UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS FÍSICAS

Departamento de Física Atómica, Molecular y Nuclear

TESIS DOCTORAL

Measurement of associated Z+charm production and search for W' bosons in the CMS experimented at the LHC

Medida de la producción asociada de Z+charm y búsqueda de bosones W' en el experimento CMS del LHC

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Alberto Escalante del Valle

Director

Juan Alcaraz Maestre

Madrid, 2017

©Alberto Escalante del Valle, 2017

CENTRO DE INVESTIGACIONES ENERGÉTICAS
MEDIOAMBIENTALES Y TECNOLÓGICAS

Measurement of associated Z+charm production and Search for W′ bosons in the CMS experiment at the LHC

by

Alberto Escalante del Valle

A thesis submitted in partial fulfillment for the
degree of Doctor of Philosophy

in the

Universidad Complutense de Madrid
Facultad de Ciencias Físicas
Departamento de Física Atómica, Molecular y Nuclear

Supervised by:
Dr. Juan Alcaraz Maestre
Dr. Juan Pablo Fernández Ramos

Madrid
February 2017

CENTRO DE INVESTIGACIONES ENERGÉTICAS
MEDIOAMBIENTALES Y TECNOLÓGICAS

Medida de la producción asociada de Z+charm y Búsqueda de bosones W′ en el experimento CMS del LHC

por

Alberto Escalante del Valle

Memoria de la tesis presentada para optar al
grado de Doctor en Filosofía

en la

Universidad Complutense de Madrid
Facultad de Ciencias Físicas
Departamento de Física Atómica, Molecular y Nuclear

Supervisado por:
Dr. Juan Alcaraz Maestre
Dr. Juan Pablo Fernández Ramos

Madrid
Febrero 2017

Abstract

Measurement of associated Z+charm production and Search for W' bosons in the CMS experiment at the LHC

by Alberto Escalante del Valle

Do we understand how elementary particles interact with each other? Are we able to predict the result of the collisions of these elementary particles at the LHC? The objective of this thesis is to investigate the validity of our current theoretical model, the Standard Model of particle physics, to explain the production of two low rate processes in proton-proton collisions at the LHC.

The first half of the thesis studies the potential production of new types of interactions, mediated by new W' bosons, present in many extensions of the Standard Model. This is the first search of this kind at a centre of mass energy of 13 TeV in CMS and uses a dataset of 2.3 fb^{-1} to search for an excess in the production of events with a muon and large missing transverse energy. As a result of this study, the analysed data is found to be in agreement with predictions of the Standard Model and disfavours the production of new $\mathrm{W}' \to \mu\nu$ processes. This analysis alone is not able to exclude the production of all kind of W' interactions. However, it puts strong constrains to its potential cross section and decay to a muon and a neutrino, $\sigma_{\mathrm{W}'}\mathcal{B}(\mathrm{W}' \to \mu\nu) < 3$ fb^{-1} for W' masses larger than 2 TeV. In the context of the Sequential Standard Model model, this result excludes $\mathrm{W}'_{\mathrm{SSM}}$ bosons with masses smaller than 3.9 TeV.

The second part of the manuscript presents the first measurement in CMS of the associated production of Z bosons with charm quarks, $\mathrm{Z} + \mathrm{c}$, using 19.7 fb^{-1} of data collected at 8 TeV. This is a rare process in the Standard Model and its cross section in the studied fiducial region is measured to be $\sigma_{fid}(\mathrm{Z} + \mathrm{c}) = 8.6 \pm 0.5$ (stat) ± 0.7 (syst) pb. Additionally its relative production to that of bottom quarks, $\mathrm{Z} + \mathrm{b}$, is measured to be $\sigma(\mathrm{Z} + \mathrm{c})/\sigma(\mathrm{Z} + \mathrm{b}) = 2.0 \pm 0.2$ (stat) ± 0.2 (syst). Both measurements are also determined differentially as a function of transverse momentum of the Z boson and of the heavy flavour jet. Finally these measurements are compared with different theoretical predictions.

Acknowledgements

¿Cuánto tiempo se tarda en escribir una tesis doctoral? En mi caso ∼3650 días. Este documento cierra mi etapa de formación como físico que empecé hace aproximadamente 10 años cuando me matriculé en Ciencias Físicas en la UCM con 18 años. Y claro, tantos años dan para muchas experiencias y para estar en deuda con mucha gente.

En primer lugar me gustaría dar las gracias a todas aquellas personas que me dieron una oportunidad sin conocerme de nada, a todas aquellas personas que tuvieron abierta la puerta de su despacho para escucharme y me regalaron un minuto de su tiempo, y sobre todo a aquellas personas que fueron capaces de motivarme y sacar lo mejor de mí mismo. Todos ell@s saben quienes son, y por ello se merecen estar en el primer párrafo. Si hoy estoy aquí escribiendo estas líneas es gracias a todos vosotros. Muchas gracias.

Me gustaría agradecer a Juan Pablo su infatigable apoyo siempre que se lo he pedido. Por poner un ejemplo, poca gente sabe que Juan Pablo se sentó durante casi 6 meses 8 horas diarias a mi lado y me enseñó a ser un físico de partículas en tiempo récord. Durante este tiempo fue capaz de motivarme, sacar lo mejor de mi mismo y dotarme de iniciativa para no tener miedo a proponer cosas nuevas. Gracias por enseñarme a volar.

Juan Pablo y yo tenemos algo en común, somos muy impulsivos y muchas veces necesitamos alguien que nos guie y nos ayude a a centrarnos cuando nos aceleramos. Por ello, he tenido una suerte increíble de poder trabajar con Isabel Josa Mutuberría. Isabel, al margen de ser una excelente investigadora, ha tenido la paciencia de escucharme, hacerme siempre las preguntas correctas y de esta forma organizar los conceptos en mi mente. Al mismo tiempo, Isabel ha demostrado una fuerza mental sobrenatural para poder sacar adelante Z+c cuando más complicado parecía. Muchas gracias por todo.

Muchas veces en una empresa se dice que si alguien se va, la empresa colapsa. Esto es cierto y en W' esa persona es Begoña de la Cruz. Begoña fue capaz de coordinar W' en tiempos de mucha presión y carga de trabajo. Al mismo tiempo Begoña me enseñó a ser ambicioso y me animó a estudiar la resolución y la escala de los muones de alto p_T. Finalmente Begoña se aseguró que W' y yo tuviésemos un peso y una visibilidad mayor de la que nos correspondería por el tamaño del grupo. No todos los héroes llevan capa.

Una tesis es una maratón y hay veces que el mayor enemigo es uno mismo. Óscar González López, al margen de ser el físico más polémico que conozco, me ha demostrado ser un apoyo continuo tanto personal como profesional cuando más lo necesitaba. Al mismo tiempo darle unas gracias enormes por haberse leído la tesis. Para los futuros estudiantes de doctorado del CIEMAT, no dejen de hablar y aprender con Óscar ("el tío del CERN"), no se arrepentirán.

Finalmente voy a cerrar este primer grupo de agradecimientos con una persona con la cual da verdadero placer hablar de física, Juan Alcaraz. Lo que más me gusta de Juan es su versatilidad para explicar un concepto y recalcar su importancia en un análisis concreto. Muchas de sus explicaciones me han hecho abrir los ojos y han dado lugar a nuevas ideas. Me gustaría agradecerle el apoyo que me dió en cada proyecto que empecé, por dejarme equivocarme para aprender de ello y sobre todo dejarme ser investigador durante estos últimos cuatro años. Gracias.

En un segundo lugar me gustaría agradecer a los CMSeros que me han acompañado en este viaje: Cruz, Mara, Jesús, Chema, Pablo, Cristina, Ignacio, Silvia, Nica, Marcos, Conchita y Marta. Gracias. In this part I would like to thank to the people at CERN that helped me during this thesis, namely: Ivan, Slava, Daniele, Petra, Caroline, Kostas, Simon, Carlo, María, Mauro, Luca, Shahram, Roberto, Alice... and especially to the $Z' \to \mu\mu$ and $W' \to e\nu$ analysis teams. El primer párrafo va dedicado a todos vosotros.

El siguiente párrafo va dedicado a mis compañeros de despacho/aventuras en el CIEMAT. Eduino un gustazo haber debatido largo y tendido de física contigo. Me llevo un amigo. Marisma, te dejo como mamá de CMS, cuida de los nuevos y demuestra lo que vales (¡muchísimo!). Javismo, simplemente: crack. Daniel D., un gustazo conversar contigo y compartir experiencias. Mariano, casi tan buen portero como bético. Gran Manu, tu grandeza esta al nivel de tu altura (¡taponazo incluido!). Manuelinho, (hostias al teclado=ON) *gran persona y mejor cantante* (hostias al teclado=OFF). Diana, por tu culpa perdí un calcetín, no te lo perdonaré jamás. Miguel Ángel y Adrián Q., os mando todas las fuerzas que me sobran, ¡ánimo para la recta final de la tesis!. Irene, juventud y energía, ¡lo tienes todo para triunfar!. A los más veteranos, Bruno, Rafa, José, Jorge y Don Nacho, un placer haber compartido batallitas con ustedes. Bárbara, mi "salvadora" en Benasque y a la par un tesoro de persona. Juanjo y Jaime, maestros de la informática. Don Antonio, ahora en versión doctor y padrazo, enhorabuena. Finalmente Carlos, todo que intentas le pones un empeño envidiable: investigación, deporte, debatir, salir de fiesta... Seguiré tu ejemplo en el futuro. ¡Gracias a todos por acompañarme en esta aventura!

¿Qué hay que hacer para completar una tesis doctoral sin volverse loco? En mi caso he recargado las pilas gracias a mis compañeros de equipo. Quiero aprovechar este último párrafo para agradecer a todas las personas con las que he compartido vestuario durante todos estos años: ESA Trainees*, Cross Físicas UED*, Físicas (Sala, Siete, Once y Rugby), Gran Potra, Olímpicos FC, Melvilles*, Hors Categorie, Solteros FC**, Saint-Genis-Ferney-Crozet, La Cañada FC, Sons of Pitches*, CERN XV**... All MVPs.

Y en último lugar agradecer a mi familia su apoyo durante todos estos años, especialmente cuando he estado más lejos física y mentalmente. Soy un afortunado, GRACIAS.

Contents

List of Figures

List of Tables

Abbreviations

BSM	**B**eyond **S**tandard **M**odel
CL	**C**onfidence **L**evel
CMS	**C**ompact **M**uon **S**olenoid
CRAFT	**C**osmic **R**un **A**t **F**our **T**esla
CRUZET	**C**osmic **RU**n **A**t **Z**ero **T**esla
CSC	**C**athode **S**trip **C**hamber
DIS	**D**eep **I**nelastic **S**cattering
DT	**D**rift **T**ube
DYT	**DY**namical **T**runcation Fit
EB	**E**lectromagnetic calorimeter **B**arrel
ECAL	**E**lectromagnetic **CAL**orimeter
EE	**E**lectromagnetic calorimeter **E**ndcap
EW	**E**lectro**W**eak
GSF	**G**aussian **S**um **F**ilter
GUT	**G**rand **U**nified **T**heory
HB	**H**adronic calorimeter **B**arrel
HCAL	**H**adronic **CAL**orimeter
HE	**H**adronic calorimeter **E**ndcap
HF	**H**eavy **F**lavour
HLT	**H**igh **L**evel **T**rigger
HO	**H**adronic calorimeter **O**uter
IP	**I**mpact **P**arameter
IVF	**I**nclusive **V**ertex **F**inder
JEC	**J**et **E**nergy **C**orrections
JER	**J**et **E**nergy **R**esolution

JES	**J**et **E**nergy **S**cale
L1	**L**evel **O**ne trigger
LEP	**L**arge **E**lectron-**P**ositron collider
LHC	**L**arge **H**adron **C**ollider
LO	**L**eading **O**rder
LR	**L**eft **R**ight
MC	**M**onte **C**arlo
MET	**M**issing **T**ransverse **E**nergy
MSTW2008NNLO	**M**artin-**S**tirling-**T**horne-**W**att
NLO	**N**ext to **L**eading **O**rder
NNLO	**N**ext to **N**ext to **L**eading **O**rder
PDF	**P**arton **D**istribution **F**unction
PDG	**P**article **D**ata **G**roup
PS	**P**roton **S**ynchrotron
PU	**P**ile**U**p
PV	**P**rimary **V**ertex
QCD	**Q**uantum **C**hromo**D**ynamics
QED	**Q**uantum **E**lectro**D**ynamics
QFT	**Q**uantum **F**ield **T**heory
RPC	**R**esistive **P**late **C**hamber
SC	**S**uper **C**luster
SM	**S**tandard **M**odel
SPS	**S**uper **P**roton **S**ynchrotron
SSB	**S**pontaneous **S**ymmetry **B**reaking
SSM	**S**equential **S**tandard **M**odel
SV	**S**econdary **V**ertex
TEC	**T**racker **E**nd**C**ap
TIB	**T**racker **I**nner **B**arrel
TID	**T**racker **I**nner **D**isc
TOB	**T**racker **O**uter **B**arrel
TPFMS	**T**racker **P**lus **F**irst **M**uon **S**tation

A mis abuelos, la tía y Margarita ...

Introduction

Who ordered that? said the physicist Isidor Isaac Rabi after he learned that the particle that we call today the muon was simply an obese version of the electron. In 1936 S. Neddermeyer and E.C. Stenvenson [1, 2], while studying cosmic rays in the atmosphere, obtained the first experimental evidence of a new charged particle with a mass between the electron and the proton that they called at the time *mesotron*[1]. One year earlier, H. Yukawa had published his meson theory of strong interactions that predicted the existence of a new light charged state responsible of the nuclear force with approximately 200 times the electron mass [3]. After the discovery of the *mesotron*, physicists were not surprised and during many years believed that Yukawa's particle had been observed. It was not until 1947 that posterior measurements of the *mesotron* interaction with matter showed that it does not interact as Yukawa predicted. At the same time, Yukawa's prediction was confirmed experimentally with the discovery by C. Powell of a new charged particle that interacts strongly called pion. After a short crisis, physicists realized that the *mesotron* was nothing else than a heavy cousin of the electron, and the first particle of a second generation of new particles yet to be discovered! The discovery of the muon was completely unexpected and leptons were introduced as a new family of particles that do not interact strongly. Today we know that there is also an obese version of the muon, the tau lepton. However physicists still wonder why we have exactly two copies of the first generation... I. I. Rabi question is still not fully resolved.

The discovery of the muon is an example that shows how experiments allow physicists to travel to the frontier of knowledge, and test whether our current understanding of nature can explain the outcome of the experimental observations. In some rare cases, the result of the experiments is so unexpected (*Who ordered that?*), that we need to refine or rebuild our view of the universe. A wonderful thing is that, in either case, no matter what the outcome of the experiment is, we will always learn something new about the nature and the world we live in.

[1] from the Greek *mes* intermediate

Readers of this thesis will travel to the frontier of knowledge thanks to the Large Hadron Collider (LHC) that collides protons close to the speed of light, at an energy only achieved 10^{-12} seconds after the birth of the universe. The results of such collisions are recorded in the Compact Muon Solenoid Experiment (CMS), offering to physicists the possibility of observing the first instants of the universe and probe the truthfulness of many theories. A short introduction to the LHC and the CMS experiment is given in Chapter 1.

Nowadays, our current understanding of the universe is the culmination of a extraordinary 20^{th} century of progress in particle physics, both in theory and experiment. As a result, the Standard Model of Particle Physics (SM) (1960-1970) [4–6] was developed providing a theoretical framework based on a local gauge symmetry principle. This theory is able to describe three of the four fundamental interactions in nature: strong, weak and electromagnetism via the exchange of spin-1 gauge bosons fields: gluons, $\mathrm{W}^{\pm}$, Z and the photon respectively. In addition, the Higgs mechanism (1964-1967) [7–12] is responsible of giving mass to the the weak vector bosons and give rise to the appearance of a massive spin-0 state, the Higgs boson.

All visible matter content in the universe is described by fermion spin-1/2 constituents grouped into quarks and leptons. Up to date, three generation of fermions have been discovered experimentally. Each generation is made of two quarks, a negatively charged lepton (e , μ, τ) and a corresponding neutral neutrino (ν_e, ν_μ, ν_τ). One of the two quarks is an up-type quark (u, c, t) with charge +2/3 and the other one is a down-type (d, s, b) with charge -1/3. The first generation is responsible for everyday matter whereas the second and third are replicas of the first generation with higher masses. A theoretical introduction to the Standard Model of Particle Physics and the Higgs mechanism is provided in Appendix A.

Since its original formulation, the main experimental evidences that verified the predictions of the Standard Model were the discovery of the $\mathrm{W}^{\pm}$ and Z bosons (1983) [13, 14] with masses as predicted by the SM, the top quark (1995) [15, 16], the tau neutrino (2000) [17], the Higgs boson (2012) [18, 19] and a variety of experimental results in different particle accelerators (LEP, Tevatron, LHC ...) in the past 50 years.

This thesis presents two analyses that further test where the SM is able to predict the results of the collisions at the LHC. The W' analysis presents a search for a new process not described in the SM and the $\mathrm{Z}+\mathrm{c}$ measurement tests a prediction of the strong sector of the theory.

Search for new interactions: W' bosons

The first analysis covered in this thesis challenges the local gauge symmetry group of the Standard Model by searching for potential new symmetries that could have played an important role in the early universe. For example, Grand Unified Theories (GUT) predict the existence of a deeper underlying symmetry in nature and try to unify the spectrum of particles and interactions observed at the weak scale. This is achieved by enlarging the symmetry group of the Standard Model. As consequence, if Nature follows the predictions of GUT theories and they provide a better description of particle interactions, new vector bosons denoted as W' and Z' should exist and potentially be produced at particle accelerators.

The analysis presents a search for extremely rare events coming from potential new charged heavy vector bosons with a mass in the range 1-5 TeV, produced in the first 2.3 fb^{-1} ever recorded in CMS at a centre of mass energy of 13 TeV. Particularly, it focuses in the second lepton generation by searching for events consistent with the production of a very high momentum muon plus a neutrino from a $W' \to \mu\nu$ decay. Experimentally this is an extremely clean signature in which only few events are needed for a discovery. Due to this signal topology, most of the studies are dedicated to understand the performance of high momentum muons. The W' search analysis is described in Chapter 2 and was made public in December 2015 [20] and published in 2016 [21].

Standard model against associated $Z + c$ and $Z + b$ production

The second analysis presented in this thesis is a precise measurement of a rare process predicted by the Standard Model: the production of a Z boson in association with heavy flavour quarks. Particularly it describes the rate in which the Z boson is produced with at least one charm quark and its relative production to bottom quarks.

This measurement will challenge the predictions of the strong sector of the SM, as well as the modelling of the charm quark content inside the proton. Additionally, this analysis serves as benchmark for many searches of new physics that have to deal with associated production of Z bosons plus heavy flavour jets as main background.

Experimentally, the analysis selects leptonic decays of Z bosons into dimuon or dielectron pairs associated with heavy flavour jets. The latter will be identified by the presence of a displaced vertex consistent with the decay of a heavy flavoured hadron in one of the following three independent channels:

- A displaced muon from a semileptonic decay of B or D hadron.
- Reconstruction of a $D^{\pm} \to K^{\mp}\pi^{\pm}\pi^{\mp}$ decay vertex.
- Full reconstruction of a $D^{*-}(2010) \to D^0\pi_s^-(D^{*+}(2010) \to \bar{D}^0\pi_s^+)$ decay chain.

This measurement is experimentally challenging and needs a very well calibrated and understood dataset with large integrated luminosity. For this reason, the analysis uses 19.7 fb^{-1} of data collected at 8 TeV and is the first time that this analysis is done in CMS. This analysis is described in Chapter 3 and originally was made public in Summer 2016 [22].

Interplay between both analysis

This manuscripts describes two analysis at different centre of mass energies targeting different physics objectives. The $\text{W}' \to \mu\nu$ analysis is a search for new physics with leptons at 13 TeV, whereas the $\text{Z} + \text{c}$ analysis is a precise measurement of a SM process at 8 TeV that additionally involves jets and heavy flavour tagging.

The rationale behind the order in which analysis are presented in the manuscript is purely pedagogical. Indeed, common concepts such as physics objects, analysis techniques or calibrations are introduced in the $\text{W}' \to \mu\nu$ analysis in Chapter 2, which is simpler in terms of physics objects. In order to avoid text repetition, those concepts are not described again in Chapter 3 for the $\text{Z} + \text{c}$ analysis. Instead, the latter focuses on the complicated and particular aspects of a precision measurement.

Chapter 1

LHC and the CMS experiment

1.1 The accelerator: Large Hadron Collider

The Large Hadron Collider [23] is a particle physics accelerator located near Geneva at the Swiss-French border with a diameter of 27 km and between 45 m to 170 m underground. The LHC is a versatile machine that depending on the physics program of the experiments provides three different type of collisions: proton-proton, lead-proton and lead-lead. There are four main different experiments placed at the collision points of the LHC: CMS, ATLAS, ALICE and LHCb. The data analysed in this manuscript corresponds only to proton-proton collisions recorded in CMS at centre of mass energies of $\sqrt{s} =$ 8 TeV and 13 TeV.

Protons are confined inside the vacuum tube of the accelerator thanks to 1232 dipole superconducting magnets made of NbTi each of them with length 14.3 m. They operate at a temperature of 1.9 K and provide a magnetic field strength of 8.4 Tesla. Additionally there are 392 focusing quadrupole magnets and 3700 multipole magnets. At the LHC, protons are accelerated up to 6.5 TeV thanks to 16 radio frequency cavities (8 per beam) that operate at 400 MHz each of them delivering 2 MV.

The initial energy of the protons when they are injected at the LHC is 450 GeV achieved thanks to the LHC pre-accelerator chain. The latter is shown in Figure 1.1, and it consists of a set of linear and circular particle accelerators that at every step increase the beam energy by a significant amount, ΔE. The first linear accelerator, LINAC2 ($\Delta E = 50$ MeV) is followed by three circular accelerators: BOOSTER (157 m diameter, $\Delta E = 50$ MeV $\rightarrow$ 1.4 GeV), Proton Synchrotron (628 m diameter, $\Delta E = 1.4$ GeV $\rightarrow$ 28 GeV) and the Super Proton Synchrotron (7 km diameter, $\Delta E = 28$ GeV $\rightarrow$ 450 GeV).

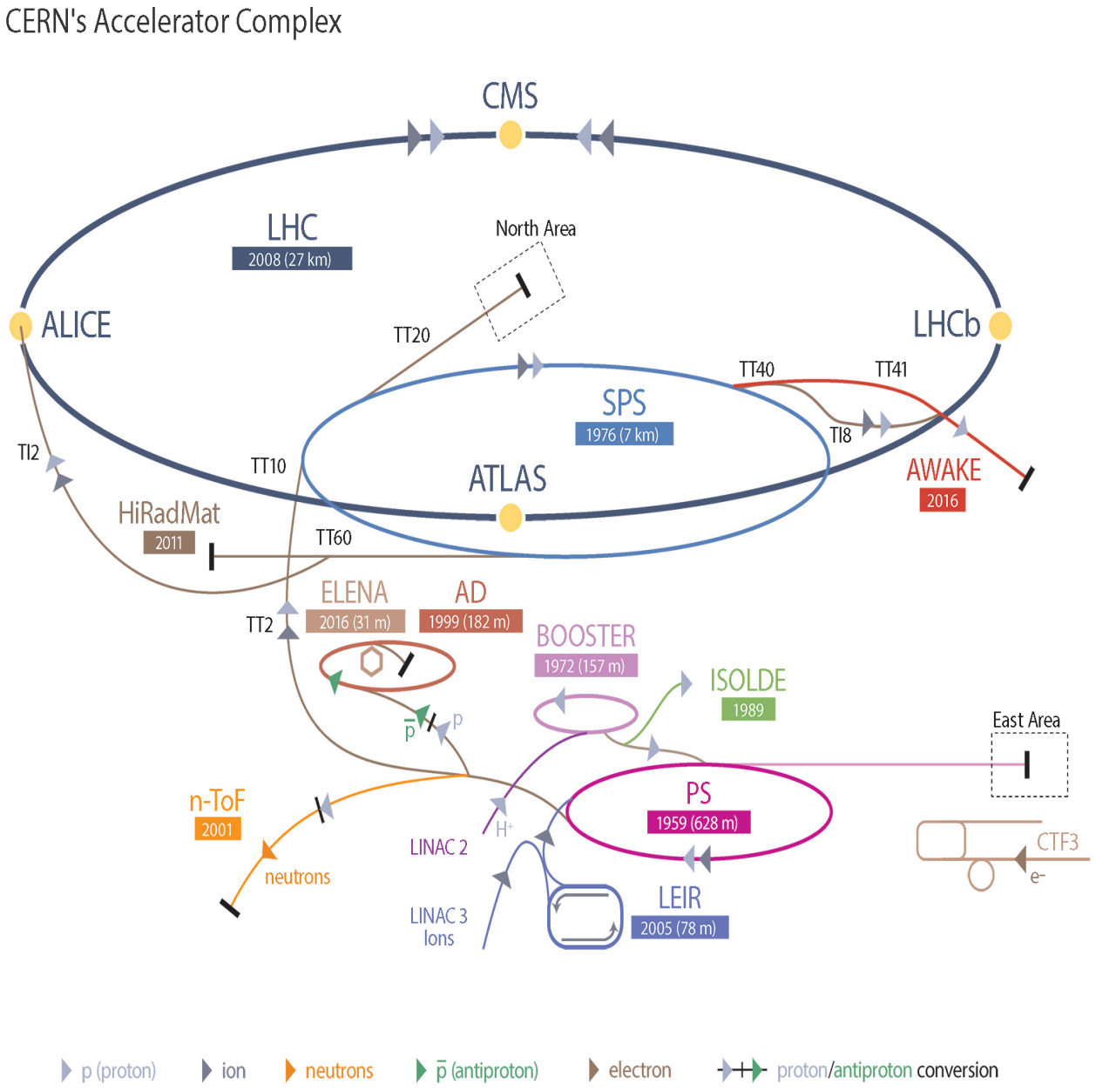

FIGURE 1.1: Sketch of the LHC pre-accelerator chain. Protons start their travel to the LHC at the linear accelerator LINAC2 and then follow three circular accelerators: BOOSTER, Proton Synchrotron (PS) and Super Proton Synchrotron (SPS) prior their injection to the LHC. The energy of the protons when they are injected at the LHC is 450 GeV and they are accelerated up an energy of 6.5 TeV before the collisions.

Protons are injected in the LHC accelerator in proton bunches. For example, during the 2015 data taking period, a good fill was made of about $n_b = 2244$ bunches, each of them made of about $N_p = 1.1 \cdot 10^{11}$ protons at the beginning of the fill. Consecutive bunches are separated by at least 7.5 m, leading to collisions every 25 ns. A maximum of $n_b = 2808$ bunches is envisage in the future. The size of the bunches was about $\sigma_z = 9$ cm along the beam direction and $\sigma_x \times \sigma_y = 20 \times 20\ \mu m^2$ in the transverse plane. The latter could be optimized and reduced in the future to maximize the amount of collisions delivered to the experiments.

The speed of the protons inside the 27 km tunnel is close to the speed of light, $3 \cdot 10^8\ m/s$, and turn about $f_{rev} = 11245.5$ laps every second. Since protons are injected inside the LHC from the SPS it takes about few minutes to several hours, depending on the conditions, to reach stable beams at the collision energy of 6.5 TeV. This energy corresponds to a relativistic Lorentz factor of $\gamma_r = 6930$, and leads to collisions in the interaction point of CMS at centre of mass energy of 13 TeV. The lifetime of the beams is typically several hours, however the intensity of the beams degrades with the time as the protons are consumed.

A extended version of the LHC parameters described in this section, and their typical values during the 2015 run are summarized Table 1.1. During the data taking campaign the beam parameters are adjusted to maximize the amount of collisions delivered to the experiments.

Parameter	Typical range in 2015	Parameter description
n_b	2244	Number of proton bunches
N_p	$1.1 \cdot 10^{11}$	Number of protons per bunch
f_{rev}	11245.5 s^{-1}	Bunch revolution frequency
γ_r	6930	Relativistic Lorentz factor
ϵ_n	3.4 μm	Normalized beam emittance
β^*	0.8 m	β^* function at collision point
$\sigma_{x/y} = \sqrt{\epsilon_n \beta^* / \gamma_r}$	19.81 μm	Transverse beam width at interaction point
σ_z	9.375 cm	Longitudinal bunch length
F	0.82	Geometrical correction
$\mathcal{L}$	$5 \cdot 10^{33}\ cm^{-2} s^{-1}$	Instantaneous luminosity (more in Equation 1.2)
μ	16	Average number of interactions per bunch crossing

TABLE 1.1: LHC beam parameters and their typical range of operation in 2015 at 13 TeV. Further explanations about them are documented in [23]. Typical values for the 2015 run are taken from [24].

