This is called Rochester correction [151], as the university responsible for developing it, and is applied to both MC samples and data before the selection cuts to mitigate the biases occurred during muon p_{T} reconstruction by reweighting the muon p_{T} by the corresponding SF. These SFs are calculated using the invariant mass spectrum of the dimuon decay of the Z boson to find the momentum scale bias.

During the 2016 and 2017 data-taking, a gradual shift in the timing of the inputs of the ECAL L1 trigger in the region at $|\eta| > 2.0$ caused a specific trigger inefficiency [102]. As a result, a fraction of high-η trigger primitives are being mistakenly associated to the previous bunch crossing (BX-1). Since the L1 trigger rules forbid two consecutive bunch crossing to fire, the event can veto itself if large ECAL energy is found in the region $2 < |\eta| < 3$. Around 1(2)% of the events for $\eta_\mu(\eta_e) > 2.1$ are lost when no correction is accounted for. For events containing an electron (a jet) with p_{T} larger than $\approx$50 GeV ($\approx$100 GeV), in the region $2.5 <$

$|\eta| < 3.0$ the efficiency loss is ≈10–20%, depending on p_T, η, and time. This effect is present in 2016 and 2017 data set and is not simulated in the MC. Correction factors were computed from data and applied to the acceptance evaluated by simulation. The uncertainties related to the L1 prefiring problem are at the per mil level and not considered in this analysis except for the last bin in η of the μ differential analysis, where we assign a 5%.

5.2.1.4 Jet corrections

The detector response to particles is not linear and therefore it is not straightforward to translate the measured jet energy to the true particle or parton energy [131]. The jet corrections are a set of tools that allows the proper mapping of the measured jet energy deposition to the particle-level jet energy. CMS has adopted a factorized solution to the problem of jet energy corrections, where each level of correction takes care of a different effect. Each level of correction is essentially a scaling of the jet four momentum with a SF which depends on various jet related quantities (p_T, η, flavour, etc.). The levels of correction are applied sequentially, meaning that the output of each step is the input to the next, and with fixed order.

L1 Pile Up. The goal of the L1 correction is to remove the energy coming from pile-up events. In principle this will remove any dataset dependence on luminosity so that the following corrections are applied upon a luminosity independent sample. The pileup offset corrections are determined from the simulation of a sample of QCD dijet events processed with and without pileup overlay. They are parameterized as a function of offset energy density ρ, jet area A, jet η and p_T. Different L1 corrections are applied to data and MC.

L2L3 MC-truth corrections. The simulated jet response corrections are determined on a QCD dijet sample, by comparing the reconstructed p_T to the particle-level one. The corrections are derived as a function of jet p_T and η and make the response uniform over these two variables.

L5 flavour corrections. As explained, the MC-truth based corrections are extracted for a QCD flavour mixture sample. Optional jet-flavour corrections are derived from MC simulation, using Z+jet and photon+jets simulated events. These corrections are provided for the Z+jet and photon+jets mixtures, and also for pure flavours. JEC flavour uncertainties are also extracted, and can be applied to data vs simulation comparisons regardless of whether or not the jet-flavour corrections are applied in the analysis. The flavour corrections and their uncertainties for b-quark jets are checked in data with Z+b events.

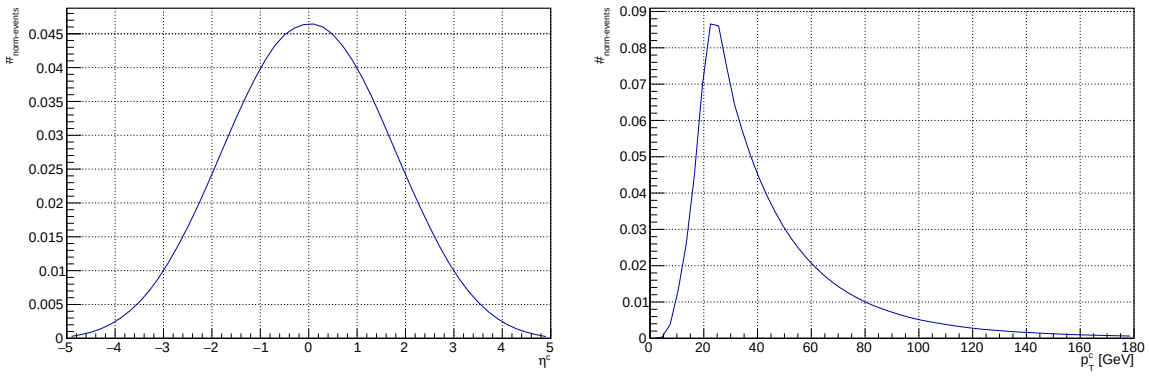
L2L3 Residuals. The L2 and L3 residuals are meant to correct for remaining small differences (of the order of %) within jet response in data and MC. The L2Residuals η-dependent corrections are determined with dijet events, relative to a jet of similar p_T in the barrel reference region. These corrections include a p_T dependence of the JES relative to the JES of the barrel jet. The L3Residuals

correct the jet absolute scale (JES vs p_T). These corrections are determined, for barrel jets, using Z($\mu\mu$, ee)+jet, photon+jet and multijet events. Relative (vs η) and Absolute (p_T) residual corrections are determined separately, but are stored in one single L2L3Residuals step for the analysers.

5.2.2 Signal kinematics

Before describing the selection requirements at the detector level, it is important to explore what kind of signal we expect to find at the generator level. The following plots represent the kinematic distributions of our process of interest, focusing on the transverse momentum and angular distribution of the leptons coming from the W boson and the c quark. The following distributions are normalized to 1 in order to show the probability density of each variable, so that we can estimate which values are most probable or what fraction of events is lost after any selection requirement. We expect very similar kinematics for each decay channel, so that we can safely unify the selection requirements.

Figure 5.5 shows the distributions of the charm quark before hadronization, with Fig. 5.5a being the η and Fig. 5.5b the p_T. Due to the symmetry of the pp collision along the z axis, we observe a symmetric distribution in η, which may allow us to later consider absolute values of this variable. About 90% of the produced c quarks lie within the detector acceptance ($|\eta| < 2.5$). The p_T distribution peaks sharply at around 25 GeV, with a smooth tail to higher values. This can be an initial consideration for our lower selection requirements for the transverse momentum of the jet containing the charm decays.

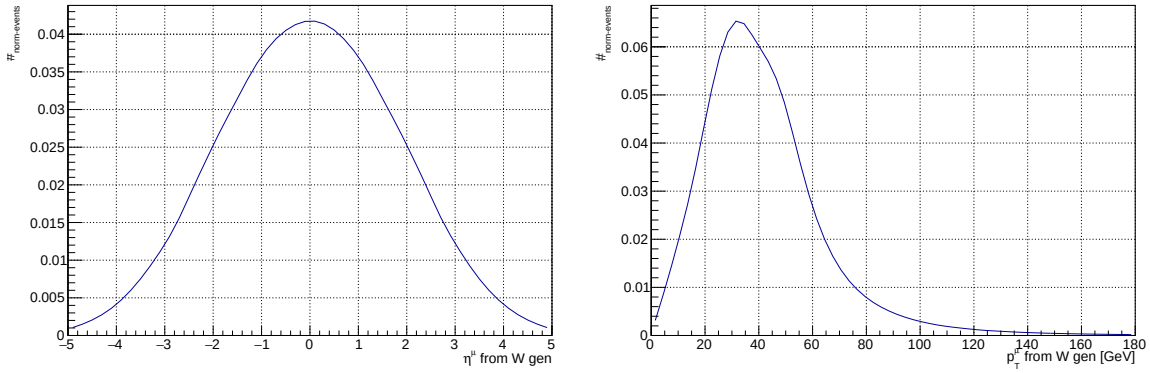


(a) (b)

Figure 5.5: Distributions of the pseudorapidity and transverse momentum of the charm quark at generator level.

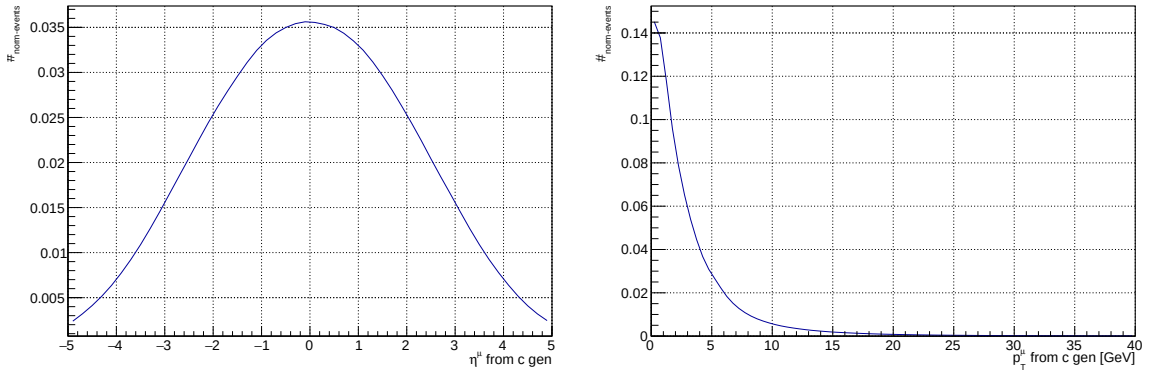
Figures 5.6a and 5.7a show the distributions of the η of the electron coming from the W boson decay and of the muon coming from the decay of the charm hadron, respectively. Due to the symmetry of the collision, we also observe symmetric distributions, which can be propagated to a selection requirement considering only absolute values. We are only showing the W $\to e\nu$ channel because the kinematical distributions will be identical for the W $\to \mu\nu$ channel, since the energy of

the process is much higher than the masses of both leptons.



(a) (b)

Figure 5.6: Distributions of the pseudorapidity and transverse momentum of the electron coming from the decay of the W boson at generator level.



(a) (b)

Figure 5.7: Distributions of the pseudorapidity and transverse momentum of the muon coming from the decay of the charm hadron at generator level for the SL channel.

Figure 5.6b shows the distribution of the p_{T} of the lepton coming from the W boson. In this case the peak is reached at around 30 GeV, and the threshold of the online trigger paths will force us to lose almost half of the events. Finally, Fig. 5.7b presents the distribution of the p_{T} of the muon coming from the semileptonic decay of the charm hadron. The distribution peaks at very low values and then quickly decays, with very few events over 25 GeV. It is worth noting that muons with p_{T} lower than around 3 GeV will not activate enough layers in the muon chambers and those events will therefore be lost. The amount of such muons is also around 50% of the total sample.

5.3. Event selection

The selection of W+c events follows closely the strategy used in the published W+c cross section analysis at $\sqrt{s} = 7\,\text{TeV}$ [7] and $\sqrt{s} = 8\,\text{TeV}$ [8]. We first apply a set of requirements to select events with a W boson that decays leptonically $\text{W} \to \ell + \nu$ (where $\ell = \text{e}, \mu$) and then select events with a c jet. The first condition that will be required for an event to be accepted is the presence of an isolated high-p_{T} electron or muon. Therefore, the event must fire the corresponding trigger paths described in the previous section.

5.3.1 $\text{W} \to \ell + \nu$ selection

The leptonic decay of a W boson into a muon or an electron and a neutrino is characterized by the presence of a high-transverse-momentum, isolated lepton. Electrons and muons are selected using tight identification criteria following the definitions provided in Refs. [101, 123], as described in Section 4.4. Although they share very similar selection requirements, we will treat these two decay channels separately throughout this analysis, since both leptons are two different objects from the detector point of view, with their specific reconstruction and systematics.

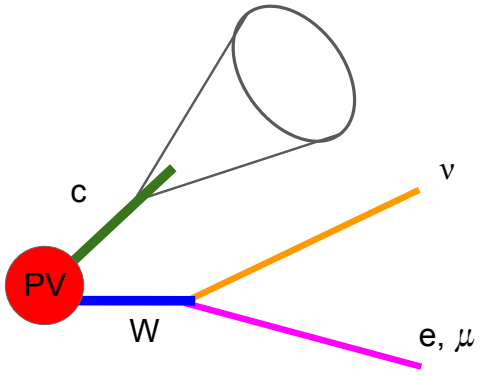


Figure 5.8: Generic W+c signature signal. The W boson decays at the PV while the decay length of the c quark, inside a c hadron, follows a exponential function with a tail extending up to some millimetres from the PV.

Electrons or muons must be in the region $|\eta^{\ell}| < 2.4$ and $p_{\text{T}}^{\ell} > 35\text{GeV}$. The combined isolation I_{comb} is used to quantify the additional hadronic activity around the selected leptons. The lepton candidate is considered to be isolated if $I_{comb}/p_{\text{T}}^{\ell} < 0.15$.

Events with an additional isolated lepton with $p_{\text{T}}^{\ell} > 20\text{GeV}$, and opposite electric charge, are rejected to suppress the contribution from Z+jets and $t\bar{t}$ events.

The transverse mass (M_{T}) of the lepton and $\vec{p}_{\text{T}}^{\,\text{miss}}$ is defined as,

$$\text{M}_{\text{T}} \equiv \sqrt{2\; p_{\text{T}}^{\ell}\; p_{\text{T}}^{\text{miss}}\; [1 - \cos(\phi_{\ell} - \phi_{p_{\text{T}}^{\text{miss}}})]}, \tag{5.3.1}$$

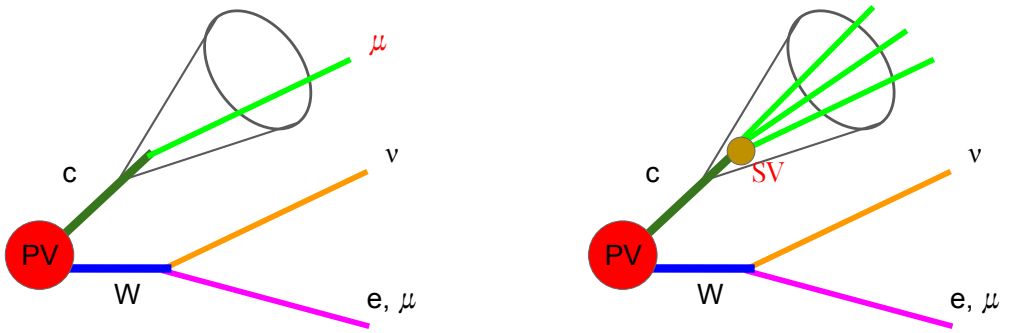
where ϕ_ℓ and $\phi_{p_{\mathrm{T}}^{\mathrm{miss}}}$ are the azimuthal angles of the lepton and the $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ vector. Events with $\mathrm{M_T} < 55\,\mathrm{GeV}$ are discarded from the analysis to reduce the contamination from QCD multijet events [138].

The electric charge of the W boson will be that of the isolated lepton selected with the above requirements. For muons with $p_{\mathrm{T}} > 10\,\mathrm{GeV}$ the charge assigned is correct in >99.9% of the cases. For electrons the charge assigned is correct in >99% of the cases and the difference between data and MC is at the per mil level. Further explanation can be found in Refs. [123, 152–154].

5.3.2 Identification of charm jets

Hadrons with c quark content decay through weak interaction with lifetimes of the order of $10^{-12}\,\mathrm{s}$ and mean decay lengths larger than $100\,\mu\mathrm{m}$ at TeV energies [17]. Secondary vertices well separated from the PV can be reconstructed from the charged particle tracks. In a sizeable fraction of the heavy flavour hadron decays (≈10–15%) there is a muon in the final state. We make use of these properties to define two independent data samples, identifying jets originating from a c quark in two different ways:

- The Semileptonic (SL) channel, where a muon coming from the semileptonic decay of a c hadron is identified inside of a jet (Fig. 5.9a).
- The Secondary Vertex (SV) channel, where a displaced SV is reconstructed inside of the jet (Fig. 5.9b).



(a) SL channel. (b) SV channel.

Figure 5.9: Graphic representation of the SL (a) and SV (b) charm identification channels.

If an event fulfils the selection requirements of both topologies, it is assigned to the SL channel. Thus, the SL and the SV channels are mutually exclusive, i.e., the samples selected in each channel are statistically independent.

We require the presence of at least one jet with $p_{\mathrm{T}}^{\mathrm{jet}} > 30\mathrm{GeV}$ and $|\eta|^{\mathrm{jet}} < 2.4$. Jets with an angular separation between the jet axis and the selected isolated lepton $\Delta R(\mathrm{jet}, \ell) < 0.4$ are not considered.

5.3.2.1 Charm jet tagging in the SL channel

The W+charm SL sample is selected using semileptonic decays of charm hadrons into a muon plus anything, identified by a reconstructed muon among the constituents of any of the selected jets. Semileptonic decays of c-quarks into electrons are not considered because of the complexity of identifying electrons inside jets.

Muons from semileptonic decays of charm hadrons are usually very close to the jet axis and non-isolated from hadron activity. Therefore, the following selection requirements enrich the SL sample:

- Muon candidates must have been identified with the Tight identification working point, following the same quality criteria as with the muons from the W decay.

- The transverse momentum of the muon candidate must be $p_{\mathrm{T}}^{\mu} < 25\,\mathrm{GeV}$ and the absolute value of the pseudorapidity must be $|\eta^{\mu}| < 2.4$. This cut is introduced in order to remove background with prompt muons overlapped with jets. Thus, we further reduce the Z+jets contamination. No minimum p_{T} threshold is explicitly required, but the muon reconstruction algorithm sets a natural threshold around 3 GeV (2 GeV) in the barrel (endcap) region since the muon must traverse the material in front of the muon detector and penetrate deep enough into the muon system to be reconstructed and satisfy the identification criteria.

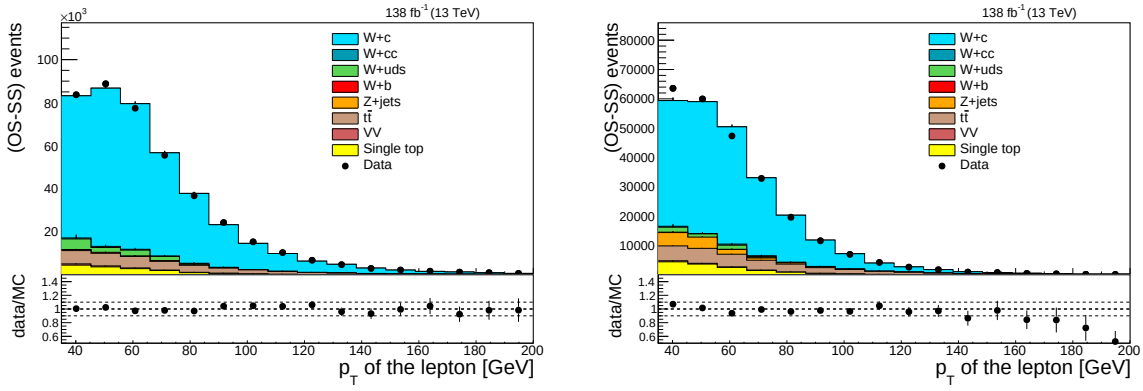
- Muon candidates must be non-isolated from hadron activity, i.e., the combined isolation variable built from the sum of transverse energies or momenta of reconstructed particles, $I_{comb}/p_{\mathrm{T}}^{\mu}$, in a cone of $\Delta R < 0.4$ must be greater than 0.2.

- The track of the muon coming from a semileptonic decay of a charm hadron tends to have a considerable transverse impact parameter with respect to the PV. We require the transverse impact parameter significance (IPS) of the muon in the jet, defined as the muon transverse impact parameter divided by its uncertainty, to be larger than 2.

- In addition, muon pair candidates (one isolated and one inside of a jet) with an invariant mass around the Z boson mass peak $70 < m_{\mu\mu} < 110\,\mathrm{GeV}$ are rejected in order to reduce $\mathrm{Z} \to \mu^{+}\mu^{-}$ contamination. Moreover, any muon pair with an invariant mass $m_{\mu\mu} < 12\,\mathrm{GeV}$ is rejected in order to reduce contamination coming from low-mass resonances $\mathrm{J}/\Psi, \Psi'$ or Υ mesons.

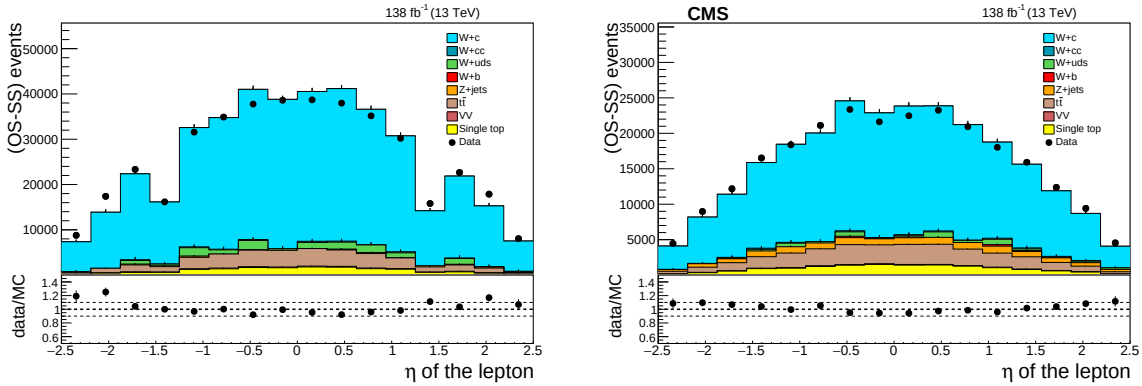
- If more than one such muon is identified, the one with the highest p_{T} is selected.

Figures 5.10 and 5.11 illustrate the difference between the background populations for the $\mathrm{W} \to e\nu$ and $\mathrm{W} \to \mu\nu$ decay channels after OS-SS substraction. It is straightforward to recognize the abundance of Z+jets (in orange) in the muon channel and its absence in the electron channel.



(a) W $\to e\nu$ channel. (b) W $\to \mu\nu$ channel.

Figure 5.10: Distribution of the p_{T} of the lepton coming from the decay of the W boson in the SL channel.



(a) W $\to e\nu$ channel. (b) W $\to \mu\nu$ channel.

Figure 5.11: Distribution of the η of the lepton coming from the decay of the W boson in the SL channel.

5.3.2.2 Charm jet tagging in the SV channel

An independent W+charm sample is selected using hadronic decays of charm hadrons by looking for secondary decay vertices within the reconstructed jets, described in Section 4.3.2. Tracks used for the reconstruction of SVs must have $p_{\mathrm{T}} > 1\,\mathrm{GeV}$ to avoid mis-reconstructed or poorly reconstructed tracks. Vertices reconstructed with the IVF algorithm are considered first. If no IVF vertex is found, then we look for SSV vertices, thus providing additional (∼3%) event candidates. If more than one SV within a jet are reconstructed, the one with the highest p_{T}, computed from its associated tracks, is considered. If there are several jets with a SV, only the SV associated to the jet of highest p_{T} is selected.

In order to suppress the background contributions (as explained below), only SVs with at least three tracks and well separated from the PV are considered. The displacement significance of the SV, defined as the three-dimensional distance between the primary and the secondary vertices divided by its uncertainty is required to be larger than 8.

5.3.2.3 Charm jet tagging efficiency

The efficiency of the SL and SV charm tagging methods has been evaluated with the W+jets simulation for different generated jet flavours. The efficiency values are shown in Table 5.4 before OS-SS substraction.

Table 5.4: Charm tagging efficiency for each jet flavour, after all selection requirements but before OS-SS substraction.

Jet flavour	c	uds	g	b
Efficiency SL(%)	1.57 ± 0.03	0.29 ± 0.04	0.07 ± 0.005	4.4 ± 0.3
Efficiency SV(%)	4.5 ± 0.1	1.0 ± 0.1	0.15 ± 0.01	24 ± 1

The tagging efficiency of charm jets is small, 1.6% for the SL method and 4.5% for the SV method. Only about 10% of the charm hadrons decay semileptonically to a muon, and a small fraction of them have enough momentum to reach the muon detector and get reconstructed. Similarly, only a small fraction of the charm hadrons produce an SV with at least three tracks that is well separated from the PV.

The suppression of jets initiated by light quarks and gluons is very large. Less than 1% of the light-quark jets are mistagged as c jets, while the mistag rate for gluon jets is about 0.1%.

The SL and SV tagging methods also tag jets coming from the hadronization of b quarks as c jets. The b jets have similar characteristics as the c jets, containing muons and SVs from the decay of the b hadrons. The contribution of b jets comes from the processes $\mathrm{W} + \mathrm{b\bar{b}}$ and $\mathrm{t\bar{t}}$. In both cases, it is equally likely to identify the b jet with the same or opposite charge than that of the W boson, and therefore these contributions are very efficiently removed by the OS-SS subtraction.

5.3.3 Determination of the charm quark electric charge

In order to apply the OS-SS technique to reduce the backgrounds of our analysis, we have to correctly identify the sign (positive or negative) of the electric charges of the W boson and the charm quark. As mentioned earlier, the charge of the W boson is carried away by the isolated high-p_{T} electron or muon. The charge of the charm quark will be determined differently for the SL and SV channels.

For the SL channel, the sign of the electric charge of the charm quark is that of the low-p_{T} non-isolated muon found inside of the jet, that comes from the semileptonic decay of a charmed hadron.

The charge determination of the charm quark in the SV channel is not as straight forward as the previous case. For charged charm hadrons, the sign of the sum of the charges of the decay products reflects that of the c quark. For neutral charm hadrons, the sign of the charge of the closest hadron produced in the fragmentation process can indicate that of the c quark [155, 156]. Hence, we assign a charge equal to the sign of the sum of the charges of the particle tracks associated with the SV. If the SV charge is zero, we take the charge of the PV track

that is closest in angular separation to the SV. We only consider PV tracks with $p_{\mathrm{T}} > 0.3\,\mathrm{GeV}$ and within an angular separation with the SV direction of 0.1 in the (η, ϕ) space. If nonzero charge can be assigned, the event is accepted.

According to the simulation, the charge assignment procedure provides a nonzero charge for 99% of the selected SVs. In 13% of the SVs the c quark charge is assigned from the PV track, and the charge assignment is correct in 65% of the cases. In the remaining 87% of events, the c quark charge is given by the SV charge with a 86% true charge assignment rate. Overall, the correct charge assignment efficiency is therefore 83%.

The modeling of the SV charge assignment in the simulation has been validated with data. Events passing both the SL and SV selection criteria are used to compare the charges of the muon inside the jet and the SV. In 95% of these events the charges agree. The difference in the charge assignment efficiency between data and simulation, around 1%, is taken as a systematic uncertainty in the cross section measurements, as detailed in Section 5.7.

The event selection requirements are summarized in Table 5.5 for the four selection channels of the analysis, the Wboson decay channels to electron or muon, and the SL and SV charm identification channels.

Table 5.5: Summary of the selection requirements for the four selection channels of the analysis.

	SL $\mathrm{W} \to e\nu$	SL $\mathrm{W} \to \mu\nu$	SV $\mathrm{W} \to e\nu$	SV $\mathrm{W} \to \mu\nu$
Lepton p_{T}^{ℓ}	$> 35\,\mathrm{GeV}$			
Lepton $\lvert\eta^{\ell}\rvert$	< 2.4			
Lepton isolation $I_{comb}/p_{\mathrm{T}}^{\ell}$	< 0.15			
Transverse mass $\mathrm{M_T}$	$> 55\,\mathrm{GeV}$			
Jet $p_{\mathrm{T}}^{\mathrm{jet}}$	$> 30\,\mathrm{GeV}$			
Jet $\lvert\eta^{jet}\rvert$	< 2.4			
$\Delta R(\mathrm{jet}, \ell)$	> 0.4			
Muon in jet p_{T}^{μ}	$< 25\,\mathrm{GeV}$			
Muon in jet $p_{\mathrm{T}}^{\mu}/p_{\mathrm{T}}^{\mathrm{jet}}$	< 0.6			
Muon in jet $\lvert\eta^{\mu}\rvert$	< 2.4			
Muon in jet $I_{comb}/p_{\mathrm{T}}^{\mu}$	> 0.2			
Muon in jet IPS		> 2		
Muon in jet $m_{\mu\mu}$		$> 12\,\mathrm{GeV}$ & $\notin [70, 110\,\mathrm{GeV}]$		
SV number of tracks			> 2	
SV displacement signific.			> 8	
SV charge			$\neq 0$	

5.4. Background determination and suppression

In this section we show how we increase the signal proportion from about 30% to 80% by further suppressing the backgrounds after the selection cuts discussed in the previous section. The OS-SS subtraction is the main cause of this reduction. In addition, we show below how we further suppress the W+udsg and Z+jets backgrounds using properties of the SVs and muons inside the jets. We also show how we set the absolute normalization of the backgrounds using control regions where we can directly compare the detector data with the simulations.

The SL and SV signatures are also featured by weakly decaying b hadrons. Events from processes involving the associated production of W bosons and b quarks will be abundantly selected in the two categories. As stated previously, a key property of the qg $\rightarrow$ W + c process is the presence in the final state of a charm quark and a W boson with opposite sign electric charges (OS). By contrast, most of the background processes deliver evenly OS and same sign (SS) events.

- The dominant background contribution stem from $t\bar{t}$ production, where a pair of W bosons and two b jets are produced in the decay of the top quark-antiquark pair. This final state mimics the analysis topology when at least one of the W bosons decays leptonically, and there is an identified muon or a reconstructed SV inside one of the b jets. However, this background is effectively suppressed by the OS-SS subtraction. The chance to identify a muon or a SV inside the b jet with opposite or same charge than the charge of the W candidate is expected to be the same, thus delivering an equal amount of OS and SS events (see Fig. 5.2a).

- Top quark-antiquark events where one of the W bosons decays hadronically into a $c\bar{s}$ (or $\bar{c}s$) pair may result in additional event candidates if the SL or SV signature originates from the c jet. This topology produces genuine OS events, which contribute to the remaining background contamination after OS-SS subtraction.

- Similarly, single top quark production also produces OS events, but at a lower level because of the smaller production cross section (see Fig. 5.2b).

- The production of a W boson and a single b quark through the process $qg \rightarrow W + b$ produces OS events, but it is heavily Cabibbo-suppressed and its contribution is found to be negligible.

- The other source of a W boson and a b quark is $W + b\bar{b}$ events where the $b\bar{b}$ pair originates from a gluon splitting mechanism and only one of the two b jets is identified. These events are also charge symmetric as it is equally likely to identify the b jet with the same or opposite charge than that of the W boson. Its contribution cancels out after the OS-SS subtraction.

Figure 5.12 shows the procedure explained above, illustrated with the distribution of the ϕ of the c-tagged jet. The first two plots present the event yields after the selection requirements for the SL and SV charm tagging channels without applying the OS-SS technique, where we can observe the large backgrounds described

previously. The second row shows a selection of only OS event yields, with a larger population of signal events. Following that, we present a selection of only SS event yields, which shows that for a fraction of the signal W+c events the charm hadron is not correctly identified resulting in a SS configuration of the event. Finally, the substraction technique is applied and we obtain a very pure sample of W+c event yields. The reason for the appearance of SS events for the signal is different for the SL and SV channels. In the SL channel, the muon electric charge identification is very accurate, so the main source of contamination may be the presence of muons from in flight decays of pions and kaons not eliminated by the reconstruction algorithm. In the case of the SV channel, the main source of contamination may be the incorrect assignment of the electric charge of the SV and the contribution of fake vertices.

5.4.1 Suppression of the W+usdg background

The contribution of the W+udsg background after OS-SS subtraction is still substantial in the SV sample, mostly from the processes $ug \to W^+ + d$ and $dg \to W^- + u$, which produce OS events. Looking at the distribution of the number of tracks forming the SVs in Fig. 5.13, we observe that most of the W+udsg events correspond to two-track SVs. We therefore select only SVs with at least three tracks to largely suppress this contribution. This also reduces the systematic uncertainty associated to the SV reconstruction efficiency, since vertices reconstructed using two tracks may be result of the casual encounter of uncorrelated tracks.

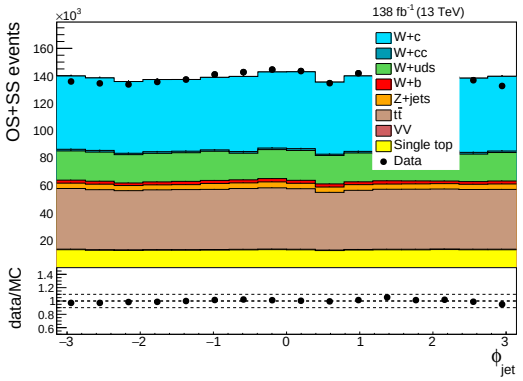
We are aware that by doing so, we are also losing a large sample of signal events. According to the simulation, the selection of SVs with at least three tracks reduces the signal sample in ≈38%. However, we are not discouraged because the measurements will not be dominated by the statistical uncertainty, so we can afford this reduction of the signal for the benefit of better background control and the reduction of its associated systematic uncertainty.

In order to further reduce the contribution from light hadron decays, we make sure that the SV is well separated from the PV and avoid light hadrons of short lifetime, we require the displacement significance, defined as the three dimensional distance between the primary and the secondary vertices divided by its uncertainty, to be larger than 8. We show the distribution of this variable in Fig. 5.14 before this requirement.

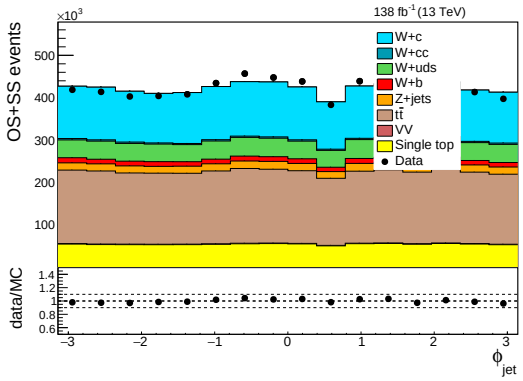
5.4.2 Suppression of the Z+jets background

The decay of a W boson into a muon and a neutrino plus a c jet containing a muon from the semileptonic decay of the charmed hadron can be mimicked by the Z+jets production process where the Z boson decays into two opposite-sign muons, such that one is reconstructed isolated and the other one is found inside a jet (see Figs. 5.3a and 5.9a). In the $W \to \mu\nu$ SL channel, this background is sizeable and must be reduced.

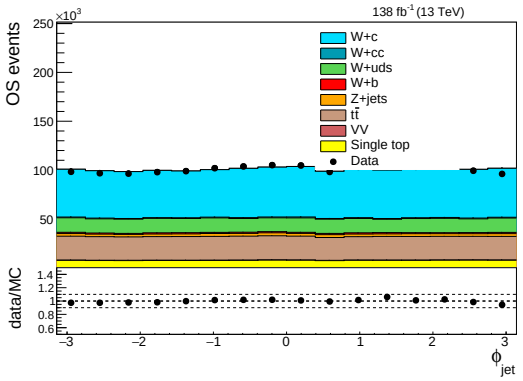
The track of the muon coming from a semileptonic decay of a charm hadron



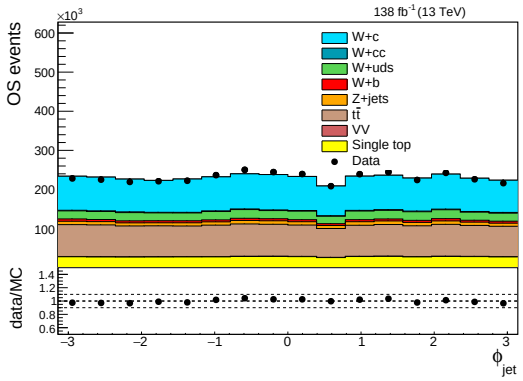
(a) OS+SS events in SL channel.



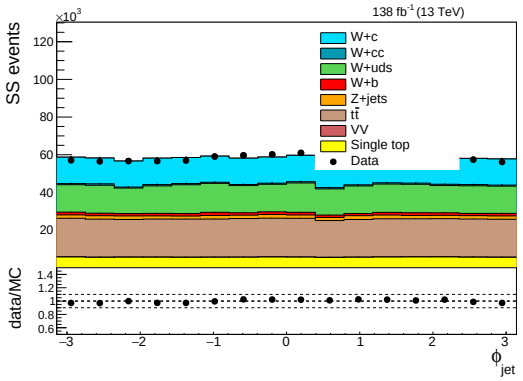
(b) OS+SS events in SV channel.



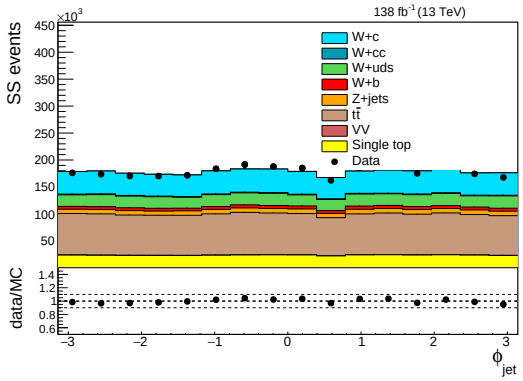
(c) OS events in SL channel.



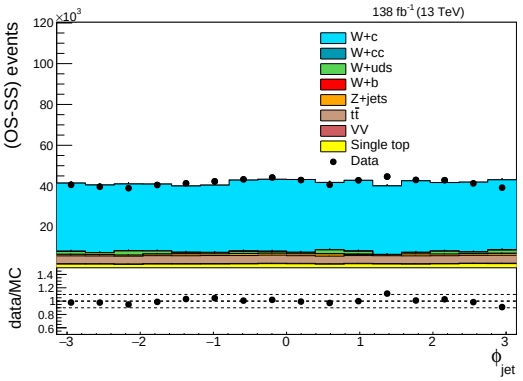
(d) OS events in SV channel.



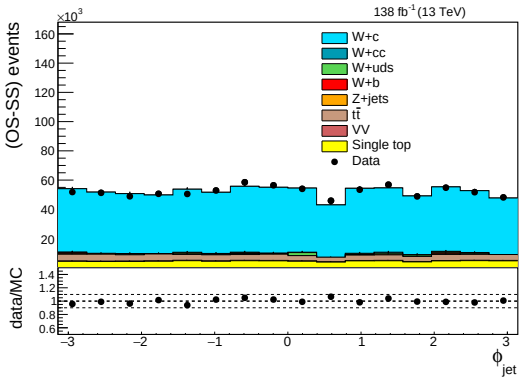
(e) SS events in SL channel.



(f) SS events in SV channel.



(g) OS-SS in SL channel.



(h) OS-SS in SV channel.

Figure 5.12: Distribution of the ϕ of the c-tagged jet before and after OS-SS subtraction.

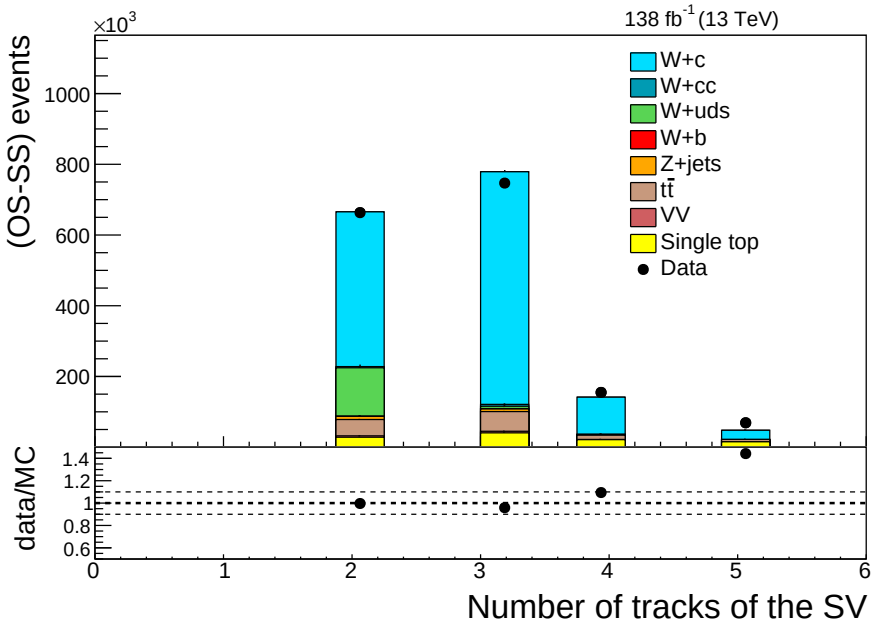


Figure 5.13: Distribution of the number of tracks of the SV.

tends to have a larger transverse impact parameter than that of a prompt muon coming from the decay of a Z boson. We therefore require the transverse impact parameter significance of the muon to be greater than 2, defined as the 2D impact parameter in the transverse plane to the beam axis of the track considered to be the muon divided by its uncertainty, in order to reduce the Z+jets contamination. This is only required in the W $\rightarrow \mu\nu$ channel, because the W $\rightarrow e\nu$ decay would produce a pair e-μ, which cannot be originated from a Z boson. A comparison between data and simulation of this variable can be seen in Fig. 5.15.

5.4.3 Evaluation and normalization of the dominant backgrounds

In order to extract the signal content of the data, the contributions of the remaining backgrounds after all selection requirements and OS-SS subtraction must be properly normalized. For this purpose, we use control samples, where we can directly compare data and simulations, to avoid relying in the normalization provided by the simulations. This is the case for the $t\bar{t}$ and Z+jets backgrounds.

The $t\bar{t}$ data control sample is established selecting events with at least three high-p_{T} jets, two of which are tagged as b jets using the loose working point of the DeepCSV b-tagging algorithm, and the third jet contains a muon. A normalization factor of 0.92 ± 0.02 is required to bring into agreement data and $t\bar{t}$ simulation, as seen in Fig. 5.16. The uncertainty in the background normalization factors reflects the statistical uncertainty of the data and the simulations in the control sample.

A Z+jets data control sample is defined using the same selection criteria as in the analysis in the SL channel but inverting the dimuon invariant mass requirement

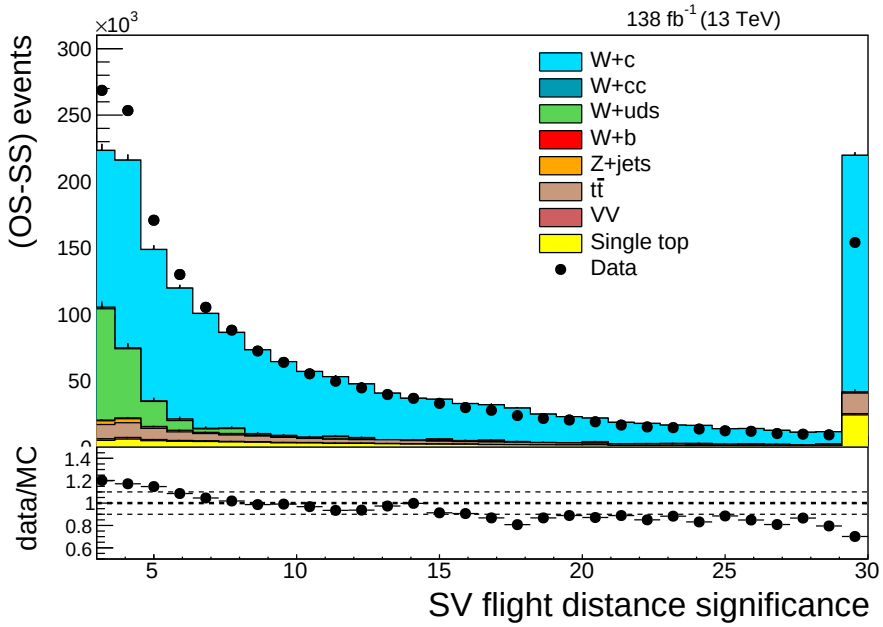


Figure 5.14: Distribution of the flight distance significance of the SV before applying the selection requirement. The last bin is cumulative for the overflow.

to select events close to the Z boson mass peak ($70 < m_{\mu\mu} < 110\,\text{GeV}$). A normalization factor of 1.08 ± 0.01 is required to match the Z+jets simulation with data, as seen in Fig. 5.17.

Regarding the single top background, we use the distribution of the corrected secondary vertex mass to normalize its contribution, together with the already normalized $t\bar{t}$ sample (see Fig. 5.22). We define the corrected SV mass, $M_{\text{SV}}^{\text{corr}}$, as the invariant mass of all charged particles associated with the SV, assumed to be pions, M_{SV}, corrected for additional particles, either charged or neutral, that may have been produced but were not reconstructed [157]:

$$M_{\text{SV}}^{\text{corr}} = \sqrt{M_{\text{SV}}^2 + p_{\text{SV}}^2 \sin^2\theta} + p_{\text{SV}} \sin\theta, \tag{5.4.1}$$

where p_{SV} is the modulus of the vectorial sum of the momenta of all charged particles associated with the SV, and θ is the angle between the momentum vector sum and the vector from the PV to the SV. The corrected SV mass is thus the minimum mass the long-lived hadron can have that is consistent with the direction of its momentum. Single top quark events populate the tail of the $M_{\text{SV}}^{\text{corr}}$ distribution. A normalization factor of 1.5 ± 0.2 for the single top quark contribution was found to be required to match data and simulation predictions.

Finally, after the normalizations of the $t\bar{t}$, Z+jets and single top backgrounds have been fixed with the data control samples, the contribution of the W+jets simulation is scaled so that the sum of the events from all predicted contributions is equal to the number of events in the selected data sample. The normalization factor of the W +jets simulation (about 5%) has only a minor effect in the contribution of the (small) predicted W+udsg background. The resulting overall normalization

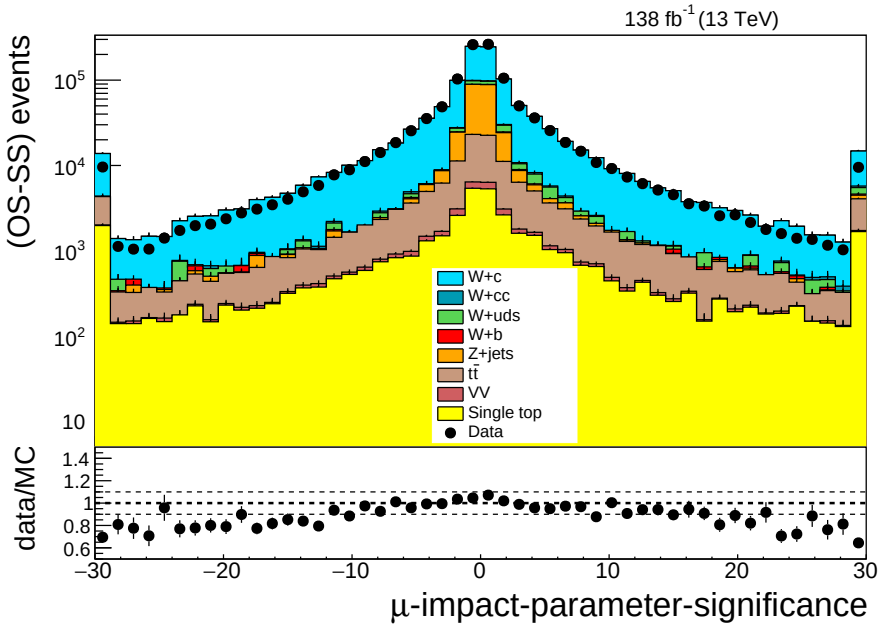


Figure 5.15: Distribution of the impact parameter significance of the muon inside the c-tgged jet before the selection requirement. The last bin is cumulative for the overflow.

of the W+c signal simulation is irrelevant for the analysis, as it is only used for acceptance and efficiency calculations and the normalization term cancels out in those ratios.

5.5. Signal extraction

This section presents the yields and flavour composition after event selection, including OS-SS subtraction, for each W lepton decay channels in each of the charm decay channels. Table 5.6 shows the event and background yields in both W decay channels for the SL channel, after the selection requirements described above and after OS-SS subtraction. The signal and background composition of the selected sample according to simulation are shown in Table 5.7.

Table 5.6: Data and background event yields, with statistical uncertainties, after selection and OS-SS subtraction for the SL channels.