1.2 Introduction to LHC physics

The LHC is responsible to collide proton beams and produce the particles that physicists detect and study at experiments such as CMS. In searches for new particles it is important to produce the maximum number of candidates and maximize the probability to detect them. The total number of processes of a certain type, N, produced in a certain period of time, dt, in a particle accelerator is given by

$$N = \sigma \int \mathcal{L} dt \tag{1.1}$$

where σ is called the cross section[1] and corresponds to the effective area for a given process to occur. $\mathcal{L}$ is called instantaneous luminosity and only depends on the parameters of the LHC (see Table 1.1 for a description of the parameters)

$$\mathcal{L} = \frac{N_p n_b f_{rev}}{4\pi} \frac{N_p}{\epsilon_n} \frac{\gamma_r}{\beta^*} F. \tag{1.2}$$

For a given process at a hadron collider, the cross section grows with the centre-of-mass energy of the accelerator and depends on the production mechanism that initiated the interaction proceess: gluon-gluon, quark-antiquark, quark-gluon. Processes such as potential $\mathrm{W}' \to \mu\nu$ production at the TeV scale, if they exist, are rare and their production cross section is smaller than a few femtobarns (fb) at 13 TeV according to previous searches [25, 26]. On the other hand, precision cross section measurements such as associated $\mathrm{Z} + \mathrm{c}$ production, are allowed in the Standard Model and its cross section at 8 TeV is expected to be around a few picobarn (pb) according to simulations. As an example, Figure 1.2 shows the impressive level of agreement between the measured cross sections at CMS for relevant processes at the LHC at three different centre-of-mass energies: $\sqrt{s} = 7, 8, 13$ TeV with the corresponding theoretical predictions.

Cross sections for the different processes at a given centre of mass energy of the accelerator are constant and cannot be modified. In other words, they cannot be adjusted to maximize their production rate during the data taking campaign. In contrast, the instantaneous luminosity depends on parameters of the accelerator that in some cases can be modified. Maximizing the instantaneous luminosity over long periods of time represents a technological challenge for the accelerator. For example, in 2015 at 13 TeV (2012 at 8 TeV) the instantaneous luminosity reached its maximum value at $\mathcal{L} = 5.13 \cdot 10^{33}$ cm^{-2}s^{-1} ($7.67 \cdot 10^{33}$ cm^{-2}s^{-1}). In 2016 the LHC surpassed for the first time the design instantaneous luminosity of the accelerator, 10^{34} cm^{-2}s^{-1}. Every increase in $\mathcal{L}$ and in general of the LHC performance, is great news for physicists because

[1] The cross section is usually measured in Barns, $1\mathrm{b} = 10^{-28}$ m^2

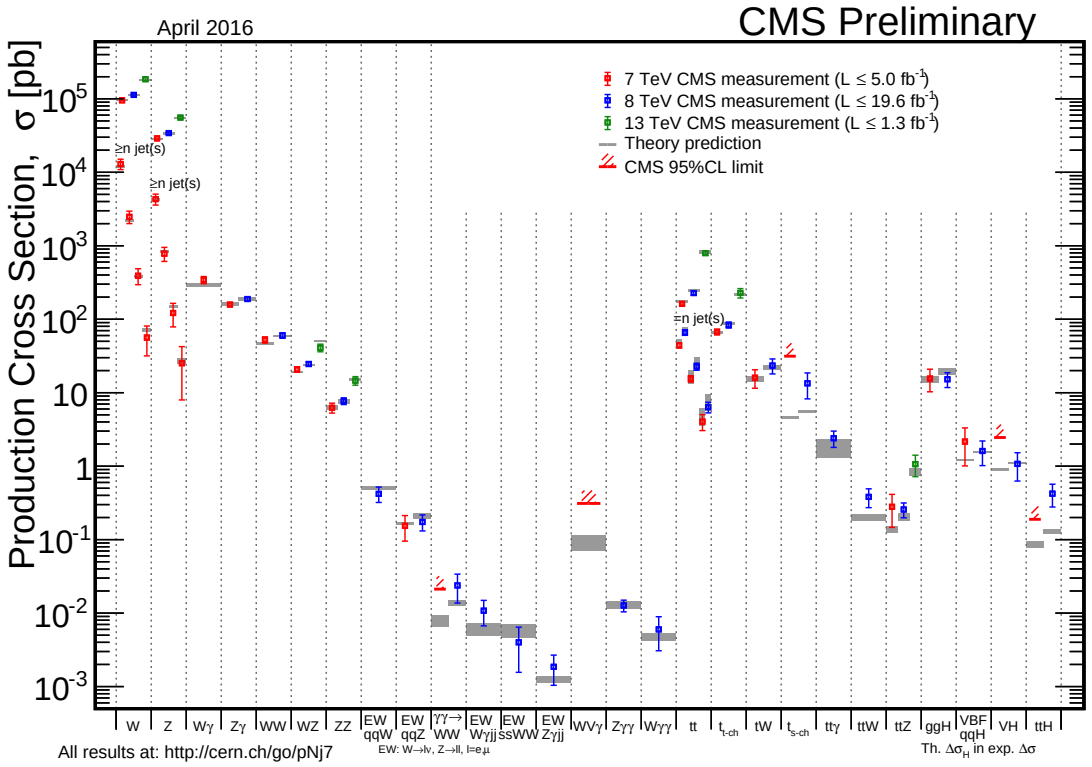

FIGURE 1.2: Measured cross sections (when available) at three different center of mass energies, compared to the predictions of the Standard Model. The figure covers well known processes such as W and Z production to processes that are more than 6 or 7 order of magnitude rarer, such $\gamma\gamma WW$. The level of agreement between the prediction with the measurements is impressive, and reinforces the predictive power of the Standard Model. This figure is sometimes called the Stairway to Heaven.

the higher the instantaneous luminosity, the higher the probability to produce processes with low cross section.

The integral of $\mathcal{L}$ over time, $L = \int \mathcal{L} dt$, is denoted as integrated luminosity or just luminosity, and is a measure of the total amount of collisions delivered by the LHC. For a detector of the size and complexity of CMS, recording all the luminosity delivered by the LHC is a complicated task that involves hundreds of people. In 2012 (2015) the LHC delivered 23.30 (4.22) fb^{-1} and the CMS detector recorded 93.5% (90.2%) of them. Prior to physics analysis, all the data used needs to pass a stringent certification procedure that ensures that only those with the highest possible quality are used. The total amount of luminosity certified for analysis at 8 TeV was about 19.7 fb^{-1} and 2.3 fb^{-1} at 13 TeV. The difference between the recorded luminosity and certified luminosity at 13 TeV in 2015 is due to a problem with the cooling of the superconducting magnet that caused that about 1 fb^{-1} was taken without magnetic field. A summary of the delivered and recorded luminosities at both centre of mass energies is shown in Figure 1.3.

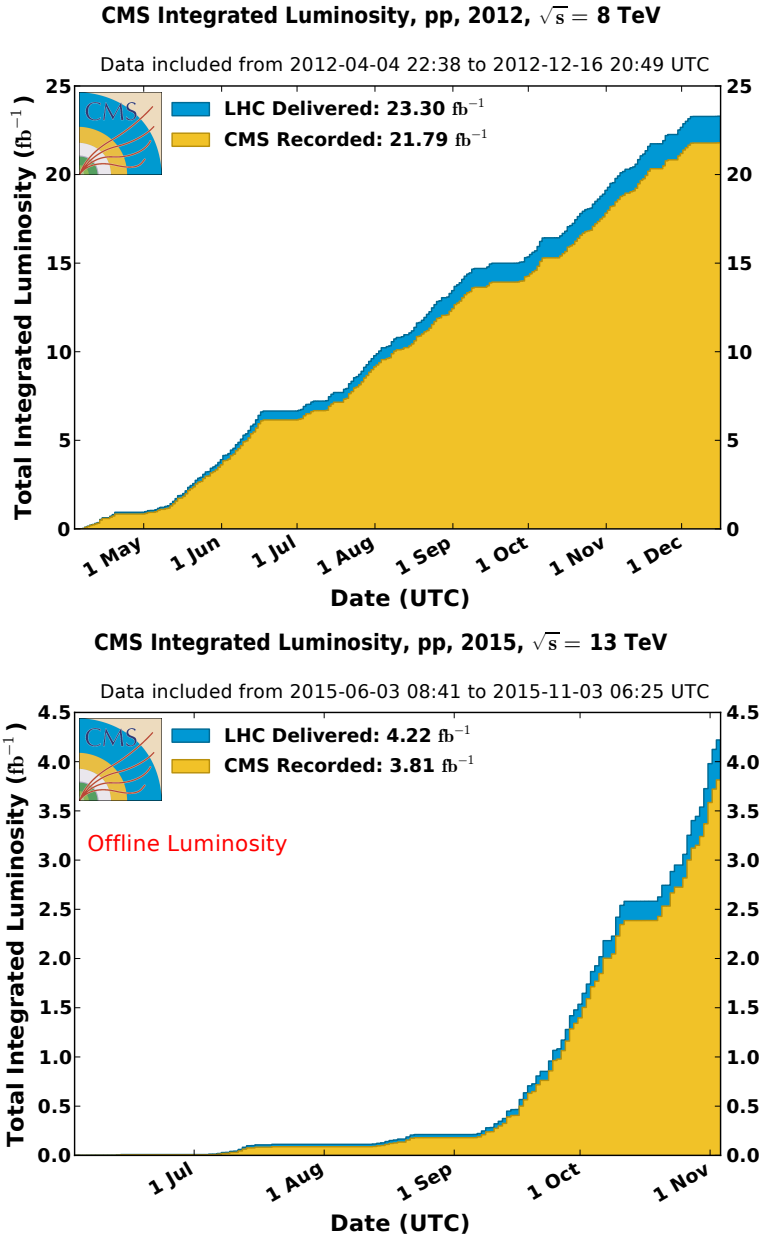

FIGURE 1.3: Integrated luminosity as a function of the month of the year corresponding to the Run I of the LHC at 8 TeV in 2012 (top) and Run II at 13 TeV in 2015 (bottom). The 2012 dataset corresponds to the Z+charm cross section measurement, whereas 2015 was used for the W′ search in the manuscript.

Pileup

From 1989 until 2000, electrons and positrons collided in a former particle accelerator called Large Electron Collider (LEP), located in the tunnel that today hosts the LHC. Collisions in a lepton collider do not suffer from the large QCD background produced from parton interactions. Additionally, the vertex of interaction, defined as the geometrical point initiating the collision, is unique and the energy deployed corresponds exactly to the centre of mass energy of the accelerator. In other words, leptonic collisions offer the simplest environment to study particle collisions. Unfortunately, in order to reach

the production of particles with highest masses and smallest cross sections, the highest centre of mass is required. The latter is technically easier to achieve in a hadron collider, therefore the LHC became the best option to explore the TeV scale in the LEP tunnel.

At the LHC, for each bunch crossing several protons may interact resulting into several parallel proton-proton collisions created in a single event. As an example, in the 2012 Run there were on average $< \mu >= 21$ proton-proton interactions per bunch crossing as it is shown in Figure 1.4. The reason of the multiple interactions per bunch crossing, also known as pileup interactions, is the proton-proton inelastic cross section, $\sigma_{inel} \approx$ 68(78) mb at 8(13) TeV, that is about twelve orders of magnitude higher than the cross section of processes relevant for the CMS physics program. The pileup rate is sensitive to the instantaneous luminosity (Equation 1.2) and raising this quantity in order to collect more data, leads necessarily to an increase of the number of proton-proton collisions per bunch crossing.

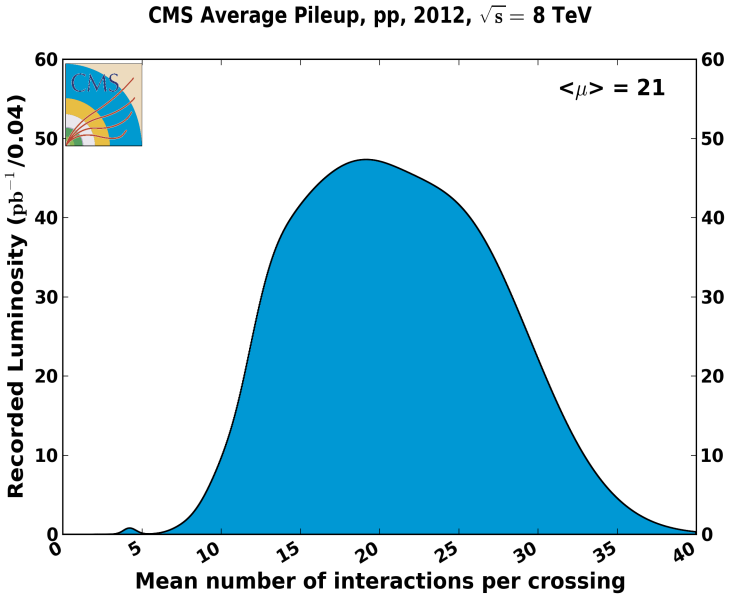

FIGURE 1.4: Mean number of interactions per bunch crossing in 2012

Unlike leptonic collisions, during hadronic collisions the initial scattering occurs between constituents that are not elementary. At first order, only one parton from each proton participates to the collision, and the energy scale of the interaction vertex of a proton-proton collision, Q^2, is defined relatively to the fraction of energy along the beam axis, x, carried by the two partons i and j that participate to the interaction: $Q^2 = \hat{s} = x_i x_j s$. In the latter formula, it is convenient to define an effective centre of mass energy of the collisions, $\hat{s}$, different from the centre of mass energy of the accelerator, s. The study of how the energy of the proton is shared among its constituents is a full area of research and an introduction is provided in the next Section 1.2.1.

1.2.1 Parton distribution functions

For each type of parton in the proton a, the Parton Density Function (PDF), $f_a(x, Q^2)$, describes the probability density function of the fraction of longitudinal momentum of the proton, x, carried by the parton at a given energy scale Q^2. They are used in a hadron collider to compute the cross section of any process, $\sigma_{ab\to X}(s, Q^2)$, using the master equation at the LHC, the factorization theorem:

$$\sigma_{ab\to X}(s, Q^2) = \sum_{ab} \int_{x_{min}}^{1} dx_1 dx_2 f_a(x_1, Q^2) f_b(x_2, Q^2) \hat{\sigma}_{ab\to X}(x_1 x_2 s, Q^2). \tag{1.3}$$

The term $\hat{\sigma}_{ab\to X}(x_1 x_2 s, M_X^2)$ is called the partonic cross section and is evaluated directly from theory using Feynman rules. Radiative corrections from initial state emissions are absorbed into PDFs and separated from the partonic cross section calculation. This procedure introduces a factorization scale in the calculation, μ_F, that is set to the energy scale that governs the process $\sigma_{ab\to X}$, $\mu_F = Q^2$.

PDFs are universal, this means that once they are determined they can be used for any process in Equation 1.3. Different theoretical constrains are applied to the PDF when they are determined. The following sum extended to all possible momentum fractions and flavours in the proton, must be satisfied

$$\sum_{a=u,d,s...} \int_0^1 dx \; x f_a(x, Q^2) = 1, \tag{1.4}$$

the proton is made of three valence quark: two up quarks and one down quark,

$$\int_0^1 dx \; (f_u(x) - f_{\bar{u}}(x)) = 2 \qquad \int_0^1 dx \; (f_d(x) - f_{\bar{d}}(x)) = 1 \tag{1.5}$$

the remaining sea quarks satisfy

$$\int_0^1 dx \; (f_s(x) - f_{\bar{s}}(x)) = 0 \qquad f_c(x) = f_{\bar{c}}(x) \qquad f_b(x) = f_{\bar{b}}(x) \tag{1.6}$$

The functional form of the x dependence in $f_a(x, Q^2)$ cannot be determined only from theory, and needs to be calculated using input from experimental measurements carried out at different energy scales. Once PDFs are determined at a given scale, they are recomputed at any other energy scale by using the DGLAP evolution equation [27].

Before the LHC and Tevatron era, the most precise measurements of the proton PDFs came from deep inelastic experiments (DIS) at HERA. For example, most of the PDF measurements were done at the DIS scale of $Q^2 \approx 10$ GeV^2 and extracted PDFs are extrapolated to the LHC physics scale $Q^2 \approx M_Z^2 \approx 10^4$ GeV^2. Different PDF groups

(NNPDF [28], CT[29], MSTW[30]...), use different methodologies and data samples to evaluate the measurements and parametrize the functional evolution of $f(x, Q^2)$ in their PDF fits. Figure 1.5 shows the PDF determination from the MSTW group using deep inelastic data (left) and its evolution to the LHC scale with with the DGLAP evolution equation (right). One of the main effects of this evolution to the LHC scale is the enhancement of gluon and sea quark PDFs at low x.

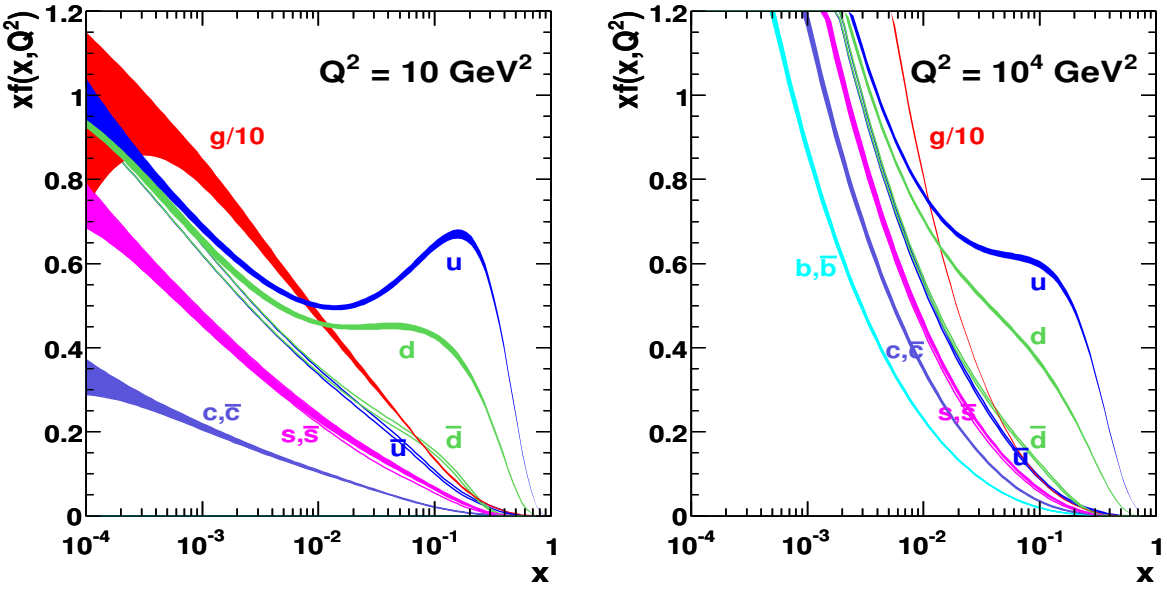

FIGURE 1.5: Parton distribution fuction for different partons at the DIS energy scale (left), $Q^2 = 10$ GeV^2 and its DGLAP evolution to the LHC scale $Q^2 = 10^4$ GeV^2. Picture taken from [31]

In the cases in which a process is dominated by a particular parton-initiated process ab, then its cross section measurement can be used to constrain the PDFs of partons a or b. Different measurements allow to constrain different PDF at different x. Latest versions of the PDFs sets from groups start to include a new generation of measurements in their PDF determination from hadron colliders such as Tevatron and LHC. For example, measurements of $t\bar{t}$ production at the LHC are used to constrain the gluon PDF whereas measurements of the $W \to l\nu$ asymmetry are used to constrain the up and down quark PDFs.

There is also a new generation of measurements that aim to study the content of the second generation quarks in the proton. The W + c cross section measurement is a unique process that probes the strange quark content. Finally, the Z + c cross section measurement and the Z + c/Z + b ratio presented in this thesis, introduces novel measurements that are sensitive to the charm and bottom PDFs. An introduction to the charm PDF is given inside the Z + c analysis in Section 3.1.1.

1.2.2 The hard interaction

The hard interaction refers to proton-proton interactions that involve the creation of a hard process characterized by a large momentum exchange Q^2. The energy scale of the hard scattering is high enough to allow the calculation of the production rates using perturbative QCD. Figure 1.6 shows a schematic example of the ingredients of a hard scattering process between two protons A and B.

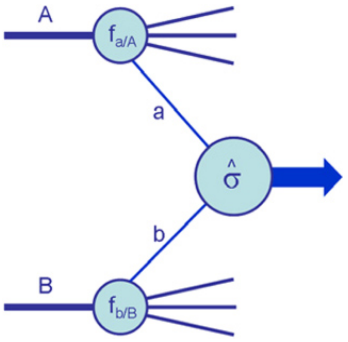

FIGURE 1.6: Example of hard scattering process. Picture taken from [31]

The hard interaction for a process $ab \to X$ is characterized by the partons a and b directly involved in the interaction, carrying a longitudinal momentum fraction x_1 and x_2, respectively. The latter is described in the proton PDFs, $f_{a/A}$ and $f_{b/B}$, as explained in Section 1.2.1.

The rate in which occurs is given by its partonic cross section, $\hat{\sigma}_{ab\to X}(x_1 x_2 s, M_X^2)$. Thanks to the Feynman formalism, for each process the partonic cross section is computed by the phase-space integral of the corresponding matrix element squared $|\mathcal{M}_{ab\to X}|^2$. The matrix element, $\mathcal{M}_{ab\to X}$, is expressed as an infinite sum of over all the possible ways to perform the process $ab \to X$.

$$\mathcal{M}_{ab\to X} = \sum_{i}^{\infty} \mathcal{F}^{(i)}_{ab\to X} \tag{1.7}$$

Each of the terms in this expansion, $\mathcal{F}^{(i)}_{ab\to X}$, corresponds to one allowed Feynman diagram that is computed using the Feynman rules. The subset of the lowest order Feynman diagrams in the expansion are called Leading Order (LO) terms and they are denoted in the text as $\mathcal{F}^{(0)}_{ab\to X}(g_s^j g'^k g^\ell)$. For a given process, their amplitude is proportional to $g_s^j g'^k g^\ell$ and they are defined as the set of diagrams with the minimum number of QCD interacting vertices j with coupling constant g_s and the minimum number of electromagnetic, ℓ, and electroweak vertices, k, with coupling constant g' and g. As an example, Figures 2.1 and 3.2 show one of the LO Feynman diagram associated to $\mathrm{W}' \to \mu\nu$ and $\mathrm{Z} + \mathrm{c}$ production respectively.

Higher order correction terms, $\mathcal{F}^{corr}_{ab\to X}$, refer to the set of allowed Feynman diagrams with higher number of interacting vertices with respect to the LO terms. Particularly, one of the most important corrections are the so called QCD corrections terms, in which the number of strong vertices, j, is increased by n. Due to the presence of n extra vertices, the effect of the QCD correction terms, is suppressed by additional powers of $\alpha_s = g_s^2/4\pi$ in the theoretical predictions. In practice, the infinite terms in Equation 1.7 are not considered and the sum is truncated:

$$\mathcal{M}_{ab\to X} = \mathcal{F}^{(0)}_{ab\to X}(g_s^j g'^k g^\ell) + \sum_{n=0}^{n'} \mathcal{F}^{corr}_{ab\to X}(g_s^{j+n} g'^k g^\ell) \tag{1.8}$$

There are several computational obstacles in the expansion used in Equation 1.8. When higher order terms $n = 1$, $n = 2$... are included in the expansion, the number of allowed Feynman diagrams used to construct the matrix elements increases factorially as well as its difficulty to evaluate them. Depending on the Feynman diagrams included in the calculation, there are different classifications for the cross section predictions:

- Leading Order (LO) prediction include terms up to $n = 0$.
- QCD Next-to-leading order calculations (QCD NLO) include up to $n = 1$
- QCD Next-to-next-to-leading order (QCD NNLO) up to $n = 2$.
- QCD Next-to-next-to-next-to leading order QCD (N^3LO) up to $n = 3$
- ....

LO calculations provide the fastest and simplest calculation. However, they are not meant to provide a precise enough prediction and usually are not compared directly with measurements. NLO computations include hard parton emissions and their accuracy is currently the minimum required to compared with data in most of the processes. A practical problem is that integrals in NLO calculations suffer from ultraviolet (UV) and soft/collinear divergences (IR). They appear when integrating high momentum additional partons or low momentum/with small angular separation additional partons. Usually dimensional regularization methods are used to remove UV divergences and introduce a re-normalization scale, μ_R, in the calculation. The re-normalization scale is set at the scale in which $\alpha_s(Q^2)$ is evaluated, and is often set to $\mu_R = \mu_F = Q^2$.

Currently there are theoretical calculations available for almost any process at QCD NLO, some at NNLO and few cases at N^3LO. In practice, there are automated (LO and NLO) matrix element software calculators that are designed to generate simulated events for LHC processes. Depending on the generator, the explicit calculation of the matrix elements is performed by the generator, implemented in their libraries from an independent

calculation, or a combination of both. Generated events in samples used in this thesis have been produced using one of the following software-packages: PYTHIA (LO)[32, 33], POWHEG (NLO)[34], MADGRAPH (LO)[35], MADGRAPH5_AMC@NLO (NLO)[35]... In all of them, their predicted cross sections are computed with an integration of the matrix element over the phase space of all the particles using a Monte Carlo (MC) sampling method. Due to this, they are often called MC event generators. In the cases in which a LO generator has been used for event generation, missing higher order QCD corrections (when available) are included via multiplicative correction factors, denoted as *k-factors*, defined as the ratio between the predicted $\sigma_{(\mathrm{N})\mathrm{NLO}}/\sigma_{\mathrm{LO}}$ from an independent calculation. As an example, electroweak processes sometimes are re-normalized to predictions from FEWZ (NNLO)[36] or in the case of diboson processes to MCFM (NLO)[37].

In the same spirit than the QCD corrections introduced in Equation 1.8, it is also possible to include in the calculation diagrams with higher number of EW vertices that involve W, Z, H or massless photons. In general the size of the NLO EW corrections is small and at the order of NNLO QCD corrections since at the LHC energy scale, $g_s^2 \approx g'$. However in some processes EW corrections could be enhanced. This occurs, for example, in electroweak processes where the invariant mass of the process Q^2, is much larger than the mass of the particles in the loop of the Feynman digram ($Q >> M_W$, M_Z). In this case, electroweak corrections are dominated by large negative Sudakov corrections of the type $g' \log^2 Q/M_W^2$ that could enhance the EW corrections at high masses, and in some cases affecting the background in searches of new particles. Due to this, NNLO QCD corrections as well as NLO EW corrections are included via dedicated *k-factors* in the evaluation of the SM W background in the W′ search.

1.2.3 Parton shower and hadronization

The transition between partons produced in the hard scattering and its evolution into physical hadrons such as kaons, pions or protons... is called parton showering and hadronization. Partons produced in the hard scattering do no propagate freely and are never observed isolated in Nature. In the other hand, hadron are colourless physical particles made of 2 or 3 valence quarks that are detected at the experiments. Figure 1.7 shows a simplified sketch the parton shower and hadronization step.

In the showering process, coloured particles formed in the hard scattering subsequently lose their energy following higher order QCD emissions. Particularly, quarks loose energy radiating gluons in processes of the type $q \to gq$ and gluons branch into lower energy partons $g \to q\bar{q}$ and $g \to gg$. This step is repeated sequentially creating a *shower* of multiple low energy and collinear partons for every parton produced in the collision.

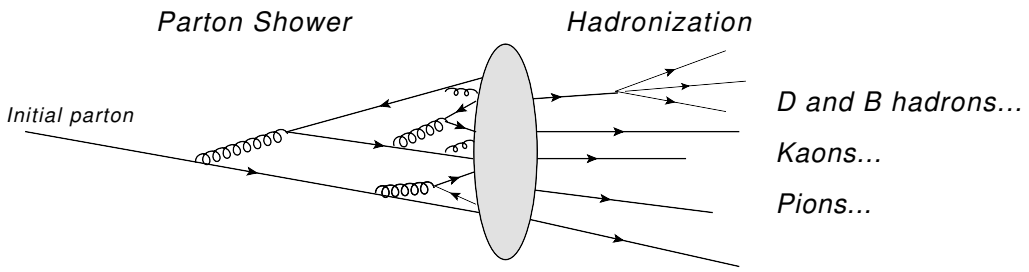

FIGURE 1.7: Simplified sketch of the parton shower process from an initial parton until its evolution into detectable physical hadrons made of 2 or 3 quarks.

This showering process finishes when the energy scale of each process in the shower approaches an energy of about $Q^2 \sim 1 - 2$ GeV^2. Below this energy scale, the shower does not continue its expansion since it is energetically more favourable that a pair of quark-antiquark is spontaneously created from the vacuum forming colourless hadrons out of the partons in the shower.

In practice, the calculations of the hard scattering and the parton shower employ different techniques and even in some cases they are carried out by different generators. The combination of both calculations is not simple since a potential double counting might occur. In order to avoid this, different merging/matching strategies have been developed to combine the hard scattering with the parton shower calculation. One particular example of matching procedure, called MLM, is used in this manuscript in samples generated with MADGRAPH for the hard scattering plus PYTHIA for the parton shower. This solution introduces a matching scale in the event to separate both calculations.

Below the QCD confinement energy, the size of the colour coupling strength is sufficiently high that the perturbative approach for calculations does not work anymore. As a consequence, the QCD hadronization is one of the least understood process in particle physics nowadays, and currently there are only phenomenological models such as the Lund String model [38] employed in PYTHIA that are used for its description.