SL channel	Data	Background
$W \to e\nu$	424 047 ± 1286	80 646 ± 933
$W \to \mu\nu$	263 669 ± 918	68 108 ± 917

After OS-SS substraction, the fraction of signal W + c events in the $W \to e\nu$ channel is above 80%, while in the $W \to \mu\nu$ channel it drops to 74% because of the additional Z+jets background (around 6%). In contrast, the purity would be 31(38)% without the OS-SS technique, showing how powerful it is. The dominant

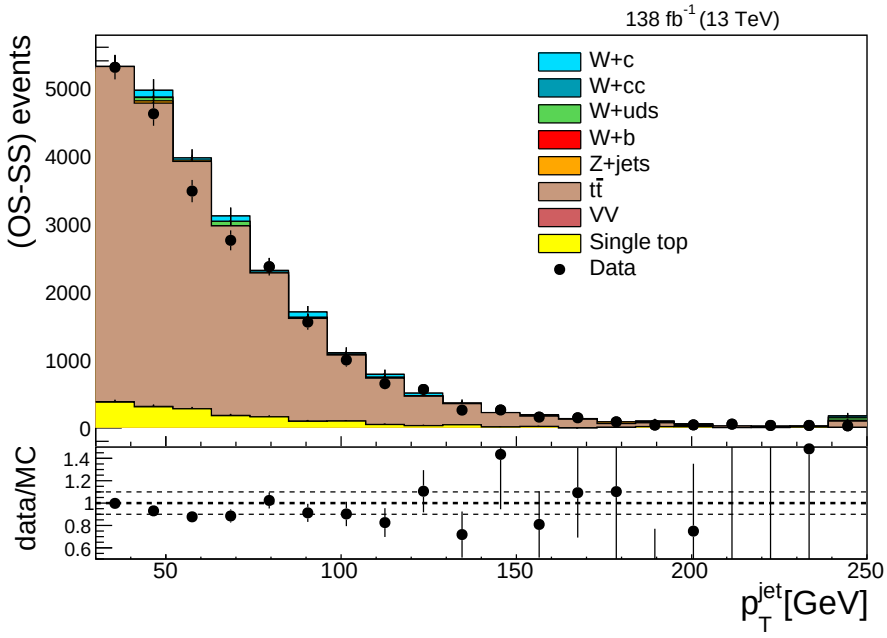


Figure 5.16: Distribution of the p_T of the jet for the $t\bar{t}$ control sample in the $W \to e\nu$ and SL channel.

background, $t\bar{t}$ production, where one of the W bosons decays leptonically and the other hadronically with a charm quark in the final state, amounts to approximately 10%.

Table 5.7: Simulated signal and background composition (in percentage) of the SL sample after event selection, before and after OS-SS subtraction. The $W + Q\bar{Q}$ stands for the sum of the contributions of $W + c\bar{c}$ and $W + b\bar{b}$. The uncertainties shown as 0.1% mean <0.1%.

SL		W + c	W + Q$\bar{Q}$	W + udsg	Z+jets	t$\bar{t}$	single t	VV
W $\to e\nu$	OS+SS	30.9 ± 0.1	2.5 ± 0.1	22.3 ± 0.2	3.4 ± 0.1	31.2 ± 0.1	9.3 ± 0.1	0.4 ± 0.1
	OS−SS	81.0 ± 0.6	0.5 ± 0.3	3.1 ± 0.5	0.4 ± 0.1	10.0 ± 0.1	4.4 ± 0.1	0.6 ± 0.1
W $\to \mu\nu$	OS+SS	37.8 ± 0.1	2.0 ± 0.1	9.5 ± 0.1	4.4 ± 0.1	35.0 ± 0.1	10.9 ± 0.1	0.3 ± 0.1
	OS−SS	74.2 ± 0.5	0.5 ± 0.3	2.0 ± 0.4	5.5 ± 0.2	11.6 ± 0.1	5.8 ± 0.1	0.4 ± 0.1

Table 5.8 shows the event and background estimation yields in both W decay channels for the SV channel, after the selection requirements and OS-SS subtraction. The signal and background composition of the selected sample, as predicted by the simulation, are shown in Table 5.9. The purity of signal W + c events is above 80% after OS-SS substraction, as opposed to less than 30% without that technique. The dominant backgrounds come from $t\bar{t}$ (8%) and single top (9%) production.

5.6. Data modeling

Once the selection criteria have been fixed and the remaining background contributions have been carefully evaluated, we present in this section the comparison of data and simulation of relevant observables to show that we have an appropriate

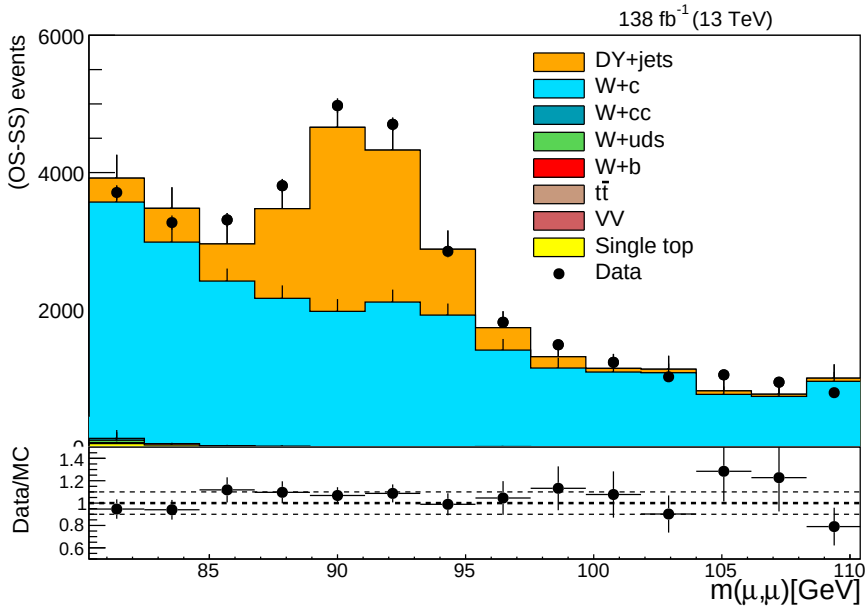


Figure 5.17: Distribution of the invariant mass of the muon pair candidates $m_{\mu\mu}$ for the $W \to \mu\nu$ and SL channel, with the same selection as in the analysis but this reverting the $m_{\mu\mu}$ mass requirement.

Table 5.8: Data and background event yields, with their statistical uncertainties, after selection and OS-SS subtraction for the SV channels.

SV channel	Data	Background
$W \to e\nu$	338 504 ± 1717	60 565 ± 1577
$W \to \mu\nu$	494 264 ± 1876	94 356 ± 2002

modeling of the data. Distributions for the charm decay channels are presented separately, with the cumulative contribution of both W boson decay channels.

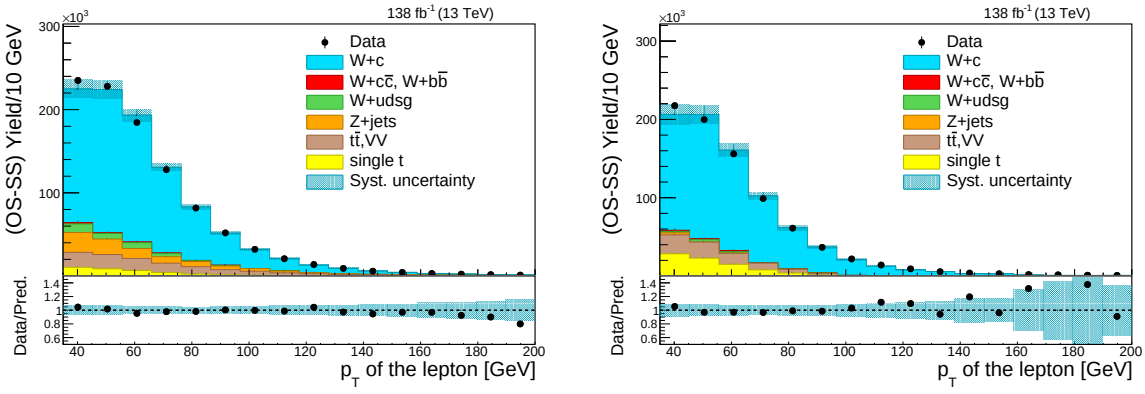
Figures 5.18 and 5.19 show the distribution of the p_T and the η of the lepton coming from the decay of the W boson, respectively. Figures 5.20 and 5.21 show the distribution of the p_T and the η of the c-tagged jet, respectively. Figure 5.22 shows the distribution of the mass of the SV inside the c-tagged jet. Figure 5.23 shows the distribution of the p_T of the muon inside the c-tagged jet. Finally, Figs. 5.24 and 5.25 show the distribution of the M_T of the reconstructed W boson, and the missing transverse of the event, respectively. Every distribution includes a shaded band that represents the systematic uncertainties, that will be discussed in the next section.

5.7. Systematic uncertainties

The impact of various sources of uncertainty in the measurements presented in Chapter 6 is estimated by recalculating the cross sections with the relevant pa-

Table 5.9: Simulated signal and background composition (in percentage) of the SV sample after event selection, before and after OS-SS subtraction. The $W + Q\bar{Q}$ stands for the sum of the contributions of $W + c\bar{c}$ and $W + b\bar{b}$. The uncertainties shown as 0.1% mean <0.1%.

SV		W + c	W + Q$\bar{Q}$	W + udsg	Z+jets	t$\bar{t}$	single t	VV
W $\to e\nu$	OS+SS	21.6 ± 0.1	3.3 ± 0.1	11.7 ± 0.1	5.4 ± 0.1	43.0 ± 0.1	14.6 ± 0.1	0.3 ± 0.1
	OS−SS	82.1 ± 0.8	0.7 ± 0.4	1.0 ± 0.6	0.1 ± 0.2	7.2 ± 0.1	8.4 ± 0.1	0.5 ± 0.1
W $\to \mu\nu$	OS+SS	29.9 ±0.1	3.0 ± 0.1	8.3 ± 0.1	3.9 ± 0.1	40.6 ± 0.1	0.4 ± 0.1	14.0 ± 0.1
	OS−SS	80.9 ± 0.6	0.7 ± 0.3	0.5 ± 0.4	0.5 ± 0.2	8.0 ± 0.1	0.5 ± 0.1	8.9 ± 0.1



(a) SL channel.

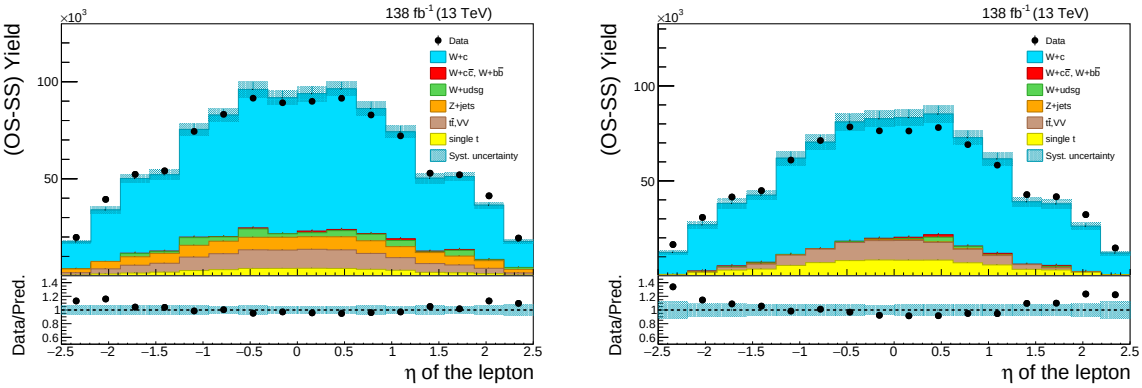
(b) SV channel.

Figure 5.18: Distribution of the p_T of the lepton coming from the decay of the W boson.

rameters varied up and down by one standard deviation of their uncertainties.

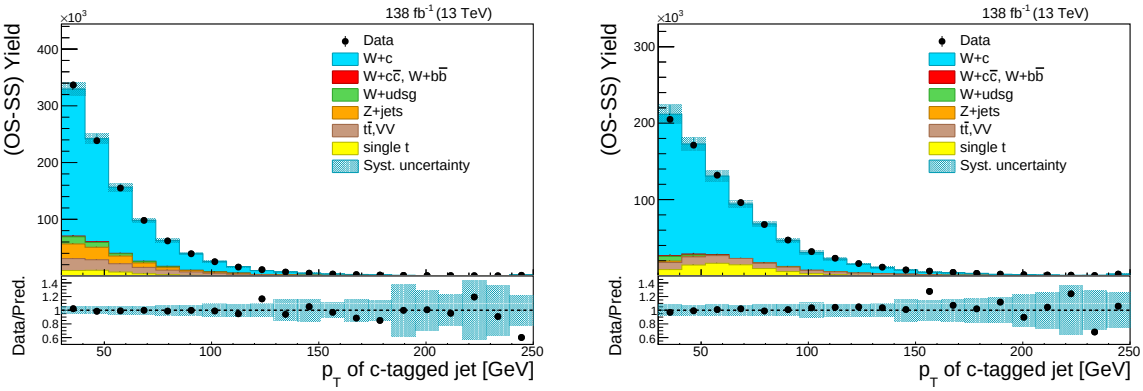
The combined uncertainty in the trigger, reconstruction, and identification efficiencies for isolated leptons results in an uncertainty in the cross section measurements of about 2 (1)% for the W $\to e\nu$ (W $\to \mu\nu$) channel. The uncertainty in the identification efficiency of non-isolated muons inside jets is approximately 3%, according to dedicated studies with Z+jets events. We required an isolated muon and a non-isolated muon, and then reweighted the p_T distribution of the Z study to the equivalent one of the W $\to \mu\nu$ SL channel of our analysis. We observed a difference between data and MC of the mentioned 3%. This uncertainty only affects the SL channel.

The effects of the uncertainty in the JES and JER are assessed by varying up and down the p_T values of jets with the corresponding uncertainty factors. The JES and JER uncertainties are also propagated to $\vec{p}_T^{\,miss}$. Measurements show that the JER in data is worse than in the simulation and the jets in MC need to be smeared to describe the data. With the scaling method, the corrected four-momentum of a reconstructed jet is rescaled with a factor calculated using its transverse momentum, the transverse momentum of the corresponding jet clustered from generator-level particles, and a data-to-simulation core resolution scale factor. The resulting uncertainty in the cross section is about 2 (1)% for the SL (SV) channel. The uncertainty from a $\vec{p}_T^{\,miss}$ mismeasurement in the event is estimated by varying within its uncer-



(a) SL channel. (b) SV channel.

Figure 5.19: Distribution of the η of the lepton coming from the decay of the W boson.



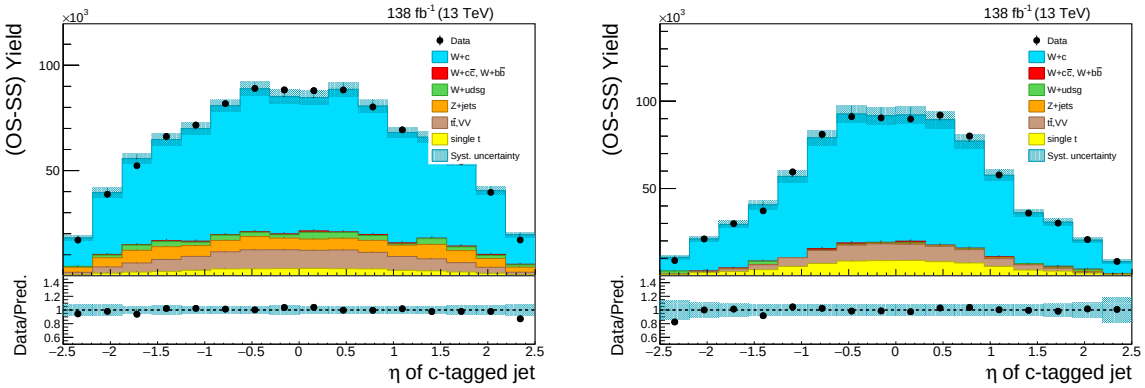
(a) SL channel. (b) SV channel.

Figure 5.20: Distribution of the p_{T} of the c-tagged jet.

tainty the contribution of the energy unassociated with reconstructed particle-flow objects. The effect in the cross section measurement is <0.5%. Uncertainties in the pileup modeling are calculated using a modified pileup profile obtained by changing the mean number of interactions by ≈5%. This variation covers the uncertainty in the pp inelastic cross section [158] and in the modeling of the pileup simulation. It results in less than 0.5% uncertainty in the cross section measurements.

The integrated luminosities of the 2016, 2017, and 2018 data-taking periods are individually known with uncertainties in the 1.2–2.5% range [141–143], while the total 2016–2018 integrated luminosity has an uncertainty of 1.6%. The improvement in precision arises from the (uncorrelated) time evolution of some systematic effects.

The uncertainty in the scale factor correcting the SV reconstruction efficiency in the simulation propagates into a systematic uncertainty of 3% in the cross section. The uncertainty in the SV charge determination is estimated as the difference (1%) in the rate obtained in data and simulation of correct SV charge assignment in the validation test described in Section 5.3.3.



(a) SL channel. (b) SV channel.

Figure 5.21: Distribution of the η of the c-tagged jet.

Given the dependence of the SV reconstruction efficiency on the SV displacement, we have evaluated the effect produced by an imperfect modeling of this observable by reweighting the SV displacement significance distribution of the simulation to match that of the data. The resulting uncertainty in the cross section measurement is 1–2%. In addition, the stability of the results with the minimum SV displacement significance requirement has been checked by changing the threshold from 8 to 7. The effect in the results is also at the 1% level.

The background contributions are evaluated with the simulations validated in data control samples, as discussed in Section 5.4.3. The uncertainty in the predicted background levels has an effect of 1% in the cross section measurements.

The signal samples used for the acceptance and efficiency calculations were generated with MadGraph + pythia8 using the NNPDF 3.0 and NNPDF 3.1 PDF sets. The envelope of the systematical variations (replicas) of the nominal PDF is taken as the systematic uncertainty due to an imperfect knowledge of the PDFs, as recommended in Ref. [159]. The effect is approximately 1%. The statistical uncertainty in the determination of the selection efficiency using the simulated samples is 1%, and is propagated as an additional systematic uncertainty.

To estimate the effect originated by the uncertainties in the corrected values used in the simulation for the charm fragmentation and decay branching fractions [17, 148], we have varied those values within their uncertainties. The impact in the cross section measurements is 1–2%, both for the fragmentation fractions and decay branching ratios.

The main systematic uncertainties are summarized in Table 5.10 for the four selection channels of the analysis. Overall, the total systematic uncertainty in the W + c fiducial cross section is approximately 5% in all channels.

Most of the reconstruction and selection efficiencies cancel out in the measurement of the cross section ratio $R_c^{\pm}$. Possible efficiency differences between positive and negative leptons and SVs are taken into account as systematic uncertainties. We evaluate effects stemming from charge confusion and charge-dependent recon-

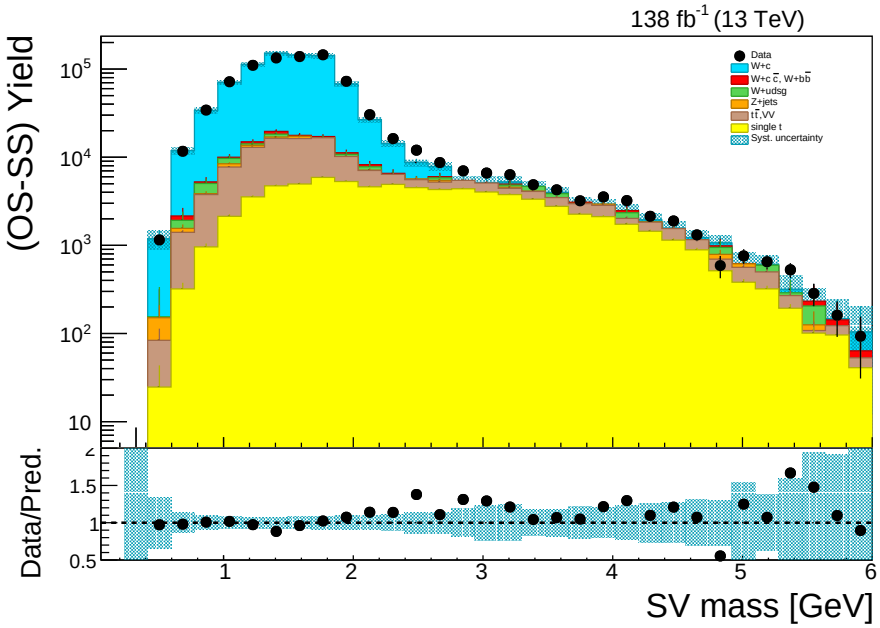


Figure 5.22: Distribution of the mass of the secondary vertex inside the c-tagged jet.

struction efficiencies.

The probability of mistakenly assigning the incorrect charge to a lepton is studied with data using $Z \to \ell^+\ell^-$ events reconstructed with SS or OS leptons. For the muons, the charge misidentification probability is found to be negligible ($< 10^{-3}$). For the electrons, the effect is around 1% but propagates into a negligible uncertainty in the cross section ratio. The charge confusion rate for the SVs is significantly larger, 17%, as described in Section 5.3.3. However, assuming that the charge confusion probability is the same for positive and negative SVs, the effect in the cross section ratio cancels out.

Potential differences in the reconstruction efficiencies of positive and negative leptons or SVs are studied with the W + c MC simulation. Efficiency ratios are calculated independently for the four channels of the analysis and are found to be consistent with unity within the statistical uncertainty (1.2–1.4%). No corrections are made in the $R_c^{\pm}$ measurements but the statistical uncertainties in the efficiency ratios are treated as systematic uncertainties.

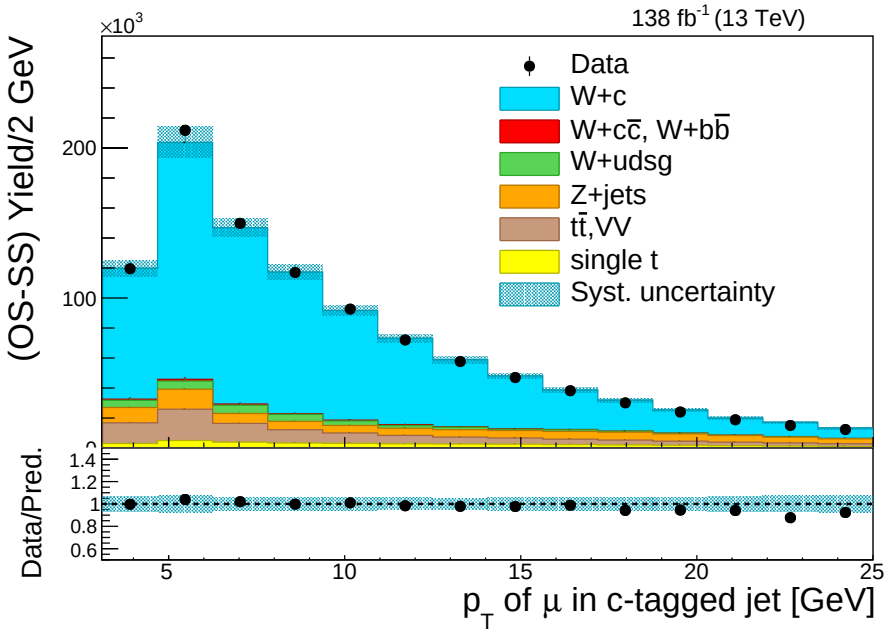
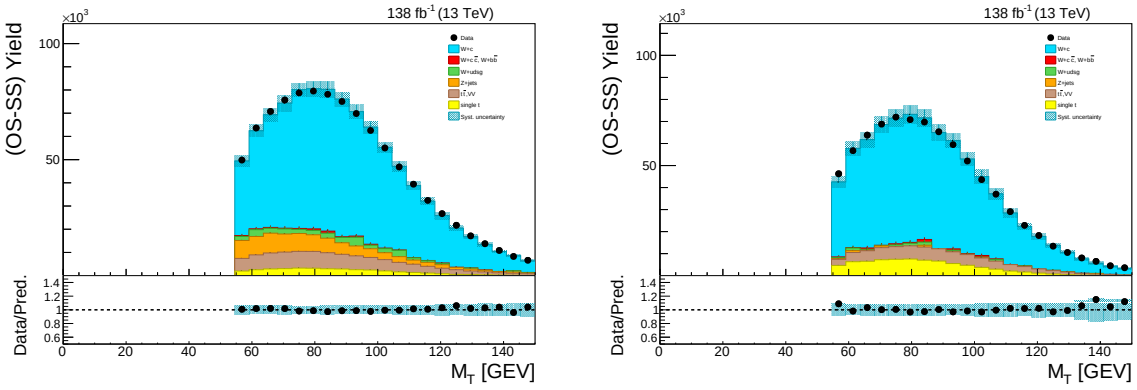


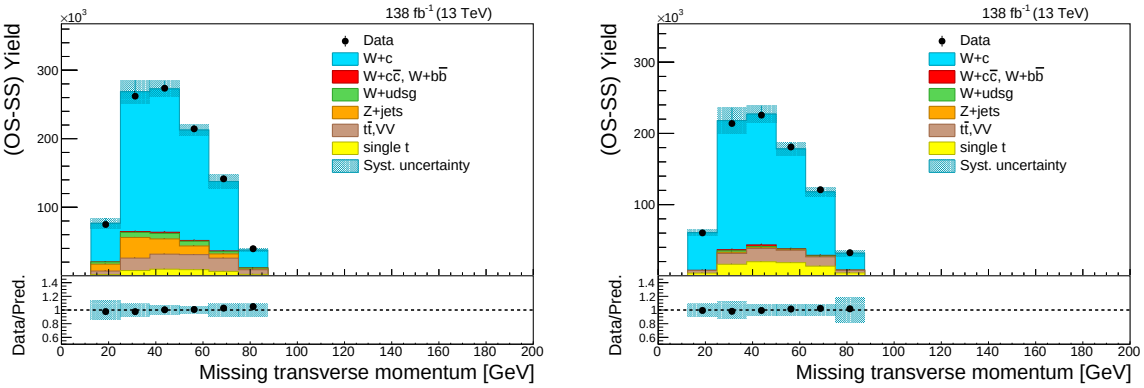
Figure 5.23: Distribution of the transverse momentum of the muon inside the c-tagged jet.



(a) SL channel.

(b) SV channel.

Figure 5.24: Distribution of the reconstructed transverse mass of the W boson.



(a) SL channel.

(b) SV channel.

Figure 5.25: Distribution of the missing transverse momentum of the event.

Table 5.10: Summary of the main systematic uncertainties, in percentage of the measured fiducial cross section, for the four selection channels of the analysis.

	SL $W \to e\nu$	SL $W \to \mu\nu$	SV $W \to e\nu$	SV $W \to \mu\nu$
Source	Uncertainty [%]			
Isolated lepton identification	1.6	0.9	1.6	0.9
Jet energy scale and resolution	2.0	2.0	1.0	1.0
Muon in jet identification	3.0	3.0	-	-
SV reconstruction	-	-	3.7	3.7
Charm fragmentation and decay	2.6	2.6	2.4	2.4
PDF in MC samples	1.0	1.0	1.0	1.0
Stat. uncert. selection efficiency	0.9	1.2	0.9	0.8
Background contributions	0.6	0.9	1.3	1.3
Integrated luminosity	1.6	1.6	1.6	1.6
Total	5.2	5.1	5.4	5.2

Chapter 6

Measurement of the W+c production cross sections

"Living is worthwhile if one can contribute in some small way to this endless chain of progress."
— P. Dirac, Notes.

The main results of the analysis will be the focus of this chapter. We will first show the inclusive cross section measurement within the fiducial phase space region and then the differential cross section as a function of p_{T}^{ℓ} and $|\eta^{\ell}|$. Each measurement is first performed independently in the four decay channels described previously, i.e., the two W boson decay channels and the two charm identification channels. These four measurements are then combined to improve the precision of the overall result. The cross section ratio, $R_c^{\pm}$, and comparisons of the measurements with theoretical predictions are also presented.

6.1. Cross section definition and fiducial volume

We will restrict our measurements to a phase space region that is close to the experimental volume with optimized sensitivity for the signal process. The fiducial region is then defined following the selection cuts in Section 5.3. We select a lepton at generator level coming from the decay of a W boson with $p_{\mathrm{T}}^{\ell} > 35\,\mathrm{GeV}$ and $|\eta^{\ell}| < 2.4$, together with a generator-level c-jet with $p_{\mathrm{T}}^{\mathrm{jet}} > 30\,\mathrm{GeV}$ and $|\eta^{\mathrm{jet}}| < 2.4$. Processes where a pair of charm quark and antiquark is produced in the hard interaction are removed from the signal definition. The c-jet should be well separated from the lepton by an angular distance of $\Delta\mathrm{R}(\text{c-jet}, \ell) > 0.4$.

The cross section is defined as,

$$\sigma(\mathrm{W}+\mathrm{c}) = \frac{Y_{\mathrm{sel}} - Y_{\mathrm{bkg}}}{\mathcal{C}\,\mathcal{L}}, \tag{6.1.1}$$

where Y_{sel} is the OS-SS event yield, and Y_{bkg} the background yield in data after OS-SS subtraction, estimated from simulation and normalized using the data control samples described in Section 5.4. $\mathcal{L}$ is the integrated luminosity of the data sample. The factor $\mathcal{C}$ corrects for acceptance and efficiency losses in the selection process of W + c events produced in the fiducial region at the generator level. It also

subtracts the contributions from W + c events outside the kinematic region of the measurements (around 12%) and from W+c events with W $\to \tau\nu$, where $\tau \to$ e+X or $\tau \to \mu + X$. It is calculated, using the sample of simulated signal events, as the ratio between the event yield of the selected W+c sample (according to the procedure described in Section 5.3 and after OS-SS subtraction) and the number of W+c events satisfying the phase space definition at the generator level. Independent correction factors $\mathcal{C}$ are computed at the particle and parton levels, and for the four selection channels.

Cross sections of the W+c associated production will be calculated using Eq. 6.1.1, unfolded to the particle and parton levels. The difference between these two resides in the nature of the simulated c-jets. At particle level, jets are formed using generator particles produced after the hadronization process. This measurement will be used to compare with MC simulations, that include parton shower and hadronization models. At parton level, jets are constructed from the hard interaction partons, so this measurement can be compared with fixed-order QCD theoretical calculations and can be used in a QCD analysis to extract the strange quark PDF of the proton.

6.2. Measurements at particle level

In this section we are presenting the W+c cross section measurements unfolded to the particle level. The measurements for the four channels are shown separately in Table 6.1, as well as the $\mathcal{C}$ correction factors. The different $\mathcal{C}$ values reflect the different reconstruction and selection efficiencies in the four channels.

Table 6.1: Measured production cross sections σ(W + c) unfolded to the particle level in the four channels (electron and muon W decay modes, SL and SV charm tagging modes) together with statistical (first) and systematic (second) uncertainties. The acceptance times efficiency values ($\mathcal{C}$) are also given.

Channel	$\mathcal{C}$(%)	σ(W + c) [pb]
W $\to e\nu$, SL	1.568 ± 0.014 ± 0.077	158.7 ± 0.6 ± 8.3
W $\to \mu\nu$, SL	0.946 ± 0.011 ± 0.044	149.8 ± 0.7 ± 7.7
W $\to e\nu$, SV	1.389 ± 0.013 ± 0.068	145.0 ± 0.9 ± 7.6
W $\to \mu\nu$, SV	1.966 ± 0.015 ± 0.093	147.4 ± 0.7 ± 7.5

The measurements of the W + c cross sections in the four different channels are consistent within uncertainties, and are combined using the best linear unbiased estimator (BLUE) [160] that takes into account individual uncertainties and their correlations. Each systematic uncertainty is considered to be uncorrelated with the others but those arising from a common source and affecting several measurements are considered to be fully correlated. In particular,

- Electron (muon) efficiencies are considered correlated (uncorrelated) between the two W $\to e\nu$ channels and uncorrelated (correlated) for the W $\to \mu\nu$ channels.

- The PU, JES, MET and luminosity uncertainties are considered correlated among the four channels.
- Secondary vertex reconstruction efficiency and secondary vertex charge determinaton are considered to be correlated in the two SV channels.
- Monte Carlo statistics is considered to be uncorrelated among the four channels.
- Semileptonic fragmentation fractions and branching ratios are taken as correlated in the two SL channels.
- The SV fragmentation fractions and branching ratios are taken as correlated in the two SV channels.

The combined measured cross section unfolded to the particle level is:

$$\sigma(\mathrm{W}+\mathrm{c}) = 148.7 \pm 0.4\,(\mathrm{stat}) \pm 5.6\,(\mathrm{syst})\ \mathrm{pb}. \tag{6.2.1}$$

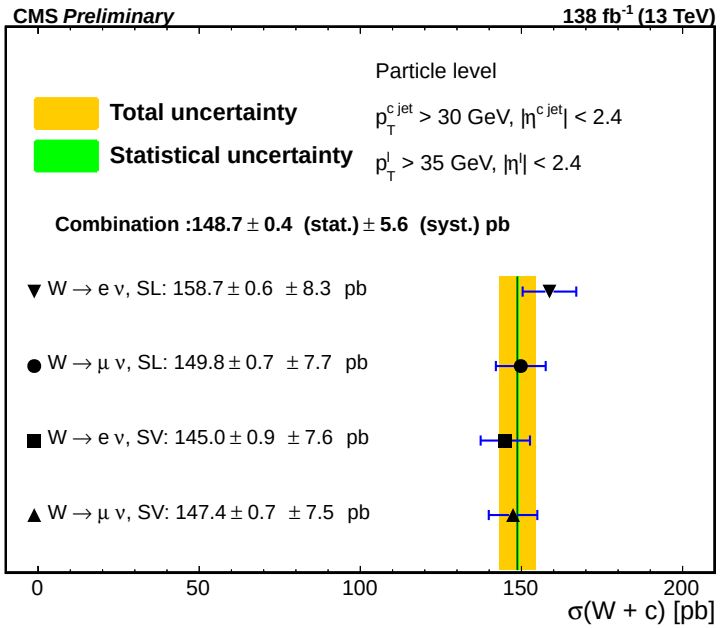


Figure 6.1: Comparison of the combined measurement of the fiducial $\sigma(\mathrm{W}+\mathrm{c})$ cross section unfolded to the particle level with the individual measurements obtained for each of the selection channels.

The combined cross section measurement is shown in Fig. 6.1 together with the four individual measurements obtained for each selection channel. As a cross-check, we have measured the fiducial W+c production cross section separately for each year of data taking (2016-2018), see Fig. 6.2. The measurements are compatible within uncertainties indicating that they can be combined.

The measured cross section unfolded to the particle level is then compared with the predictions from the MadGraph5_aMC@NLO MC generator, as shown in Fig. 6.3. The predictions are computed using two different PDF sets, NNPDF 3.0

NLO and NNPDF 3.1 NNLO. It is worth noting that, as described in Section 5.2, the two predictions differ as well in the tune used in PYTHIA8 for the parton showering, hadronization and underlying event modelling (CUETP8M1 and CP5). The predicted cross sections are about 10% (using NLO NNPDF 3.0) and 20% (NNLO NNPDF 3.1) higher than the measured value, with relative uncertainties close to 10%. The uncertainty associated with the MC predictions takes into account the uncertainties associated with the renormalization and factorization scales, as well as the uncertainty related to the PDFs used in the simulation. The scale uncertainties are estimated using a set of weights provided by the generator that corresponds to independent variations of the scales by factor of 0.5, 1, and 2. The prediction is obtained for all combinations (excluding the cases where one scale is reduced and the other is increased at the same time) and their envelope is quoted as the uncertainty. The uncertainty in the PDFs is estimated using different Hessian eigenvectors of each PDF set.

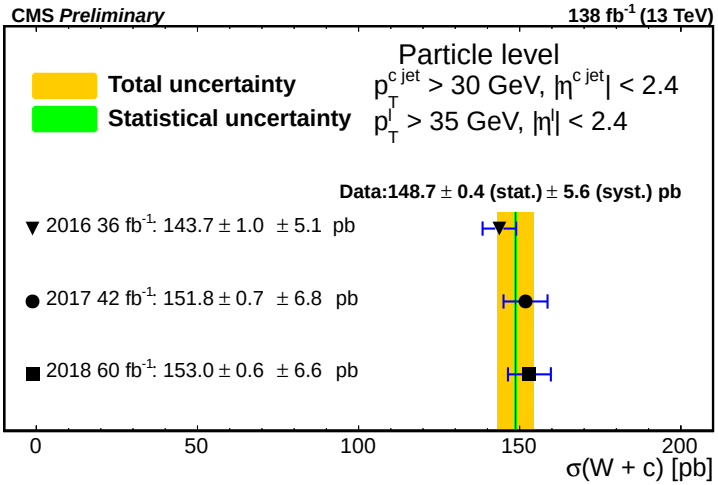


Figure 6.2: Comparison of the combined measurement of the fiducial $\sigma(\mathrm{W}+\mathrm{c})$ cross section unfolded to the particle level with the individual measurements obtained for each of the three years of data taking.

The $\sigma(\mathrm{W}+\mathrm{c})$ production cross section is also measured differentially as a function of $|\eta^{\ell}|$ and p_{T}^{ℓ}, dividing the total sample into subsamples of these variables and computing the cross section with Eq. (6.1.1) as well. The binning of the differential distributions is chosen in such a way that each bin is sufficiently populated to perform the measurement.

We have evaluated the possibility of event migration between neighbouring bins caused by detector resolution effects with the simulated signal sample and concluded that it is negligible, as seen in Figures 6.4 for p_{T}^{ℓ} and 6.5 for $|\eta^{\ell}|$. We compare the reconstructed value of a given variable minus its generated one, divided by the generated one, which results in a normal distribution of events. Its mean value would show a bias on the detector measurement, and the standard deviation corresponds to the resolution. We show in these figures the bias (point) and resolution (error bar) for each bin in which we calculate the differential cross section, and conclude that we do not need to correct for migrations in p_{T}^{ℓ} or $|\eta^{\ell}|$ because the size of each

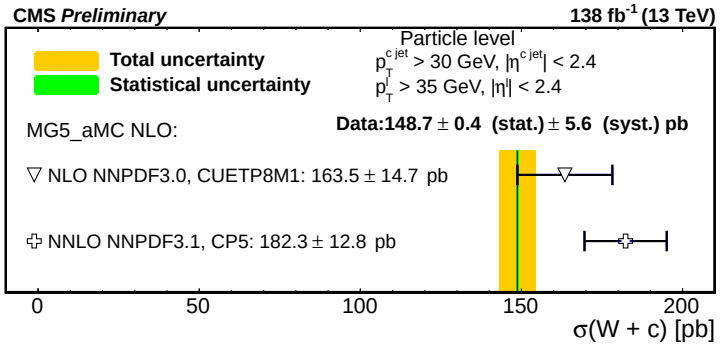
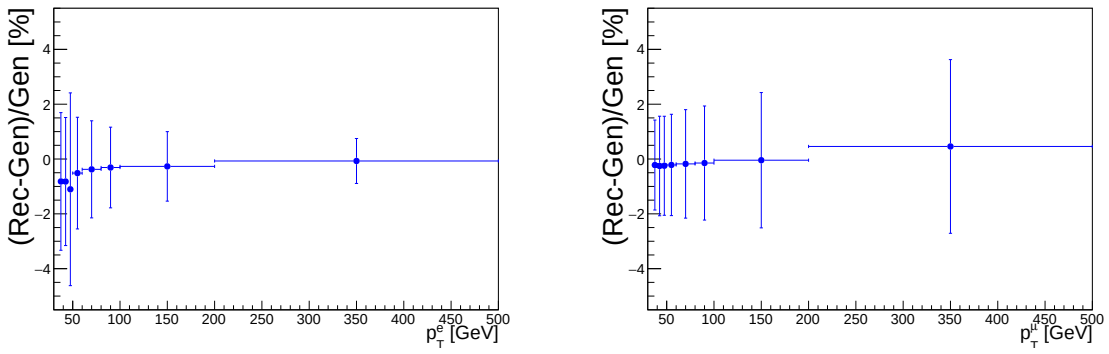


Figure 6.3: Comparison of the measured fiducial $\sigma(\mathrm{W}+\mathrm{c})$ cross section unfolded to the particle level with the predictions from the MadGraph5_aMC@NLO simulation using two different PDF sets (NLO NNPDF 3.0 and NNLO NNPDF 3.1). Two different tunes (CUETP8M1 and CP5) for the parton showering, hadronization and underlying event modelling in Pythia8 are also used.

bin is larger than the corresponding resolutions. Respectively, resolutions are about 2% in p_{T}^{ℓ} and 0.5% in $|\eta^{\ell}|$, with biases much smaller than 1%.

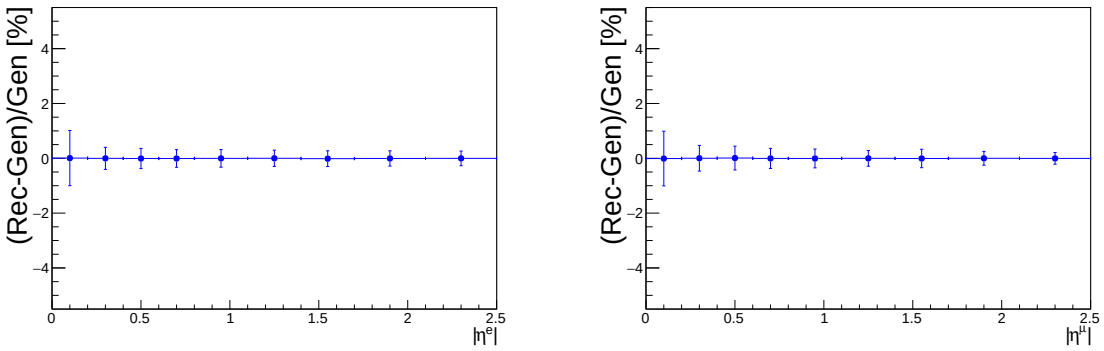


(a) Electron SL channel.

(b) Muon SL channel.

Figure 6.4: Migration studies in bins of p_{T}^{ℓ}. Each point is the mean value of the distribution of the ratio of the difference between the reconstructed and generated events over the generated ones, and shows any bias of the detector. The error bars are the standard deviation of that distribution and show the resolution of the detector.

The charm identification efficiency and its description in simulation vary with the p_{T} of the jet containing the c quark. Furthermore, in $\mathrm{W}+\mathrm{c}$ events there is a correlation between the transverse momentum of the c jet and that of the lepton from the W boson decay. Thus, for the determination of the differential cross sections as a function of p_{T}^{ℓ}, we must apply jet p_{T}-dependent charm identification SFs to the simulated samples. These are determined using the same procedure described in Section 5.3.2 and dividing the SL sample into subsamples depending on the jet-p_{T}, and computing data/MC SFs for the efficiency of charm identification through the reconstruction of a SV for each of them. The value of the SFs range from 0.9 to 1.



(a) Electron SL channel.

(b) Muon SL channel.

Figure 6.5: Migration studies in bins of $|\eta^{\ell}|$. Each point is the mean value of the distribution of the ratio of the difference between the reconstructed and generated events over the generated ones, and shows any bias of the detector. The error bars are the standard deviation of that distribution and show the resolution of the detector.

Systematic uncertainties in the differential $\sigma(\mathrm{W}+\mathrm{c})$ cross section measurements are in the range of 4–6%. The main sources of systematic uncertainty, as discussed in Section 5.7, are related to the charm hadron fragmentation and decay fractions in the simulation (2%), and the efficiency of identifying a SV or a muon inside a jet (3%).

The $\sigma(\mathrm{W}+\mathrm{c})$ differential cross section as a function of $|\eta^{\ell}|$ and p_{T}^{ℓ}, obtained after the combination of the measurements in the four measurement channels, is shown in Fig. 6.6, compared with the predictions from the MadGraph5_aMC@NLO simulation. Observed shape differences are within 10%.

6.3. Measurements at parton level

In this section we present the measurements unfolded to the parton level including an additional correction to account for the c-quark fragmentation and hadronization processes.

Results of the fiducial cross sections in the four selection channels are presented in Table 6.2 and the combined measurement is:

$$\sigma(\mathrm{W}+\mathrm{c}) = 163.4 \pm 0.5\,(\mathrm{stat}) \pm 6.2\,(\mathrm{syst})\ \mathrm{pb}. \tag{6.3.1}$$

The combined cross section measurement is shown in Fig. 6.7 together with the four individual measurements. The measurements in each channel are compatible with each other within uncertainties. The fiducial cross section measured at parton level is 10% larger than that at particle level. During the hadronization and jet clustering processes, the momentum of the c quark gets smeared and biased towards slightly smaller values, as seen in Fig. 6.12. A fraction of charm quarks (about 16%) near the $p_{\mathrm{T}}^{\mathrm{c}} > 30\,\mathrm{GeV}$ threshold of the fiducial region of the measurement do not

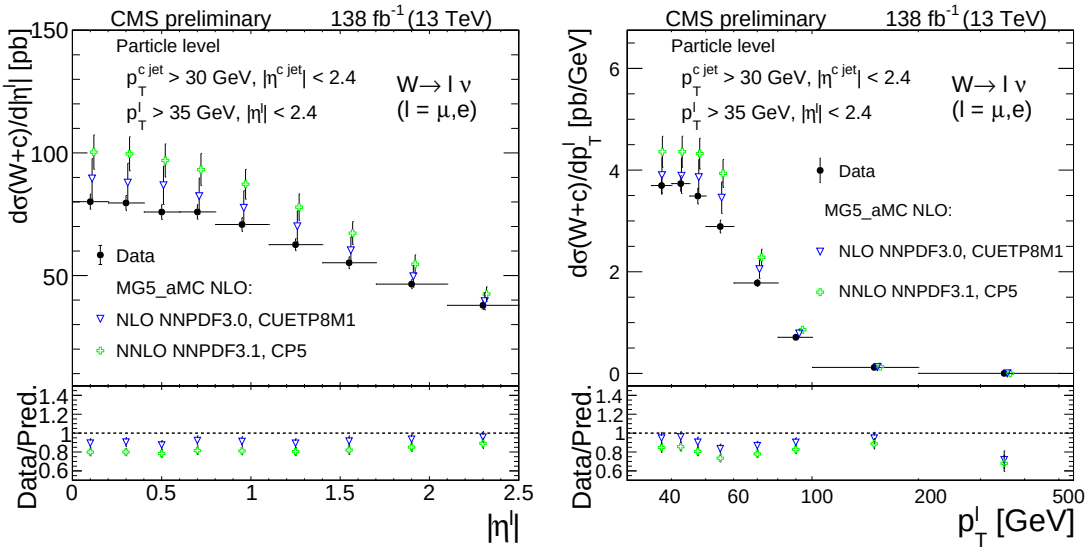


Figure 6.6: Measured differential cross sections $\mathrm{d}\sigma(\mathrm{W}+\mathrm{c})/\mathrm{d}|\eta^{\ell}|$ (left) and $\mathrm{d}\sigma(\mathrm{W}+\mathrm{c})/\mathrm{d}p_{\mathrm{T}}^{\ell}$ (right) unfolded to the particle level, compared with the predictions of the MadGraph5_aMC@NLO simulation. The lower panels are the individual ratios of the measurement over each prediction.

Table 6.2: Measured production cross sections $\sigma(\mathrm{W}+\mathrm{c})$ unfolded to the parton level in the four channels (electron and muon W decay modes, SL and SV charm tagging modes) together with statistical (first) and systematic (second) uncertainties. The acceptance times efficiency values ($\mathcal{C}$) are also given.

Channel	$\mathcal{C}$(%)	$\sigma(\mathrm{W}+\mathrm{c})$ [pb]
$\mathrm{W}\to e\nu$, SL	$1.419 \pm 0.012 \pm 0.069$	$175.3 \pm 0.7 \pm 9.2$
$\mathrm{W}\to \mu\nu$, SL	$0.856 \pm 0.010 \pm 0.040$	$165.4 \pm 0.8 \pm 8.5$
$\mathrm{W}\to e\nu$, SV	$1.261 \pm 0.012 \pm 0.062$	$159.6 \pm 1.0 \pm 8.4$
$\mathrm{W}\to \mu\nu$, SV	$1.786 \pm 0.014 \pm 0.084$	$162.3 \pm 0.8 \pm 8.2$

result in c jets with $p_{\mathrm{T}}^{\mathrm{c-jet}} > 30\,\mathrm{GeV}$ (Fig. 6.11a). On the other hand, a number of W + c events with a c quark with $p_{\mathrm{T}}^{\mathrm{c}} < 30\,\mathrm{GeV}$ get reconstructed with a generator level jet with $p_{\mathrm{T}}^{\mathrm{c-jet}} > 30\,\mathrm{GeV}$ (Fig. 6.11b). The net effect is the reduction of the cross section at the particle level by the mentioned amount of 10%.

As a similar cross-check as in the previous section, we have computed the cross section separately for the three data-taking years (2016-2018). As can be seen in Fig. 6.8, the measurements are compatible within uncertainties and the combination makes sense.