In a similar fashion than the PDF inside the proton, the probability $D_i^h(z)$ that a parton i, hadronizes into a parton h carrying a fraction of the momentum z, cannot be computed directly from the theory and needs to be calibrated with measurements. These functions receive the name of fragmentation functions and satisfy $\sum_h \int dz D_i^h(z) = 1$. The modelling of the hadronization of charm and bottom quarks into D and B hadrons is for instance one of the dominant theoretical uncertainties in the Z + c measurement.

Underlying event

The underlying event (UE) refers to the additional activity produced in parallel with the hard scattering process. At the LHC, its contribution is mostly due to the hadronization of the reminder partons that did not participate in the hard scattering and potential multiple parton interactions (MPI). The latter represents secondary hard scattering processes between other incoming partons. Due to the theoretical complexity of describing the UE activity, current generators rely in phenomenological models and constrain their parameters using collision data. The calibration of the UE activity with data is referred as the tuning of the generator. For example, in the W′ analysis the tuning was called CMS Underlying Event Tune Pythia 8 Monash Version 1 (CUETp8M1), whereas in the Z + c measurement is called Z2.

Detector simulation and calibration factors

Generated Monte Carlo events are processed through a full detector simulation of the CMS detector based on the GEANT4 software package [39], and measurable objects are reconstructed using the official CMS software package (CMSSW). This step includes (among many others) simulation of the detector geometry and its alignment, magnetic field in each part of the detector, event trigger emulation, event reconstruction ... For a detector of the size and complexity of CMS, it is a real challenge to create simulations that are 100% accurate in describing its performance. Due to this, there are dedicated performance groups in CMS to study and minimize the small differences in performance between simulations and CMS measurements. In most of the cases, these differences are small and are corrected with multiplicative data-to-simulation calibration factors. As an example of this, Section 2.3 contains a description of some of these steps in the CMS reconstruction chain such as muon triggering and identification, and how they are calibrated in the simulation.

1.2.4 13 TeV vs 8 TeV physics

The analyses presented in this manuscript employ two datasets that amount to different integrated luminosities, recorded at different centre of mass energies. In Section 1.2, Figure 1.3 showed that the total integrated luminosity is about ten times lower at 13 TeV compared to 8 TeV. However, the relative sensitivity of the 13 TeV dataset with respect to the 8 TeV dataset, for a given process, strongly depends on the signal gain, G, defined as the ratio between the cross section at 13 TeV and 8 TeV.

Figure 1.8 shows the gain in parton luminosity ratio from the MSTW2008NLO[30] PDF group, a measure of the reach of a hadron collider that gives similar information as G.

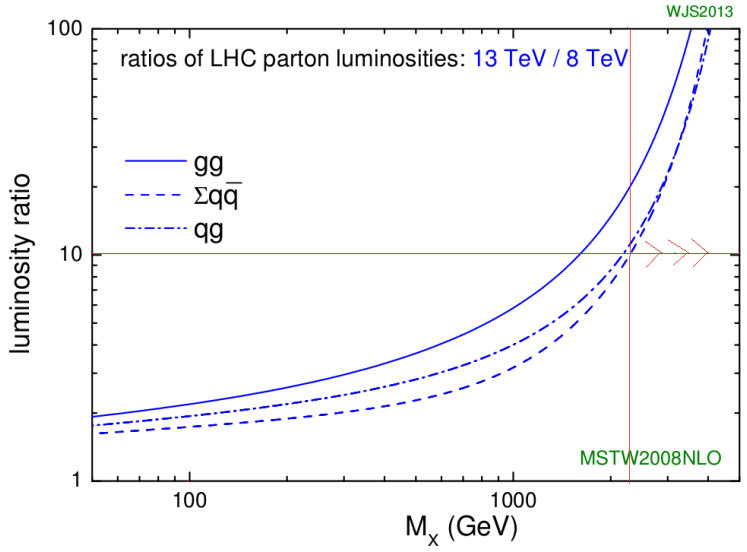

FIGURE 1.8: Ratio of parton luminosities at 13 TeV over 8 TeV from MSTW2008NLO for different production mechanisms $gg, q\bar{q}, qg$ as a function of the scale of the interaction, M_X. The ratio grows rapidly above the TeV scale and leads to a higher discovery potential at 13 TeV compared to 8 TeV for the highest W′ masses.

The several curves in the figure refer to different production mechanism, qq', qg, gg, that govern the production of a new particle X with mass M_X as a function of the scale of the interaction, $Q = M_X$. The higher the scale of the interaction, the higher the gain in signal cross section at 13 TeV. Particularly, for most of the studies in which typical scale has a luminosity ratio larger than ≈ 10 (red line), the 13 TeV dataset with only 2.3 fb^{-1} is expected to have at least the same sensitivity as the 8 TeV dataset.

In the manuscript, the Z + c cross section measurement is based on the 8 TeV dataset whereas the W′ search focuses on the 13 TeV dataset. One of the motivations for this choice is that heavy mass searches probe high M_X, and therefore profit of a large gain in signal cross section by going from 8 TeV to 13 TeV. Particularly, with only 2.3 fb^{-1} at 13 TeV, the W′ analysis is already more sensitive above $M_X \approx 2$ TeV than the previous W′ search at 8 TeV with 19.7 fb^{-1}. In the other hand, for lower M_X processes like precision measurement of Z + c production, the luminosity in the 2015 13 TeV dataset is still too low compared to 19.7 fb^{-1} at 8 TeV.

1.3 The detector: Compact Muon Solenoid

The Compact Muon Solenoid is a particle detector located in Cessy (France) at the Interacting Point 5 (P5) of the LHC. After the discovery of a Higgs boson consistent with the one predicted by the SM in 2012, the search for physics beyond the Standard Model is the new main goal in the CMS physics program. Particularly, the detection of potential new states such as W′ bosons or deviations with respect to the SM predictions in precise measurements, are more important than ever. Before entering into details of the detector or analysis techniques for the W′ and Z + c analysis presented in this thesis, it is convenient to introduce the coordinate system employed for both analyses in the document.

The CMS Cartesian coordinate system is defined with the origin at the centre of the detector. The z axis is defined along the proton beam in the way that the CMS magnetic field points towards positive $+z$. The (x, y) plane or transverse plane is defined perpendicularly to the proton beam. Quantities are often referred to this plane and denoted in the text with a $\mathrm{X_T}$ subscript. The y axis is perpendicular to the LHC plane and the x axis points to the center of the LHC. However it is convenient for most of CMS analysis to use a cylindrical coordinate system (z, ϕ, θ). In this case, the polar coordinate $0 < \theta < \pi$ is measured with respect to the z axis and the azimuthal coordinate $-\pi < \phi < \pi$ is measured in the (x, y) plane. In hadronic collisions, it is useful to define the pseudo-rapidity, η, as a re-parametrization of the polar coordinate θ,

$$\eta \equiv -\ln[\tan(\frac{\theta}{2})]. \tag{1.9}$$

The minus sign is added in the pseudo-rapidity definition to make η positive when points to the same direction as the magnetic field. As an example of this, the polar coordinate $\theta = 0, \frac{\pi}{2}, \pi$, leads to $\eta = \infty, 0, -\infty$ in terms of the pseudo-rapidity. Often in the thesis the angle, ΔR, between two objects (η_1, ϕ_1) and (η_2, ϕ_2) is used, being $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, and $\Delta\eta = \eta_1 - \eta_2$, $\Delta\phi = \phi_1 - \phi_2$.

Going back to CMS, the detector is placed 100 m underground and it has a cylindrical shape with 28.7 m in length and 14.6 m in diameter with a total weight of 14 000 tons. Along the beam line, the detector is divided into 5 wheels (-2,-1,0,+1,+2) in the barrel region that can be opened independently, plus two endcaps. Figure 1.9 shows a real picture of CMS in which one of the wheels and one of the endcaps are separated.

The innermost part of the detector is made of three concentric layers of silicon pixel plus a strip tracker detector surrounding beam spot. Their objective is to measure the momentum and trajectory of charged particles. In the next layer, a crystal electromagnetic

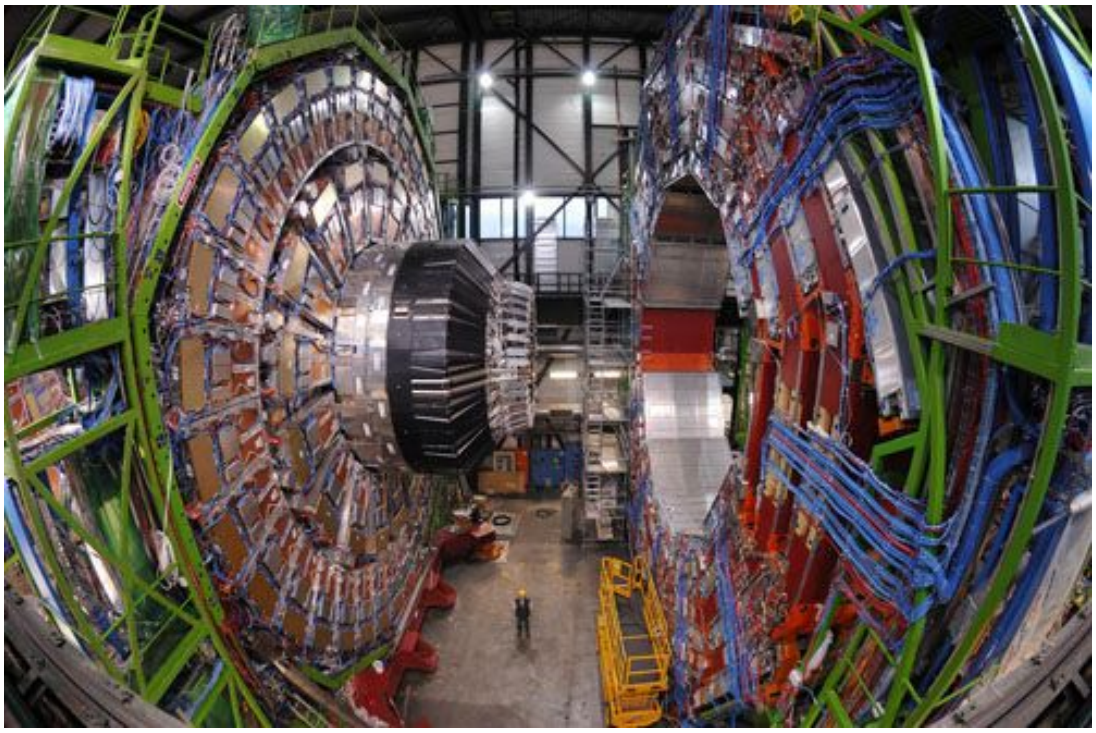

FIGURE 1.9: CMS detector in one of the latest stages of its construction. In the figure one of the wheels and one of the endcaps are opened. (Photo courtesy of CERN/Maximilien Brice).

calorimeter (ECAL) made of lead tungstate, $PbWO_4$, measures the energy of electrons and photons. Between the ECAL and the CMS magnet, a hadron calorimeter (HCAL) made of brass plates inter-laid with plastic scintillator tiles measures the energy and direction of hadronic jets produced in the hadronization of quarks and gluons. Additional calorimeters are placed in the large pseudo-rapidity range $3 < |\eta| < 5$ to guarantee a good closure and coverage of the detector. A liquid helium cooled solenoid magnet generates a magnetic field of 3.8 T along the positive $+z$ axis that bends charged particle trajectories and allows the measurement of their track momenta. Outside the magnetic field, and embedded in the steel return yoke, gas-ionization muon detectors are placed for muon trigger, identification and momentum measurement purposes.

Most of the sub-detectors in the $|\eta| < 1.2$-1.6 region are placed surrounding the beam pipe, whereas in the forward region $|\eta| > 1.2$-1.6 are oriented orthogonally with respect to the beam pipe. This configuration maximizes the hermeticity of the detector around the interaction point. Figure 1.10 shows a illustrative sketch of the different sub-detectors in CMS. Technical details about each of the sub-detectors are provided in the CMS Technical Design Report [40], and a short introduction to each of them is provided in the following sections.

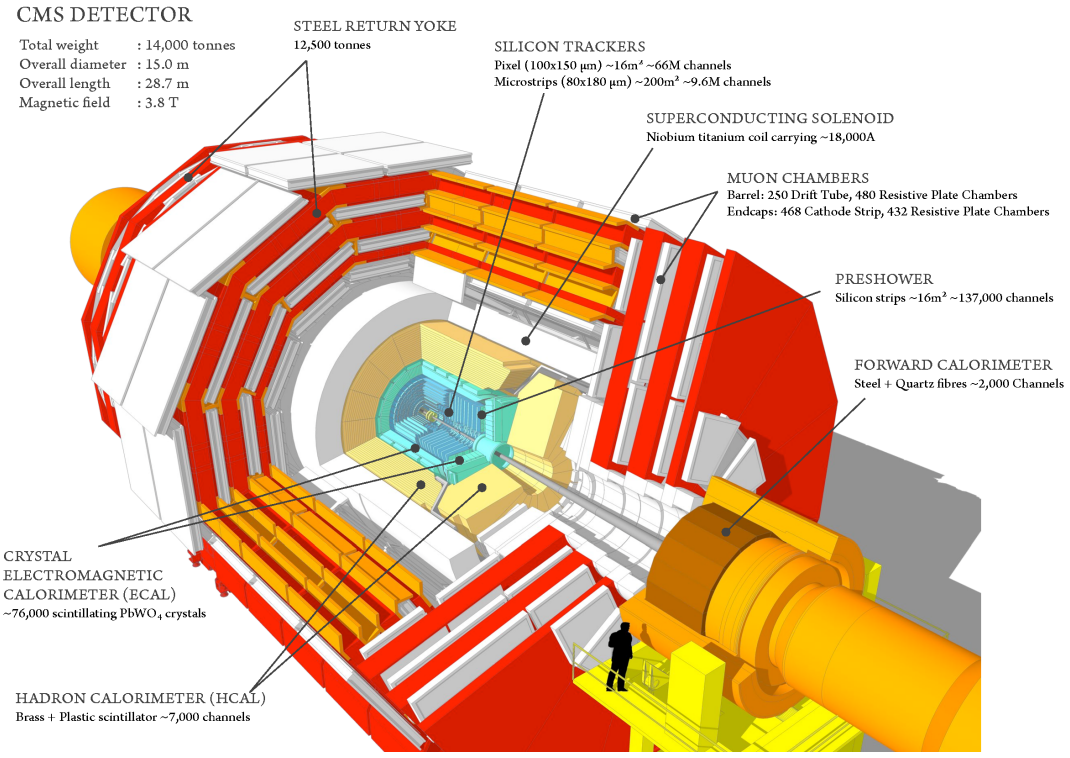

FIGURE 1.10: 3-Dimensional sketch of the CMS detector and its different sub-detectors. Figure taken from [41].

1.3.1 Tracker system

The main purpose of the tracker system is to reconstruct trajectories of charged particles and measure their momenta. A schematic view of the CMS tracker is provided in Figure 1.11, and a detailed description of the detector and its performance is provided in [42–44]. The tracker has cylindrical shape with a 2.5 m diameter and 5.8 m length. It is divided into two silicon subsystems: pixel and strip detectors. They have to provide good spatial and temporal resolution while operating efficiently under high instantaneous luminosities ($\mathcal{L} \approx 1.0^{34}\text{cm}^{-2}\text{s}^{-1}$) and bunch crossing every 25 ns, that create a high occupancy and high radiation environment.

The innermost detector, the pixel, is distributed over three cylindrical layers with radii 4.4 cm, 7.3 cm and 10.2 cm respectively and two disks in the forward region in $z = \pm 34.5$ cm and $z = \pm 46.5$ cm covering the pseudo-rapidity $|\eta| < 2.5$. It is segmented into 66 million of single pixels with size of 100x150 μm^2. The spatial resolution of the pixel detector is 10 μm in the transverse plane and $20 - 40$ μm in the longitudinal coordinate. The strip detector that surrounds the pixel detector is made of 9,3 million strips and it is composed of four different sub-detectors. In the innermost part, the tracker inner barrel (TIB) has four layers with 20-55 cm radius and $|z| < 60$ cm plus three tracker inner disks (TID) between $z = \pm[60-118]$ cm providing a spatial resolution

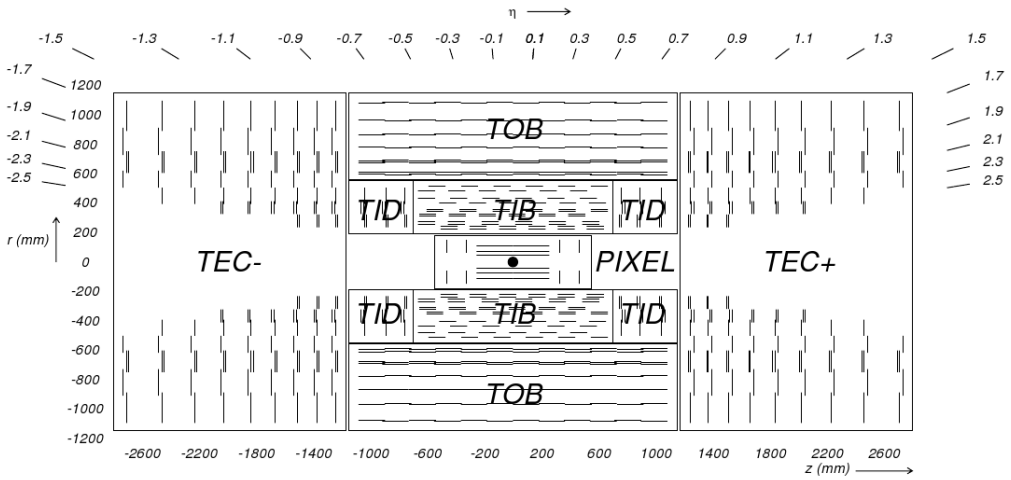

FIGURE 1.11: Schematic view of the CMS Tracker in the r-z plane and its different sub-detectors where each line represents a different detector module. More details are provided in Section 1.3.1. Figure taken from [42].

of $13 - 38$ μm in the transverse plane. The tracker outer barrel (TOB) with six layers with $55 - 116$ cm radius and $|z| < 118$ cm has a spatial resolution of $18 - 47$ μm. Each endcap consists of nine tracker endcaps (TEC) between z = $\pm[124 - 282]$ cm with similar resolution as the TOB.

The good spatial resolution of the tracker allows to distinguish tracks for different proton-proton interaction vertices along the z axis in each event [42]. This is important to separate tracks coming from the vertex in which the hardest interaction happened, denoted as primary vertex (PV), with respect to other tracks from all the other vertices, denoted as pileup vertices (PU). The PV assignment in the context of the W′ search is discussed in Section 2.3.2. Additionally, the tracker spatial resolution leads to an impact parameter (IP) resolution, defined as the minimum distance in the transverse plane between a track and the PV, of ~ 15 μm in the barrel. The latter allows CMS to identify displaced tracks from decays of B/D hadrons and reconstruct secondary vertices displaced with respect to the PV. The reconstruction of secondary vertices from semileptonic decays of B/D hadrons, $D^{\pm}$ or $D^{*\pm}(2010)$ is an important ingredient in the manuscript because they are used to disentangle associated Z + c production over the overwhelming background of Z + jets. More details about this are provided in the heavy flavour tagging Section 3.3.3 and in Section 3.4.2.

The momentum resolution of the tracker is about $2\% - 3\%$ for 100 GeV particles in the $|\eta| < 1.5$ region. In the case of TeV muons from a potential W′, the performance slightly degrades due to the small bending of the track combined with the limited size of the tracker. The inclusion of the muon detector, described in Section 1.3.4, enlarges the tracking volume and improves the resolution. More details about the track momentum assignment for muons are given in Section 2.3.4.

1.3.2 Electromagnetic calorimeter

The main goal of the electromagnetic calorimeter (ECAL) is to trigger and identify electrons and photons as well as measure their energy. The fine granularity of the calorimeter reduces the misidentification of jets as photons or electrons, and when combined with the tracker, helps to distinguish between electrons and photons. The ECAL calorimeter covers the pseudo-rapidity region $|\eta| < 3$ and is made of two different sub-detectors, both made of $PbWO_4$ scintillating crystals. This material has high-density, 8.28 g/cm^3, short radiation length, $X_0 = 0.89$ cm and small Molière radius, $R_M = 2.2$ cm. These properties are convenient for an efficient electron and photon reconstruction. Figure 1.12 shows a sketch of the different parts of the ECAL.

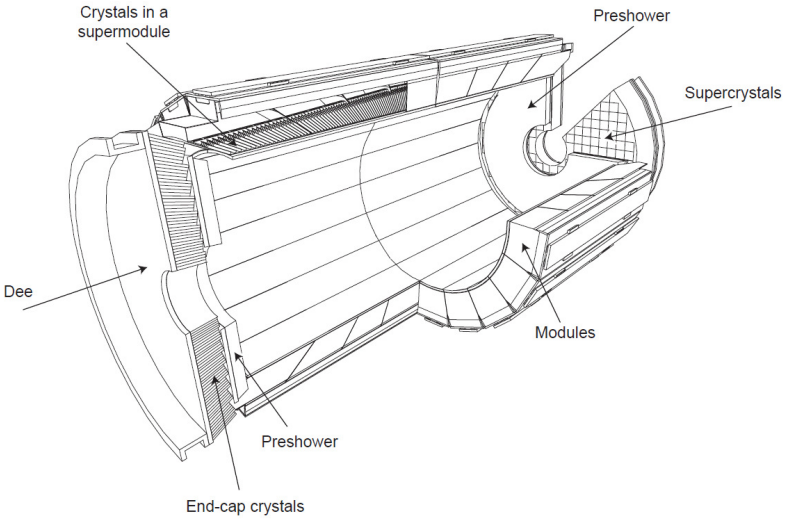

FIGURE 1.12: Sketch of the CMS ECAL showing the barrel supermodules, the two endcaps and the preshower detectors. The ECAL barrel coverage is up to $|\eta| = 1.48$; the endcaps extend the coverage to $|\eta| = 3.0$; the preshower detector fiducial area is approximately $1.65 < |\eta| < 2.6$. Figure taken from [45].

The barrel (EB) is made of 61 200 crystals each of them 23 cm long and 2.2 x 2.2 cm^2 in the transverse plane, covering the pseudo-rapidity $|\eta| < 1.48$. Additionally two endcaps (EE), each of them made of 7 324 crystals of 23 cm long and 2.86 x 2.86 cm^2 in the transverse plane, extend the pseudo-rapidity up to $|\eta| = 3.0$. Finally preshowers detectors, made of lead absorbers and silicon strip sensors are placed before the endcaps to improve the photon identification and cover the pseudo-rapidity range $1.65 < |\eta| < 2.6$. One important property is that at high transverse momentum, contrary to the tracker that relies in the detector alignment, the electron and photon energy resolution are dominated by the energy deposited in the calorimeter and the performance is expected to improve. Due to this, high energy photons and electrons are excellent probes for searches of new high mass particles. For example, at 1 TeV the electron energy resolution is expected to be better than 2% according to simulations.

1.3.3 Hadronic calorimeter

The main goal of the hadronic calorimeter (HCAL) is to measure the energy of jets produced in the hadronization of quarks and gluons. The constituents of the jets are typically particles such as protons, neutrons, kaons, pions, electrons, muons etc... The measurement of energies for all the jets in each bunch crossing up to large η is combined with information from other detectors giving valuable information of the activity in the event. This combined measurement allows CMS to measure indirectly the total amount of transverse momenta of invisible particles, E_{T}^{miss}, such as neutrinos or dark-matter (if produced) [46, 47] that leave no trace in the detector. This estimation is based on the assumption of conservation of the momentum in the transverse plane before and after the collision and is discussed in Section 2.3.8. The layout of the detector is shown in Figure 1.13. A detailed description of the detector is documented [48, 49] and its performance to measure the energy of jets is documented here [50].

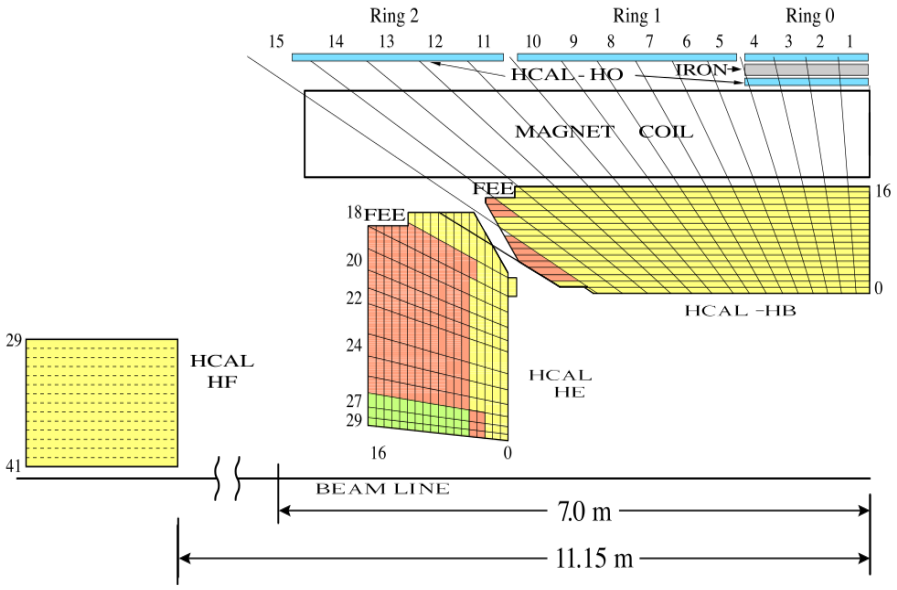

FIGURE 1.13: Schematic layout of the CMS HCAL. The HCAL barrel coverage is up to $|\eta| = 1.3$; the endcaps extend the coverage to $|\eta| = 3.0$; forward calorimeters are placed outside the CMS wheels in the region $2.5 < |\eta| < 5.0$. Figure taken from [49]

The HCAL is made of an interplay of layers of a dense brass absorber and tiles of plastic scintillator. Three different sub-detectors form the calorimeter: barrel (HB) $|\eta| < 1.3$ and endcap (HE) for $1.3 < |\eta| < 3.0$ placed in the space between the ECAL and the magnet, plus the outer detector (HO) covering $|\eta| < 1.3$ outside the magnet. Finally, forward calorimeters (HF) are placed outside the CMS wheels covering the region $2.5 < |\eta| < 5.0$ and are located at $z = \pm 12.5$ m. The HF measures Cherenkov light produced in quartz fibres and ensures a good hermeticity of the calorimeter.

1.3.4 Muon detectors

Muons have a mass that is about about 200 times the mass of the electron and only interact electromagnetically and weakly. They typically behave as minimum ionizing particles in the calorimeters and are the only interacting particle[2] that leave the detector without being stopped. The muon system is located outside the magnetic solenoid and embedded in the steel return yoke. The main tasks of the muon detector are: muon triggering, muon identification and the improvement of the muon momentum resolution for high momentum muons. Three different type of gas-ionization muon detectors are placed to meet this goals: drift tubes (DT), $|\eta| < 1.2$, cathode strip chambers (CSC), $0.9 < |\eta| < 2.4$ and resistive plate chambers (RPC), $|\eta| < 1.8$. A detailed description of the muon detectors and its performance can be found in [51, 52] and a schematic view of the different subsystems in Figure 1.14.

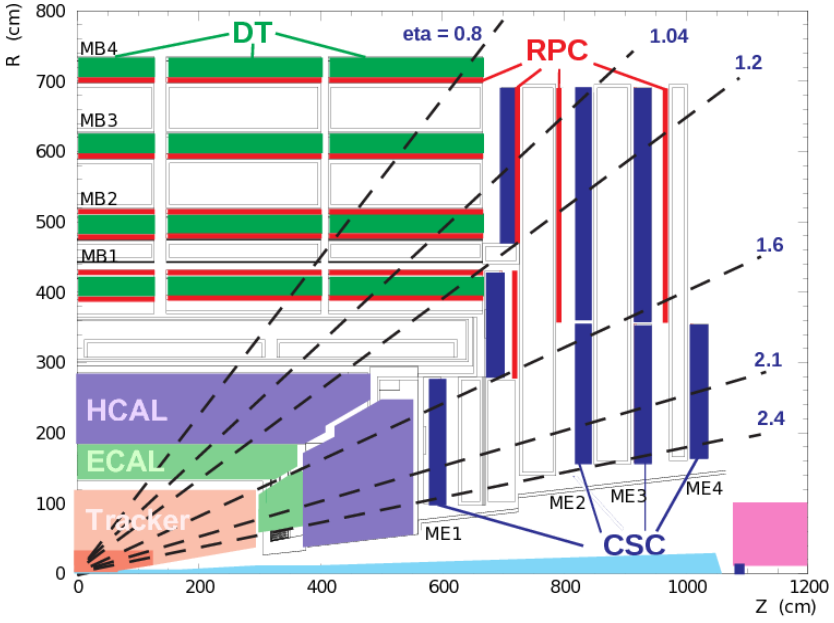

FIGURE 1.14: Sketch of the CMS Muon System. Additional details of the differents sub-detectors are given in Section 1.3.4. Figure taken from [52].

The DTs, shown in grey in Figure 1.15 (left), are organized in 12 sectors per wheel along the ϕ coordinate. Each sector is composed of at least four stations at different radii amounting a total of 250 chambers surrounding the beam pipe. Each of the chambers is made of two super-layers that measure the position along $r\phi$ and one along rz, where each the super-layers is made of 4 layers of drift tubes. The basic element of the drift tube is the drift cell with size 42 $\times$ 13 mm^2 and with a 50 μm diameter gold-plated stainless-steel anode wire at the center. The designed spatial resolution of the DTs are 100 μm along $r\phi$ and 150 μm along rz.

[2] In the group of non-interacting particles I refer to neutrinos and hypothetical new particles such as: neutralinos, dark-matter, etc...

In the forward region CSC detectors, shown in Figure 1.15 (right), are placed transversally with respect to the beam pipe and operate in harder conditions such as higher muon and background rates as well as stronger and non-uniform magnetic fields. The CSCs are divided in two endcaps sandwiching the five wheels, each of them divided in 4 disks made of 2-3 rings located at different radii. The spatial resolution is $75 - 150$ μm along $r\phi$ and $1.9 - 6$ mm in r.

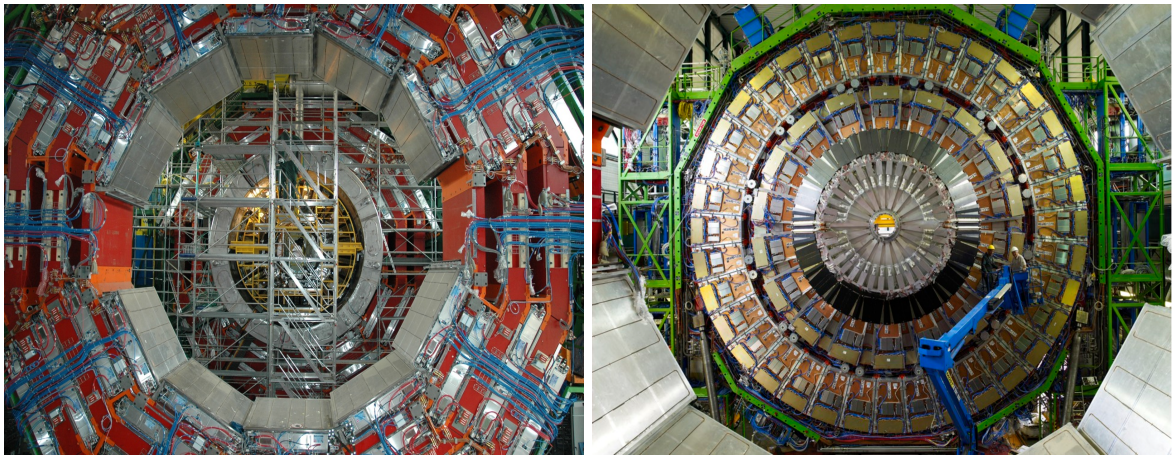

FIGURE 1.15: CMS Muon system during its installation, DT muon chambers are displayed in grey in left and CSCs in the right. RPCs are located behind the CSC in the right pad.

A key property of DT and CSCs systems is that each of them is able to measure the exact collision bunch crossing, among the sea of LHC collisions every 25 ns, that generated the muon that was used to trigger the event. In addition to the DT and CSC detectors, the RPC system is a complementary system with excellent time resolution better than 3 ns that is used to for trigger purposes and reinforces the correct measurement of the bunch crossing at the highest LHC luminosities.

Muon triggering and identification are crucial ingredients in a large fraction of CMS analysis. This is because isolated muons are a powerful tool to select events with potential signatures of new physics over the overwhelming rate of QCD processes at the LHC. Both aspects are discussed in Section 2.3.3 in the context of the W$'$ search.

At high transverse momentum, $p_{\mathrm{T}}^{\mu} > 200$ GeV, the muon resolution is largely improved when DT's, CSC's and tracker system are aligned with respect to each other and muon hits are included in the muon track reconstruction. The designed requirements for the CMS muon system originally specified a muon resolution for the tracker plus muon system track of the order of $6 - 17\%$ for 1 TeV muons. Measurements with cosmic muons in 2012 determined that the muon resolution is better than 10% at 1 TeV in the pseudo-rapidity $|\eta| < 1.2$. Muon momentum performance and detector alignment with the 2015 conditions in the context of W$'$ search are discussed in Section 2.3.4.

Chapter 2

Search for $\mathrm{W}' \to \mu\nu$ bosons

2.1 Is the Standard Model the ultimate theory?

The Standard Model is without any doubt, the most sucessful theory at describing strong, weak and electromagnetic interactions with high precision up to $\mathcal{O}(1$ TeV) including the latest results from the LHC. As an example, in Figure 1.2 it was shown the impressive level of agreement between many predictions of the SM for several processes and the experimental measurements. So far, there is not clear evidence of any new particle after the observation of a new scalar with a mass of 125 GeV [18, 19].

Latest measurements of the properties of the newly discovered boson with the full 8 TeV dataset [53], and the evidence of the same observation at 13 TeV with the recent Run II dataset [54, 55], reinforced the observation that the recently discovered scalar, is in fact a SM Higgs-like boson[1]. Higgs apart, after many dedicated searches by the ATLAS and CMS collaborations, there is no strong evidence of any other new state found with respect to the predictions of the Standard Model. After the 4^{th} of July of 2012, day of the announcement of the discovery of the Higgs boson, all the particles in the Standard Model have been detected experimentally.

Now the question is, after the Higgs discovery, should we declare the end of particle physics? The answer is no. There are already conceptual tensions and experimental observations that indicate that the SM is not the final theory. This section discusses some of the open questions in high energy physics that point to some new physics Beyond the Standard Model (BSM).

In the original formulation of the SM neutrinos are massless particles. However, the observation of neutrino flavour oscillation [56], and its confirmation by many experiments

[1]The jargon used here is because there is experimental hope to find deviations with respect to the predictions of the SM Higgs boson.

since then, causes a major breach in the SM. Indeed, in order to oscillate, neutrinos need to have a non-zero mass. Recent measurement point to a light mass of the neutrino which would require a modification of the SM to explain how neutrinos acquire mass and why its small value. There are different theoretical formulations that try to give mass to the neutrinos. Among the most popular ones, there are theories that predict the existence of heavy Majorana neutrinos to explain the smallness of the neutrino mass with the so called see-saw mechanism [57]. The latter would imply the existence of heavy neutrinos that could be produced at the LHC. So far, there are not evidences of such a kind of heavy neutrino and, as today, the nature of the neutrino mass remains unknown.

Another evidence of the limitations of the SM is its failure to explain why our universe is dominated by matter over anti-matter. According to the Shakarov conditions [58], in order to evolve into a universe dominated by only matter starting from a balanced situation, one would need violation of barion number, CP violation and interactions out of thermal equilibrium in the early universe. Within the SM frame, CP violation was observed for the first time in kaon decays [59] and also recently in B meson decays [60, 61]. However the rarity of such events is not enough to explain the amount of matter in the universe. A hope of new sources of CP violation in the coming years is foreseen in the neutrino sector after the discovery of neutrino oscillations... Unfortunately, as of today there is not a convincing answer to explain why not all the matter annihilated with anti-matter in the early universe.

Probably the strongest evince of a new phenomena not explained in the SM comes from studies of the anisotropies in the Cosmic Microwave Background (CMB) temperature. Recent measurements with data collected with the Planck satellite reported that about 27% of the energy content in the universe is in form of dark matter and 68% in form of dark energy [62]. The SM cannot provide an explanation for their nature and nowadays their study is a very active front of research in both, theory and experiment. Dark matter, if it couples to Standard Model particles, could be produced and studied at the LHC. The strategy followed by the ATLAS and CMS collaborations is to search for dark matter in the so called Mono-X channels. These analyses search for direct production of invisible particles (dark matter) produced in association with detectable initial state radiated particles, e.g jets (mono-jets), vector bosons (mono-vector bosons) or photons (mono-photons)... Unfortunately, as today, there is no evidence of production of dark matter at the LHC, nor convincing explanation of the nature of dark energy.

The next motivations for BSM physics are two conceptual problems in the theory that are related with the Higgs boson itself. They are called: *technical hierarchy problem* and the *vacuum stability of the Higgs potential*. The first case is related with the fact that the Higgs boson is the only scalar in the SM and suffers from divergent radiative corrections

to its mass. The main source of corrections comes from the one-loop Higgs correction with positive sign plus the one-loop top and gauge boson correction with opposite sign. The problem arises because the size of the corrections to the value of the Higgs mass is proportional to the cut-off, Λ_{NP}, defined by the scale in which of new physics is suppose to play a role. If there is no BSM at the TeV scale and the next scale of new physics is at the Planck scale $\Lambda_{Planck} \sim 10^{19}$ GeV, the low value of the Higgs mass, 125 GeV, would be a result of a cancellation of large positive and negative corrections several orders of magnitude larger than the mass itself. This picture is unnatural for physicists and a mechanism is expected that protects the Higgs mass against large radiative corrections. However such mechanism, if existing, is still unknown. The size of the technical hierarchy problem could be reduced if the cut-off scale occurs at the TeV scale instead that at the Planck scale. This would imply the presence of new states with masses of the order of TeV. There are alternative models that predict that the Higgs is not an elementary particle and does not suffer from radiative corrections [63]. Finally a third solution predicts the existence of extra dimensions that move the Planck scale to the TeV scale and eliminate the large corrections to the Higgs mass [64]. In all the cases, new states could emerge at the new Run of the LHC to give more light to this conceptual problem.

The vacuum stability [65] is another undesired feature in the theory motivated by a potential instability in the Higgs potential (see Equation A.29) due to the running of the Higgs quartic coupling, λ. At the electroweak scale, $v = 246$ GeV, the value of the Higgs quartic coupling is positive and depends on the Higgs mass and the Higgs vacuum expectation value, $\lambda = \frac{1}{2}(\frac{M_H}{v})^2 \approx 0.13$. However, it is not guaranteed that λ will remain positive up to the Planck scale since its evolution strongly depends on its radiative corrections. The problem arises because with the current measurements of the top quark and W boson mass, the evolution of λ points to small and possibly negative value of λ before the Planck scale [66]. If this was true, the Higgs potential could have a metastable or unstable minima that compromises the existence of our universe. The fact that your are reading this paragraph suggests that probably this is not true, and the simplest explanation for this conceptual problem is to assume that there are new particles, *yet-to-be-discovered*, that affect the evolution of λ and keeps it positive all the way up to the Planck scale.

Moving to more speculative scenarios of new physics, historically physicists always aimed to find the deepest underlying principles in each phenomenal studied. However in the SM there are almost 20 free parameters (excluding the neutrino sector), including three gauge couplings with completely different magnitudes and 100% parity violation in weak decays... This picture seems not very simple and physicists, Einstein included, tried to find an unified, simple and beautiful theory that could explain and unify the spectrum of particles and interactions. These are the basis of the Grand Unified Theories (GUT).

In these theories, the underlying gauge symmetry principle in Nature is in fact a simple and larger than the SM gauge symmetry group. The Standard Model as we know would be the result of a broken symmetry at a low energy scale of this unified theory. Next section will briefly introduce some generalities of GUT theories, and most importantly their prediction of additional W' and Z' gauge bosons and interactions that can be directly searched at particle accelerators such as the LHC.

Finally, the Standard Model does not include gravity in its original formulation. In the process of writing this manuscript, gravitational waves have been firstly detected by the Ligo and Virgo collaborations [67], opening a new era and field of research, and perhaps a new experimental bridge where to attempt to include gravity and particle physics under a common formulation.

2.1.1 New interactions and extended gauge symmetries

Extended gauge symmetries are inspired by GUT theories that aim to unify all the known forces in nature except gravity, namely: electromagnetic, weak and strong, into a single force with an unique coupling. In the same spirit that above the electroweak energy scale, the electromagnetic and the weak theory are two different manifestations of the electroweak force (EW), some GUT theories defend the possibility that the three fundamental forces were no longer distinguishable in the early universe above a given energy scale, $\Lambda_{GUT} < \Lambda_{Plank}$, called the GUT scale.

An important requirement for any of these BSM models is that they should recover the SM at low energies. For this, normally it is assumed that the vacuum expectation values (VEVs) in the first and second Spontaneous Symmetry Breaking (SSB) are such that $v_1 >> v_2$, where v_1 is the scale in which the unbroken SM is recovered and $v_2 = 246$ GeV corresponds to the Higgs VEV. Contrary to the SM model, where mass relation of the W and Z give little freedom in the choices of the Higgs representation to break the SM, different GUT theories can have can different symmetry breaking pattern possibilities at different energy scales, that give rise to different predictions [68].

There is no unique and accepted GUT theory, but the commonality in all of them is that this unification is achieved by enlarging the symmetry group of the Standard Model into a larger symmetry group called the GUT symmetry group. It is important to remark that the choice of the gauge structure of the SM, $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$, is not a prediction of the theory, and it is rather adopted because is the simplest that explains all the interactions observed in Nature. When the GUT group contains a subgroup in addition to $SU(3) \otimes SU(2) \otimes U(1)$, it means that new gauge bosons exist and interactions associated to those also exist.

As an example of one of these extended gauge symmetries, one of the minimal extensions of the electroweak sector, $SU(2)_L \otimes U(1)_Y$, that incorporates new W and Z bosons, hereafter denoted as W′ and Z′ bosons are the G_{221} models which present a extended gauge symmetry: $SU(2)_1 \otimes SU(2)_2 \otimes U(1)$ [68]. Among the GUTs theories that may contain the G_{221} as a subgroup, $SO(10)$ is one of the most popular. This section briefly introduces the case of the G_{221} group, focusing on the possible W′ signals in the dilepton (lepton+neutrino) final state that may appear in an extension of this kind.

Case I: identify $SU(2)_1$ as $SU(2)_R$

The parity violation in weak interaction in the SM has been tested experimentally with high level of precision in different experiments. However its origin in Nature is still not well understood. Due to this, many extensions of the SM predict that the Left-Right (LR) symmetry in weak interactions is recovered at higher energies. One of the simplest ways to do this is to extend the electroweak part in the SM, $SU(2)_L \times U(1)_Y$, to a new Left-Right symmetric $SU(2)_R \times SU(2)_L \times U(1)_{B-L}$ group, and assume that the symmetry breakdown occurs in two steps:

$$SU(2)_R \times SU(2)_L \times U(1)_{B-L} \xrightarrow{\Phi} SU(2)_L \times U(1)_Y \xrightarrow{H} U(1)_{EM} \tag{2.1}$$

Where Φ and H fields give mass to the new W′/Z′ and SM gauge bosons respectively. In this extension, one of the predictions is the existence of a new W′ gauge boson, denoted as W_R, that couples to right handed fermions. If the mass of the W_R is much larger than the mass of the SM W, its existence does not contradict the current observations. The existence of a potential W_R can be probed at accelerators if the centre of mass energy is high enough. At the LHC, this possible W_R with mass at the TeV scale could be produced via s-channel and potentially discovered in $W_R \to \ell_R \nu_R$ final states if $m_{\nu_R} < m_{W_R}$, where $\ell = e, \mu, \tau$.

In principle a $W_R \to \ell_R \nu_R$, does not necessarily produce a significant amount of signal in the dilepton final state because it only couples to right handed fermions. However the nature of the right handed neutrino, if exist, is unknown and its mass is not constrained as long as it does not couple to SM particles. A possible massive right handed neutrino that decays outside the detector, or decays into invisible particles, or passes undetected trough the detector, will produce a clear signal in $\ell + E_T^{miss}$ final state and is searched in the analysis presented in this chapter.

Finally, if the hypothetical right handed neutrino is heavier than the W_R, $m_{\nu_R} > m_{W_R}$, then limits into leptonic final states decays do not apply, and searches in other final states such as: $W' \to qq'$, $W' \to WZ$ or $W' \to WH$ are probably better places to look for a W_R.

Case II: identify $U(1)$ as U_Y

In this second scenario, the W′ couples directly to left handed fermions and interference between the W and the W′ is expected. The symmetry breaking pattern is the following:

$$SU(2)_{L_1} \times SU(2)_{L_2} \times U(1)_Y \xrightarrow{\Phi} SU(2)_L \times U(1)_Y \xrightarrow{H} U(1)_{EM} \tag{2.2}$$

In this scenario the W′ directly decays into a lepton and a SM neutrino such as, $W' \to \mu\nu$. This possibility clearly produces a signal in $\ell + E_T^{miss}$ final state and is also searched in this chapter.

In practice, the real experimental and theoretical picture is more complicated. It is important to remark that if there is ever a experimental discovery of a W′-like signature, it is not guaranteed that it will correspond to any of the GUT theories or even to a extension of the symmetry group of the Standard Model. There are other theories that also predict W′-like signals and could produce a similar experimental signature. For example, Kaluza-Klein excitations of the W boson that appear in extra-dimension theories, predict the existence of interactions of the type $W_{KK} \to \mu\nu$ at the TeV scale. These theories propose that the Plank scale could be as low as the TeV scale, thus solving the *hierarchy problem* of the SM in a different way [69–72]. In conclusion, in the hypothetical case of the discovery of a W′-like signal, further studies on the potential signal properties and in other decays modes will be needed to discern between different theoretical explanations.

Due to the large variety of different W′ signals that one could build in the zoo of GUT theories (or any other model), historically the experimental collaborations use the so-called benchmark models to simulate the W′ signals in their searches. Benchmark models do not aim to be complete theories, they rather use phenomenological Lagrangians in which the most relevant parameters are encoded in a simple way, leaving the details of the full model aside. In other words, in case of discovery it will never be the one predicted by the benchmark model, though it is expected that the main kinematical properties of the observed W′ resemble the ones described by the benchmark model.

In summary, the study presented in this chapter aims to discover a trace of an extended gauge sector of the SM. This is done by searching for new interactions mediated via new heavy vector bosons that could have played a role in the early universe. Particularly the search is designed to discover a new hypothetical $W' \to \mu\nu$ boson that could produce a signal with one high momentum muon plus an invisible particle for the detector, such as a neutrino. This final state is denoted in the text as $\mu + E_T^{miss}$.

2.1.2 Sequential Standard Model W′

The benchmark signal model used in the $W' \to \mu\nu$ search corresponds to the Sequential Standard Model (SSM) W′ that is described in the "Extended gauge model" in [73]. In this scenario, the W′ has the same left-handed couplings to fermions as the W of the SM described in Appendix A.4. Figure 2.1 shows the leading order Feynman diagram for $W' \to \ell\nu$ at the LHC.

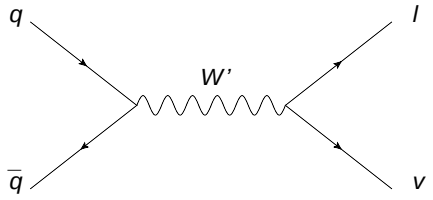

FIGURE 2.1: Leading order Feynman diagram describing SSM W' production at the LHC.

Taking the notation from the Particle Data Group (PDG)[74], the new Lagrangian terms that describe the potential interaction of the W'^{+} boson with fermions are given by a generalization of Equation A.26 to three generations and right and left handed couplings:

$$\frac{W'^{+}_{\mu}}{2\sqrt{2}}\left[\bar{u}_i(C^R_{q_{ij}}P_R + C^L_{q_{ij}}P_L)\gamma^\mu d_j + \bar{\nu}_i(C^R_{\ell_{ij}}P_R + C^L_{\ell_{ij}}P_L)\gamma^\mu e_j\right] \tag{2.3}$$

Where u, d, ν, e are the SM fermions in the mass eigenstate basis, and $i = 1, 2, 3$ refers to each of the three fermion generations. The C^R_X and C^L_X are couplings to right and left handed fermions respectively, and $P_R = \frac{1+\gamma^5}{2}$ and $P_L = \frac{1-\gamma^5}{2}$ are the chirality projector matrices that satisfy $P_L + P_R = 1$, and are given by the Dirac gamma matrices for which $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$.

If a $C^R \neq 0$ term is included in the theory, a W'_R with coupling to right handed fermions is predicted and right handed neutrinos are expected. The latter, would correspond to something close to the Case I described in Section 2.1.1.

Alternatively if the W′ couplings are $C^L_{q_{ij}} = gV_{CKM_{ij}}$, $C^L_\ell = g$ and C^R_X=0. The new W′ is the one predicted in the SSM W′ and is rather similar to a SM W with higher mass. This scenario is something closer to the Case II in Section 2.1.1.

In the SSM, if the mass of the W′ is larger than the $t\bar{b}$ threshold, $m_{W'} > 180$ GeV, the process $W \to t\bar{b}$ opens up yielding to a reduction of the $\mathcal{B}(W' \to \mu\nu) \approx 1/12 = 8.3\%$ for each of the three leptonic decay modes $\ell = e, \mu, \tau$, and $\mathcal{B}(W' \to q\bar{q}) \approx 25\%$ for each of the hadronic decay modes $u\bar{d}'$, $c\bar{s}'$, $t\bar{b}'$, thus slightly modifying the branching ratios of the SM W boson (see Table A.1).

Potential W' couplings with other boson such as: $W' \to WZ$ or $W' \to WH$ are in general highly model dependent, and particularly in the SSM they are assumed to not exist. With this choice, the SSM maximizes the $W' \to \mu\nu$ signal that one could expect in the final state $\mu + E_T^{miss}$.

Interference effects between the SM W and the W' are not considered in the signal modelling. The reason for this choice is that generally they are model dependent and a second-order effect. For discovery purposes this effect can be ignored with the current luminosity at 13 TeV. However, in case of discovery with high enough statistics, interference effects would give valuable information about the W' properties with left handed couplings.

If one neglects the interference effect with the SM W, the kinematics of the decay products of a W' with right handed and left handed couplings do not differ and produce the same signature in the fermionic decay channel. In this case, the W' can be reinterpreted as W_R in a Left-Right model if the potential right handed neutrino leaves no trace in the detector and is light compared to m_{W_R}.

Hypothetical new resonances that decay into leptonic final states produce clean signatures in the detector making them "golden channels" for early discoveries. Due to this, W' a Z' bosons have been searched in previous experiments, and those analysis were particularly important at the early days of the accelerators after an increase in the centre-of-mass energy.

Current experimental limits at $\sqrt{s} = 8$ TeV for W' at 2 TeV

The search of a new particle is always a marathon that requires a dedicated program of searches targeting different final states. Table 2.1 shows the current observed limits to the W' production from the CMS experiment at 8 TeV for all the possible final states. For this, a W' with mass 2 TeV is taken as common reference in the table.

Channel	Observed cross section limit	Reference
$\sigma_{W'} \times \mathcal{B}(W' \to \mu\nu)$	< 1 fb^{-1}	EXO-12-060 [25]
$\sigma_{W'} \times \mathcal{B}(W' \to e\nu)$	< 0.5 fb^{-1}	EXO-12-060 [25]
$\sigma_{W'} \times \mathcal{B}(W' \to \tau\nu)$	< 3 fb^{-1}	EXO-12-011 [75]
$\sigma_{W'} \times \mathcal{B}(W' \to qq)$	< 200 fb^{-1}	EXO-12-059 [76]
$\sigma_{W'} \times \mathcal{B}(W' \to tb)$	< 20 fb^{-1}	B2G-12-010 [77]
$\sigma_{W'} \times \mathcal{B}(W' \to WZ)$	< 10 fb^{-1}	EXO-12-024 [78]
$\sigma_{W'} \times \mathcal{B}(W' \to WH)$	< 7 fb^{-1}	EXO-14-009 [79]

TABLE 2.1: Observed cross section limit to a W' of 2 TeV from CMS measurements done at 8 TeV in different final states.

2.2 Data samples and simulation

2.2.1 13 TeV dataset in 2015

2015 was a historical year in particle physics. The LHC restarted its program of collisions in Summer at a centre of mass energy record of 13 TeV and delivered 4.2 fb^{-1} in CMS. The dataset analysed for the W' search in this thesis corresponds to 2.3 fb^{-1} certified good for physics. The main difference between the recorded luminosity and the certified dataset is due to a problem with the cooling of the superconducting magnet that caused that about 1 fb^{-1} of data were taken without magnetic field and is not used for the search.

Only in the muon performance Section 2.3 of the document, where studies did not require the explicit usage of the calorimeters, an extended dataset with a less stringent certification for the calorimeters was used and amounts up to 2.8 fb^{-1}.

The 2015 data taking was carried out with two different bunch crossing time conditions. An early and short run in June 2015 with bunch crossings every 50 ns, and the nominal run at 25 ns for the rest of the year. The data analysed in this thesis corresponds only to the 25 ns period. A detector performance note was presented with 48 pb^{-1} of data at the 2015 LHCP conference collected at the end of the 50 ns run [80].

Finally cosmic muons were collected in April 2015 in a dedicated campaign called CRAFT15 (Cosmic Ray at Four Tesla 2015). This dataset amounts to about 10 days of cosmic muon data, and are used for dedicated studies of the performance of the muon momentum measurement using high p_T cosmic muons.

A preliminary version of this analysis was first presented at CERN during the 2015 end of the year Jamboree [81], and documented in the Physics Analysis Summary [20]. After the 2015 Jamboree, calorimeters were recalibrated and detector re-aligned giving as a result an updated 2015 dataset with improved calibrations called 2015 re-calibrated dataset. The results presented in this manuscript correspond to an updated version of Ref. [20] using this re-calibrated data sample and published in December of 2016 [21].

2.2.2 Monte Carlo simulations

Backgrounds

There are several processes whose final state is the same or may be mistaken as that of the W' signal. This section describes how the contribution from those processes is estimated in the analysis. Prior to read this chapter it is convenient to familiarize with the concepts associated to the LHC proton-proton phenomenology introduced in Chapter 1, particularly Section 1.2.1 about PDFs and Section 1.2.2 about the hard scattering.

The dominant and indistinguishable process from the W' signal is SM $W \to \mu\nu$, especially when the W is produced with high mass and at rest. The latter is sometimes denoted as W^* to distinguish both topologies. The $W^* \to \tau_{\text{lep}}(\mu\nu)\nu$ channel contribution to the high mass region is reduced compared to that from muons due to the presence of two neutrino and the analysis selection (described in Section 2.4.1). Other background processes are $Z/\gamma^* \to \mu\mu$ in which one of the leptons is lost or outside the detector acceptance, $t\bar{t}$ in their semileptonic or dileptonic decay channel, single top production and diboson processes (mainly high mass WW and to a less extent WZ and ZZ).

Figure 2.2 shows a sketch of these backgrounds that could mimic a W' signal. In the figure, red lines correspond to high momentum muons and green lines to neutrinos. Yellow lines correspond to potential additional particles that contribute to the final state, and black and blue lines correspond to intermediate particles. Section 2.4.1 gives further details about each of the backgrounds and describes how their contribution is evaluated, as well as how they can be suppressed.

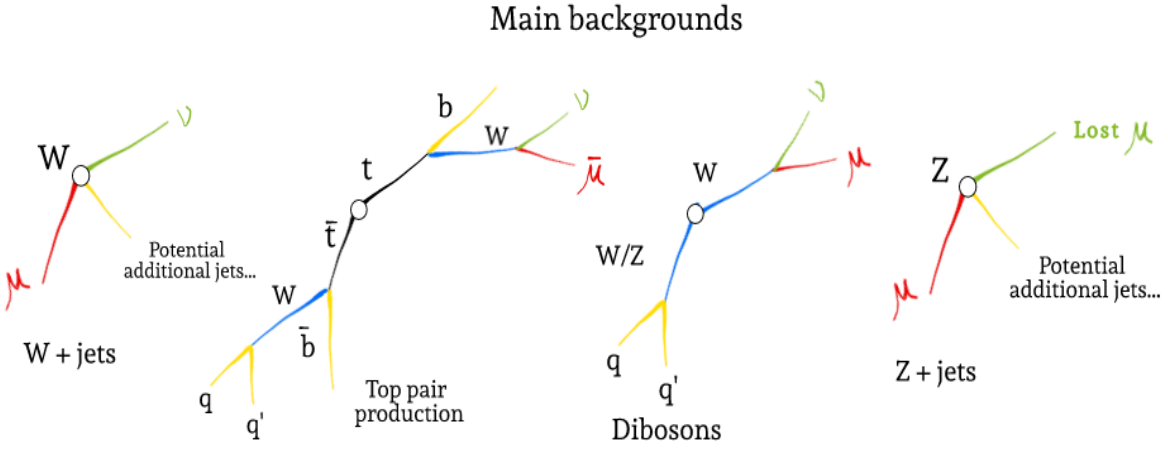

FIGURE 2.2: Sketch of the main processes that could mimic a $W' \to \mu\nu$ signal at the LHC. Details about of each of the processes and the colour code of the lines in the sketch are provided in Section 2.2.2.

All samples of simulated events are generated with MC event generators. A sample of inclusive W boson plus jets is generated with the PDFs from the NNPDF 3.0 [28] group and MadGraph5_aMC@NLO [35] for the hard scattering interfaced with pythia8 [32] for parton showering and hadronization.

Regarding the remaining background processes, a sample of $t\bar{t}$ is generated with Powheg [34] and uses CT10 [29] for the PDF. Generated events are interfaced with pythia8 for parton showering and hadronization. Z/γ^* electroweak backgrounds are generated also with Powheg, including a dedicated high-mass sample with the same generator. Contributions from diboson processes such as WW, WZ and ZZ are generated with pythia8 and CTEQ6L1 for the PDF.