The measurement unfolded to the parton level can be compared with analytical calculations of W + c production. We have used the MCFM 9.1 program [161] to evaluate the cross section predictions in the phase space of the analysis: $p_{\mathrm{T}}^{\ell} > 35\,\mathrm{GeV}$, $|\eta^{\ell}| < 2.4$, $p_{\mathrm{T}}^{\mathrm{c-jet}} > 30\,\mathrm{GeV}$ and $|\eta^{\mathrm{c-jet}}| < 2.4$, with jets clustered using the anti-k_{T} jet algorithm with a distance parameter $R = 0.4$. The W + c process description is available in MCFM up to $\mathcal{O}(\alpha_s{}^2)$ with a massive charm quark

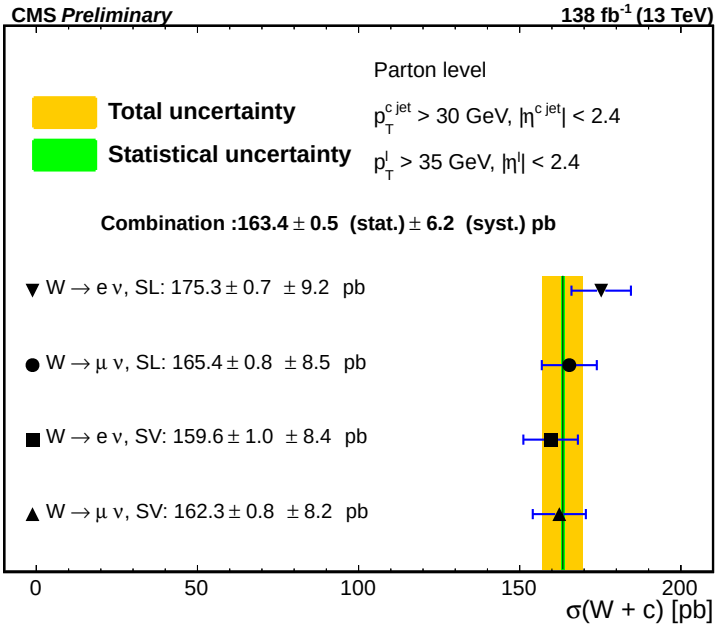


Figure 6.7: Comparison of the combined measurement of the fiducial $\sigma(\mathrm{W}+\mathrm{c})$ cross section unfolded to the parton level with the individual measurements obtained for each of the selection channels.

($m_\mathrm{c} = 1.5\,\mathrm{GeV}$). We only consider charm production via the coupling of a strange or down quark to the W boson and the contributions from gluon splitting into $\mathrm{c\bar{c}}$ are not included. We have computed predictions using the following NLO PDF sets: MSHT20 [162], CT18 [163], CT18Z [163], ABMP16 [164], NNPDF 3.0 [165] and NNPDF 3.1 [145].

The LHAPDF library [166] was used to access the PDF sets, which were derived using strangeness-sensitive experimental data, including LHC W/Z and jet production cross-section measurements. The NNPDF and MSHT20 sets additionally incorporate the CMS W+c production at $\sqrt{s} = 7\,\mathrm{TeV}$ data. CT18Z differs from CT18 in that the former includes the ATLAS W/Z 7 TeV precision measurements [167] leading to an enhancement of the strange PDF. The PDF parameterizations of the MSHT20 and NNPDF groups allow for strangeness asymmetry.

The factorization and the renormalization scales are set to the value of the W boson mass [149]. The uncertainty from missing higher perturbative orders is estimated by computing cross section predictions varying independently the factorization and renormalization scales to twice and half of their nominal values, with the constraint that the ratio of scales is never larger than 2. The envelope of the resulting cross sections with these scale variations defines the theoretical scale uncertainty. The value in the calculation of the strong coupling constant (Eq. 2.1.39) at the energy scale of the mass of the Z boson, $\alpha_\mathrm{s}(m_\mathrm{Z})$, is set to the recommended values by each of the PDF groups. Uncertainties in the predicted cross sections associated with $\alpha_\mathrm{s}(m_\mathrm{Z})$ are evaluated as half the difference in the predicted cross sections evaluated with a variation of $\Delta(\alpha_\mathrm{s}) = \pm 0.002$.

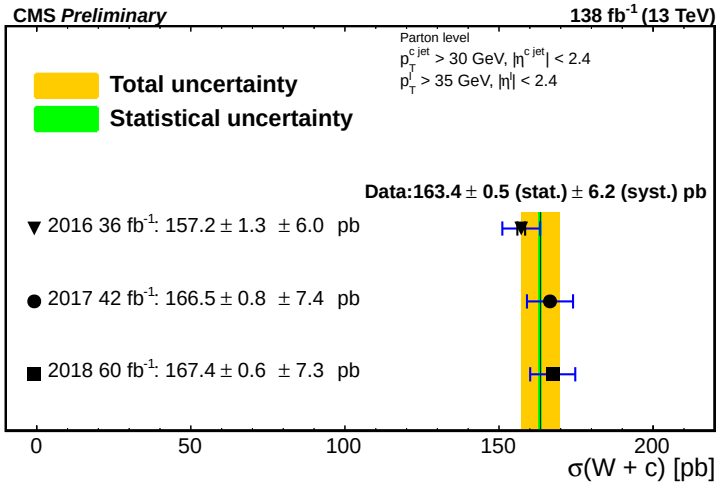


Figure 6.8: Comparison of the combined measurement of the fiducial $\sigma(\mathrm{W}+\mathrm{c})$ cross section unfolded to the parton level with the individual measurements obtained for each of the three years of data taking.

The theoretical predictions for the fiducial $\mathrm{W}+\mathrm{c}$ cross section in the phase space of the measurements are summarized in Table 6.3, along with the measured cross section for comparison. The central value of each prediction is provided together with the relative uncertainties arising from the PDF variations within each set, the choice of scales and α_s. The size of these uncertainties depends on the different input data and methodology used by the various groups. In particular, they depend on the parameterization of the strange quark PDF and on the definition of the one standard deviation uncertainty band. The maximum difference between the central values of the various PDF predictions is $\sim$10%, which is smaller than the total uncertainty in each of the individual predictions. Theoretical predictions lie slightly above the measured cross section but are in agreement within the uncertainties, as depicted in Fig. 6.9.

The $\sigma(\mathrm{W}+\mathrm{c})$ production cross section is also measured differentially as a function of $|\eta^{\ell}|$ and p_{T}^{ℓ}. The comparison of the measured values with the predictions calculated with MCFM is displayed in Fig. 6.10. The predictions are generally consistent with the measurements within uncertainties.

6.4. Measurements of the $\sigma(W^+ + \bar{c})/\sigma(W^- + c)$ cross section ratio

The cross section ratio $\sigma(W^+ + \bar{c})/\sigma(W^- + c)$ is measured in the four channels as the ratio of the event yields in which the lepton from the W-boson decay is positively or negatively charged:

$$R_c^{\pm} = \frac{\sigma(W^+ + \bar{c})}{\sigma(W^- + c)} = \frac{Y_{\mathrm{sel}}^{+} - Y_{\mathrm{bkg}}^{+}}{Y_{\mathrm{sel}}^{-} - Y_{\mathrm{bkg}}^{-}}, \tag{6.4.1}$$

Table 6.3: Predictions for $\sigma(\mathrm{W}+\mathrm{c})$ production from MCFM at NLO for the phase space of the analysis. For every PDF set, the central value of the prediction is given, together with the uncertainty as prescribed from the PDF set, and the uncertainties associated with the scale variations and with the value of α_s. The total uncertainty is given in the last column. The last row in the table gives the experimental results presented in this document.

PDF set	$\sigma(\mathrm{W}+\mathrm{c})$ [pb]	Δ_{PDF} [pb]	Δ_{scales} [pb]	Δ_{α_s} [pb]	Total uncert. [pb]
MSHT20	176.3	$^{+6.8}_{-6.3}$	$^{+6.8}_{-7.4}$	±0.01	$^{+9.6}_{-9.7}$
CT18	164.9	$^{+11.1}_{-8.7}$	$^{+6.1}_{-6.8}$	$^{+0.9}_{-0.8}$	$^{+12.7}_{-11.1}$
CT18Z	176.4	$^{+13.5}_{-10.5}$	$^{+7.0}_{-7.4}$	$^{+0.6}_{-0.5}$	$^{+15.2}_{-12.8}$
ABMP16	183.6	±3.3	$^{+7.2}_{-7.8}$	$^{+1.5}_{-0.9}$	$^{+7.9}_{-8.4}$
NNPDF 3.0	161.9	±6.2	$^{+5.8}_{-6.7}$	±0.01	$^{+8.5}_{-9.1}$
NNPDF 3.1	175.2	±6.1	$^{+6.6}_{-7.3}$	±0.01	$^{+9.1}_{-9.5}$
CMS	$163.4 \pm 0.5\,(\mathrm{stat}) \pm 6.2\,(\mathrm{syst})$ pb				

where it is assumed that the acceptance $\times$ efficiency factor $\mathcal{C}$ is the same for both processes.

The background contributions, Y^{+}_{bkg} and Y^{-}_{bkg}, estimated with the simulations, are subtracted from the selected event yields Y^{+}_{sel} and Y^{-}_{sel}. The statistical uncertainty in the background contributions in the four analysis channels is treated as a source of systematic uncertainty (0.5–0.8%) in the cross section ratio.

The $R_c^{\pm}$ measurements in the four channels are presented in Table 6.4. The four measurements are combined considering as fully correlated the systematic uncertainties of electron, muon and SV reconstruction efficiencies affecting several channels. The combined cross section ratio measurement is:

$$R_c^{\pm} = 0.950 \pm 0.005\,(\mathrm{stat}) \pm 0.010\,(\mathrm{syst}). \tag{6.4.2}$$

Figure 6.13 shows the comparison of the combined measurement with those obtained with each selection channel. Individual measurements are compatible with each other within uncertainties. The precision in the $R_c^{\pm}$ measurement has been improved by factor of two with respect to previous CMS measurements [7–9], leading to the most precise measurement of $R_c^{\pm}$ to date. Following the same reasoning as before, we have also computed the cross section ratio separately for the three data-taking years (2016-2018). The measurements are shown in Fig. 6.14, and are compatible within uncertainties.

The $R_c^{\pm}$ measurement is compared in Fig. 6.15 with the MCFM calculations using various PDF sets. Theoretical predictions for $\sigma(\mathrm{W}^{+}+\bar{\mathrm{c}})$ and $\sigma(\mathrm{W}^{-}+\mathrm{c})$ are computed independently under the same conditions explained in Section 6.3 and for the same $|\eta^{\ell}|$ and p_{T}^{ℓ} ranges used in the analysis. Expectations for $R_c^{\pm}$ are derived from them and presented in Table 6.5. All theoretical uncertainties are significantly reduced in the cross section ratio prediction.

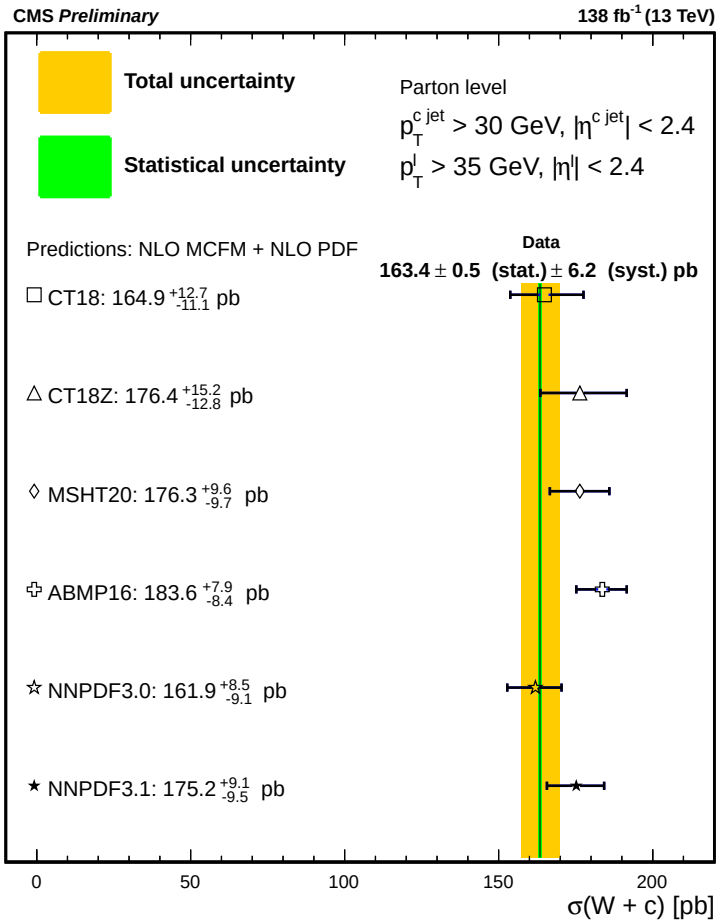


Figure 6.9: Comparison of the experimental measurement of $\sigma(\mathrm{W}+\mathrm{c})$, unfolded to the parton level, with the predictions from the MCFM NLO calculations using different NLO PDF sets.

The $R_c^{\pm}$ observable is sensitive to a potential strangeness asymmetry in the proton but also to the down quark and antiquark asymmetry through the Cabibbo-suppressed down quark contribution to the W+c production. In the absence of strangeness asymmetry, as in the PDF sets CT18 and ABMP16, the predicted $R_c^{\pm}$ value in the kinematical region of the analysis ranges from 0.955 to 0.964 with a small uncertainty of about 2 per mille. The predictions calculated using PDF sets that allow for strangeness asymmetry in the proton (MSHT20 and NNPDF) are about 2% lower, ranging from 0.935 to 0.948 with a 2% uncertainty as a result of the larger uncertainty associated with the difference between the strange quark and antiquark PDFs. Within experimental and theoretical uncertainties, the measured $R_c^{\pm}$ value is consistent with both sets of predictions.

The cross section ratio $R_c^{\pm}$ is also measured differentially as a function of $|\eta^{\ell}|$ and p_{T}^{ℓ}. The measurements are compared with MCFM predictions in Fig. 6.16. The predictions are generally consistent with the measurements, with some modest deviations in shape within 5%.

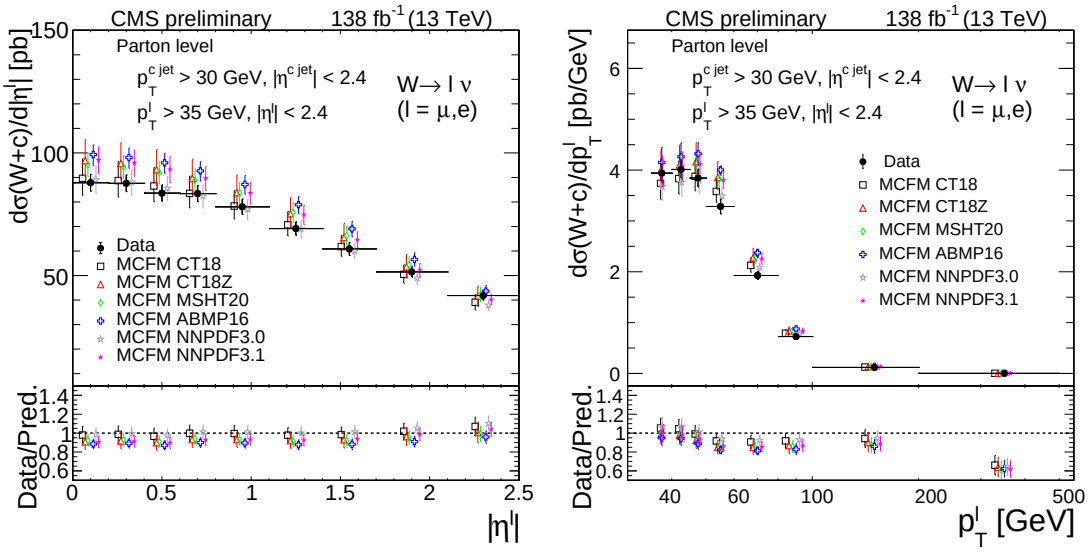


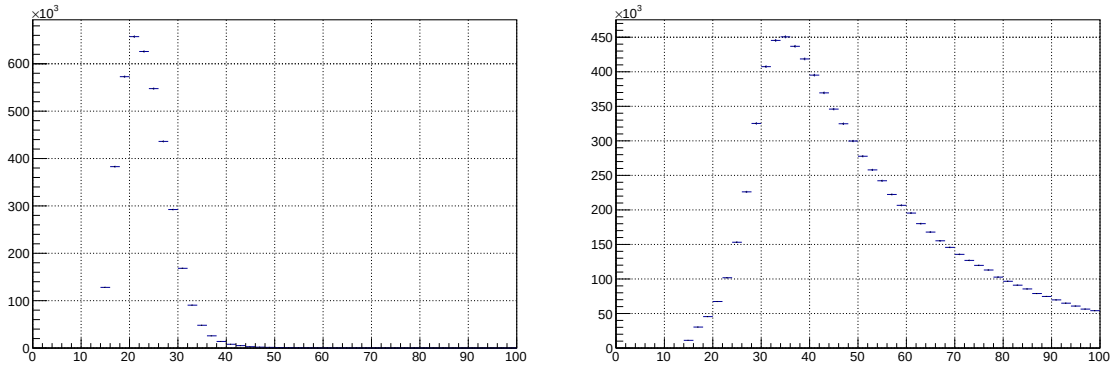
Figure 6.10: Measured differential cross sections $\mathrm{d}\sigma(\mathrm{W}+\mathrm{c})/\mathrm{d}|\eta^{\ell}|$ (left) and $\mathrm{d}\sigma(\mathrm{W}+\mathrm{c})/\mathrm{d}p_{\mathrm{T}}^{\ell}$ (right) unfolded to the parton level, compared with the predictions from the MCFM NLO calculations using different NLO PDF sets. The lower panels are the individual ratios of the measurement over each PDF set.

Table 6.4: Measured production cross section ratio $R_c^{\pm}$ in the four channels (electron and muon W decay modes, SL and SV charm tagging modes). Statistical (first) and systematic (second) uncertainties are also given.

Channel	$R_c^{\pm}$
$\mathrm{W} \to e\nu$, SL	$0.934 \pm 0.006 \pm 0.013$
$\mathrm{W} \to \mu\nu$, SL	$0.940 \pm 0.006 \pm 0.014$
$\mathrm{W} \to e\nu$, SV	$0.961 \pm 0.008 \pm 0.013$
$\mathrm{W} \to \mu\nu$, SV	$0.974 \pm 0.006 \pm 0.015$

6.5. Comparison with NNLO QCD NLO EW predictions

The NLO QCD cross sections for $\mathrm{W}+\mathrm{c}$ production at the Tevatron [168] and at the LHC [169] have been known for a long time. The first computation of NNLO QCD corrections was only recently presented [170]. In that work, off-diagonal CKM elements were included only at LO and the flavoured k_{T} jet clustering algorithm was used. In addition, EW corrections for $\mathrm{W}+\mathrm{c}$ production were not known at that moment and were not included in the calculations. The $\mathrm{W}+\mathrm{c}$ NNLO QCD calculations have been extended [171] with full CKM dependence and including the dominant NLO EW corrections. Furthermore, predictions have been computed using the infrared-safe flavoured anti-k_{T} jet algorithm recently proposed [172]. This is important for a fair comparison between theory predictions and experimental measurements, since experimental results are derived using the anti-k_{T} jet algorithm.



(a) (b)

Figure 6.11: Distribution of the transverse momentum of the charm jet at particle level when the matched charm quark is not in the acceptance region (left), and when the matched charm quark is in the acceptance region (right).

Predictions corresponding to the phase space of the measurements, $p_{\mathrm{T}}^{\ell} > 35\mathrm{GeV}$, $|\eta^{\ell}| < 2.4$, $p_{\mathrm{T}}^{\mathrm{c-jet}} > 30\mathrm{GeV}$, $|\eta^{\mathrm{c-jet}}| < 2.4$, $\Delta R(\mathrm{jet}, \ell) > 0.4$, have been specifically computed by the authors of Ref. [171] for the purpose of this comparison. The input parameters of the calculations are the same as those used in that reference. The charge-dependent flavoured anti-k_{T} jet algorithm with distance parameter $a = 0.1$ is used [172]. The theoretical cross sections are provided at LO, NLO, and NNLO QCD accuracies. At LO, the W + c process is defined at order $\mathcal{O}(\alpha_{\mathrm{s}}\alpha^2)$ in the strong and EW couplings. At NLO, the QCD corrections include all virtual and real contributions of order $\mathcal{O}(\alpha_{\mathrm{s}}^2\alpha^2)$. In the same way, at NNLO accuracy all double-virtual, double-real, and real-virtual contributions of order $\mathcal{O}(\alpha_{\mathrm{s}}^3\alpha^2)$ are taken into account. The calculation is carried out in the 5-flavour scheme with massless bottom and charm quarks, which can therefore be found in both initial and final states. NLO EW corrections of order $\mathcal{O}(\alpha_{\mathrm{s}}\alpha^3)$ are calculated including all virtual corrections and the real corrections involving single real photon emission to cancel the corresponding IR divergences appearing in the EW one-loop amplitude.

The nominal renormalization (μ_{R}) and factorization (μ_{F}) scales are set both to $\frac{1}{2}(E_{\mathrm{T,W}} + p_{\mathrm{T}}^{\mathrm{c-jet}})$, where $E_{\mathrm{T,W}} = \sqrt{M_{\mathrm{W}}^2 + (\vec{p}_{\mathrm{T}}^{\,\ell} + \vec{p}_{\mathrm{T}}^{\,\nu})^2}$. To estimate missing higher-order QCD corrections, the scale uncertainty is obtained by taking the envelope of the 7-point variations of the renormalization and factorization scales $\{(\frac{1}{2}\mu_{\mathrm{R}}, \frac{1}{2}\mu_{\mathrm{F}}), (\frac{1}{2}\mu_{\mathrm{R}}, \mu_{\mathrm{F}}), (\mu_{\mathrm{R}}, \frac{1}{2}\mu_{\mathrm{F}}), (\mu_{\mathrm{R}}, \mu_{\mathrm{F}}), (\mu_{\mathrm{R}}, 2\mu_{\mathrm{F}}), (2\mu_{\mathrm{R}}, \mu_{\mathrm{F}}), (2\mu_{\mathrm{R}}, 2\mu_{\mathrm{F}})\}$.

The NNLO QCD NNPDF 3.1 PDF set was used for computing the predictions for all orders, following the PDF4LHC recommendation [159]. To evaluate the PDF uncertainty of the NNPDF 3.1 sets, specialized minimal PDF sets [173] which contain only 8 replicas were used. The PDF uncertainty is calculated as the square root of the quadratic sum of the differences between the cross section obtained with the nominal PDF and that obtained with each replica.

In Table 6.6, the theoretical predictions for the OS, SS, and OS-SS inclusive fiducial cross section are given at LO, NLO, and NNLO QCD accuracy. The QCD

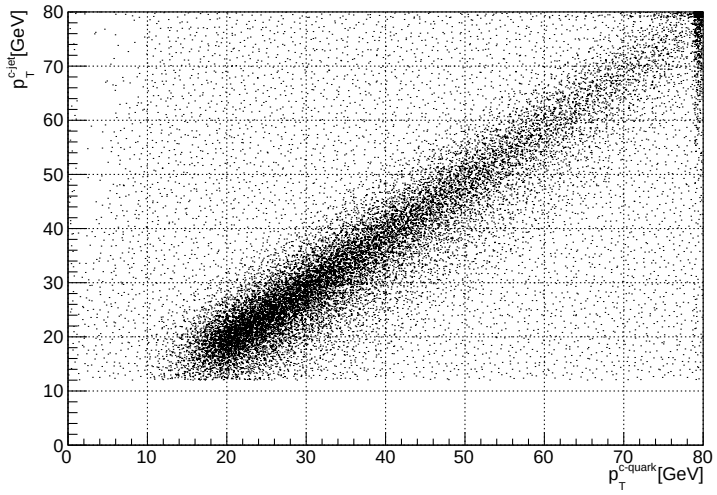


Figure 6.12: Distribution of the transverse momentum of the charm jet at particle level versus the transverse momentum of the matched charm quark.

corrections show good perturbative convergence, the NNLO QCD corrections being significantly smaller than the NLO ones. The NNLO correction for the OS-SS cross section is negative, about -2%. This is due to the fact that the NNLO QCD corrections to SS are larger than those for OS. Note that at LO there is no SS contribution to the $\mathrm{W}+\mathrm{c}$ process and the first SS contribution enters at NLO. The last line in Table 6.6 corresponds to the cross section calculated at NNLO QCD including the NLO EW corrections. The EW corrections amount to -2%. It has been included as a multiplicative factor with negligible statistical uncertainty.