All the distributions from the simulated samples are normalized to an equivalent luminosity of 2.3 fb^{-1} for the W' search and to 2.8 fb^{-1} in the muon performance section. The cross section for the W plus jet sample is taken from a NNLO prediction from FEWZ [36], using MSTW08 [30] for the PDF. In diboson processes the cross sections are normalized to NLO predictions from MCFM [37] and MSTW08 for the PDF. Top pair production, $t\bar{t}$, is taken at Next-to-Next-to-Leading-Logarithm (NNLL).

Additionally, it is important to have an accurate description of the mass distribution of the W^* production. This is particularly important at the highest masses where the analysis is more sensitive due to the lower backgrounds. Due to this, the W boson plus jets sample is combined with a dedicated set of W^* samples generated with pythia8 in different bins in mass of the W. Higher order QCD and EW corrections are taken into account using dedicated differential *k-factors* as a function of the mass.

At the time in which this analysis was done, there was not any complete NNLO QCD + NLO EW calculation for the W^* process. However there were available separate QCD and EWK corrections from different generators. Due to this, the practical approach is to assume that their combination is additive (Equation 2.4) or multiplicative (Equation 2.5), and consider their difference as an additional systematic uncertainty. NLO EW corrections were computed with the MCSANC [82] event generator while NNLO QCD corrections were derived with FEWZ [36], both at 13 TeV. Figure 2.6 shows the two possible combinations that give rise to two sets of *k-factors*.

$$[\frac{d\sigma}{d\mathcal{O}}]_{QCD\oplus EW} = [\frac{d\sigma}{d\mathcal{O}}]_{QCD} + [\frac{d\sigma}{d\mathcal{O}}]_{EW} - [\frac{d\sigma}{d\mathcal{O}}]_{LO} \tag{2.4}$$

$$[\frac{d\sigma}{d\mathcal{O}}]_{QCD\otimes EW} = \frac{[\frac{d\sigma}{d\mathcal{O}}]_{QCD}}{[\frac{d\sigma}{d\mathcal{O}}]_{LO}} \cdot [\frac{d\sigma}{d\mathcal{O}}]_{EW} \tag{2.5}$$

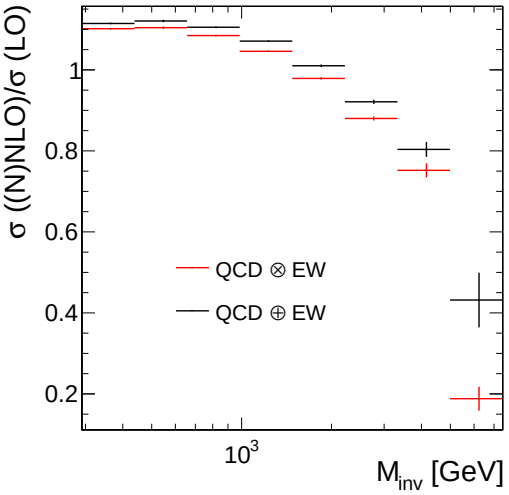

FIGURE 2.3: Combination of the NLO electroweak and NNLO QCD corrections to the $W^* \to \mu\nu$ production as a function of the invariant mass of the W^*. Two different combinations are shown, additively (red) and multiplicatively (black). Credit for this study: Collaborators from the $W' \to e\nu$ search analysis team.

The final *k-factors*, defined in Equation 2.6, are computed as a function of the invariant mass and corresponds to the average of the additive and multiplicative combinations (Equation 2.4 and 2.5).

$$K(\mathrm{M_{inv}}) = \frac{\Delta\sigma(\mathrm{NLO})/\Delta\mathrm{M_{inv}}}{\Delta\sigma(\mathrm{LO})/\Delta\mathrm{M_{inv}}} \tag{2.6}$$

The dependence of the *k-factors* is flat and constant as a function of the mass with a value around 1.3 up to $M_W = 1$ TeV. For higher masses this value slightly drops due to a higher impact of negative NLO EW corrections. An additional uncertainty of 5% in the *k-factors* is considered in the analysis to cover the differences between the two possible combinations and the individual QCD and EW corrections.

Finally, the modelling of the additional activity in parallel with the hard scattering, known as underlying event, was modelled using PYTHIA8 with the tuning CUETP8M1 [83, 84]. All the generated events are processed through the full GEANT4 [39] detector simulation, trigger emulation and event reconstruction chain of CMS.

W' Signals

$W' \to \mu\nu$ signals samples for several W' mass hypothesis are generated using PYTHIA8 and NNPDF 2.3 LO, following the predictions of the Sequential Standard Model (see Section 2.1.2 for a description of the SSM). Signal cross sections are renormalized to include a *k-factor* that includes NNLO QCD corrections derived with FEWZ and its size is between 1.1 and 1.3. These corrections are computed assuming that the W' is a copy of the W with different mass. Table 2.2 shows the LO cross section from PYTHIA8 and NNLO from FEWZ.

Sequential SM W' boson			
W' mass [TeV]	$\sigma_{\mathrm{LO}}\,\mathcal{B}$ (fb)	$\sigma_{\mathrm{NNLO}}\,\mathcal{B}$ (fb)	*k-factor*
1.0	2699.0	3623.5	1.343
1.2	1275.8	1708.9	1.340
1.4	657.28	877.23	1.335
1.6	360.15	478.18	1.328
1.8	206.47	272.38	1.319
2.0	122.55	160.43	1.309
2.2	74.726	96.957	1.298
2.4	46.584	59.813	1.284
2.6	29.560	37.510	1.269
2.8	19.035	23.852	1.253
3.0	12.412	15.348	1.237
3.2	8.1952	9.9844	1.218
3.4	5.4675	6.5682	1.201
3.6	3.6865	4.3708	1.186
3.8	2.5158	2.9432	1.170
4.0	1.7337	2.0092	1.159
4.2	1.2108	1.3926	1.150
4.4	8.5649×10^{-1}	9.8188×10^{-1}	1.146
4.6	6.1476×10^{-1}	7.0621×10^{-1}	1.149
4.8	4.4871×10^{-1}	5.1886×10^{-1}	1.156
5.0	3.3227×10^{-1}	3.8928×10^{-1}	1.172
5.2	2.5146×10^{-1}	2.9832×10^{-1}	1.186
5.4	1.9300×10^{-1}	2.3371×10^{-1}	1.211
5.6	1.5128×10^{-1}	1.8659×10^{-1}	1.233
5.8	1.2029×10^{-1}	1.5148×10^{-1}	1.259

TABLE 2.2: Signal production cross sections (at LO and NNLO) times branching fraction, as well as the *k-factor*, for SSM Wprime bosons decaying to lepton plus neutrino, at $\sqrt{s} = 13$ TeV. Credit for the *k-factor* study: Collaborators from the $W' \to e\nu$ search analysis team.

2.3 Physics objects: muons and MET

This section introduces the physics objects that are built from measurements in CMS, giving special attention to the ones that are relevant for the muon object used in the W' search such as:

- Event triggering.
- Primary vertex assignment.
- Muon reconstruction and identification.
- Muon performance.
- Jets.
- Missing E_T performance.

Without any doubt, the physics objects and their calibration described in this section are as important as the W' search itself. The reason for this can be explained with simple a argument:

There is only one observable that is measured experimentally in a $W' \to \mu\nu$ *search, a high transverse momentum muon.*

In 2015, CMS has re-started its operation after two years of upgrade that allowed the detector to cope with the LHC collisions at 13 TeV. After such a long break all the detector calibrations need to be re-derived, simulations re-calibrated and physics objects revised. The goal of this chapter is to show that predictions from simulation and measurements are under control with the new conditions of the LHC and CMS detector for the Run II.

Having the best possible understanding of the physics objects and their calibrations, is what gives an analysis the best sensitivity, reduces systematic uncertainties and maximizes the discovery potential. Before claiming any discovery, it is necessary to show that the predictions from simulations reproduce the observations in regions where the contribution from a W' signal is expected to be low. It is always important to keep in mind that in searches with very low backgrounds, a badly calibrated detector could easily create a fake signal.

Finally, most of the physics objects and techniques for calibration described in this section (except for the muon performance, that is dedicated to high momentum muons in the W' search), were also employed in a similar fashion at 8 TeV for the $Z + c$ cross section measurement in Chapter 3.

2.3.1 Trigger

In order to discover a W′ or any other new particle, events containing potential candidates need to be recorded. However, due to the high collisions rate of 40 MHz at 25 ns and the large cross section of QCD processes, it is impossible and highly inefficient to record all the collisions in CMS. As an example of this, at least 10^{12} proton-proton interactions are needed before producing one $W' \to \ell\nu$ high mass candidate[2].

The CMS trigger is a system designed to keep interesting events for physics analysis, while discarding low energy events based on physics driven choices. A potential operation without a trigger system is not possible and any problem with it could potentially prevent CMS to take data. Thus, the trigger is one of the most critical aspects to monitor during the data taking and has to be adapted to the data taking conditions and instantaneous luminosity (see Equation 1.2 and Table 1.1 for details). Among the several requirements of the CMS trigger system, there are two necessary characteristics that the trigger has to fullfill in order to allow CMS to operate: the data acquisition rate has to be reduced to 1 kHz, five orders of magnitude less than the collision rate, and the time to take the decision has be less that few hundreds of millisecond. This is achieved in the CMS Trigger system in two sequential steps: the Level 1 Trigger (L1) and the High Level Trigger (HLT). For example, in the case of triggering a potential $W' \to \mu\nu$ signal candidate, the most efficient way is to reconstruct a high p_T muon at the trigger level and decide to record the event in terms of the measured muon p_T.

The first step towards a positive triggered event is the so called Level-1 Trigger (L1). The L1 must provide an extremely fast response in few microseconds. Due to this, the L1 trigger is purely hardware based and the decision is carried out in the detector front-end electronics. In order to achieve a fast response, the trigger decision is based only on ECAL, HCAL and muon system information. A positive L1 triggered event creates a L1 seed. The latter in the case of a high p_T muon is called L1_Mu25, which means that the Level-1 Trigger track built in the muon system had a p_T larger than 25 GeV. After this step, the total fraction of total events with a positive L1 trigger reduces the rate from 40 MHz to about 100 kHz.

Only L1 seeds passing the L1 trigger decision will enter into the second step in the chain: the High Level Trigger system (HLT). In this second step, the decision is made using a global overview of the event combining inputs from different sub-detectors. At this point a more sophisticated decision is used involving selection criteria as close as possible to the one used in the final analysis. Each set of selection criteria associated to a L1 seed is called an HLT path. The CMS trigger consist a set of several HLT paths that aim

[2]Assuming $\sigma_{W'} Br(W' \to \mu\nu) < 1$ fb^{-1} at 13 TeV.

to select events for physics analysis with different topologies implemented with different selection criteria. After this step the CMS output rate has to be of the order of 1 kHz.

In the case of the W' search, the HLT trigger combines muon system information with tracker information improving the muon p_T resolution. The associated HLT path receives the name of HLT_Mu50 and is seeded by the L1_Mu25. The HLT_Mu50 path selects only an event if the p_T of the muon is larger than 50 GeV and if it was reconstructed in the pseudo-rapidity region $|\eta| < 2.4$. Due to resolution effects associated to the muon momentum measurement in the trigger system, only reconstructed muons with momentum larger than 53 GeV are selected for search. After a positive HLT decision, $W' \to \mu\nu$ candidates should be recorded on tape and ready to start the analysis.

The HLT muon trigger efficiency has been estimated with W' signal samples (see Section 2.2.2) up to high transverse momentum. This efficiency, shown in Figure 2.4 (left), is about 90% at 1 TeV for reconstructed and identified muons[3] as the ones used in the search. Figure 2.4 (right) shows the same efficiency for muons with $p_T > 53$ GeV as a function of the pseudo-rapidity, η.

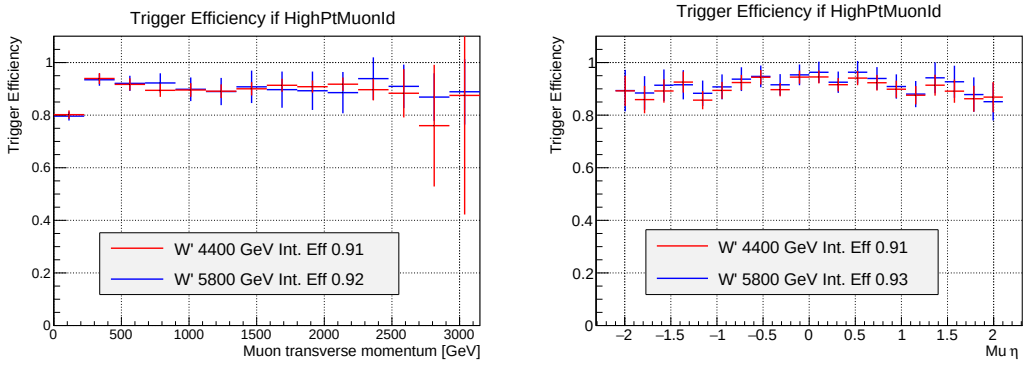

FIGURE 2.4: Muon trigger efficiencies as a function of the generated p_T for a couple of W' samples (left) for well identified High-p_T muons reconstructed in $|\eta| < 2.4$. Same efficiency as a function of η for muons with $p_T > 53$ GeV (right).

Simulations are not always perfect and their performance needs to be calibrated from measurements in data. This is done using di-muon pais coming from $Z \to \mu\mu$ events using the Tag & Probe technique [52].

In the Tag & Probe technique, dimuon events with opposite charge are selected with an invariant mass in the range $m_{\mu\mu} \in [70, 110]$ GeV. Additionally, both muons have to be identified as good muons for physics. Once they are selected, one of the muons is called Tag muon and the other is called the Probe muon. In the case of the Tag muon, it has to geometrically match with a reconstructed HLT_Mu_50 muon object. This way, the muon trigger efficiency is computed as the fraction of events in which the Probe

[3]Muon reconstruction and identification will be defined more precisely in Section 2.3.3

muon is also matched with a different HLT_Mu_50 object. Often these efficiencies are computed separately in data and simulation as a function of the pseudo-rapidity and p_T of the Probe muon. For each of them, a calibration factor defined as the ratio between the efficiency in data and simulation, $\epsilon_{Data}^{Trigger}/\epsilon_{MC}^{Trigger}$, is evaluated. In the case of the trigger used in the analysis: HLT_Mu_50, the calibration factors are shown in Table 2.3 as a function of the p_T and η of the Probe muon.

Probe muon p_T	Probe muon η	$\epsilon_{Data}^{Trigger}/\epsilon_{MC}^{Trigger}$
53 GeV $< p_T <$ 60 GeV	$0.9 > \lvert\eta\rvert > 0$	0.9655 ± 0.0015
	$1.2 > \lvert\eta\rvert > 0.9$	0.9533 ± 0.0040
	$2.1 > \lvert\eta\rvert > 1.2$	0.9568 ± 0.0023
	$2.4 > \lvert\eta\rvert > 2.1$	0.8902 ± 0.0072
$p_T >$ 60 GeV	$0.9 > \lvert\eta\rvert > 0$	0.9708 ± 0.0001
	$1.2 > \lvert\eta\rvert > 0.9$	0.9528 ± 0.0040
	$2.1 > \lvert\eta\rvert > 1.2$	0.9712 ± 0.0023
	$2.4 > \lvert\eta\rvert > 2.1$	0.9417 ± 0.0072

TABLE 2.3: Calibration of the trigger used in the W′ search: HLT_Mu_50. The calibration factors for the trigger are defined as the ratio of the trigger efficiency in data with respect to the same value in simulation. Efficiencies were derived in dimuon events using a Tag & Probe technique. Quoted uncertainties are only statistical. Credit of the study: preliminary results from CMS Muon Performance Group.

The HLT_Mu_50 trigger efficiencies are about $3\% - 5\%$ higher in simulation depending on the pseudo-rapidity region. Trigger efficiencies in simulation are corrected at the analysis level to improve the modelling of the simulations. An additional uncertainty of 5-8% depending on the p_T of the muon is assigned to the extrapolation to higher p_T of the calibration factors listed in Table 2.3.

2.3.2 Primary vertex

In the 2015 (2012) run at the LHC, for each bunch crossing there were on average about 10 (21) proton-proton interactions. This number of interactions was lower at 13 TeV than at 8 TeV and the explanation for this is that the instantaneous luminosity was not high enough in most of 2015 run as shown in Figure 2.5. However it is expected in 2016 that the LHC will surpass the barrier of $\mathcal{L} = 10^{34}\text{cm}^{-2}\text{s}^{-2}$ that will create a pileup higher than 40 proton-proton interactions per bunch crossing.

Fortunately the CMS tracker is able to reconstruct tracks coming from different proton-proton interactions along the z axis, and reconstruct the geometrical position of each of the interaction vertices. A minimum of 4 tracks is needed to form a reconstructed vertex, however not all the reconstructed vertices are interesting for the W′ search. Only the vertex associated to the muon that triggered the event will be considered. This vertex

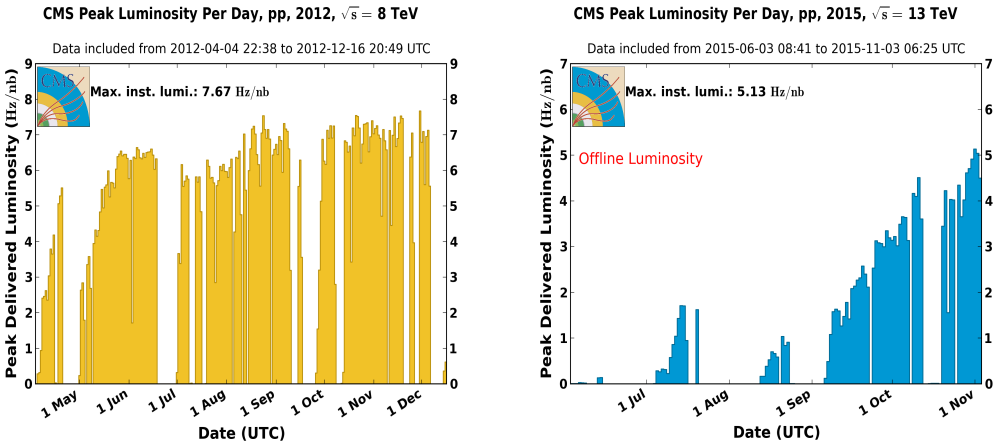

FIGURE 2.5: Left (Right) Evolution of the maximum instantaneous luminosity (reached at the beginning of the fill) along 2012(2015) in the 8(13) TeV run.

receives the name of primary vertex, and all the other reconstructed vertices are called pileup vertices. Figure 2.6 shows the number of different reconstructed vertices along the axis parallel to the beam line in a $W' \to \mu\nu$ candidate event. Monte Carlo simulations

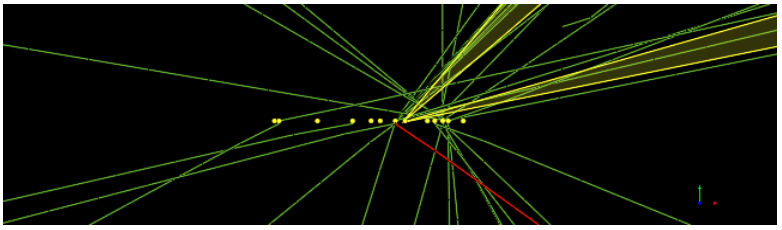

FIGURE 2.6: Event display that shows the number of reconstructed vertices (yellow dots) along the z-axis in a W' candidate event with a total pileup of $< PU >= 13$. The muon track candidate is shown in red and green lines correspond to all reconstructed tracks with $p_T > 1.5$ GeV. The yellow areas correspond to reconstructed jets.

are produced with the closest knowledge of the number of proton proton interactions at the time of the generation of the samples. Unfortunately, this variable is hard to predict since it strongly depends on the instantaneous luminosity delivered by the LHC, as well as on the proton-proton cross section that is known with a 5% uncertainty. All the simulations are corrected in order to match the number of proton-proton interactions with the ones measured in data. This correction is done using a re-weighting function denoted as $C(N_{PU})$ in the text. Figure 2.7 shows the distribution of reconstructed vertices in data and simulation in dimuon events before (left) and after (right) applying the re-weighting function.

2.3.3 Muon reconstruction and identification

Muon candidates are reconstructed independently in the inner tracker (tracker-track) and in the muon system (stand-alone track). These two reconstructions are combined

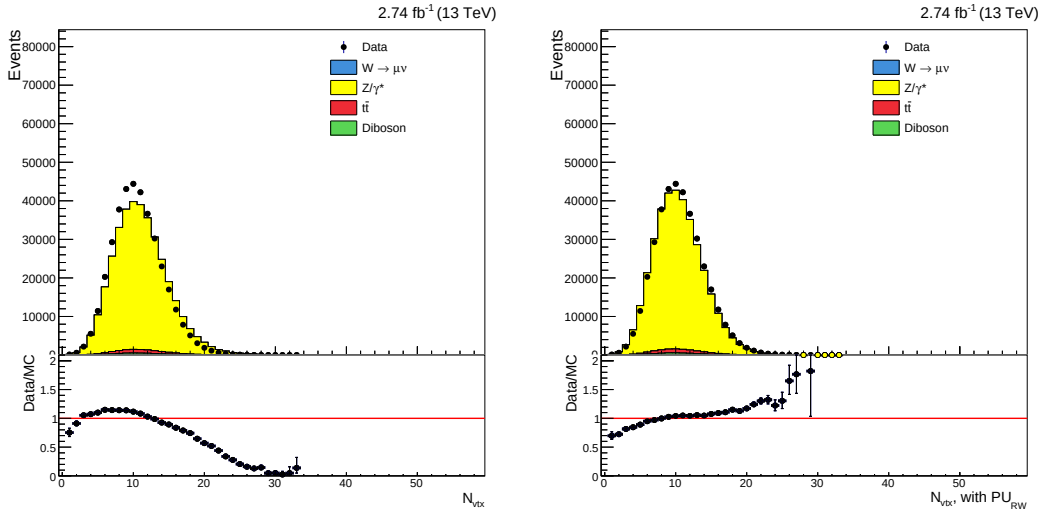

FIGURE 2.7: Number of reconstructed vertices in a preselected $\mu^+\mu^-$ sample in data and simulation, before (left) and after (right) application of the re-weight. The average number of reconstructed vertices in the sample is $< PU >= 10$.

into a common track that is called Global Muon track reconstruction. Under this reconstruction, Global Muons are built following an outside-in reconstruction where starting from a stand-alone track, a tracker track is found when propagating it back to the inner tracker detector. After both tracks candidates are geometrically matched, the Global Muon track is built using a Kalman-filtering technique [85] taking as input the list of hits from both tracks. The choice of the Global Muon reconstruction is optimal because it reduces the rate of fake tracks and at the same time brings a gain in detector length by adding the hits found in the muon system. The latter improves the muon momentum resolution for high p_{T} muons, particularly above 200 GeV.

Due to the high efficiency in the track reconstruction in the tracker system and segment reconstruction in the muon system, more than 97% of the muons are reconstructed within the acceptance of the muon system, $|\eta| < 2.4$. Figure 2.8 shows the reconstruction efficiency in simulation with respect to generated muons for different W' samples up to several TeV.

Once Global Muons are reconstructed, dedicated identification criteria are applied to ensure that the track candidate is in fact a real muon. The goal of the identification is to reduce the rate of fake, bad or misreconstruced muons while keeping the efficiency of W' candidates as high as possible.

The muon identification definition used in the W' search is optimized for high p_{T} muons, and is referred as High-p_{T} muon identification in the literature [52]. Under this definition, a High-p_{T} muon is required to have at least one hit in the pixel detector, at least six tracker layer hits and segments in two or more muon detector layers. Since segments

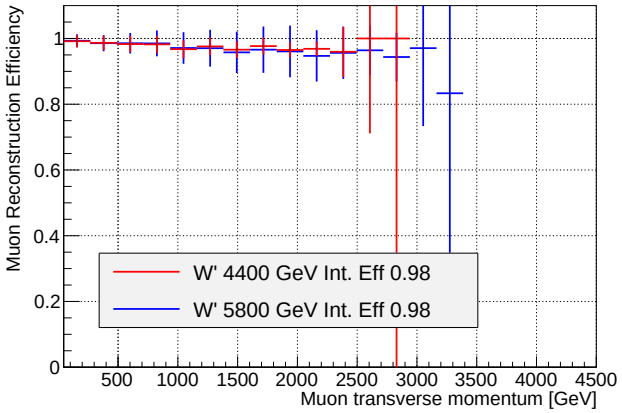

FIGURE 2.8: Global muon reconstruction efficiency for generated $W' \to \mu\nu$ muons in the acceptance as a function of the generated muon p_T for different W' samples.

are typically found in consecutive layers separated by thick layers of steel, the latter requirement significantly reduces the amount of hadronic punch-through (pions and to a less extent kaons that escape the HCAL). To reduce background from cosmic ray muons, each muon is required to have a transverse (longitudinal) impact parameter of the tracker track of less than 0.02 cm (0.5 cm). Both parameters are defined relative to the primary vertex. In order to suppress muons with mismeasured p_T, an additional requirement $\sigma_{p_T}/p_T < 0.3$ is applied, where σ_{p_T} is the uncertainty from the track reconstruction. Finally, the muon pseudo-rapidity must be $|\eta| < 2.4$ and the reconstructed muon p_T must be larger than 53 GeV, where the HLT_Mu50 trigger is already fully efficient.

Muons from W' candidates are expected to be isolated from hadronic activity. Due to this, a dedicated muon isolation criteria is applied to reduce the contribution of secondary muons produced in hadron decays. This contribution is efficiently reduced by requiring that the scalar sum of the p_T of all tracks originating from the primary vertex within a $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} < 0.3$ cone around its direction, excluding the muon itself, to be less than 10% of the muon p_T.

The High-p_T muon identification efficiency for muons from a W' in simulation is found to be larger than 97% for TeV muons, as shown in Figure 2.9. Contributions from other objects such as: jets, cosmic muons, electrons or photons that could mimic the signal of a High-p_T muon are found to be negligible compared to backgrounds that involve real muons from collision events. The low rate of QCD background processes that could mimic a High-p_T muon signal, makes final states with isolated muons golden channels to look for new physics in hadronic colliders.

As in the case of the trigger calibration in data described in Section 2.3.1, the calibration and stability of the High-p_T muon identification is done with an adapted version of the Tag & Probe technique with a slightly different definition of the Probe muon. The reason

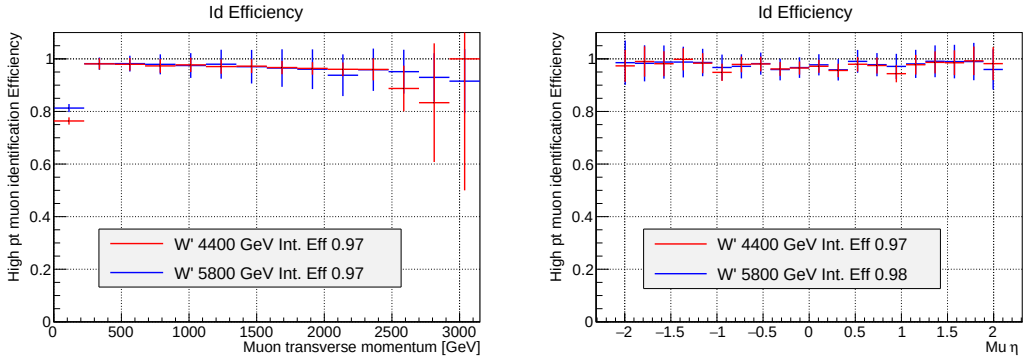

FIGURE 2.9: High-p_T muon identification efficiency for reconstructed muons as a function of the generated muon p_T for different W′ samples (left), and as a function of η for muons with > 53 GeV (right).

for this change is that the identification criteria cannot be applied since is the element to calibrate. In order to measure the muon identification in data, the Probe muon is defined as a Global Muon track with p_T more than 20 GeV. The Tag muon and the rest of the event selection remains unchanged. This way, the High-p_T muon efficiency is computed as the fraction of events in which the Probe muon is additionally identified as High-p_T muon. The High-p_T muon calibration factor, defined as the ratio between the efficiency in data and simulation, $\epsilon^{ID}_{Data}/\epsilon^{ID}_{MC}$, is shown in Table 2.4 as a function of the p_T and η of the Probe muon.

Probe muon p_T	Probe muon η	$\epsilon^{ID}_{Data}/\epsilon^{ID}_{MC}$
53 GeV $<p_T <$ 80 GeV	$\lvert\eta\rvert < 1.2$	0.9869 ± 0.0012
	$2.1 < \lvert\eta\rvert < 1.2$	0.9790 ± 0.0052
	$2.4 < \lvert\eta\rvert < 2.1$	0.958 ± 0.028
80 GeV $<p_T <$ 120 GeV	$\lvert\eta\rvert < 1.2$	0.9984 ± 0.0013
	$2.1 < \lvert\eta\rvert < 1.2$	0.9892 ± 0.0069
	$2.4 < \lvert\eta\rvert < 2.1$	1.00 ± 0.023
120 GeV $<p_T <$ 200 GeV	$\lvert\eta\rvert < 1.2$	0.9764 ± 0.0045
	$2.1 < \lvert\eta\rvert < 1.2$	0.945 ± 0.025
	$2.4 < \lvert\eta\rvert < 2.1$	1.001 ± 0.063

TABLE 2.4: High-p_T muon identification calibration factors used in the W′ search at 13 TeV. The calibration factors for the muon identification are defined as the ratio of the muon identification efficiency in data with respect to the same value in simulation. Efficiencies were derived in dimuon events using a Tag & Probe technique. The quoted uncertainties are only statistical. Credit of the study: preliminary results from CMS Muon Performance Group.

The identification efficiencies are about $1\% - 2\%$ higher in simulation depending on the pseudo-rapidity region. The High-p_T muon identification efficiencies in simulation are corrected at the analysis level to improve the modelling of the simulations. An additional uncertainty of 2-3% depending on the p_T of the muon is assigned to the extrapolation to higher p_T of the calibration factors in Table 2.4.

2.3.4 Muon p_T assignment and performance

Experimentally the muon momentum measurement is an indirect measurement that depends on the magnetic field, B, and the radius of the reconstructed track, R.

$$p_T[GeV] = 0.3 \cdot B[\mathrm{T}] \cdot R[\mathrm{m}] \tag{2.7}$$

The magnetic field is monitored with high precision in the different parts of the detector and its maximum, $B = 3.8$ T, is reached inside the tracker volume. The measurement of R depends on the physical observable sagitta, s, defined in Figure 2.10 (purple line), and on the path length of the reconstructed track, L.

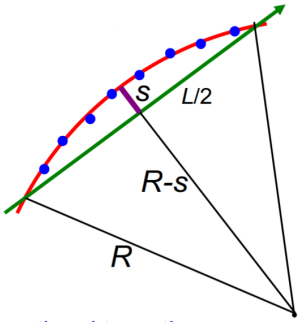

FIGURE 2.10: Sketch that illustrates the definition of the observable sagitta of a track.

Geometrically, saggita, the radius, and the path length are related as

$$R \approx L^2/2s. \tag{2.8}$$

In Equation 2.7, sagitta is proportional to $1/p_{\mathrm{T}}$. This makes the curvature, $\kappa = q/p_{\mathrm{T}}$, instead of p_{T} the relevant quantity for muon momentum performance studies.

Inspecting Equations 2.7 and 2.8, rapidly one sees the problem for high p_{T} tracks or muons. For a fixed magnetic field, the radius grows proportionally to the p_{T} of the muon and given a fixed length of the detector L, $s \to 0$ when the p_{T} is large. In other words, due to the smallness of the real physical observable sagitta at high p_{T}, the muon momentum assignment is sensitive to mismeasurements or small variations in the measurement of s.

As described in the muon reconstruction (Section 2.3.3), the Global Muon track reconstruction relies on the matching of two independent tracks reconstructed in two different sub-detectors (tracker track and stand-alone track). Both detectors are separated more than 3 meters and the best performance in the muon momentum measurement is reached when this information is combined efficiently. The remaining of this section presents relevant aspects that affect the performance of the high p_{T} muon momentum measurement.

As already stated, due to the smallness of sagitta at high p_T, the most important aspect that affects the muon p_T performance is the detector alignment. This involves: internal tracker alignment, internal muon system alignment and the interplay between both systems. Internal tracker alignment dominates at moderate $p_T < 200$ GeV. However, at the region of interest in p_T for the W' search, the interplay between tracker and muon system alignment determines the ultimate performance at the TeV scale.

Before the start of the proton-proton run in June 2015, detectors were aligned from a direct measurement of the positions of individual sub-detectors, as well as with tracks coming from special runs of cosmic muons at 0 T and 3.8 T taken in March 2015. This alignment is often referred in the literature as *start-up* alignment and corresponds to the first dataset before re-calibrations. As the data taking progresses individual detectors are further aligned using tracks from collision, for example including in the alignment procedure the input of known resonances such as $Z \to \mu\mu$ events, or adapting the alignment to the magnet cycles or movements of the individual sub-detectors. Typically a good detector alignment is achieved when about 1 fb^{-1} of collision data is used in the alignment procedure. This stage is often called *asymptotic* alignment in the literature. The dataset analysed for the W' search has been re-processed at the end of 2015 to include the best detector alignment conditions, and is close to the *asymptotic* alignment.

Additionally, as muons pass through the steel of the magnet return yoke that sustains the muon system, they could emit radiation that alters the muon trajectory and produce electromagnetic showers in the muon chambers. The latter gives rise to additional fake hits and worsen the performance of the momentum as measured directly from the Global Muon track. In order to mitigate this effect, several dedicated high p_T muon assignments, hereafter denoted as TeV re-fits, are computed for each muon and the final p_T assignment comes from a case-by-case decision. Currently four TeV re-fits are considered in the muon momentum assignment, with different techniques to mitigate the impact of fake hits coming from electromagnetic showers in the muon system:

The *Picky* algorithm starts with the full hit list of the Global-Muon track, but re-fits the extrapolated tracker trajectory excluding individual hits in chambers with anomalously large hit occupancy. This is done to avoid hits from potential electromagnetic showers.

The *Dynamical Truncation* (DYT) algorithm is designed to avoid hits after a large energy loss in the muon system. This is done computing an estimator, E_i, for every muon station, i, with reconstructed segments in the Global Track. This estimator is based on the compatibility of the propagated track from the tracker system to the muon station $i-1$, to the reconstructed segment in the muon station i. Based on this estimator, the algorithm decides to skip individual chambers in the re-fit, or if consecutive chambers present large incompatibilities, directly truncates the re-fit at a certain station.

The *Tracker Plus First Muon Station* (TPFMS) algorithm re-fits the Global Muon track excluding all the hits in the muon system except the ones in the innermost layer of the muon station. This is done aiming to reduce to the minimum the bias of using hits in the neighbouring area of a shower. Additionally, this re-fit particularly protects the muon momentum against large multiple scattering deviations in the muon system as well as potential misalignments between different muon stations.

Finally the *Tracker-only*, takes the muon p_{T} assignment as measured by the tracker detector alone. This measurement is not sensitive to muon chamber misalignments, however the performance degrades rapidly with the p_{T}. Due to this, the tracker assignment is essentially considered below 200 GeV and rarely used at higher momentum.

The best performance is finally obtained with the *Tune-P* assignment [52], an algorithm that is designed to choose in a muon-by-muon basis the best muon p_{T} among the different TeV re-fits. The choice is based in terms of a weighted comparison between the χ^2 of the different re-fits. Currently the re-fits that enter in the *Tune-P* decision are: Picky, DYT, TPFMS and the Tracker-only assignment. Despite the fact that four algorithms enter in *Tune-P* for muons with p_{T} above 200 GeV, more than 90% of the times the choice of the *Tune-P* decision is only between Picky or DYT. Searches in CMS and particularly in the presented W' search, use the muon p_{T} assignment based on the *Tune-P* decision algorithm. The *Tune-P* and individual TeV re-fits performance with the 2015 conditions are presented in the muon resolution and scale Sections 2.3.5 and 2.3.6 respectively.

2.3.5 Muon momentum resolution

The easiest and most efficient way to measure the muon resolution in data and calibrate simulations, is to select dimuon events coming from well known-resonances with well understood mass and width. At low transverse momentum, $p_{\mathrm{T}} < 15$ GeV, in the range of interest for b-physics, muons are calibrated using J/ψ and Υ resonances. For intermediate momentum, 15 GeV $< p_{\mathrm{T}} <$ 100 GeV, muons from Z decays are used instead. Different resonances at different masses are needed to calibrate muons against different physical process that could affect the muon measurement. For example, J/ψ and Υ resonances cover the region where multiple scattering is dominant and are useful to calibrate the description of the tracker material. In the other hand, muons from Z decays are less affected by the tracker material and are more efficient to calibrate potential misalignment in the tracker. Typically with these three resonances, most of the muons used in CMS are calibrated with high precision. This is the standard method to calibrate muons in CMS and it was employed in the $\mathrm{Z} + \mathrm{c}$ analysis [45, 51].

However, searches for heavy resonances such as $W' \to \mu\nu$ production employ muons with p_T larger than 200 GeV up to several TeV. At this regime, only a good internal tracker alignment is not sufficient for the best performance. The proper alignment between tracker detector plus muon system is what improves the performance at very high momentum. Unfortunately, at the TeV regime there are no resonances *(yet)* in which to calibrate the leptons and the *Tune-P* algorithm. As a consequence, the calibration of High-p_T muons is experimentally challenging because it requires new techniques that do not rely in resonances to probe the highest p_T possible. Particularly, in this manuscript two control regions are employed. The first control region uses cosmic muons and the second control region uses high p_T muons from large $p_T(Z)$ or high mass Z/γ^* events.

Cosmic muons were taken in a dedicated data-taking campaign with the magnet field at 3.8 Tesla called CRAFT and during pp collisions when there is no beam. The advantage of cosmic muons is that the slope of the falling p_T spectrum from cosmic muons is softer than the same slope from dimuons from collision. As a consequence, cosmic muons are an easy and powerful source of TeV muons in CMS for calibration.

Muons produced in collision events cross the detector starting from the primary vertex to the exterior of the detector. In contrast, cosmic muons cross the detector vertically leaving a trace in the upper-half as well as in the lower-half of CMS. Due to this, a dedicated muon reconstruction is employed in cosmic muon studies. A sketch of the cosmic muon reconstruction is shown in Figure 2.11. The cosmic muon reconstruction is called 2-cosmic-leg in which each cosmic muon is reconstructed as two separate muons that correspond to the upper-half track and the lower-half track respectively. In order to reject false cosmic muon tracks, both cosmic-legs need to match geometrically in η and ϕ, and each track separately has to be identified has a High-p_T muon. As a consequence, each muon is reconstructed, identified and its momentum measured twice independently.

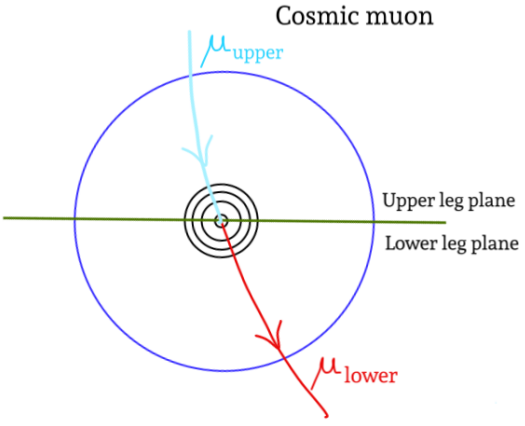

FIGURE 2.11: Sketch of the cosmic muon two leg reconstruction. Each cosmic muon is reconstructed twice (red and light blue tracks), and its transverse momentum is also measured twice independently.

This way, the resolution in the muon curvature, $1/p_{\mathrm{T}}$, is obtained from a Gaussian fit to the relative residual $R(q/p_{\mathrm{T}})$ distribution obtained from the $1/p_{\mathrm{T}}$ measurement in the upper and lower half separately.

$$R(q/p_{\mathrm{T}}) = \frac{(q/p_{\mathrm{T}})_{upper} - (q/p_{\mathrm{T}})_{lower}}{\sqrt{2}(q/p_{\mathrm{T}})_{lower}} \tag{2.9}$$

The $\sqrt{2}$ factor comes due to the fact that the two fits are independent. The lower half measurement is taken as reference since is closer to the proton-proton reconstruction. The validity of $R(q/p_{\mathrm{T}})$ to represent the curvature resolution has been validated with the simulation. Additionally, it has been checked in the simulation that the upper and lower leg have similar resolution. The drawback of this method is that probes only vertical muons in the pseudo-rapidity range of $|\eta| < 0.9$. Fortunately, the central region of the detector is where most of the muons from a heavy W' candidates are expected according to the simulated events.

The width of $R(q/p_{\mathrm{T}})$ as a function of the reconstructed p_{T} of the lower cosmic muon leg for different muon momentum assignments up to 1 TeV is shown in Figure 2.12 (left). As an example of one of the fits, the distribution of $R(q/p_{\mathrm{T}})$ for cosmic muons in the p_{T} range between 420 GeV and 2 TeV is shown in Figure 2.12 (right). Measured values were found in agreement with the performance in the simulation.

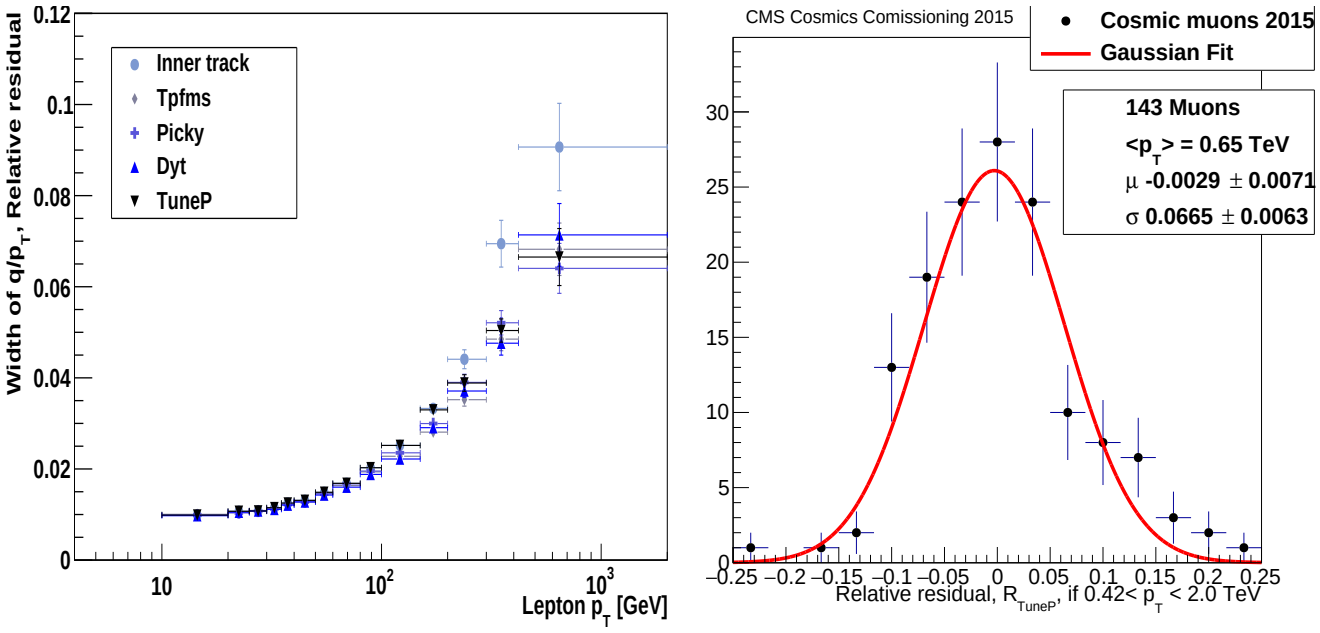

FIGURE 2.12: Width of the relative inverse p_{T} distribution, $R(q/p_{\mathrm{T}})$, evaluated with cosmic muons taking during the CRAFT15 for different the different TeV re-fits (left). Example of the Gaussian fit to $R(q/p_{\mathrm{T}})$ using *Tune-P* muon momentum assignment for muons between 420 GeV and 2 TeV (right).

2.3.6 Muon momentum scale

In situations where different sub-detectors are not perfectly aligned, the measurement of the muon curvature could be biased. Figure 2.13 (left) shows an example of a rotated geometry that could bias the muon momentum measurement. The difference between the reconstructed curvature of the muon, $\kappa_{measured}(\eta, \phi)$, and the real curvature, $\kappa_{real}(\eta, \phi)$, is called the curvature bias, $\kappa_b(\eta, \phi)$.

$$\kappa_b(\eta, \phi) = \kappa_{measured}(\eta, \phi) - \kappa_{real}(\eta, \phi) \tag{2.10}$$

In a perfectly aligned detector the average bias should be small and consistent with zero, $< \kappa_b(\eta, \phi) >= 0$. In this section, the study of the muon momentum scale refers to the study and calibration of the regions of the detector (if any) in which $< \kappa_b(\eta, \phi) > \neq 0$.

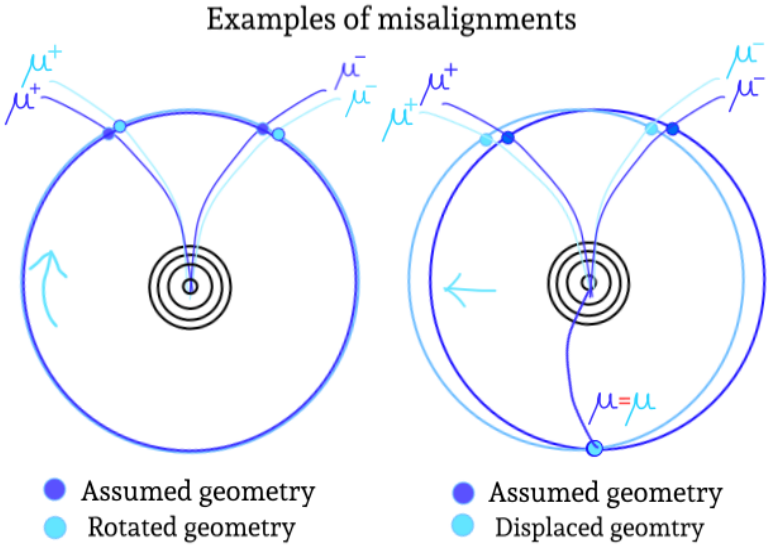

FIGURE 2.13: Sketch that shows academic examples of the effect of different misaligned geometries in a reconstructed track. In the left panel, an hypothetical muon system has been rotated in the $r\phi$ plane (light blue). This type of geometry could potentially cause a constant global bias along ϕ. The right panel, shows a shift in the transverse plane, x, y. This type of distortion could cause a potential local bias that strongly depends on the ϕ coordinate. In both examples, the misaligned geometry affects differently positive and negative tracks (μ^- and μ^+).

It is especially important at the beginning of each data-taking period, particularly after opening/closing the wheels of CMS, to quantify the size of potential biases in the reconstruction. For example, there are geometrical regions in which is more complicated to align the detector like the endcaps, where it is important to ensure $< \kappa_b(\eta, \phi) > \neq 0$. In general, biases are introduced when the assumed geometry of the detector in the reconstruction of the data, is different than the real geometry of the detector. Owing to its geometrical nature, potential biases are evaluated separately in η and ϕ regions.

At the p_T range of interest used in most of the analysis in CMS, small non-zero-biases can be corrected and marginally affect the performance of the muon momentum measurement. However, searches such as $W' \to \mu\nu$ deal with almost zero backgrounds at high mass and three or four events are sufficient to claim a discovery. In this case, the understanding of potential biases in the reconstruction is crucial to claim a discovery.

In order to put in context the typical size of biases and their potential impact at high p_T, Table 2.5 illustrates the effect in the measured momentum for different p_T ranges assuming a moderate bias of $\kappa_b = 0.05$ c/TeV. Due to the smallness of the curvature at high p_T, the bias in the reconstruction has higher impact for TeV muons originated in the decay of a heavy W'. The motivation of the choice of 0.05 c/TeV in the examples of the table, is because during Run I the precision in the determination of biases in the reconstruction was of the order of 5% at 1 TeV.

Typical Range in p_T	$< p_T >$	$< p_T >_{biased}$
SM measurement	25 GeV	25 GeV
Highest p_T at 8 TeV W' data	600 GeV	700 GeV
Run I W' exclusion	1600 GeV	1400 GeV
Expected W' sensitivity	2000 GeV	1600 GeV

TABLE 2.5: Example to illustrate the effect of a curvature bias $\kappa_b = 0.05$ c/TeV, for different p_T cases.

One important characteristic is that potential detector misalignments usually affect differently positive and negative tracks, as shown in the same Figure 2.13. In a misaligned detector, positively-charged muons tend to increase (decrease) their curvature whereas negatively-charged muons to decrease (increase).

The Generalized Endpoint (GE) method developed in this thesis, aims to quantify reconstruction biases, such as the ones described in Equation 2.10. The sample used for the study is enriched in high mass Z^* and high $p_T(Z)$ events by selecting high p_T muons from $\mu^+\mu^-$ events.

The analysis selection uses only events with at least two identified High-p_T muons, in which at least one of them has p_T larger than 200 GeV. In order to probe separately biases in barrel (DT) and each endcap (CSC), the analysis is done separately in three independent categories: muons with $|\eta| < 1.2$, $\eta > 1.2$ and $\eta < -1.2$. The strategy is simple, the method starts with a well known Z/γ^* simulation from POWHEG generated with artificially injected biases ranging from $< \kappa_b >= [-1.0\ c/\text{TeV}, 1.0\ c/\text{TeV}]$ in steps of 0.01 c/TeV. The discriminant variable for the study will be the muon curvature and the p_T assignment returned by the *Tune-P* algorithm. The latter is shown in Figure 2.14 for the selected events in data and a simulation with $\kappa_b = 0$.

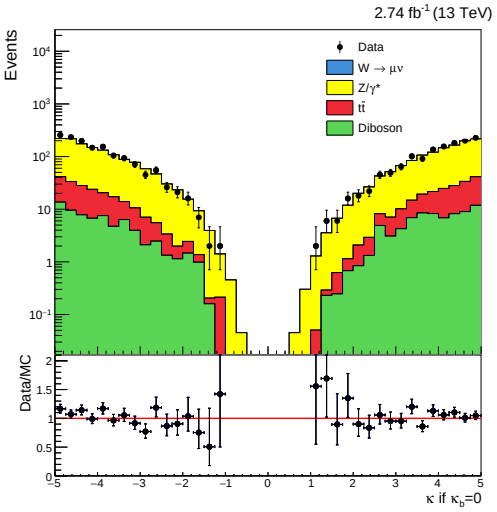

FIGURE 2.14: Data to simulation comparison of the curvature distribution in $\mu^+\mu^-$ events for muons with more than 100 GeV. Negative curvature corresponds to the μ^- and positive curvature corresponds to μ^+.

A χ^2 test is performed between the curvature distribution in data and the curvature distribution for each of the simulated samples as a function of injected bias κ_b. The result of the χ^2 test for the three η categories is shown in Figure 2.15. In the three cases, the measured κ_b is obtained at the minimum of a fitted 2^{th} order polynomial to the χ^2 distribution and the uncertainty is taken from $\Delta\chi^2 = 1$. As a result, in the three pseudo-rapidity categories, κ_b is consistent with zero (Table 2.6) and large structure misalignments that could cause a global scale bias in barrel or in the endcaps are disfavoured by the measurement.

Region in detector	$< \kappa_b >$ [c/TeV]
$\lvert\eta\rvert < 1.2$	-0.01 ± 0.03
$\eta > 1.2$	0.01 ± 0.05
$\eta < -1.2$	-0.01 ± 0.06

TABLE 2.6: Measured k_b in the in the pseudo-rapidity $|\eta| < 1.2$, $\eta > 1.2$ and $\eta < -1.2$ using muons with $p_{\text{T}} > 200$ GeV. Quoted uncertainties refer only to statistical errors.

The reach in p_{T} of the Generalized Endpoint method is statistically limited by the amount of data collected in 2015. Due to this, an independent cross-check study was performed with the comic dataset used for resolution studies using an adapted version of the Endpoint Method using cosmic muons [51]. In this case the bias κ_b is defined as the difference in curvature for positive and negative muons. The result of this study is shown in Figure 2.16 and also disfavours large scale biases in the region $|\eta| < 1.2$, thus reinforcing the results of the Generalized Endpoint method.

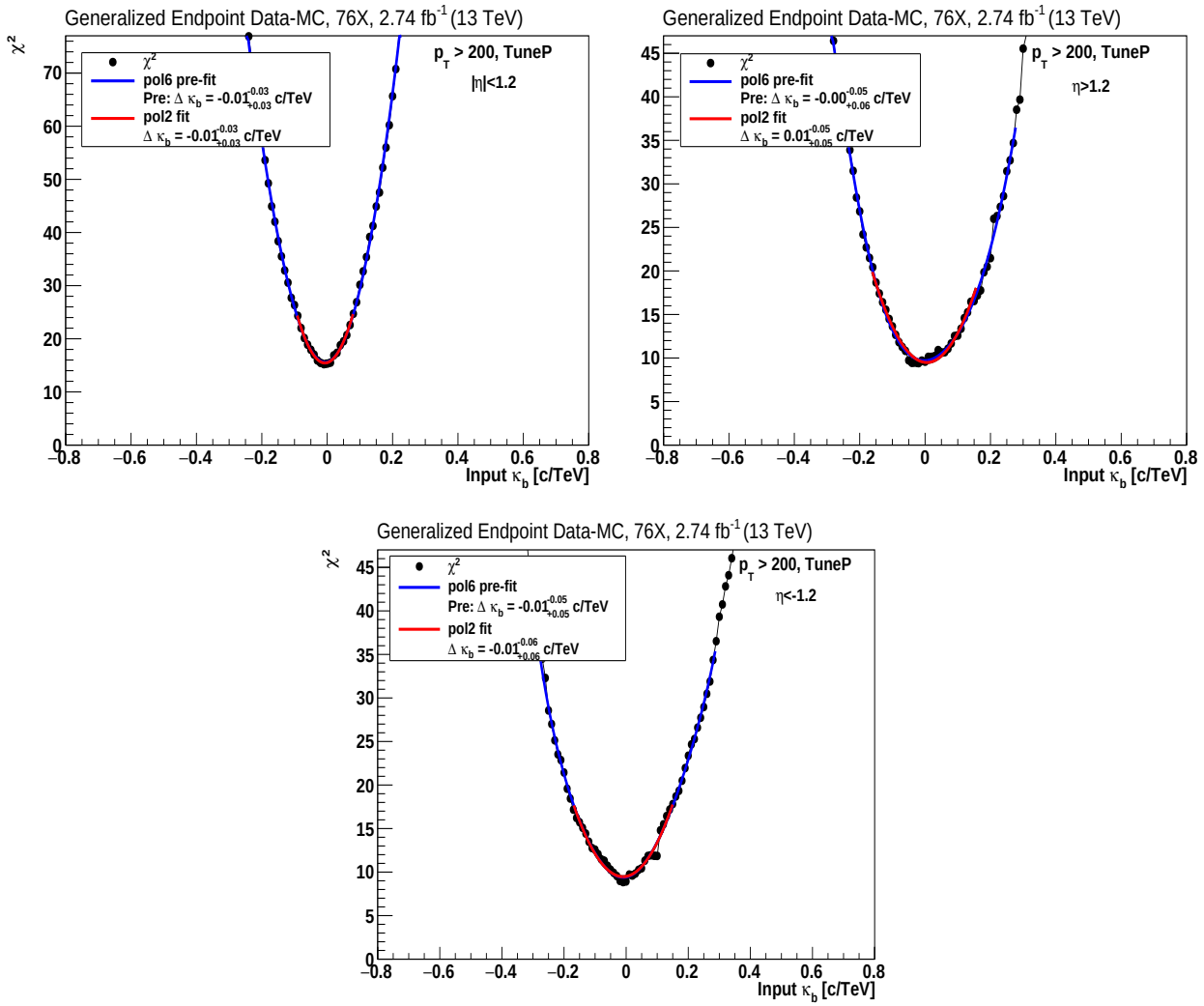

FIGURE 2.15: Measurement of the κ_b via χ^2 minimization for High p_{T} identified muons above 200 GeV in the pseudo-rapidity $|\eta| < 1.2$ (top, left), $\eta > 1.2$ (top, right) and $\eta < -1.2$ (bottom).

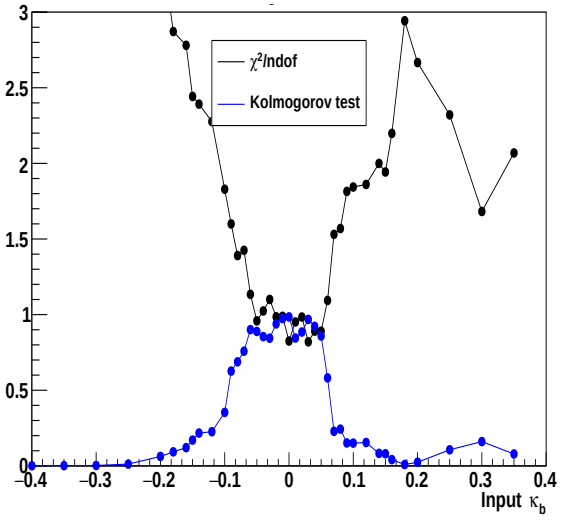

FIGURE 2.16: Measurement of κ_b via $\chi^2/n.d.of$ minimization for High p_{T} identified muons above 200 GeV using cosmic muons from CRAFT15.

Misalignments could be rare and complicated for a detector of the size of CMS, and certain types of distortions could induce large biases that strongly depend on the ϕ coordinate in a particular η region. This biases will be denoted in the text as *local* scale biases and Figure 2.13 (right) shows a sketch of one them. In order to quantify them, the same procedure is used extending the number of (η, ϕ) categories. Due to the limited statistics when enlarging the number of categories, the p_T thresholds for the study have been lowered to 110 GeV taking as reference the Picky assignment. Only after lowering the p_T thresholds, the sample can be further divided into a grid of η and ϕ bins. The measured κ_b for each of the η, ϕ regions is shown in Figure 2.17.