At LO and NLO the total uncertainty in the predictions is dominated by the scale uncertainty (around 5% at NLO) while at NNLO the scale uncertainty is reduced to 1% and the PDF uncertainty (4%) dominates. The inclusion of NNLO QCD corrections can therefore allow a more precise determination of the strange quark content of the proton from the cross section observable.

The predictions are compared to the fiducial cross section measurement in Fig. 6.17. The OS-SS predictions are used in the comparison given that the experimental measurement was performed with OS-SS subtraction. The reasons for this were to suppress the backgrounds and to enhance the sensitivity to the strange quark content of the proton by removing contributions not directly related to the strange PDF, such as gluon splitting into $\mathrm{c\bar{c}}$, which contribute equally to OS and SS final states. The OS-SS subtraction reduces the NNLO corrections, but does not remove them completely. The inclusion of the NNLO corrections decreases the uncertainty in the prediction and also brings it closer to the experimental measurement. The EW NLO corrections further improves the theory/data agreement. The theory prediction and the experimental measurement agree within uncertainties.

The predictions are also compared with the differential cross section measurements $\mathrm{d}\sigma(\mathrm{W}+\mathrm{c})/\mathrm{d}|\eta^{\ell}|$ and $\mathrm{d}\sigma(\mathrm{W}+\mathrm{c})/\mathrm{d}p_{\mathrm{T}}^{\ell}$ in Fig. 6.17. The NLO correction is approximately flat in $|\eta^{\ell}|$ while it is larger at low and high values of p_{T}^{ℓ}. The NLO

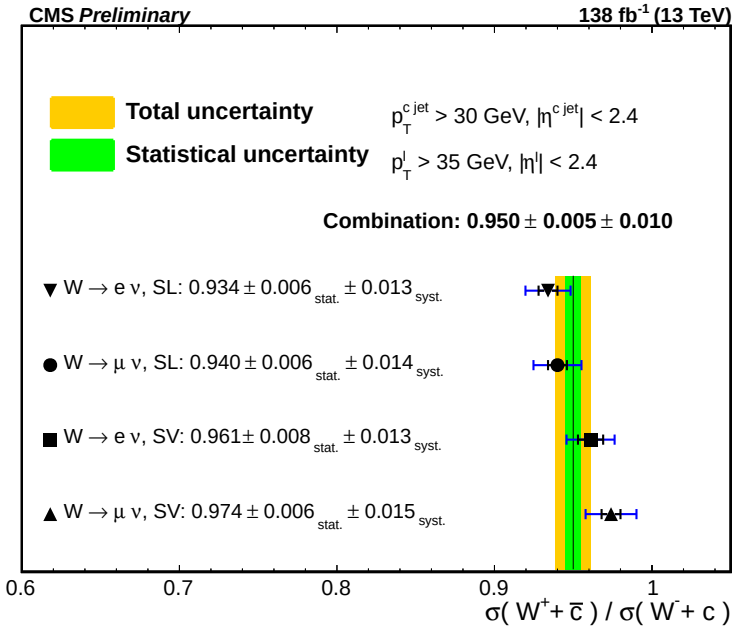


Figure 6.13: Comparison of the combined measurement of the fiducial $R_c^{\pm}$ cross section ratio with the individual measurements obtained for each of the selection channels.

predictions are very similar to those shown in Fig. 6.10 calculated with MCFM at NLO using the same PDF set (NNPDF 3.1). The NNLO correction is small and does not change the shape of the NLO predictions. The EW NLO correction is flat in $|\eta^{\ell}|$ and gets larger with p_{T}^{ℓ}, from 0.99 in the first bin to 0.90 in the highest p_{T}^{ℓ} bin.

Predictions for the OS-SS cross section ratio $R_c^{\pm}$ have also been computed and are collected in Table 6.7. In computing the scale variation of $R_c^{\pm}$, the scale uncertainty for the positive and negative signatures is taken as correlated. The $R_c^{\pm}$ observable is rather stable under perturbative QCD corrections, varying in less than 1% from LO to NNLO accuracy. The NLO EW correction does not affect $R_c^{\pm}$, the change being smaller than one per mille. The comparison of the predictions with the fiducial inclusive and differential measurements are presented in Figs. 6.19 and 6.20, showing a good agreement.

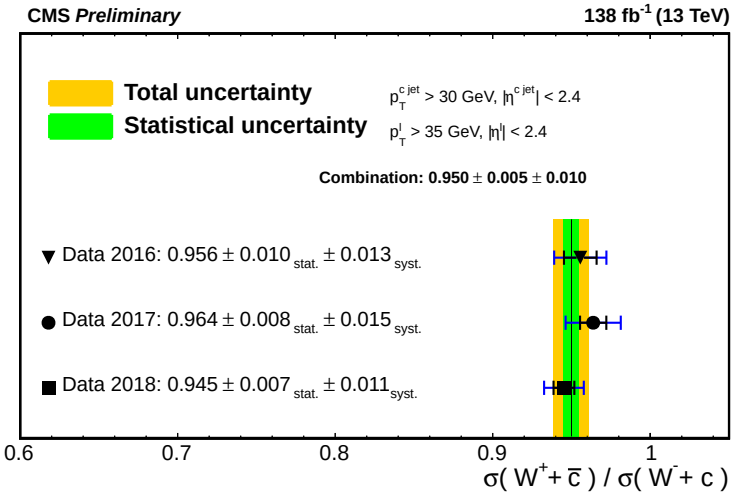


Figure 6.14: Comparison of the combined measurement of the fiducial cross section ratio $R_c^{\pm}$ with the individual measurements obtained for each of the three years of data taking.

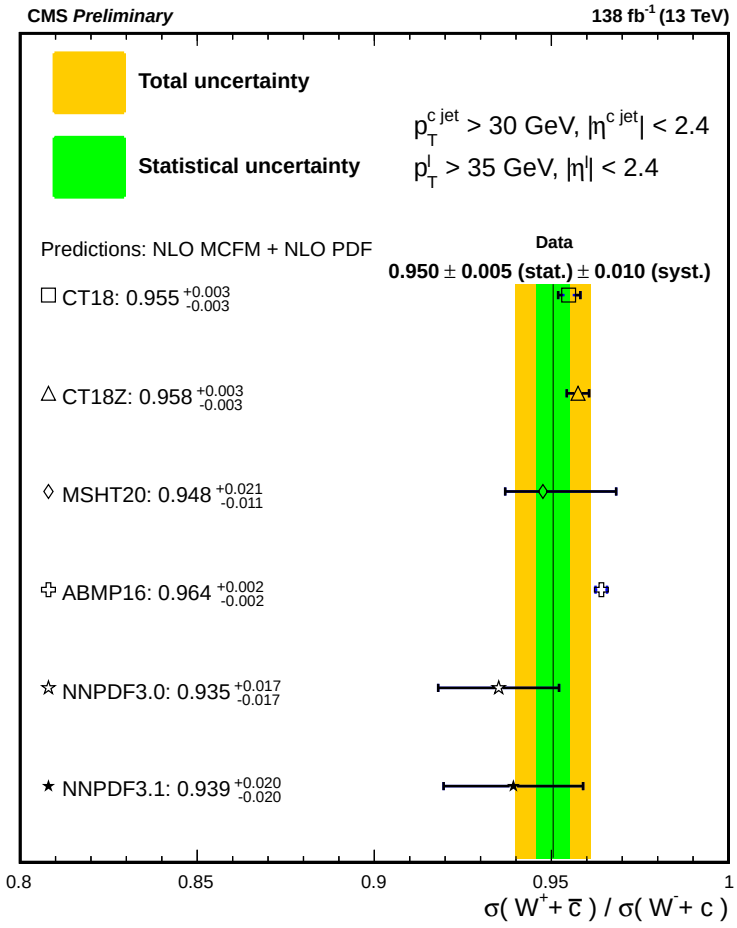


Figure 6.15: Comparison of the experimental measurement of $R_c^{\pm}$ with the MCFM NLO calculations using different NLO PDF sets.

Table 6.5: Theoretical predictions for $R_c^{\pm}$ calculated with MCFM at NLO. The kinematic selection follows the experimental requirements. For every PDF set, the central value of the prediction is given, together with the uncertainty as prescribed from the PDF set, and the uncertainties associated with the scale variations and with the value of α_s. The total uncertainty is given in the last column. The last row in the table gives the experimental results presented in this work.

PDF set	$R_c^{\pm}$	Δ_{PDF}	Δ_{scales}	Δ_{α_s}	Total uncert.
MSHT20	0.948	$^{+0.021}_{-0.011}$	± 0.001	± 0.0001	$^{+0.021}_{-0.011}$
CT18	0.955	± 0.003	± 0.003	± 0.001	± 0.004
CT18Z	0.958	± 0.003	± 0.001	± 0.001	± 0.003
ABMP16	0.964	± 0.002	± 0.001	± 0.001	± 0.002
NNPDF 3.0	0.935	± 0.017	± 0.001	± 0.0001	± 0.017
NNPDF 3.1	0.939	± 0.020	± 0.001	± 0.0001	± 0.020
CMS	$0.950 \pm 0.005\,(\text{stat}) \pm 0.010\,(\text{syst})$				

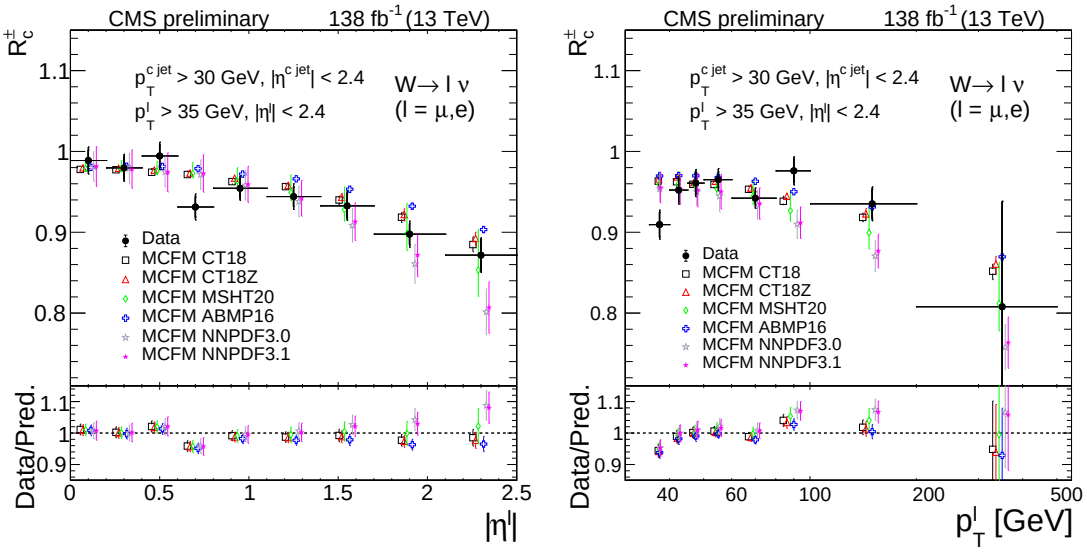


Figure 6.16: Measured cross section ratio $R_c^{\pm}$ as a function of the absolute value of η^{ℓ} (left) and p_{T}^{ℓ} (right), compared with the MCFM NLO calculations using different PDF sets. The lower panels are the individual ratios of the measurement over each PDF set.

Table 6.6: Predictions for $\sigma(\mathrm{W}+\mathrm{c})$ in the phase space of the analysis. For each QCD and EW order, the central values of the OS, SS and OS-SS predictions are given, together with the statistical, scales, PDF, and total uncertainties of the OS-SS prediction. All values are given in pb.

QCD $\mathcal{O}$	EW $\mathcal{O}$	$\sigma^{\mathrm{OS}}_{\mathrm{W+c}}$	$\sigma^{\mathrm{SS}}_{\mathrm{W+c}}$	$\sigma^{\mathrm{OS-SS}}_{\mathrm{W+c}}$	$\Delta^{\mathrm{OS-SS}}_{\mathrm{stat}}$	$\Delta^{\mathrm{OS-SS}}_{\mathrm{scales}}$	$\Delta^{\mathrm{OS-SS}}_{\mathrm{PDF}}$	$\Delta^{\mathrm{OS-SS}}_{\mathrm{Total}}$
LO	LO	137.4	0	137.4	±0.1	$^{+16.6}_{-13.3}$	±5.1	$^{+17.4}_{-14.3}$
NLO	LO	182.4	4.1	178.3	±0.3	$^{+9.3}_{-9.4}$	±6.8	$^{+11.6}_{-11.6}$
NNLO	LO	182.9	8.2	174.7	±1.0	$^{+1.2}_{-2.8}$	±6.8	$^{+7.0}_{-7.4}$
NNLO	NLO	179.1	8.0	171.1	±1.0	$^{+1.2}_{-2.8}$	±6.8	$^{+7.0}_{-7.4}$

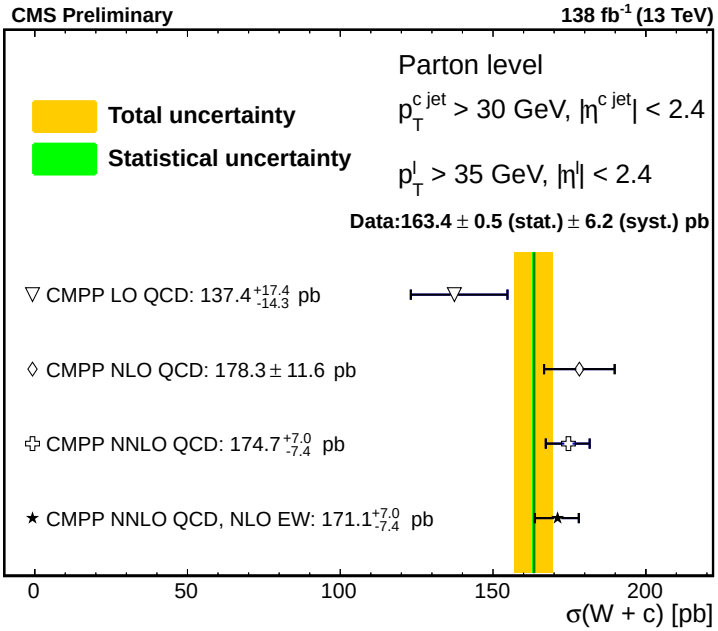


Figure 6.17: Comparison of the experimental measurement of $\sigma(\mathrm{W}+\mathrm{c})$ unfolded to the parton level with the OS-SS LO, NLO, and NNLO QCD predictions, and NLO EW corrections. The NNLO QCD NNPDF 3.1 PDF set is used for computing all the predictions. CMPP stands for the authors of the calculations.

Table 6.7: Theoretical predictions for $R_c^{\pm}$. For each QCD order, the central values are given, together with the MC statistical, scales, PDF, and total uncertainties.

QCD order	$R_c^{\pm}$	Δ_{stat}	Δ_{scales}	Δ_{PDF}	Δ_{Total}
LO	0.945	±0.001	±0.001	±0.022	±0.022
NLO	0.939	±0.004	±0.002	±0.023	±0.023
NNLO	0.936	±0.011	±0.002	±0.023	±0.026

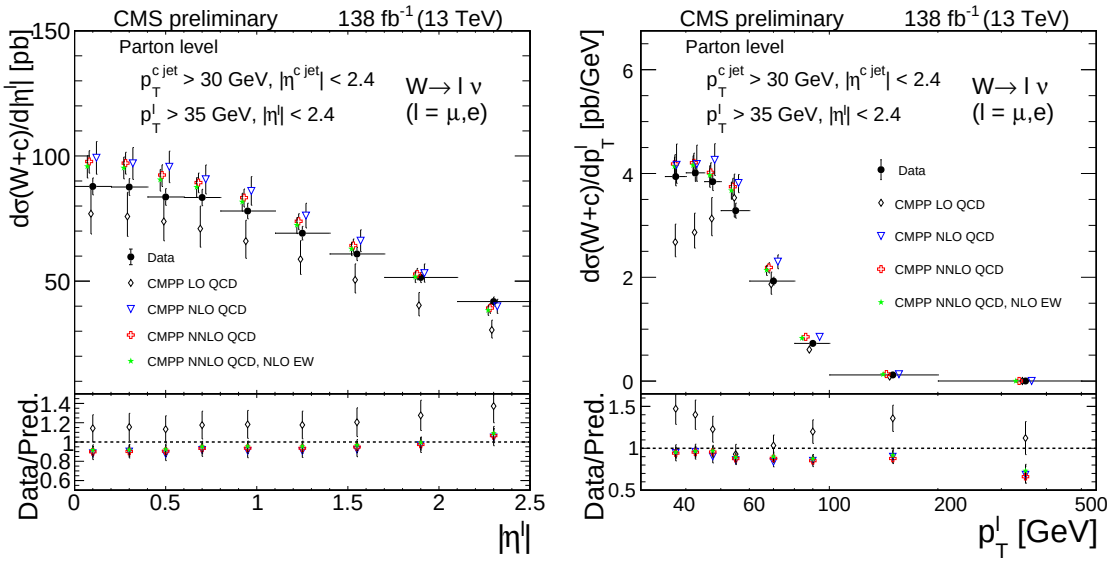


Figure 6.18: Comparison of the measured differential cross sections $\mathrm{d}\sigma(\mathrm{W}+\mathrm{c})/\mathrm{d}|\eta^{\ell}|$ (left) and $\mathrm{d}\sigma(\mathrm{W}+\mathrm{c})/\mathrm{d}p_{\mathrm{T}}^{\ell}$ (right) with the OS-SS LO, NLO, and NNLO QCD predictions, and NLO EW corrections. The NNLO QCD NNPDF 3.1 PDF set is used for computing all the predictions. CMPP stands for the authors of the calculations. The ratios of data to predictions are shown in the lower panels. The uncertainty in the ratio includes the uncertainties in the data and prediction.

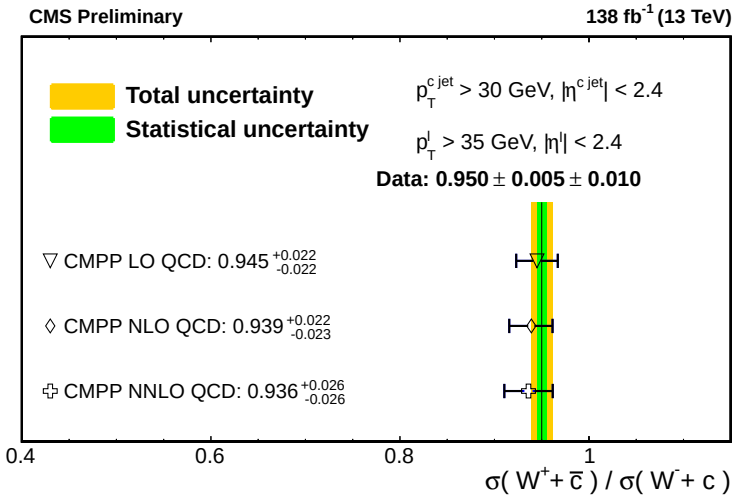


Figure 6.19: Comparison of the experimental measurement of $R_c^{\pm}$ with the OS-SS LO, NLO and NNLO QCD predictions. The NNLO QCD NNPDF 3.1 PDF set is used for computing all the predictions. CMPP stands for the authors of the calculations.

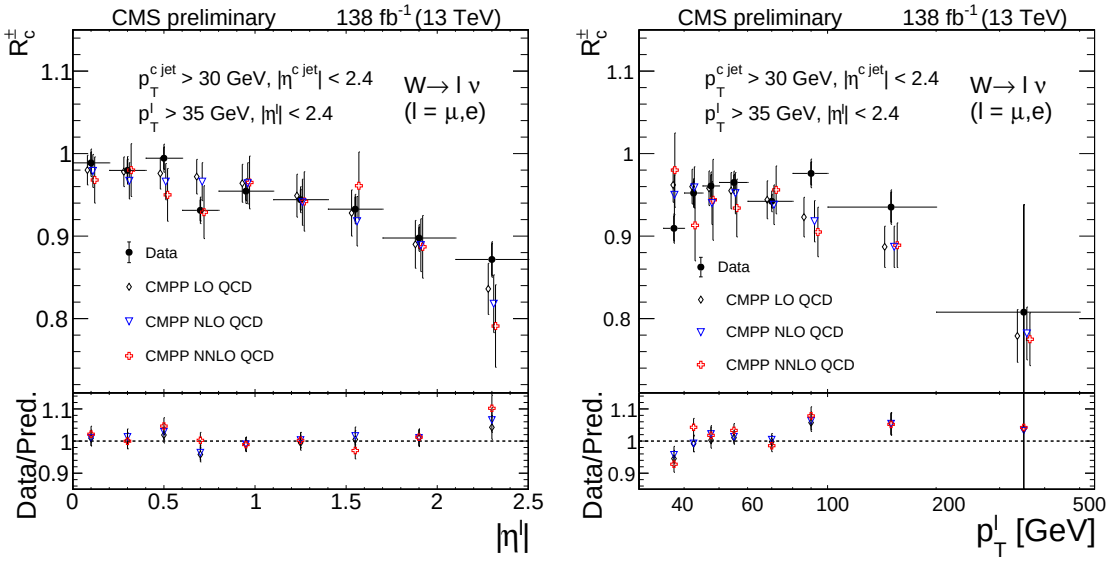


Figure 6.20: Comparison of the measured differential cross section ratio $R_c^{\pm}$ as a function of the absolute value of η^{ℓ} (left) and p_{T}^{ℓ} (right) with the OS-SS LO, NLO, and NNLO QCD predictions. The NNLO QCD NNPDF 3.1 PDF set is used for computing all the predictions. CMPP stands for the authors of the calculations. The ratios of data to predictions are shown in the lower panels. The uncertainty in the ratio includes the uncertainties in the data and prediction.

Chapter 7

Conclusions

"El que no come después de harto, no trabaja después de cansado."
- Miguel Navas Martín.

This Ph.D. thesis is the result of my research in the CMS Collaboration, as a member of the Basic Research Department at CIEMAT (Madrid). I have analyzed proton-proton collisions produced in the LHC at a centre-of mass-energy never reached before in this kind of experiment, 13 TeV, and collected with the CMS detector during the Run 2 data-taking period (2016-2018). The recorded data sample corresponds to an integrated luminosity of $138\,\text{fb}^{-1}$.

The goal of this work was to perform measurements probing the strange quark content of the proton, whose PDF is one of the least constrained. The associated production of a W boson and a single charm quark (W+c) in pp collisions at the CERN LHC is directly sensitive to the strange quark content of the colliding protons at an energy scale of the order of the W boson mass. This sensitivity comes from the fact that the W+c final state is predominantly generated by a strange quark in the initial state, coupling to a gluon from the other proton.

The W + c process is selected based on the presence of a high transverse momentum lepton (electron or muon), coming from a W boson decay, and a jet with the signature of a charm hadron decay. Charm hadron decays are identified either by the presence of a muon inside a jet or by reconstructing a secondary decay vertex within the jet. Measurements are combined from the four different channels (electron and muon W boson decay channels, muon and secondary vertex charm identification channels).

The W+c production cross section measurements are performed within a fiducial region defined by the kinematics of the lepton from the W boson decay and the jet originated by the charm quark ($p_{\text{T}}^{\ell} > 35\,\text{GeV}$, $|\eta^{\ell}| < 2.4$, $p_{\text{T}}^{\text{c-jet}} > 30\,\text{GeV}$, $|\eta^{\text{c-jet}}| < 2.4$). The cross sections are measured at two levels, the particle and the parton levels. The measurements unfolded to the particle level, using particles produced after quark fragmentation and hadronization, can be directly compared with the predictions from MC generators, which implement matrix elements of the W+c process up to certain order in QCD and simulate the contributions of higher QCD orders by means of parton showers. These MC generators also implement parton fragmentation and particle hadronization. The measurements at parton level,

defined by the partons created in the hard interaction, can be compared with fixed-order QCD theoretical calculations.

The cross sections are also measured differentially as functions of the transverse momentum p_{T}^{ℓ} and absolute value of the pseudorapidity $|\eta^{\ell}|$ of the lepton from the W boson decay, allowing for a more detailed exploration of the kinematics of the W+c production and the dependence of the strange PDF on the momentum fraction of the proton taken by the strange quark in the production process.