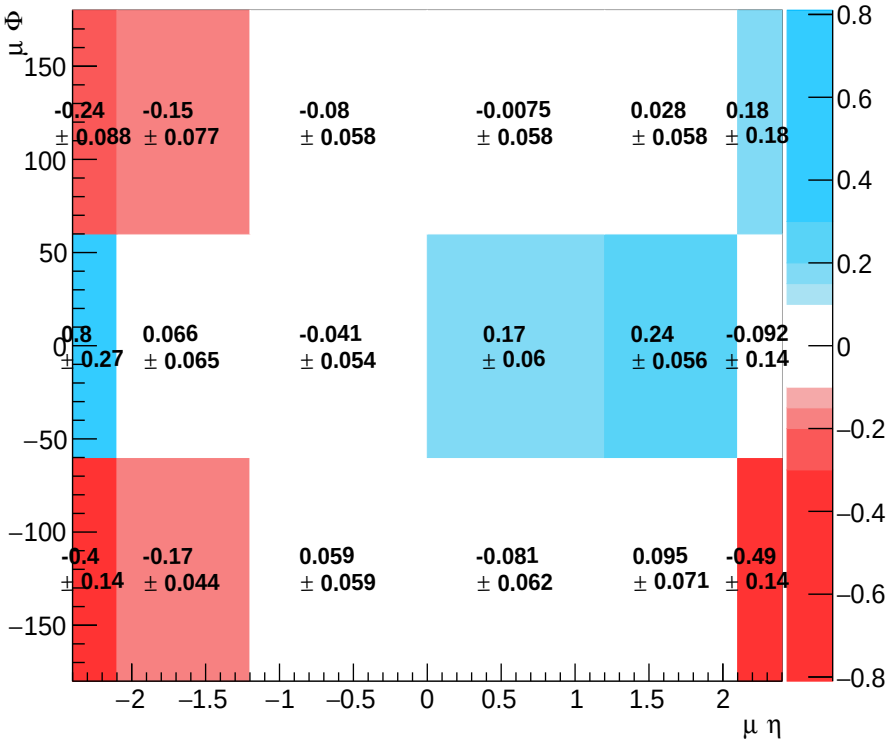

FIGURE 2.17: Measurement of potential local scale effects due to misalignment for muons above 110 GeV in a grid of (η, ϕ) regions. Quoted uncertainties refer only to statistical errors.

According to the results shown in the figure, local biases in the $|\eta| < 1.2$ region, if existing, are not large and do not show clear patterns along ϕ. Due to this, the ϕ independent measurement of $|\eta| < 1.2$ in Table 2.6 is less affected by statistical fluctuations and is considered as reference. In the forward endcaps $|\eta| > 2.0$, particularly in the negative endcap, local biases present a sinusoidal shape that could be non-negligible and affect the muon p_T measurement. The simulation was corrected to reproduce this misalignment.

As a result of the studies with the Generalized Endpoint and with cosmic muons, potential local and global scale biases as big as the ones described in this section, are included as sources of systematic uncertainties in the W' search and will be discussed in Section 2.4.6.

2.3.7 Jets

Jets are produced in the hadronization of quarks and gluons and are the most abundant objects produced in CMS. They are reconstructed by merging energy deposits obtained by combining tracker and calorimeter information using the anti-k_T algorithm [86] with radius $R = 0.4$ at 13 TeV and $R = 0.5$ at 8 TeV. In practice, the energy measured from the reconstructed jets does not correspond to the true particle-level energy and needs to be corrected. For example, the calorimeter response is not linear in p_{T} or η and pileup jets could affect the energy measurement. These effects are mitigated with a dedicated set of jet energy corrections applied to the measured energy that enable a mapping with the particle-level energy. The dominant corrections, derived separately for data and simulation, subtract the energy coming from pileup jets and flatten the jet response in η and p_{T}. Additional residual corrections are applied only in data to fix remaining differences in jet response between data and the simulation.

Often the flavour of the primary quark from which the jet originated may be determined thanks to the excellent tracking capabilities of the CMS tracker. Particularly, jets originating from b-quarks, referred as b-jets, contain tracks with different kinematical properties that can be exploited for identification purposes. For example, B-hadrons produced inside b-jets often have longer-lifetime and mass than hadrons produced in other jets, yielding to larger number of high momentum displaced tracks with respect to the primary vertex. Often the combination of displaced tracks under a common fit allows the reconstruction of a displaced vertex, called secondary vertex, originated at the decay point of the hadron.

CMS has developed a set of dedicated algorithms, called b-tagging algoritms, that exploit these differences to identify b-jets efficiently from other jets. As an example, one of the most-widely used b-tagging algorithms in CMS, the Combined Secondary Vertex (CSV) [87], uses a multivariate analysis technique that exploits the different properties of tracks and secondary vertices in jets to identify b-jets. The CSV algorithm has a 70% b-tagging efficiency with only a 1% of misidentification rate of jets originated in the hadronization of up, down or gluons. Heavy flavour tagging will be discussed in more detail in the context of the $\mathrm{Z} + \mathrm{c}$ measurement in Section 3.3.3.

Despite jets and b-tagging techniques are widely used in many analysis in CMS, the $W' \to \mu\nu$ search is expected to have little contribution from additional jets, and the hadronic activity in the event will affect mostly the MET computation as described in the next section.

Finally secondary muons originated in the semileptonic decay of hadrons are produced inside jets and not used in the search. In general those muons are surrounded by large

hadronic activity preventing them to be identified as High-p_T muons. The contribution of non-isolated muons or jets misidentified as High-p_T muons is negligible with respect to other backgrounds and is considered to have no impact for the search.

2.3.8 MET

W bosons in their leptonic decay always produce a neutrino that unfortunately is experimentally undetectable in CMS due to its low interaction cross section with matter. Due to this, an indirect measurement of neutrino-like[4] particles is employed in the analysis. In order to do this a new quantity is defined, the missing transverse energy, $\vec{p}_T^{\,miss}$, often referred as MET in the literature. The MET exploits the conservation of the quadrimomentum vector in the transverse plane, which has to be $\sum \vec{p_T} = 0$ before and after the collision. This way, the MET is defined as the missing $\vec{p_T}$ component that restores the conservation of the quadrimomentum in the transverse plane

$$\vec{p}_T^{\,miss} = -\sum \vec{p_T}, \tag{2.11}$$

where $-\sum \vec{p_T}$ sums over all reconstructed objects in the event: jets, muons, electrons... and E_T^{miss} is defined as the modulus of $\vec{p}_T^{\,miss}$.

The simple definition in Equation 2.11 implies complicated experimental challenges. Having a good MET resolution is equivalent to measure all the activity in the transverse plane with good resolution and scale. For example, since the MET uses jets in its calculations, all the corrections to the energy of the jets need to be propagated accordingly in the $\vec{p}_T^{\,miss}$ sum.

It is important to remark that a mismeasured quantity in the transverse plane could produce artificially large E_T^{miss} in the event and mimic the production of potential new signals or interactions. Particularly, in the $W' \to \mu\nu$ search, the sum in Equation 2.11 is dominated by a high p_T muon term and a mismeasurement of its p_T would translate into a mismeasurement of E_T^{miss}. Due to this, a good understanding of the muon momentum at high p_T is also needed to achieve a good MET performance. Due to this, all the discussions about muon momentum performance in previous Sections 2.3.5 and 2.3.6, directly translate to the MET resolution and scale in the context of the $W' \to \mu\nu$ search.

Events without neutrinos also present a non-zero E_T^{miss}, since due to resolution and instrumental effects, all the terms in the sum in Equation 2.11 do not exactly add up to zero.

[4] It must be noted, that any particle that leaves no trace in the detector will contribute to the MET. However the only particles in the SM that we know that fulfills this condition are the neutrinos.

Dilepton selections in which the invariant mass of the lepton pair is compatible with the Z boson mass, are naturally enriched in Z/γ^* events and are useful for E_T^{miss} studies. This is because the dilepton selection allows to check and calibrate simulations in a sample where contributions from real neutrinos are expected to be small. As an example, Figure 2.18 (left) shows the good agreement between data and simulations in describing the MET distribution in dilepton events. Figure 2.18 (right) shows the same measurement compared to a simulation generated with higher number of pileup vertices compared to data. Collisions in a high pileup environment involve the production of multiple jets and particles in the transverse plane, leading to a slight worsening of the MET resolution.

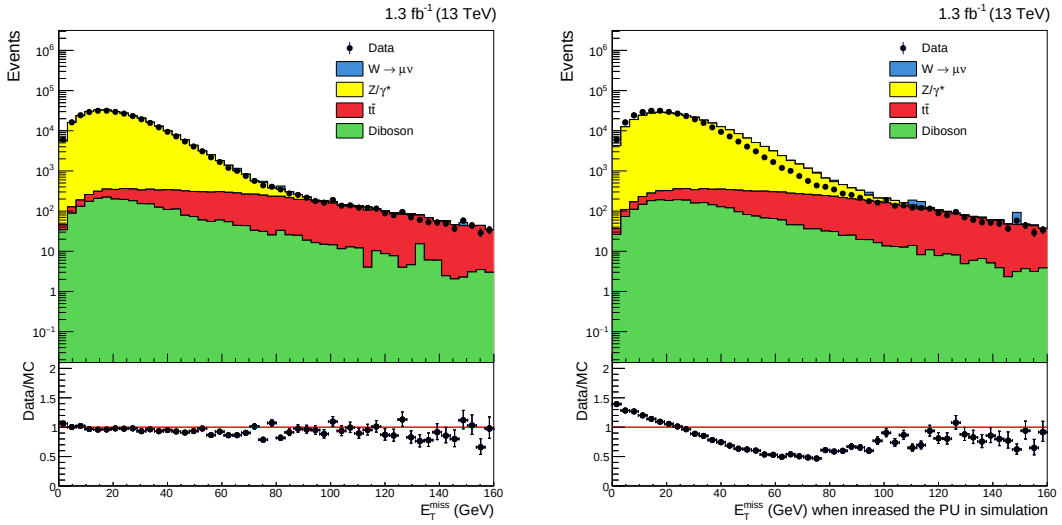

FIGURE 2.18: Missing transverse energy distribution in $\mu^+\mu^-$ events (left). The same distribution is shown (right) compared to a simulation generated with an artificially larger number reconstructed primary vertices.

Finally, a very small fraction of events could have a anomalously large MET. For example, in some cases they are produced by calorimeter noise, muons from the beam or jets near non-functioning channels in the calorimeters and others... These events are rare and are removed efficiently in the analysis using a dedicated set of filters designed to remove these pathological events.

2.4 The W' search:

The experimental signature in this search is the production of a high-momentum muon and large missing transverse energy, E_{T}^{miss}, which may flag the presence of a non-interacting particle from a potential new $W' \to \mu\nu$ boson. The selection of W' candidates follows a similar strategy to the one used in the 8 TeV version of this analysis [25].

The first set of selection cuts aims to keep event candidates containing just a single and well identified High-p_{T} muon. This selection is referred in the document as the *pre-selection*. Once potential candidates are identified, the *final* W' *selection* is applied. The latter aims to further improve the background rejection with additional kinematical cuts between the muon and the missing transverse energy.

2.4.1 Analysis pre-selection: single muon candidates

A single muon enriched *pre-selection* is defined as follows:

- Events have to be recorded by the trigger HLT_Mu_50.
- Leading muon must have p_{T} larger than 53 GeV and reconstructed in the pseudo-rapidity region $|\eta| < 2.4$, defined by the muon system coverage. The muon momentum p_{T} assignment is given by the *Tune-P* algorithm (see Section 2.3.4).
- Leading muon has to be identified as High-p_{T} and isolated from hadron activity.
- Sub-leading muon (if existing) must have p_{T} smaller than 25 GeV. This requirement reduces contribution from dilepton and cosmic muon events.
- The transverse mass built from the muon p_{T} and the $\vec{p}_T^{\,miss}$ (see Equation 2.13 for details) has to be larger than $M_{\mathrm{T}} > 50$ GeV.

A word about the SM backgrounds

Backgrounds are defined as physical processes different from a $W' \to \mu\nu$ production that could lead to a signature with a muon and missing transverse energy. The individual contribution from each of the backgrounds used in the search is taken from simulated events. The several generators used in their production was discussed in Section 2.2.2.

The number of primary vertices, muon trigger and identification efficiencies are re-calibrated in simulation in order to improve their description of the data. Details of

the calibrations were discussed in Section 2.3. The contribution of each individual background is normalized to an equivalent luminosity of $L = 2.3$ fb^{-1}. The weight, w_b, of each entry in the histograms for a given background, b, is calculated as

$$w_b = L\frac{N_b}{\sigma_b} \cdot C(N_{PU}) \cdot \frac{\epsilon_{Data}^{Trigger}}{\epsilon_{MC}^{Trigger}} \cdot \frac{\epsilon_{Data}^{ID}}{\epsilon_{MC}^{ID}} \tag{2.12}$$

where N_b and σ_b are the number of generated events and the theoretical cross section of the generated background process, b. $C(N_{PU})$ is the re-weighting function that matches the number of proton-proton interactions in simulation to the ones measured in data (Section 2.3.2). The $\epsilon_{Data}^{Trigger}/\epsilon_{MC}^{Trigger}$ factor corresponds to the calibration of the trigger efficiencies (Section 2.3.1), and the $\epsilon_{Data}^{ID}/\epsilon_{MC}^{ID}$ factor corresponds to the calibration of the muon reconstruction and identification efficiencies (Section 2.3.3).

In order of importance, the dominant background comes from $W \to \mu\nu$ production, particularly when this process occurs at high mass W^*. The W^* background is called irreducible because is kinematically identical to the $W' \to \mu\nu$ contribution. Contributions from $W \to \tau(\mu\nu)\nu$ are also included in the simulations although its contribution is expected to populate the low transverse mass region and has little impact in the search.

Top quarks decay almost 100% of the times into a bottom quark and a W boson. Subsequently, the W boson decays into a muon and a neutrino 11% of the times. This way, top pair production, $t\bar{t}$, may lead to signatures with one isolated lepton plus missing transverse energy and additional jets. This background is partially rejected in its fully leptonic decay mode with the veto of events in which there is a second muon. In the case of the highest masses, where the search is more sensitive due to the lower backgrounds, top quarks have large p_{T} and its decay products, the bottom quark and the W boson, are geometrically close to each other. As a consequence, the muon produced in the W decay is geometrically close to the jet produced in the hadronization of the bottom quark and is not identified as a High-p_{T} muon because of the isolation requirement.

Finally diboson processes such as $W(\ell\nu)W(qq)$, $W(\ell\nu)Z(\nu\nu)$ may also mimic a W' signal, although with a much reduced cross section. Decays from $Z(\ell\ell)Z(\nu\nu)$ are highly suppressed with the lepton veto and have even lower cross section. In general, diboson contribution is expected to be small at high mass in the SM due to its low cross section.

Contrary to the clean $W' \to \mu\nu$ signal, often high p_{T} W bosons produce high p_{T} muons accompanied by large hadronic activity to balance the momentum of the W in the transverse plane. Other processes such as $t(b\ell\nu)\bar{t}(\nu bqq)$, $W(\ell\nu)W(qq)$ or $W(\ell\nu)Z(qq)$ ... also present a visible amount of hadronic activity. These contributions are reduced with a dedicated kinematical selection presented in Section 2.4.3.

Figure 2.19 shows the distribution of the number of reconstructed vertices before (left) and after (right) applying the $C(N_{PU})$ re-weight described in Section 2.3.2. Figure 2.20 shows the ϕ (left) and η (right) distribution of the leading muon in the event. W' signal events are expected to be distributed symmetrically around ϕ and more preferentially at low η. The muon transverse momentum distribution, p_T, is shown in Figure 2.21 and the MET distribution in Figure 2.22. All the simulations are normalized to an equivalent luminosity of 2.3 fb^{-1} as described in Equation 2.12.

In all the cases the simulation provides a good description of the observed data and enables this analysis to proceed and find the best distribution to discriminate between a potential W' signal and the backgrounds.

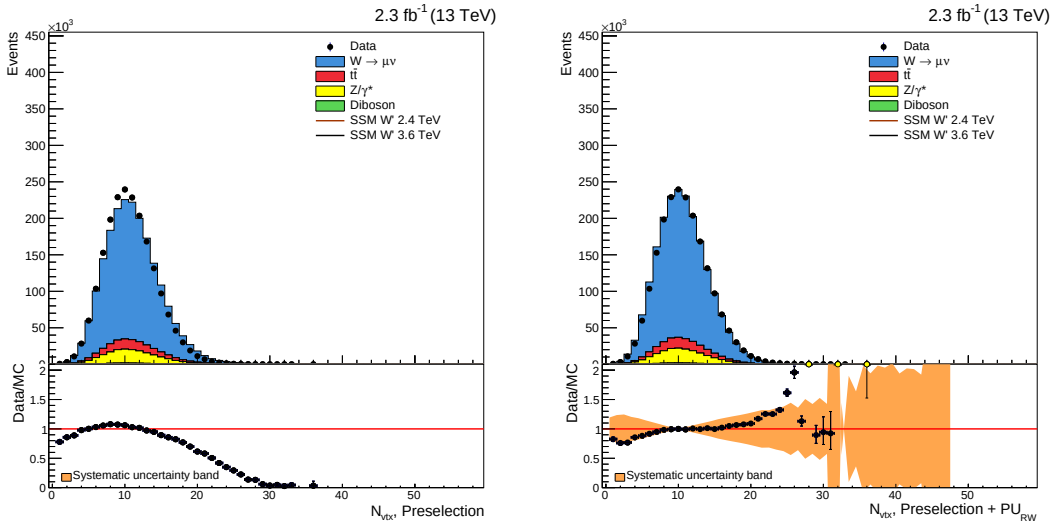

FIGURE 2.19: Number of reconstructed vertices distribution at pre-selection before (left) and after (right) applying the $C(N_{PU})$ re-weight described in Section 2.3.2.

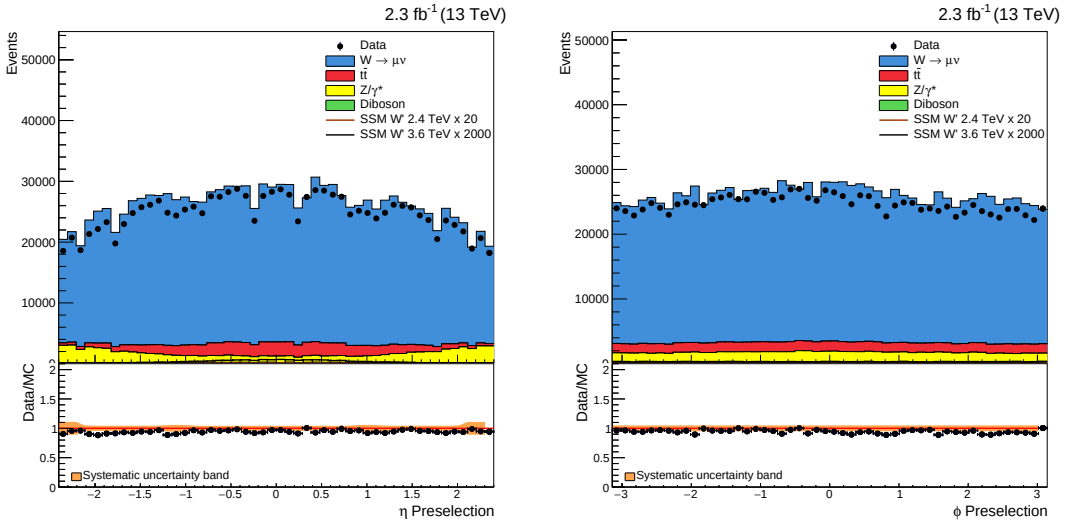

FIGURE 2.20: Muon η (left) and ϕ (right) distributions at the pre-selection stage.

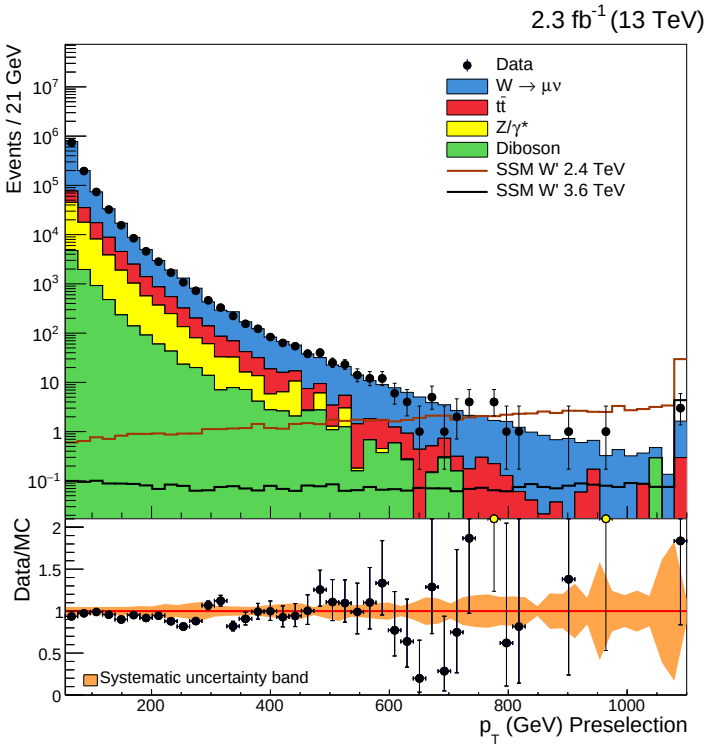

FIGURE 2.21: Muon transverse momentum p_{T} distribution at the pre-selection stage. Expectation from a SSM W′ signal with masses 2.4 TeV (brown line) and 3.6 TeV (black line) are also shown.

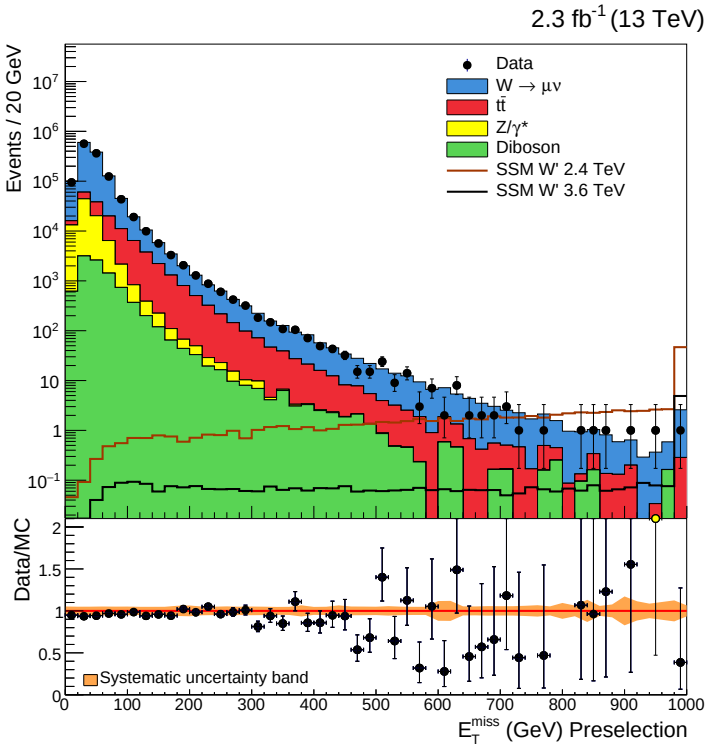

FIGURE 2.22: Missing transverse energy distribution, E_{T}^{miss}, at pre-selection. Expectation from a SSM W′ signal with masses 2.4 TeV (brown line) and 3.6 TeV (black line) are also shown.

2.4.2 Discriminant variable for the search: transverse mass

A discriminant variable in the $W' \to \mu\nu$ search needs to fullfill two important criteria. Firstly, it must have high separating power between signal and backgrounds. Secondly, in case of positive signal, it needs to be able to give information about the nature of the signal.

Owing to the two-body resonant nature of the W' signal, a natural candidate would be the muon-neutrino invariant mass that is directly sensitive to the pole mass and width of the resonance. Unfortunately in the $W' \to \mu\nu$ search, the invariant mass cannot be reconstructed since the missing transverse energy component along the beam axis, $\vec{p}_Z^{\,miss}$, cannot be measured experimentally. Instead, the transverse mass, M_{T}, defined only with quantities measured in the transverse plane is used

$$M_{\mathrm{T}} = \sqrt{2p_{\mathrm{T}}^{l} E_{\mathrm{T}}^{\mathrm{miss}} (1 - \cos[\Delta\phi(\vec{p}_{\mathrm{T}}^{\,l}, \tilde{\mathrm{p}}_{\mathrm{T}}^{\mathrm{miss}})])}. \tag{2.13}$$

A sketch of the discrimination power in the transverse mass between a potential W' signal and the background is shown in Figure 2.23. In order to avoid strong constrains from previous W' searches at the LHC, potential contributions from a new $W' \to \mu\nu$ with the new CMS 13 TeV data, if existing, should appear as anomalous events above $\mathrm{M_T} > 1$ TeV and around the W' mass.

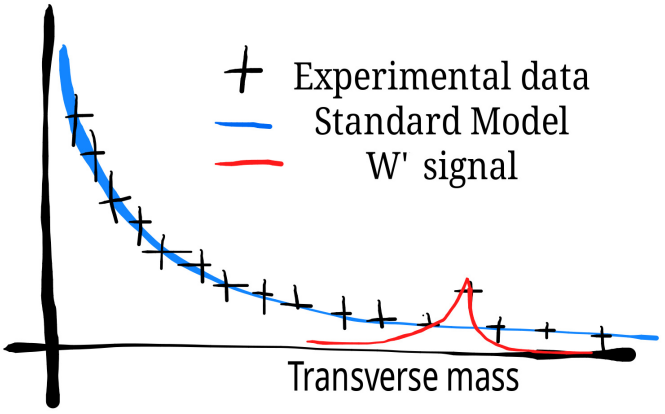

FIGURE 2.23: Blue and red lines show a sketch of the shape of the transverse mass distribution for background and signal events respectively. Individual signal events are expected to stand out over the falling background spectrum.

The shape of the transverse mass of a resonance has an end-point at the original mass: $M_T < M_{W'}$. For instance, this property has been used to determine the W mass at the Tevatron [88].

In practice, the situation is more complicated than the sharp peak shown in Figure 2.23. Reconstruction effects such as the muon momentum scale and resolution discussed in Sections 2.3.5 and 2.3.6, and their correlation with the missing transverse energy determination, E_{T}^{miss}, worsen the M_{T} resolution at higher masses. As a consequence, Figure 2.24 shows the difference between the M_{T} distribution calculated using Equation 2.13 before (red) and after reconstruction (pink) for two different SSM W′ masses.

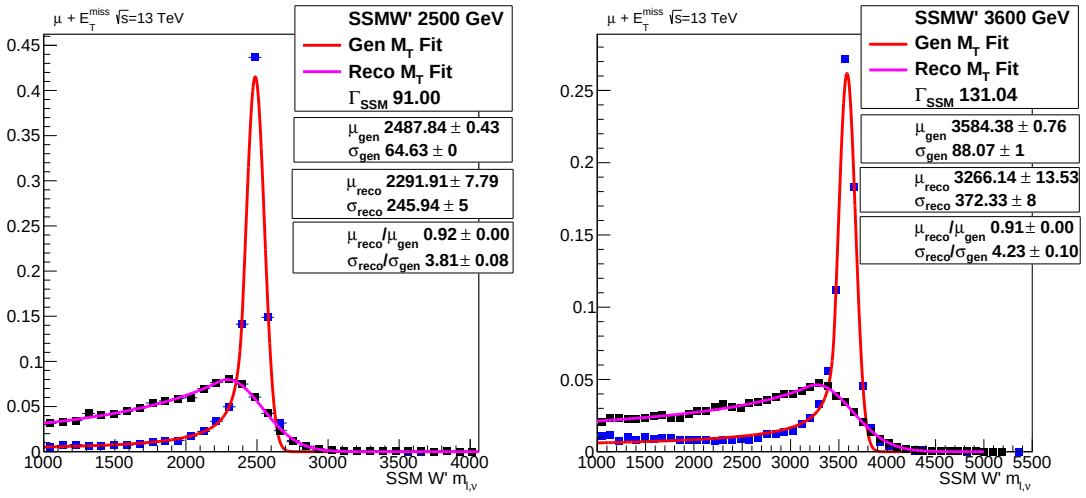

FIGURE 2.24: Comparison between the transverse mass distribution at the generator level (red) and same distribution after the CMS reconstruction (pink) for a W′ boson with mass 2.5 TeV (left) and 3.6 TeV (right). At high mass the muon and E_{T}^{miss} resolution slightly degrade and the resonant peak broadens.

Going back to measured data, Figure 2.25 shows the reconstructed transverse mass for each selected event in data and simulated samples. Potential W′ signal events are expected to appear at high M_{T} values, while the different background contributions and particularly the SM W boson decays, appear as a tail as the transverse mass increases. At this stage of the analysis, simulations reproduce the measurement and there is no hint of W′ production in the spectrum. In next section, the kinematics of the two-body decay of the W′ are exploited to further reduced backgrounds and potentially enhance contributions from new physics sources.