The W+c cross section is measured separately for each year of data taking and for each of the four selection channels. Results are compatible within uncertainties and the individual measurements are combined taking into account the correlations of the systematic effects. The combined measurement of the fiducial $\sigma(\mathrm{W}+\mathrm{c})$ production cross section, unfolded to the particle level, is:

$$\sigma(\mathrm{pp} \to \mathrm{W}+\mathrm{c}+\mathrm{X}) \times \mathcal{B}(\mathrm{W} \to \ell\nu) = 148.7 \pm 0.4\ (\mathrm{stat}) \pm 5.6\ (\mathrm{syst})\ [\mathrm{pb}].$$

The cross section measurement unfolded to the parton level yields:

$$\sigma(\mathrm{pp} \to \mathrm{W}+\mathrm{c}+\mathrm{X}) \times \mathcal{B}(\mathrm{W} \to \ell\nu) = 163.4 \pm 0.5\ (\mathrm{stat}) \pm 6.2\ (\mathrm{syst})\ [\mathrm{pb}].$$

The cross section ratio for the processes $\mathrm{W}^{+}+\bar{\mathrm{c}}$ and $\mathrm{W}^{-}+\mathrm{c}$ is sensitive to a potential asymmetry between strange quarks and antiquarks in the proton in the phase space of the measurements. The measured $\sigma_{\mathrm{W}^{+}+\bar{\mathrm{c}}}/\sigma_{\mathrm{W}^{-}+\mathrm{c}}$ cross section ratio is:

$$R_c^{\pm} \equiv \frac{\sigma(\mathrm{pp} \to \mathrm{W}^{+}+\bar{c}+\mathrm{X})}{\sigma(\mathrm{pp} \to \mathrm{W}^{-}+\mathrm{c}+\mathrm{X})} = 0.950 \pm 0.005\ (\mathrm{stat}) \pm 0.010\ (\mathrm{syst}).$$

The measurements presented in these thesis work are the most precise to date for the W+c production process. On the one hand, the inclusion of the large dataset collected during the LHC Run 2, together with the combination of four independent signal channels, have made it possible to significantly reduce the statistical uncertainty. The systematic uncertainties have also been reduced thanks to a more precise description of charm fragmentation and decay in the simulation and to a reduction of the background systematics, as a result of a more aggressive background suppression allowed by the availability of a large data sample.

The particle level measurements are compared with the predictions of the MadGraph5_aMC@NLO MC generator. The fiducial cross section measurement is somewhat lower than the prediction, but compatible within uncertainties.

The parton level measurements and the cross section ratio measurement are compared with predictions from the MCFM program that implements QCD NLO calculations. The comparison is done using different NLO PDF sets, some of them assuming symmetric contributions of the strange quark and antiquark PDFs, and the others allowing for asymmetry. The measurements are compatible with all the

predictions within experimental and theoretical uncertainties. Even though the measurements are the most precise to date, discrimination between predictions would require a further reduction of uncertainties.

The measurements unfolded to the parton level are also compared with recently available calculations that compute the W+c process at NNLO in QCD and include NLO electroweak corrections. The authors of the calculations have provided us with specific predictions for the phase space of our measurements. This is the first comparison of experimental measurements of W+c production at 13 TeV with NNLO QCD predictions. The inclusion of the NNLO QCD and NLO electroweak corrections improves the agreement data/theory for the inclusive and differential cross sections. The cross section ratio is essentially insensitive to these corrections.

The theoretical uncertainties are dominated by the uncertainty introduced by the strange PDF. This highlights the importance of our measurements: the inclusion of these high precision measurements in future PDF fits should improve the modeling of the strange PDF in the proton.

The analysis described in this thesis has been published as a CMS Physics Analysis Summary [174] and has been presented in the major conference of the field, the International Conference on High Energy Physics (ICHEP) in Bologna (Italy) [175]. The analysis has undergone a thorough review process by the CMS collaboration and it is about to be submitted for publication to the European Physical Journal.

Abstract

Scientific knowledge experienced an age of explosion during the XX century, fruitful in discoveries that radically changed the perspective we had of Nature. Our current quest is, however, to continue questioning the fundamental aspects of everything that surrounds us, to test until the very end how far the great theories can go, and to pursue with the undeniable task of human restlessness to answer the question "and why is that?". In the specific topic of this work, we will make use of the largest machine ever built to obtain information about some of the smallest elements of matter, measuring with the best precision up to date the cross section of a process that could shed some light on the mystery of fundamental asymmetry between matter and antimatter in the Universe. Our mind demands symmetry.

The associated production of a W boson and a single charm quark (W+c) in proton-proton collisions at the LHC is directly sensitive to the strange quark content of the colliding protons at an energy scale of the order of the W boson mass. Studies of this process provide valuable information on the strange quark parton distribution function (PDF), which is one of the least constrained PDFs of the proton. Accurate measurements of the W +c production cross section and of the $R_c^{\pm} = \sigma(W^+ + \bar{c})/\sigma(W^- + c)$ cross section ratio can probe the level of asymmetry between the s and $\bar{\text{s}}$ PDFs and help to measure their relative content with respect to the up and down antiquarks coming from the quark-gluon sea of the proton.

A measurement of the associated production of a W boson and a charm quark in proton-proton collisions at a centre-of-mass energy of 13 TeV is reported. The analysis uses a data sample corresponding to a total integrated luminosity of $138\,\text{fb}^{-1}$ collected by the CMS detector at the LHC. W bosons are identified through their leptonic decays to an electron or a muon, and a neutrino. Charm jets are tagged by the presence of a muon or a secondary vertex inside the jet. The W+c production cross section and the cross section ratio $R_c^{\pm}$ are measured inclusively in a fiducial region of phase space, and differentially as a function of the transverse momentum and the absolute value of the pseudorapidity of the lepton from the W boson decay.

We present the measurement of the W+c production cross section at $\sqrt{s} = 13\,\text{TeV}$ in four independent selection channels: the decay of the W boson into electron or muon and a neutrino, and the identification of jets originated from a c quark through the reconstruction of a muon (SL) or a secondary vertex (SV) inside the jet. Systematic uncertainties have been reduced with respect to previous CMS W+c analyses, resulting in the most accurate measurements up to date. Measurements are performed at the particle (hadron) and parton levels, and are compared to predictions from Monte Carlo generators as well as theoretical calculations. Specifically, the measurements unfolded to the particle level are compared to the MadGraph Monte Carlo generator, that implements calculations up to next-to-leading order (NLO) in perturbative quantum chromodynamics (QCD) interfaced with parton showering simulations, whereas the measurements unfolded to the parton level are

compared at fixed-order perturbative QCD calculations at the NLO and next-to-next-to-leading order (NNLO) accuracies. The theoretical predictions agree with the measurements within experimental and theoretical uncertainties. The measurements presented in this work will help increase the accuracy of the determination of the strange quark PDF of the proton.

Resumen

El conocimiento científico vivió una época de explosión durante el siglo XX, con numerosos descubrimientos que cambiaron radicalmente la perspectiva que teníamos de la Naturaleza. Nuestra misión actual es, sin embargo, continuar cuestiónandonos los aspectos fundamentales de todo lo que nos rodea, comprobar a fondo hasta dónde son capaces de llegar las grandes teorías desarrolladas y continuar con la innegable tarea de la inquietud humana de responder a la pregunta de "¿y eso por qué?". En el caso concreto de este trabajo, utilizaremos la máquina más grande jamás construida para obtener información de algunos de los elementos más pequeños de la materia, midiendo con la mejor precisión hasta la fecha la sección eficaz de un proceso que podría arrojar luz sobre la incógnita de la asimetría fundamental entre materia y antimateria en el Universo. Nuestra mente necesita simetría.

La producción asociada de un bosón W y un único quark encanto (charm, c) en colisiones protón-protón en el LHC es directamente sensible al contenido en quark extraño (strange, s) de los protones incidentes a la escala de energías del orden de la masa del bosón W. Los estudios de este proceso proporcionan información valiosa sobre la función de distribución de partones (PDF) del quark strange, que es la PDF medida con menor precisión del protón. Unas medidas precisas de la sección eficaz de producción de W+c y del cociente de secciones eficaces $R_c^{\pm} = \sigma(W^+ + \bar{c})/\sigma(W^- + c)$ permiten examinar el nivel de asimetría entre las PDF del s y el $\bar{\text{s}}$ y medir su contenido relativo respecto a los antiquarks up y down procedentes del mar de quarks y gluones del protón.

Se presenta una medida de la producción asociada de un bosón W y un quark c en colisiones de protón-protón a energía del centro de masas de 13 TeV. El análisis utiliza una muestra de datos correspondiente a una luminosidad integrada total de $138\,\text{fb}^{-1}$ registrados por el detector CMS en el LHC. Los bosones W se identifican a través de su desintegración leptónica a un electrón o un muon y un neutrino. Los chorros o jets procedentes del quark charm se identifican por la presencia de un muon o un vértice secundario dentro del jet. Se mide la sección eficaz de producción de W+c y del cociente $R_c^{\pm}$ de forma inclusiva en una región fiducial del espacio de fases y de forma diferencial en función del momento transverso y del valor absoluto de la pseudorrapidez del leptón de la desintegración del W.

Presentamos la medida de la sección eficaz de producción de W+c a $\sqrt{s} = 13\,\text{TeV}$ en cuatro canales de selección independientes: la desintegración de un bosón W en un electrón o un muon y un neutrino y la identificación de jets procedentes de un quark c a través de la reconstrucción de un muon (SL) o un vértice secundario (SV) dentro de un jet. Las incertidumbres sistemáticas se han reducido con respecto a anteriores análisis de W+c en CMS, dando lugar a las medidas más precisas hasta la fecha. Se realizan las medidas a los niveles de partículas (hadrones) y de partones y se comparan con predicciones de generadores de Monte Carlo y cálculos teóricos. Específicamente, las medidas a nivel de partículas se comparan con el generador

MadGraph de Monte Carlo, que implementa cálculos hasta segundo orden en teoría de perturbaciones (NLO) de cromodinámica cuántica (QCD) junto con simulación de lluvia de partones, mientras que las medidas a nivel de partones se comparan con cálculos a orden fijo de teoría de perturbaciones en QCD con precisión de segundo y tercer orden (NNLO). Las predicciones teóricas concuerdan con los datos dentro de las incertidumbres experimentales y teóricas. Las medidas realizadas en este trabajo servirán para aumentar la precisión con la que puede determinarse la PDF del quark strange dentro del protón.

Bibliography

[1] G. Arnison et al. "Experimental observation of isolated large transverse energy electrons with associated missing energy at s=540 GeV". In: *Physics Letters B* 122.1 (1983), pp. 103–116. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(83)91177-2.

[2] J. E. Augustin et al. "Discovery of a Narrow Resonance in e^+e^- Annihilation". In: *Phys. Rev. Lett.* 33 (23 1974), pp. 1406–1408. DOI: 10.1103/PhysRevLett.33.1406.

[3] J. J. Aubert et al. "Experimental Observation of a Heavy Particle J". In: *Phys. Rev. Lett.* 33 (23 1974), pp. 1404–1406. DOI: 10.1103/PhysRevLett.33.1404.

[4] U. Baur, F. Halzen, S. Keller, M. L. Mangano, and K. Riesselmann. "The charm content of W + 1 jet events as a probe of the strange quark distribution function". In: *Phys. Lett. B* 318 (1993), p. 544. DOI: 10.1016/0370-2693(93)91553-Y. arXiv: hep-ph/9308370 [hep-ph].

[5] F. Faura, S. Iranipour, E. R. Nocera, J. Rojo, and M. Ubiali. "The Strangest Proton?" In: *Eur. Phys. J. C* 80.12 (2020), p. 1168. DOI: 10.1140/epjc/s10052-020-08749-3. arXiv: 2009.00014 [hep-ph].

[6] S. Catani, D. de Florian, G. Rodrigo, and W. Vogelsang. "Perturbative Generation of a Strange-Quark Asymmetry in the Nucleon". In: *Physical Review Letters* 93.15 (2004). DOI: 10.1103/physrevlett.93.152003.

[7] S. Chatrchyan et al. "Measurement of associated W+charm production in pp collisions at $\sqrt{s} = 7$ TeV". In: *JHEP* 02 (2014), p. 013. DOI: 10.1007/JHEP02(2014)013. arXiv: 1310.1138 [hep-ex].

[8] "Measurements of the production of a W boson in association with a charm quark in proton-proton collisions at $\sqrt{s} = 8$ TeV". In: *The European Physical Journal C* (2022). DOI: 10.1140/epjc/s10052-022-10897-7. arXiv: 2112.00895 [hep-ex].

[9] A. M. Sirunyan et al. "Measurement of associated production of a W boson and a charm quark in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Eur. Phys. J. C* 79 (2019), p. 269. DOI: 10.1140/epjc/s10052-019-6752-1. arXiv: 1811.10021 [hep-ex].

[10] G. Aad et al. "Measurement of the production of a W boson in association with a charm quark in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector". In: *JHEP* 05 (2014), p. 068. DOI: 10.1007/JHEP05(2014)068. arXiv: 1402.6263 [hep-ex].

[11] R. Aaij et al. "Study of W boson production in association with beauty and charm". In: *Physical Review D* 92 (2015), p. 052001. DOI: 10.1103/PhysRevD.92.052001. arXiv: 1505.04051 [hep-ex].

[12] D. Griffiths. "Introduction to Particle Physics". In: *J. Wiley & Sons, Inc* (2014).

[13] A. Zee. *Quantum field theory in a nutshell*. Princeton university press, 2010.

[14] S. Tomonaga. "On a Relativistically Invariant Formulation of the Quantum Theory of Wave Fields*". In: *Progress of Theoretical Physics* 1.2 (1946), pp. 27–42.

[15] J. Schwinger. "Quantum Electrodynamics. I. A Covariant Formulation". In: *Phys. Rev.* 74 (10 1948), pp. 1439–1461.

[16] R. P. Feynman. "Mathematical Formulation of the Quantum Theory of Electromagnetic Interaction". In: *Physical Review* 80 (1950), pp. 440–457.

[17] P. D. Group. "Review of Particle Physics". In: *Progress of Theoretical and Experimental Physics* 2022.8 (2022). 083C01. ISSN: 2050-3911. DOI: 10.1093/ptep/ptac097. eprint: https://academic.oup.com/ptep/article-pdf/2022/8/083C01/45434166/ptac097.pdf.

[18] "Gargamelle: first neutral current." 1973. URL: https://cds.cern.ch/record/39468.

[19] F. Hasert et al. "Search for elastic muon-neutrino electron scattering". In: *Physics Letters B* 46.1 (1973), pp. 121–124. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(73)90494-2. URL: https://www.sciencedirect.com/science/article/pii/0370269373904942.

[20] E. Fermi. "Tentativo di una teoria dei raggi β". In: *Il Nuovo Cimento (1924-1942)* 11.1 (1934), pp. 1–19.

[21] S. L. Glashow and M. Gell-Mann. "Gauge theories of vector particles". In: *Annals of Physics* 15.3 (1961), pp. 437–460.

[22] S. Glashow. "Partial–symmetries of weak interactions". In: *Nuclear Physics* 22.4 (1961), pp. 579–588. DOI: 10.1016/0029-5582(61)90469-2.

[23] S. L. Glashow, J. Iliopoulos, and L. Maiani. "Weak Interactions with Lepton–Hadron Symmetry". In: *Phys. Rev. D* 2.7 (1970), p. 1285. DOI: 10.1103/PhysRevD.2.1285.

[24] S. Weinberg. "A model of leptons". In: *Physical review letters* 19.21 (1967), p. 1264. DOI: 10.1103/PhysRevLett.19.1264.

[25] A. Salam and N Svartholm. "Elementary particle theory". In: *Almquist and Wiksell, Stockholm* (1968), p. 367.

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

[27] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321.

[28] P. Higgs. "Broken symmetries, massless particles and gauge fields". In: *Phys. Lett.* 12.2 (1964), pp. 132–133. DOI: 10.1016/0031-9163(64)91136-9.

[29] P. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13.16 (1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508.

[30] P. Higgs. "Spontaneous Symmetry Breakdown without Massless Bosons". In: *Phys. Rev.* 145.4 (1966), pp. 1156–1163. DOI: 10.1103/PhysRev.145.1156.

[31] G. 't Hooft. "Renormalization of massless Yang–Mills fields". In: *Nuclear Physics B* 33.1 (1971), pp. 173–199. DOI: 10.1016/0550-3213(71)90395-6.

[32] CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716.1 (2012). arXiv:1207.7235, pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021.

[33] CMS Collaboration. "Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s}$ =7 and 8 TeV". In: *Journal of High Energy Physics* 6 (2013). arXiv:1303.4571, 81. DOI: 10.1007/JHEP06(2013)081.

[34] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020.

[35] N. Cabibbo. "Unitary Symmetry and Leptonic Decays". In: *Phys. Rev. Lett.* 10.12 (1963), pp. 531–533. DOI: 10.1103/PhysRevLett.10.531.

[36] M. Kobayashi and T. Maskawa. "CP–Violation in the Renormalizable Theory of Weak Interaction". In: *Progress of Theoretical Physics* 49.2 (1973), pp. 652–657. DOI: 10.1143/PTP.49.652.

[37] B. Pontecorvo. "Mesonium and antimesonium". In: *Soviet Phys. JETP* 6 (1958).

[38] Z. Maki, M. Nakagawa, and S. Sakata. "Remarks on the Unified Model of Elementary Particles". In: *Progress of Theoretical Physics* 28.5 (1962), pp. 870–880. ISSN: 0033-068X. DOI: 10.1143/PTP.28.870. eprint: https://academic.oup.com/ptp/article-pdf/28/5/870/5258750/28-5-870.pdf. URL: https://doi.org/10.1143/PTP.28.870.

[39] H. Fritzsch, M. Gell-Mann, and H. Leutwyler. "Advantages of the color octet gluon picture". In: *Physics Letters B* 47 (1973), pp. 365–368.

[40] J. Gao, L. Harland-Lang, and J. Rojo. "The structure of the proton in the LHC precision era". In: *Physics Reports* 742 (2018), pp. 1–121. DOI: 10.1016/j.physrep.2018.03.002.

[41] R. P. Feynman. "The Behavior of Hadron Collisions at Extreme Energies". In: *Special Relativity and Quantum Theory: A Collection of Papers on the Poincaré Group*. Ed. by M. E. Noz and Y. S. Kim. Dordrecht: Springer Netherlands, 1988, pp. 289–304. ISBN: 978-94-009-3051-3. DOI: 10.1007/978-94-009-3051-3_25.

[42] J. D. Bjorken and E. A. Paschos. "Inelastic Electron-Proton and γ-Proton Scattering and the Structure of the Nucleon". In: *Phys. Rev.* 185 (5 1969), pp. 1975–1982. DOI: 10.1103/PhysRev.185.1975.

[43] J. C. Collins, D. E. Soper, and G. Sterman. "Factorization for short distance hadron-hadron scattering". In: *Nuclear Physics B* 261 (1985), pp. 104–142. ISSN: 0550-3213. DOI: 10.1016/0550-3213(85)90565-6.

[44] J. C. Collins, D. E. Soper, and G. Sterman. "FACTORIZATION OF HARD PROCESSES IN QCD". In: *Perturbative QCD*, pp. 1–91. DOI: 10.1142/9789814503266_0001.

[45] G. T. Bodwin. "Erratum: Factorization of the Drell-yan cross section in perturbation theory". In: *Phys. Rev. D* 34 (12 1986), pp. 3932–3932. DOI: 10.1103/PhysRevD.34.3932.

[46] G. Altarelli and G. Parisi. "Asymptotic freedom in parton language". In: *Nuclear Physics B* 126.2 (1977), pp. 298–318. ISSN: 0550-3213. DOI: 10.1016/0550-3213(77)90384-4.

[47] S. Bailey, T. Cridge, L. A. Harland-Lang, A. D. Martin, and R. S. Thorne. "Parton distributions from LHC, HERA, Tevatron and fixed target data: MSHT20 PDFs". In: *The European Physical Journal C* 81.4 (2021). DOI: 10.1140/epjc/s10052-021-09057-0.

[48] S. Chatrchyan et al. "Measurement of the muon charge asymmetry in inclusive pp $\rightarrow$ W + X production at $\sqrt{s}$=7 TeV and an improved determination of light parton". In: *Phys. Rev. D* 90 (2014), p. 032004. DOI: 10.1103/PhysRevD.90.032004. arXiv: 1312.6283.

[49] M. Aaboud et al. "Measurement of the W-boson mass in pp collisions at $$\sqrt{s}=7\, \hbox {TeV}$$ s = 7 TeV with the ATLAS detector". In: *The European Physical Journal C* 78.2 (2018). DOI: 10.1140/epjc/s10052-017-5475-4.

[50] null null et al. "High-precision measurement of the ¡i¿W¡/i¿ boson mass with the CDF II detector". In: *Science* 376.6589 (2022), pp. 170–176. DOI: 10.1126/science.abk1781. eprint: https://www.science.org/doi/pdf/10.1126/science.abk1781. URL: https://www.science.org/doi/abs/10.1126/science.abk1781.

[51] and A. O. Bazarko et al. "Determination of the strange quark content of the nucleon from a next-to-leading-order QCD analysis of neutrino charm production". In: *Zeitschrift für Physik C Particles and Fields* 65.2 (1995), pp. 189–198. DOI: 10.1007/bf01571875.

[52] M. Goncharov et al. "Precise measurement of dimuon production cross sections in ν_μFe and $\overline{\nu}_\mu$Fe deep inelastic scattering at the Fermilab Tevatron". In: *Phys. Rev. D* 64 (11 2001), p. 112006. DOI: 10.1103/PhysRevD.64.112006.

[53] O. Samoylov et al. "A precision measurement of charm dimuon production in neutrino interactions from the NOMAD experiment". In: *Nuclear Physics B* 876.2 (2013), pp. 339–375. ISSN: 0550-3213. DOI: https://doi.org/10.1016/j.nuclphysb.2013.08.021.

[54] A Kayis-Topaksu et al. "Measurement of charm production in neutrino charged-current interactions". In: *New Journal of Physics* 13.9 (2011), p. 093002. DOI: 10.1088/1367-2630/13/9/093002.

[55] S. Alekhin et al. "Determination of strange sea quark distributions from fixed-target and collider data". In: *Phys. Rev. D* 91 (9 2015), p. 094002. DOI: 10.1103/PhysRevD.91.094002.

[56] N. Bartosik. "Associated top-quark-pair and b-jet production in the dilepton channel at $\sqrt{s} = 8$ TeV as test of QCD and background to tt+Higgs production". Universität Hamburg, Diss., 2015. Dr. Hamburg: Universität Hamburg, 2015, p. 312. DOI: 10.3204/DESY-THESIS-2015-035.

[57] T. Stelzer and W. Long. "Automatic generation of tree level helicity amplitudes". In: *Computer Physics Communications* 81.3 (1994), pp. 357–371. DOI: 10.1016/0010-4655(94)90084-1.

[58] T. Sjöstrand et al. "An introduction to PYTHIA 8.2". In: *Comput. Phys. Commun.* 191 (2015), p. 159. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].

[59] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer. "MadGraph 5 : Going Beyond". In: *JHEP* 1106 (2011), p. 128. DOI: 10.1007/JHEP06(2011)128. arXiv: 1106.0522 [hep-ph].

[60] O. M. P. Artoisenet R. Frederix and R. Rietkerk. "Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations". In: *JHEP* 03 (2013), p. 015. DOI: 10.1007/JHEP03(2013)015. arXiv: 1212.3460 [hep-ph].

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

[62] S. Frixione and B. R. Webber. "Matching NLO QCD computations and parton shower simulations". In: *Journal of High Energy Physics* 2002.06 (2002), pp. 029–029. DOI: 10.1088/1126-6708/2002/06/029.

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

[64] S. Frixione, P. Nason, and C. Oleari. "Matching NLO QCD computations with parton shower simulations: the POWHEG method". In: *Journal of High Energy Physics* 2007.11 (2007), pp. 070–070. DOI: 10.1088/1126-6708/2007/11/070.

[65] S. Alioli, P. Nason, C. Oleari, and E. Re. "NLO vector-boson production matched with shower in POWHEG". In: *JHEP* 0807 (2008), p. 060. DOI: 10.1088/1126-6708/2008/07/060. arXiv: 0805.4802 [hep-ph].

[66] S. Alioli, P. Nason, and E. Re. "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX". In: *JHEP* 06 (2010), p. 04. DOI: 10.1007/JHEP06(2010)043. arXiv: 1002.2581 [hep-ph].

[67] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjöstrand. "Parton fragmentation and string dynamics". In: *Physics Reports* 97.2 (1983), pp. 31–145. ISSN: 0370-1573. DOI: https://doi.org/10.1016/0370-1573(83)90080-7.

[68] J. M. Campbell and R. K. Ellis. "Update on vector boson pair production at hadron colliders". In: *Physical Review D* 60.11 (1999). DOI: 10.1103/physrevd.60.113006.

[69] J. M. Campbell and R. Ellis. "MCFM for the Tevatron and the LHC". In: *Nucl. Phys. B - Proc. Suppl.* 205 (2010), p. 10. DOI: 10.1016/j.nuclphysbps.2010.08.011. arXiv: 1007.3492 [hep-ph].

[70] P. Skands, S. Carrazza, and J. Rojo. "Tuning PYTHIA 8.1: the Monash 2013 Tune". In: *Eur. Phys. J. C* 74.8 (2014), p. 3024. DOI: 10.1140/epjc/s10052-014-3024-y. arXiv: 1404.5630 [hep-ph].

[71] V. Khachatryan et al. "Event generator tunes obtained from underlying event and multiparton scattering measurements". In: *Eur. Phys. J. C* 76 (2016), p. 155. DOI: 10.1140/epjc/s10052-016-3988-x. arXiv: 1512.00815 [hep-ex].

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

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

[74] L. Evans and P. Bryant. "LHC Machine". In: *JINST* 3 (2008), S08001. DOI: 10.1088/1748-0221/3/08/S08001.

[75] LEP. "LEP Design Report Vol.2: The LEP main ring". In: *CERN-LEP-84-01* 2 (1984). URL: https://cds.cern.ch/record/102083.

[76] "The Super Proton Synchrotron". In: *Technical Report* (2012). URL: http://cds.cern.ch/record/1997188.

[77] M. Benedikt, P. Collier, V. Mertens, J. Poole, and K. Schindl. "LHC Design Report: the LHC Injector Chain". In: *CERN Yellow Reports: Monographs* (2004). DOI: 10.5170/CERN-2004-003-V-3. URL: http://cds.cern.ch/record/823808.

[78] C. Lefèvre. "The CERN accelerator complex. Complexe des accélérateurs du CERN". 2008. URL: https://cds.cern.ch/record/1260465.

[79] *LHC The Guide*. Public Results. URL: https://cds.cern.ch/record/2809109/files/CERN-Brochure-2021-004-Eng.pdf.

[80] The ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *JINST* 3 (2008), S08003. DOI: 10.1088/1748-0221/3/08/S08003.

[81] The CMS Collaboration. "The CMS experiment at the CERN LHC". In: *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.

[82] The LHCb Collaboration. "The LHCb Detector at the LHC". In: *JINST* 3 (2008), S08005. DOI: 10.1088/1748-0221/3/08/S08005.

[83] The ALICE Collaboration. "The ALICE experiment at the CERN LHC". In: *JINST* 3 (2008), S08002. DOI: 10.1088/1748-0221/3/08/S08002.

[84] S van der Meer. *Calibration of the effective beam height in the ISR*. Tech. rep. Geneva: CERN, 1968. URL: https://cds.cern.ch/record/296752.

[85] *CMS Luminosity*. Public Results. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults.

[86] CMS Collaboration. "CMS slice raw illustrator files". In: (2016). URL: https://cds.cern.ch/record/2204899.

[87] A. Tumasyan et al. "Strategies and performance of the CMS silicon tracker alignment during LHC Run 2". In: *Nucl. Instrum. Meth. A* 1037 (2022). DOI: 10.1016/j.nima.2022.166795. arXiv: 2111.08757 [physics.ins-det].

[88] *The CMS magnet project: Technical Design Report*. Technical design report. CMS. Geneva: CERN, 1997. DOI: 10.17181/CERN.6ZU0.V4T9. URL: http://cds.cern.ch/record/331056.

[89] S. Chatrchyan et al. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". In: *JINST* 9 (2014), P10009. DOI: 10.1088/1748-0221/9/10/P10009. arXiv: 1405.6569 [physics.ins-det].

[90] V Karimäki et al. *The CMS tracker system project: Technical Design Report*. Technical design report. CMS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/368412.

[91] W. Adam et al. "The CMS Phase-1 Pixel Detector Upgrade". In: *JINST* 16 (2021), P02027. DOI: 10.1088/1748-0221/16/02/P02027. arXiv: 2012.14304 [physics.ins-det].

[92] CMS Collaboration. *Track impact parameter resolution for the full pseudo rapidity coverage in the 2017 dataset with the CMS Phase-1 Pixel detector*. CMS Detector Performance Summary CMS-DP-2020-049. 2020. URL: https://cds.cern.ch/record/2743740.

[93] *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical design report. CMS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/349375.

[94] Q. Ingram. "Energy resolution of the barrel of the CMS Electromagnetic Calorimeter". In: *JINST* 2.04 (2007), P04004. DOI: 10.1088/1748-0221/2/04/P04004. URL: http://cds.cern.ch/record/1009081.

[95] *The CMS hadron calorimeter project: Technical Design Report*. Technical design report. CMS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/357153.

[96] A. M. Sirunyan et al. "Particle-flow reconstruction and global event description with the CMS detector". In: *JINST* 12 (2017), P10003. DOI: 10.1088/1748-0221/12/10/P10003. arXiv: 1706.04965 [physics.ins-det].

[97] CMS Collaboration. "Calibration of the CMS hadron calorimeters using proton–proton collision data at $\sqrt{s} = 13\,\text{TeV}$". In: *JINST* 15.05 (2020), P05002. DOI: 10.1088/1748-0221/15/05/P05002. arXiv: 1910.00079 [physics.ins-det].

[98] C. Collaboration. "The very forward CASTOR calorimeter of the CMS experiment". In: *JINST* 16 (2021), P02010. 65 p. DOI: 10.1088/1748-0221/16/02/P02010. arXiv: 2011.01185. URL: https://cds.cern.ch/record/2743313.

[99] O. A. Grachov et al. "Status of Zero Degree Calorimeter for CMS Experiment". In: *AIP Conference Proceedings*. AIP, 2006. DOI: 10.1063/1.2396962. URL: https://doi.org/10.1063%2F1.2396962.

[100] J. G. Layter. *The CMS muon project: Technical Design Report*. Technical design report. CMS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/343814.

[101] A. M. Sirunyan et al. "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *JINST* 13 (2018). DOI: 10.1088/1748-0221/13/06/P06015. arXiv: 1804.04528 [physics.ins-det].

[102] A. M. Sirunyan et al. "Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$". In: *JINST* 15 (2020), P10017. DOI: 10.1088/1748-0221/15/10/P10017. arXiv: 2006.10165 [hep-ex].

[103] V. Khachatryan et al. "The CMS trigger system". In: *JINST* 12 (2017), P01020. DOI: 10.1088/1748-0221/12/01/P01020. arXiv: 1609.02366 [physics.ins-det].

[104] C. Collaboration. "Performance of the CMS drift–tube chamber local trigger with cosmic rays". In: *JINST* 5 (2009), T03003. 31. DOI: 10.1088/1748-0221/5/03/T03003. arXiv: 0911.4893 [hep-ex].

[105] W. Adam et al. "The CMS High Level trigger". In: *Eur. Phys. J. C* 46 (2006), pp. 605–667. DOI: 10.1140/epjc/s2006-02495-8. arXiv: hep-ex/0512077.

[106] K. Bos et al. "LHC Computing Grid". In: *Technical Design Report. LCG* (2005). URL: https://cds.cern.ch/record/840543.

[107] G. L. Bayatyan, M. Della Negra, Foà, A Hervé, and A. Petrilli. *CMS computing: Technical Design Report*. Technical design report. CMS. Submitted on 31 May 2005. Geneva: CERN, 2005. URL: http://cds.cern.ch/record/838359.

[108] S. Dal Pra et al. "Evolution of monitoring, accounting and alerting services at INFN-CNAF Tier-1". In: *EPJ Web of Conferences* 214 (2019), p. 08033. DOI: 10.1051/epjconf/201921408033.

[109] R. Brun and F. Rademakers. "ROOT – An object oriented data analysis framework". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 389.1 (1997). New Computing Techniques in Physics Research V, pp. 81 –86. DOI: 10.1016/S0168-9002(97)00048-X.

[110] D. Lange. *The CMS Reconstruction Software*. Tech. rep. Geneva: CERN, 2011. URL: https://cds.cern.ch/record/1322136.

[111] C. Collaboration. *Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET*. CMS Physics Analysis Summary CMS-PAS-PFT-09-001. 2009. URL: http://cdsweb.cern.ch/record/1194487.

[112] W. Adam, B. Mangano, T. Speer, and T. Todorov. *Track Reconstruction in the CMS tracker*. Tech. rep. Geneva: CERN, 2006. URL: http://cds.cern.ch/record/934067.

[113] S. Chatrchyan et al. "Alignment of the CMS tracker with LHC and cosmic ray data". In: *JINST* 9 (2014), P06009. DOI: 10.1088/1748-0221/9/06/P06009. arXiv: 1403.2286 [physics.ins-det].

[114] P. Billoir. "Progressive track recognition with a Kalman–like fitting procedure". In: *Computer Physics Communications* 57.1 (1989), pp. 390–394. ISSN: 0010-4655. DOI: 10.1016/0010-4655(89)90249-X. URL: https://www.sciencedirect.com/science/article/pii/001046558990249X.

[115] R. Mankel. "A concurrent track evolution algorithm for pattern recognition in the HERA-B main tracking system". In: *Nuclear Instruments and Methods in Physics Research Section A* 395.2 (1997), pp. 169–184. ISSN: 0168-9002. DOI: 10.1016/S0168-9002(97)00705-5.

[116] R. Fruhwirth. "Application of Kalman filtering to track and vertex fitting". In: *Nucl. Instrum. Methods Phys. Res., A* 262.HEPHY-PUB-503 (1987), p. 444. DOI: 10.1016/0168-9002(87)90887-4.. URL: https://cds.cern.ch/record/178627.

[117] P. Azzurri. "Track Reconstruction Performance in CMS". In: *Nuclear Physics B - Proceedings Supplements* 197.1 (2009), pp. 275–278. DOI: 10.1016/j.nuclphysbps.2009.10.084.

[118] K. Rose. "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems". In: *Proceedings of the IEEE* 86.11 (1998), pp. 2210–2239. DOI: 10.1109/5.726788.

[119] W. Waltenberger, R. Frühwirth, and P. Vanlaer. "Adaptive vertex fitting". In: *J. Phys. G* 34 (2007), N343. DOI: 10.1088/0954-3899/34/12/N01. URL: https://cds.cern.ch/record/1027031.

[120] S. Chatrchyan et al. "Identification of b-quark jets with the CMS experiment". In: *JINST* 8 (2013), P04013. DOI: 10.1088/1748-0221/8/04/P04013. arXiv: 1211.4462 [hep-ex].

[121] V. Khachatryan et al. "Measurement of $B\bar{B}$ Angular Correlations based on Secondary Vertex Reconstruction at $\sqrt{s} = 7$ TeV". In: *JHEP* 03 (2011), p. 136. DOI: 10.1007/JHEP03(2011)136. arXiv: 1102.3194 [hep-ex].

[122] S. Chatrchyan et al. "Measurement of the cross section and angular correlations for associated production of a Z boson with b hadrons in pp collisions at $\sqrt{s} = 7$ TeV". In: *JHEP* 12 (2013), p. 039. DOI: 10.1007/JHEP12(2013)039. arXiv: 1310.1349 [hep-ex].

[123] A. M. Sirunyan et al. "Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC". In: *JINST* 16 (2021), P05014. DOI: 10.1088/1748-0221/16/05/P05014. arXiv: 2012.06888 [hep-ex].

[124] CMS Collaboration. *ECAL 2016 refined calibration and Run2 summary plots*. CMS Detector Performance Summary CMS-DP-2020-021. CMS, 2020. URL: https://cds.cern.ch/record/2717925.

[125] W Adam, R Frühwirth, A Strandlie, and T Todorov. "Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC". In: *Journal of Physics G: Nuclear and Particle Physics* 31.9 (2005), N9. DOI: 10.1088/0954-3899/31/9/N01.

[126] M. Cacciari and G. P. Salam. "Pileup subtraction using jet areas". In: *Phys. Lett. B* 659 (2008), p. 119. DOI: 10.1016/j.physletb.2007.09.077. arXiv: 0707.1378 [hep-ph].

[127] M. Cacciari, G. P. Salam, and G. Soyez. "The anti-k_t jet clustering algorithm". In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ex].

[128] M. Cacciari, G. P. Salam, and G. Soyez. "FastJet user manual". In: *Eur. Phys. J. C* 72 (2012), p. 1896. DOI: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ph].

[129] *The Jet Plus Tracks Algorithm for Calorimeter Jet Energy Corrections in CMS*. Tech. rep. 2009. URL: http://cds.cern.ch/record/1190234.

[130] V. Khachatryan et al. "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV". In: *JINST* 12 (2017), P02014. DOI: 10.1088/1748-0221/12/02/P02014. arXiv: 1607.03663 [hep-ex].

[131] CMS Collaboration. "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS". In: *JINST* 6.11 (2011), P11002–P11002. DOI: 10.1088/1748-0221/6/11/P11002. arXiv: 1107.4277 [physics.ins-det].

[132] E. Bols, J. Kieseler, M. Verzetti, M. Stoye, and A. Stakia. "Jet Flavour Classification Using DeepJet". In: *JINST* 15 (2020), P12012. DOI: 10.1088/1748-0221/15/12/P12012. arXiv: 2008.10519 [hep-ex].

[133] 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". In: (2018). URL: http://cds.cern.ch/record/2646773.

[134] C. Collaboration. *Performance of b tagging at $\sqrt{s} = 8$ TeV in multijet, ttbar and boosted topology events*. CMS Physics Analysis Summary CMS-PAS-BTV-13-001. 2013. URL: http://cds.cern.ch/record/1581306.

[135] D. Guest et al. "Jet flavor classification in high-energy physics with deep neural networks". In: *Physical Review D* 94.11 (2016). DOI: 10.1103/physrevd.94.112002. URL: https://doi.org/10.1103%2Fphysrevd.94.112002.

[136] A. M. Sirunyan et al. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV". In: *JINST* 13 (2018), P05011. DOI: 10.1088/1748-0221/13/05/P05011. arXiv: 1712.07158 [physics.ins-det].

[137] CMS Collaboration. "Performance of MET reconstruction and pileup mitigation techniques in CMS". In: *Nuclear and Particle Physics Proceedings* 273-275 (2016), pp. 2512–2514. DOI: 10.1016/j.nuclphysbps.2015.09.442. arXiv: 1502.05207 [hep-ex].

[138] A. M. Sirunyan et al. "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ using the CMS detector". In: *JINST* 14 (2019), P07004. DOI: 10.1088/1748-0221/14/07/P07004. arXiv: 1903.06078 [hep-ex].

[139] C. Collaboration. "Measurement of associated Z + charm production in proton-proton collisions at $\sqrt{s}$= 8 TeV". In: *Eur. Phys. J. C* 78 (2018), p. 287. DOI: 10.1140/epjc/s10052-018-5752-x. arXiv: 1711.02143 [hep-ex].

[140] CMS Collaboration. *Search for Higgs boson decay to a charm quark-antiquark pair in proton-proton collisions at $\sqrt{s} = 13$ TeV*. Accepted for publication in Phys. Rev. Lett. 2022. DOI: 10.48550/ARXIV.2205.05550.

[141] A. M. Sirunyan et al. "Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS". In: *Eur. Phys. J. C* 81 (2021), p. 800. DOI: 10.1140/epjc/s10052-021-09538-2. arXiv: 2104.01927 [hep-ex].

[142] 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. CMS, 2018. URL: https://cds.cern.ch/record/2621960/.

[143] 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. CMS, 2019. URL: https://cds.cern.ch/record/2676164/.

[144] R. Frederix and S. Frixione. "Merging meets matching in MC@NLO". In: *JHEP* 12 (2012), p. 061. DOI: 10.1007/JHEP12(2012)061. arXiv: 1209.6215 [hep-ph].

[145] R. D. Ball et al. "Parton distributions from high-precision collider data". In: *Eur. Phys. J. C* 77 (2017), p. 663. DOI: 10.1140/epjc/s10052-017-5199-5. arXiv: 1706.00428 [hep-ph].

[146] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt. "Parton distributions for the LHC". In: *Eur. Phys. J. C* 63 (2009), p. 189. DOI: 10.1140/epjc/s10052-009-1072-5. arXiv: 0901.0002 [hep-ph].

[147] R. Gavin, Y. Li, F. Petriello, and S. Quackenbush. "W Physics at the LHC with FEWZ 2.1". In: *Comput.Phys.Commun.* 184 (2013), pp. 208–214. DOI: 10.1016/j.cpc.2012.09.005. arXiv: 1201.5896 [hep-ph].

[148] M. Lisovyi, A. Verbytskyi, and O. Zenaiev. "Combined analysis of charm-quark fragmentation-fraction measurements". In: *The European Physical Journal C* 76.7 (2016). DOI: 10.1140/epjc/s10052-016-4246-y. arXiv: 1509.01061 [hep-ex].

[149] P. Zyla et al. "Review of Particle Physics". In: *Prog. Theor. Exp. Phys.* (2020), p. 083C01. DOI: 10.1093/ptep/ptaa104.

[150] S. Chatrchyan et al. "Measurement of the Inclusive W and Z Production Cross Sections in *pp* Collisions at $\sqrt{s} = 7$ TeV". In: *JHEP* 10 (2011), p. 132. DOI: 10.1007/JHEP10(2011)132. arXiv: 1107.4789 [hep-ex].

[151] A. Bodek, A. Van Dyne, J. Y. Han, W. Sakumoto, and A. Strelnikov. "Extracting muon momentum scale corrections for hadron collider experiments". In: *The European Physical Journal C* 72.10 (2012). DOI: 10.1140/epjc/s10052-012-2194-8. arXiv: 1208.3710 [hep-ex].

[152] C. Collaboration. "Performance of CMS Muon Reconstruction in Cosmic-Ray Events". In: *JINST* 5 (2010), T03022. DOI: 10.1088/1748-0221/5/03/T03022. arXiv: 0911.4994 [physics.ins-det].

[153] C. Collaboration. "Performance of the reconstruction and identification of high-momentum muons in proton-proton collisions at $\sqrt{s}$= 13 TeV". In: *JINST* 15 (2020), P02027. DOI: 10.1088/1748-0221/15/02/P02027. arXiv: 1912.03516 [physics.ins-det].

[154] C. Collaboration. "Measurements of production cross sections of WZ and same-sign WW boson pairs in association with two jets in proton-proton collisions at $\sqrt{s}$= 13 TeV". In: *Phys. Lett. B* 809 (2020), p. 135710. DOI: 10.1016/j.physletb.2020.135710. arXiv: 2005.01173 [hep-ex].

[155] A. Ali and F. Barreiro. "The Final States $\ell^{\mp}K^{\pm}K^{\pm}$ X in Jets as Signatures of $B(s)0$ $\bar{B}(s)0$ Mixings". In: *Z. Phys. C* 30 (1986). Ed. by M. Markytan, W. Majerotto, and J. Macnaughton, p. 635. DOI: 10.1007/BF01571814.

[156] M. Gronau, A. Nippe, and J. L. Rosner. "Method for flavor tagging in neutral B meson decays". In: *Phys. Rev. D* 47 (1993), p. 1988. DOI: 10.1103/PhysRevD.47.1988. arXiv: hep-ph/9211311 [hep-ph].

[157] R. Aaij et al. "Identification of beauty and charm quark jets at LHCb". In: *JINST* 10 (2015), P06013. DOI: 10.1088/1748-0221/10/06/P06013. arXiv: 1504.07670 [hep-ex].

[158] A. M. Sirunyan et al. "Measurement of the inelastic proton-proton cross section at $\sqrt{s}$ = 13 TeV". In: *JHEP* 07 (2018), p. 161. DOI: 10.1007/JHEP07(2018)161. arXiv: 1802.02613 [hep-ex].

[159] J. Butterworth et al. "PDF4LHC recommendations for LHC Run II". In: *Journal of Physics G: Nuclear and Particle Physics* 43 (2016), p. 023001. DOI: 10.1088/0954-3899/43/2/023001. arXiv: 1510.03865 [hep-ex].

[160] L. Lyons, D. Gibaut, and P. Clifford. "How to combine correlated estimates of a single physical quantity". In: *Nucl. Instrum. Meth. A* 270 (1988), pp. 110–117. ISSN: 0168-9002. DOI: 10.1016/0168-9002(88)90018-6. URL: https://cds.cern.ch/record/183996/files/OUNP-88-05.pdf.

[161] J. Campbell and T. Neumann. "Precision Phenomenology with MCFM". In: *JHEP* 12 (2019), p. 034. DOI: 10.1007/JHEP12(2019)034. arXiv: 1909.09117 [hep-ph].

[162] S. Bailey, T. Cridge, L. A. Harland-Lang, A. D. Martin, and R. S. Thorne. "Parton distributions from LHC, HERA, Tevatron and fixed target data: MSHT20 PDFs". In: *The European Physical Journal C* 81.4 (2021). DOI: 10.1140/epjc/s10052-021-09057-0. arXiv: 2012.04684 [hep-ph].

[163] T.-J. Hou et al. "New CTEQ global analysis of quantum chromodynamics with high-precision data from the LHC". In: *Phys. Rev. D* 103 (1 2021), p. 014013. DOI: 10.1103/PhysRevD.103.014013. arXiv: 1912.10053 [hep-ex].

[164] S. Alekhin, J. Blümlein, and S. Moch. "NLO PDFs from the ABMP16 fit". In: *Eur. Phys. J. C* 78 (2018), p. 477. DOI: 10.1140/epjc/s10052-018-5947-1. arXiv: 1803.07537 [hep-ph].

[165] R. D. Ball et al. "Parton distributions for the LHC Run II". In: *JHEP* 04 (2015), p. 040. DOI: 10.1007/JHEP04(2015)040. arXiv: 1410.8849 [hep-ph].

[166] A. Buckley et al. "LHAPDF6: parton density access in the LHC precision era". In: *Eur. Phys. J. C* 75 (2015), p. 132. DOI: 10.1140/epjc/s10052-015-3318-8. arXiv: 1412.7420 [hep-ph].

[167] ATLAS Collaboration. "Precision measurement and interpretation of inclusive W^+, W^- and Z/γ^* production cross sections with the ATLAS detector". In: *Eur. Phys. J. C* 77 (2017), p. 367. DOI: 10.1140/epjc/s10052-017-4911-9.

[168] W. T. Giele, S. Keller, and E. Laenen. "QCD corrections to W boson plus heavy quark production at the Tevatron". In: *Physics Letters B* 372.1-2 (1996), pp. 141–149. DOI: 10.1016/0370-2693(96)00078-0.

[169] W. J. Stirling and E. Vryonidou. "Charm production in association with an electroweak gauge boson at the LHC". In: *Phys. Rev. Lett.* 109 (2012), p. 082002. DOI: 10.1103/PhysRevLett.109.082002. arXiv: 1203.6781 [hep-ph].

[170] M. Czakon, A. Mitov, M. Pellen, and R. Poncelet. "NNLO QCD predictions for Wc-jet production at the LHC". In: *Journal of High Energy Physics* 2021.6 (2021). DOI: 10.1007/jhep06(2021)100.

[171] M. Czakon, A. Mitov, M. Pellen, and R. Poncelet. *A detailed investigation of W+c-jet at the LHC*. 2022. DOI: 10.48550/ARXIV.2212.00467.

[172] M. Czakon, A. Mitov, and R. Poncelet. *Infrared-safe flavoured anti-k_T jets*. 2022. DOI: 10.48550/ARXIV.2205.11879.

[173] S. Carrazza, S. Forte, Z. Kassabov, and J. Rojo. "Specialized minimal PDFs for optimized LHC calculations". In: *The European Physical Journal C* 76.4 (2016). DOI: 10.1140/epjc/s10052-016-4042-8.

[174] *Measurement of the production cross section of a W boson in association with a charm quark in proton-proton collisions at $\sqrt{s} = 13$ TeV*. Tech. rep. Geneva: CERN, 2022. URL: http://cds.cern.ch/record/2815162.

[175] P. Gras. *Vector boson associated with jets in CMS*. 2022. URL: https://agenda.infn.it/event/28874/contributions/171112/.