Finally it is not discarded that the W′ width could be larger than expected, or that new physics could show up as non-resonant tail in the M_{T} distribution. In this case, the excess should be visible mostly in the cumulative M_{T} distribution shown in Figure 2.26. Each bin of this distribution corresponds to the integral from that bin to infinity. Again, no excess in data with respect to the simulations is present in the distribution.

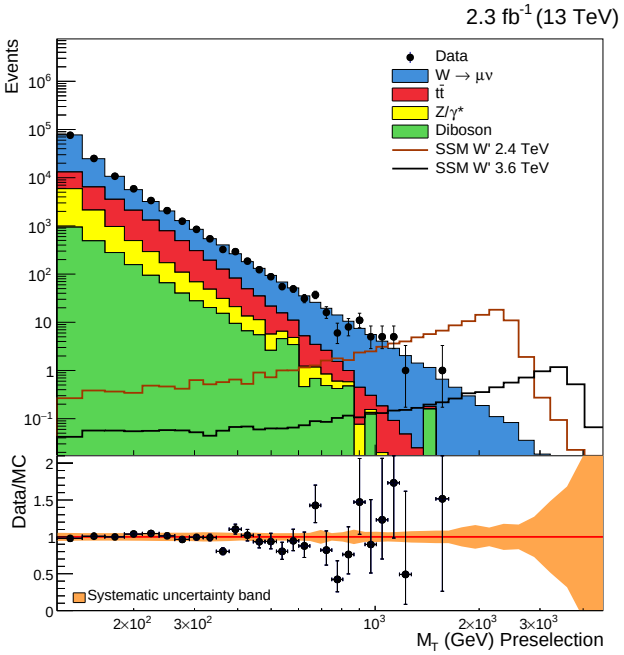

FIGURE 2.25: Transverse mass at pre-selection stage. Expectation from a SSM W′ signal with masses 2.4 TeV (brown line) and 3.6 TeV (black line) are also shown.

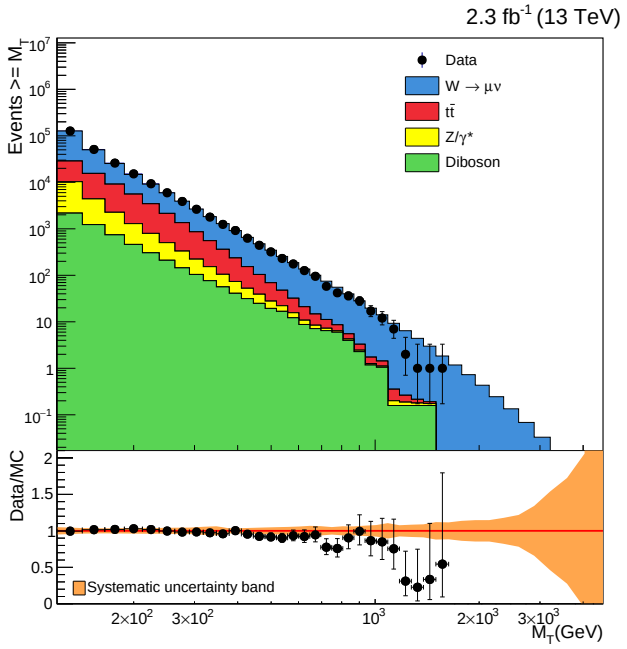

FIGURE 2.26: Cumulative transverse mass distribution at the pre-selection stage. The content of each bin of the distribution corresponds to the integral starting in that bin to infinity.

2.4.3 Kinematic selection

In the SSM W' and in any other model that predicts a new heavy $W' \to \mu\nu$ resonance, the decay products have the kinematical properties of a $2 \to 2$ body decay. Under this assumption, the following kinematical relations are expected:

- Conservation of the momentum in the transverse plane implies that $p_{\mathrm{T}}^{\mu} \approx p_{\mathrm{T}}^{\nu}$, where p_{T}^{ν} and p_{T}^{μ} correspond to the transverse momentum of the muon and neutrino.
- If additionally the daughters have almost no boost along the longitudinal axis, $p_{Z}^{\nu} \approx p_{Z}^{\mu} \ll p_{\mathrm{T}}^{\mu}$, then $M_{\mathrm{W}'} \approx 2p_{\mathrm{T}}^{\mu}$.

In case of $W' \to \mu\nu$ production at the TeV scale, one can approximate the p_{T}^{ν} with the $\vec{p}_{T}^{\,miss}$ component and exploit the kinematical properties of the muon and the missing transverse energy to reduce backgrounds. For example, a fraction of events containing high p_{T} muons and large hadronic activity can be distinguished from W' signal with two simple and loose kinematic criteria:

- The difference in azimuthal angle between the muon and the missing transverse energy is required to be: $\Delta\phi(\vec{p}_{\mathrm{T}}^{\,l}, \vec{p}_{T}^{\,miss}) > 2.5$. Figure 2.27 (left).
- The ratio of muon lepton transverse momentum and the missing transverse energy, E_{T}^{miss}, must lie in the region $0.4 < p_{\mathrm{T}}/E_{\mathrm{T}}^{miss} < 1.5$. Figure 2.27 (right).

Typically these set of kinematical cuts are useful to reduce events with high p_{T} muons coming from W with large p_{T}(W) or $t\bar{t}$.

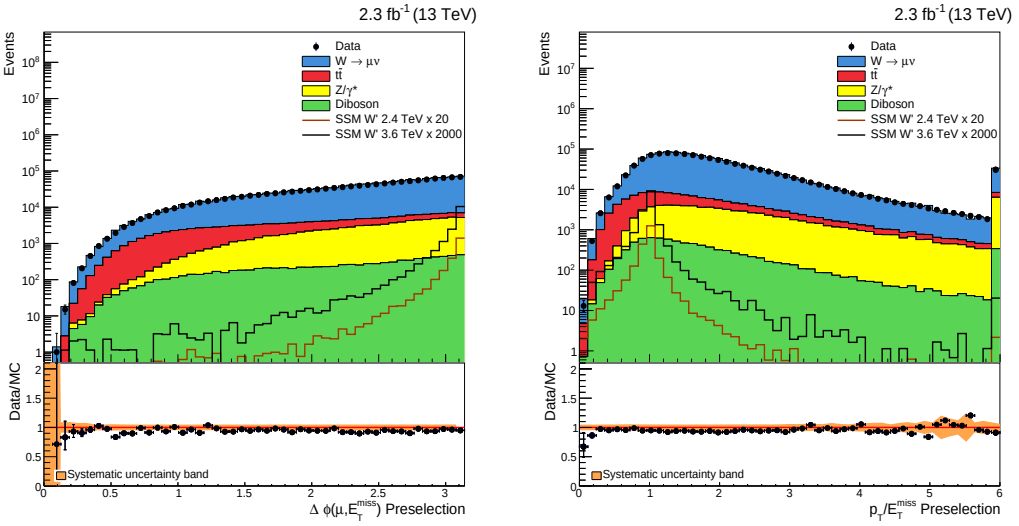

FIGURE 2.27: Distribution of the kinematical cuts at the pre-selection stage. Only events in which $\Delta\phi(\vec{p}_{\mathrm{T}}^{\,l}, \vec{p}_{T}^{\,miss}) > 2.5$ (left) and $0.4 < p_{\mathrm{T}}/E_{\mathrm{T}} < 1.5$ (right) are selected.

In order to further understand the effect of the kinematical cuts, it is useful to compare the distribution of the vectorial sum of the lepton momentum and the missing transverse energy $|\vec{p}_{\mathrm{T}}^{\,l} + \vec{p}_T^{\,miss}|$, before and after they are applied. This distribution is highly correlated with the $p_{\mathrm{T}}(\mathrm{W})$ and is shown in Figure 2.28, before (top) and after (bottom) the kinematical selection. As consequence of the kinematical selection, the selected events have lower boost in the transverse plane and are closer to be produced at rest.

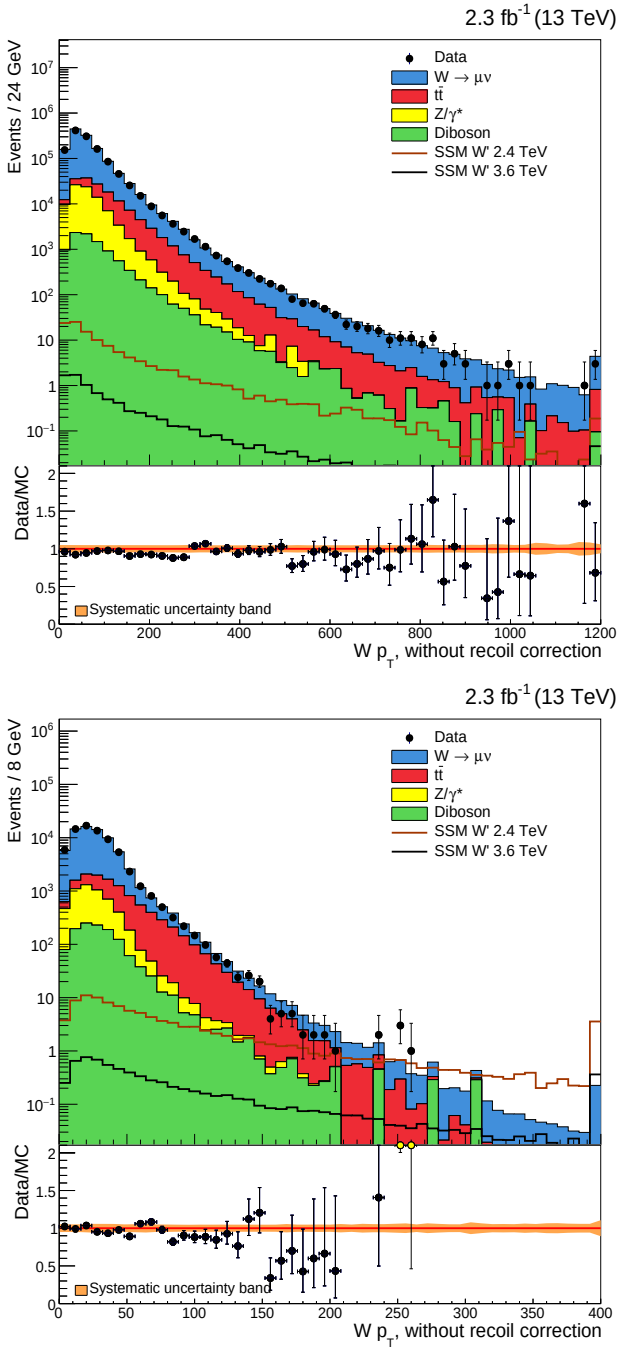

FIGURE 2.28: Distribution of the module of the vectorial sum of the lepton momentum and the missing transverse energy, $|\vec{p}_{\mathrm{T}}^{\,l} + \vec{p}_T^{\,miss}|$ before (up) and after (down) the kinematical selection shown in Figure 2.27. After kinematical selection, background events are more balanced in the transverse plane (note the different x-axis range).

2.4.4 Final selection

The final analysis selection is obtained with the pre-selection cuts introduced in Section 2.4.1 and the kinematical cuts motivated in Section 2.4.3 to further reduced backgrounds. Additionally, only events with $M_{\mathrm{T}} > 120$ GeV will be considered to avoid contamination from on-shell W.

The signal efficiency times acceptance, $A\epsilon$, defined with respect to generated SSM W' events as a function of the W' mass is shown in Figure 2.29. About 75 $\pm$ 8% of the W$'$ candidates, if produced, are expected to be selected with the full analysis selection. The uncertainties in $A\epsilon$ are dominated by the extrapolation of the trigger, reconstruction and identification efficiencies to high p_{T}, where no measurements are available to validate the simulations. Systematical uncertainties will be discussed in Section 2.4.6.

Above 3 TeV in mass, the drop in $A\epsilon$ is due to the $M_T > 120$ GeV cut that removes highly off-shell W$'$. The explanation for this is that the low mass contribution becomes non-negligible for such heavy W$'$ due to its larger resonance width and the suppression of parton PDFs at very high x. However, in order to allow other reinterpretations less dependent on the modelling of the W$'$ production towards lower masses, an $A\epsilon$ is also determined excluding the M_{T} cut. The resulting value is a bit higher and about 83$\pm$7%.

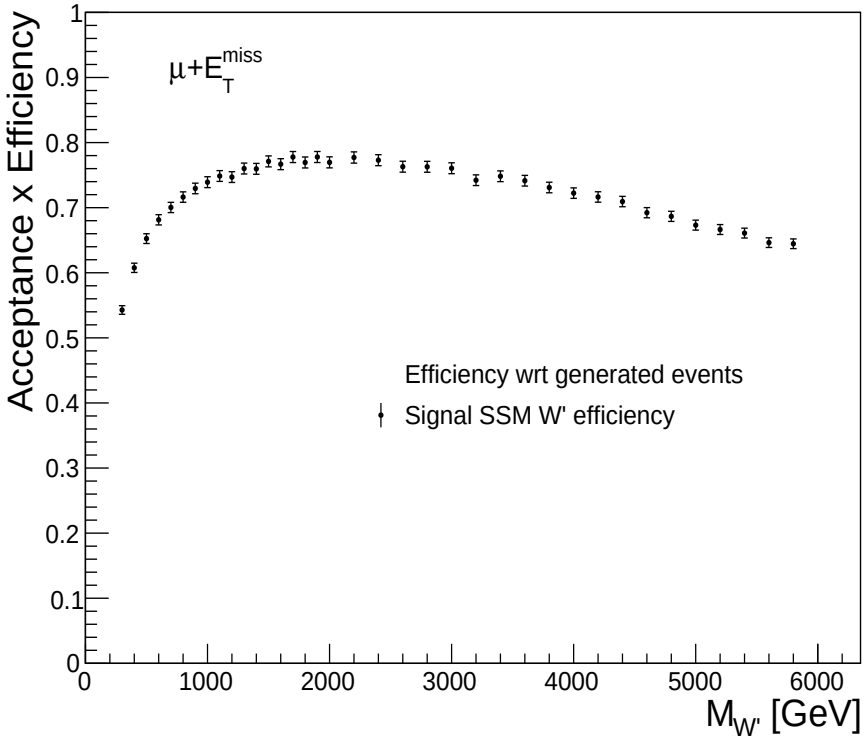

FIGURE 2.29: Acceptance times efficiency, $A \cdot \epsilon$, for SSM W$'$ as a function of the W$'$ mass. Numbers are computed with respect to generated W$'$.

Finally, Figure 2.30 shows the final transverse mass distribution, M_T, with the 2015 dataset and the cumulative M_T distributions in Figure 2.31. Unfortunately, none of the distributions suggest a potential excess consistent with the production of new $W' \to \mu\nu$ bosons with respect to the predictions of the Standard Model.

It is interesting to note that, due to the evolution of the proton parton distribution functions, the M_T distribution is also sensitive to the momentum fraction and nature of the partons involved in the interaction. The ratio of W^+/W^- measured at the W mass is about $1.29 \pm 0.01 \pm 0.01$ at 13 TeV [89], while at higher mass the contribution of W^+ is expected to be enhanced. Due to this, the consistency of the simulations with respect to the observations has been also checked in the transverse mass distribution for positive and negative muons separately. Both results are shown in Figure 2.32. As expected, more W^+ are measured in agreement with the simulations.

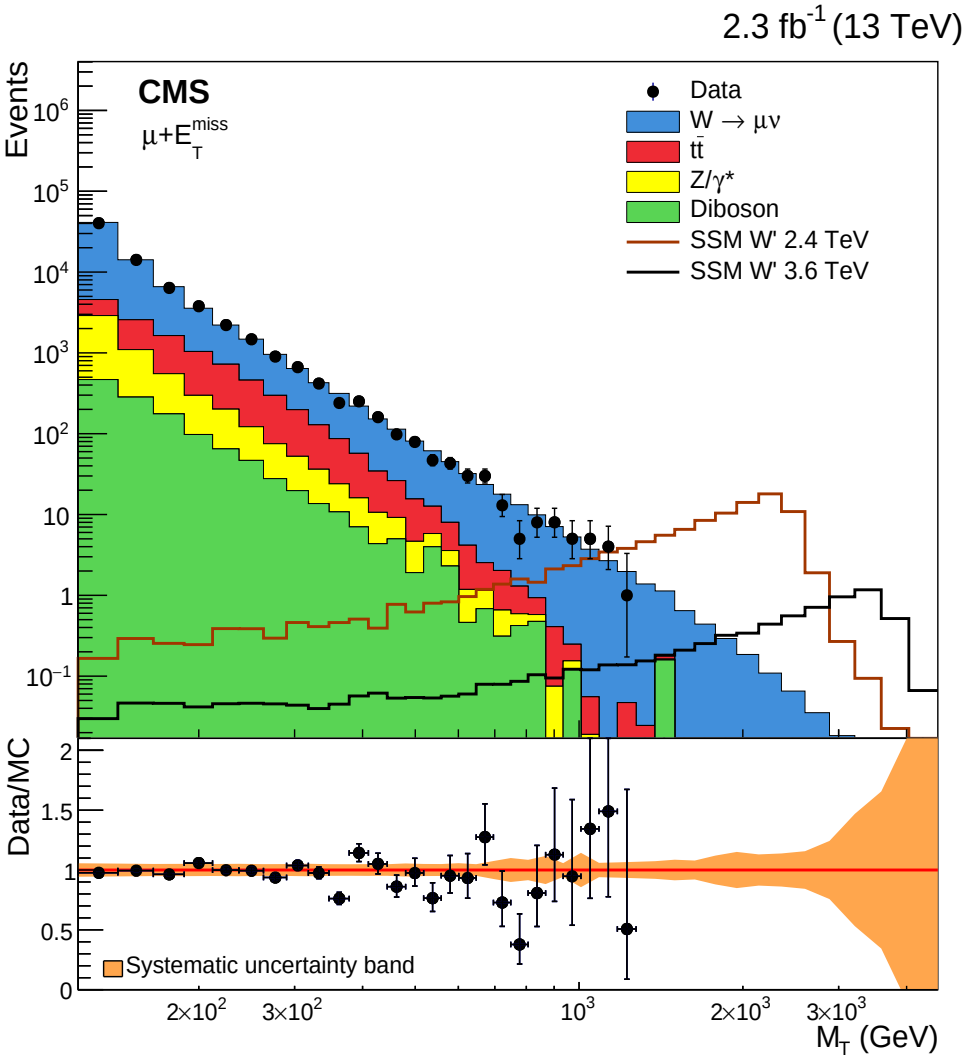

FIGURE 2.30: Final transverse mass distribution, M_T, after full analysis selection. Expectation from a SSM W' signal with masses 2.4 TeV (brown line) and 3.6 TeV (black line) are also shown.

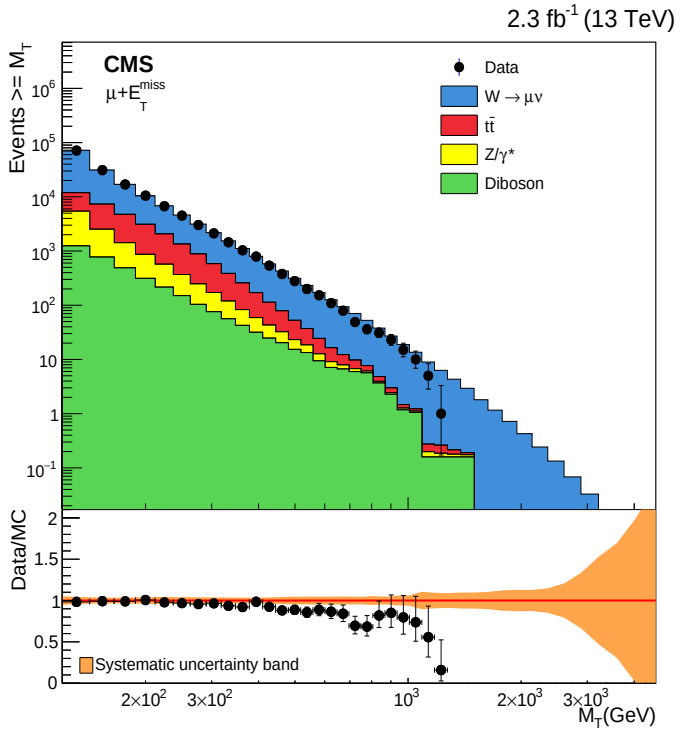

FIGURE 2.31: Cumulative transverse mass distribution after the full analysis selection. In the figure the content of each bin of the distribution corresponds to the integral between that bin to infinity.

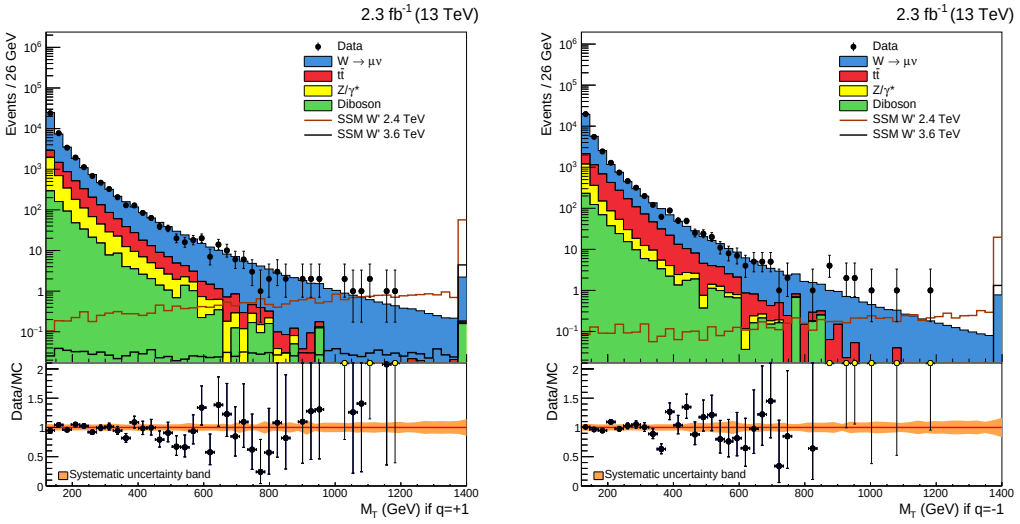

FIGURE 2.32: Transverse mass distribution separately for positive (left) and negative (right) muons after full analysis selection.

2.4.5 Scrutiny of the highest M_{T} events

Due to the low (almost zero) expectations from different backgrounds at the TeV scale (see Figure 2.31), a couple of mis-reconstructed events could create a fake or hide a signal. In order to avoid this, a manual inspection of the highest mass events in a

$W' \to \mu\nu$ search is needed. Examples of pathological signatures that could suggest a bad-reconstruction and create a fake signal are the following:

- In case of alignment problems, the muon-p_T measurement could have large inconsistencies among different TeV re-fits described in Section 2.3.4, and also with the p_T assignment from the tracker detector alone.
- Especial attention to muons passing near boundaries of the detector, for example at the intersection between two wheels.
- Checking potential mis-matches between a tracker-track with a stand-alone track coming from a different muon leading to a fake-high p_T track.
- Problems in the determination of the MET, such as calorimeter noise, jets or electrons near non-functioning channels...
- Potential non-colliding backgrounds such as: cosmic muons, fake tracks, muons from the beam or other type of anomalous activity in the detector...

All events analysed involving high-p_T muons are likely to come from real proton-proton collisions, and measured track parameters are consistent with studies in Sections 2.3.5 and 2.3.6. Nothing worrisome has been found in the manual inspection. As an example, the event display of one of the highest M_T events is shown in the $r-\phi$ plane in Figure 2.33 and in the $r-z$ plane in Figure 2.34. Both images have been pre-processed and cleaned from non-relevant detector activity such as low momentum tracks or noise.

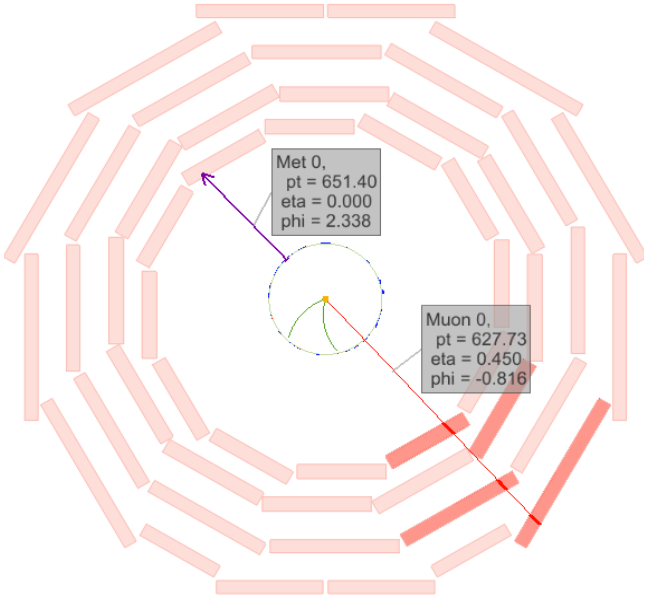

FIGURE 2.33: Event display of the r-ϕ view of one of the highest M_T events.

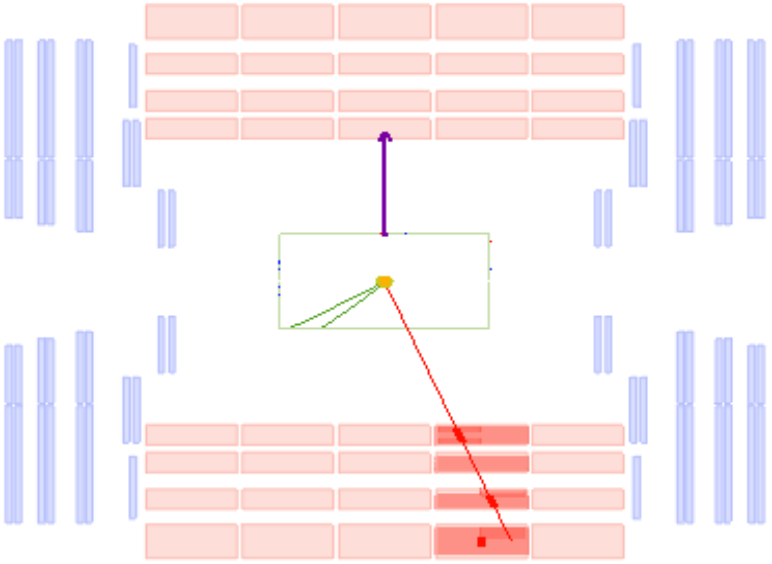

FIGURE 2.34: Event display of the r-z view of one of the highest M_T events.

2.4.6 Systematic uncertainties

Different effects, denoted as systematic uncertainties, could affect the modelling of the shape and normalization of the M_T distribution. One example of this kind is the value of the integrated luminosity that is difficult to measure with a precision better than 2%. As a result, this uncertainty directly translates to the expected normalization of the M_T distribution. This section describes the main sources of systematic uncertainties, how they are estimated, and their impact in the modelling of the transverse mass distribution/templates. Often in the manuscript the word template is used and refers to the distribution of one individual component in simulation of a particular variable. For example, the $t\bar{t}$ template of the M_T distribution.

In principle, each systematic uncertainty could affect differently signal and backgrounds. Depending on their effect on the M_T distribution, they are grouped into two different categories: those that affect only the overall normalization (multiplicative uncertainties) and the ones that modify both, shape and normalization (shape uncertainties). For each multiplicative uncertainty, their effect in the normalization is provided whereas in the case of shape uncertainties, two alternative M_T templates were derived that correspond to a variation of plus or minus one standard deviation of that systematic uncertainty. Often in the text and for illustration, the size of a shape uncertainty is quantified by providing the uncertainty in the expected number of events above a minimum $M_{T_{min}}$ threshold.

A common experimental uncertainty that affects both, the normalization of W' signals and backgrounds, is the estimation of the recorded luminosity in CMS. The measured value is $L = 2.3$ fb^{-1} with an uncertainty of 2.7% [90]. This uncertainty, evaluated in a dedicated Van der Meer scan [91], is taken into account as a common multiplicative uncertainty that directly translates to the expected background and signal events.

An additional source of uncertainty comes from the description in simulation of the number of proton-proton interaction per bunch crossing as explained in Section 2.3.2. Two alternative M_{T} templates were derived using alternative re-weight function obtained varying $\pm 5\%$ the proton-proton inelastic cross section. The effect of this uncertainty in the number of reconstructed vertices was shown in Figure 2.19.

The calibration factors, $\epsilon^{Trigger}_{Data}/\epsilon^{Trigger}_{MC}$ and $\epsilon^{ID}_{Data}/\epsilon^{ID}_{MC}$, described in Section 2.3.3 and Section 2.3.1 were derived using intermediate p_{T} muons produced in dilepton events. The limitation of the method comes from the finite statistics of high p_{T} muons of the data sample. The calibration factors used to correct the simulations need to be extrapolated to higher p_{T}. Due to this, a p_{T} dependent uncertainty that ranges between $5\% - 8\%$ was assigned to their extrapolation to the TeV scale. The size of the uncertainty corresponds to the maximal variation in the range 200 GeV $< p_{\mathrm{T}} <$ 3 TeV of the trigger and identification efficiencies as estimated from the simulation.

The muon scale and resolution uncertainties, when propagated to the muon momentum measurement, do not affect the overall normalization of the M_{T} distribution. However, they may largely affect the slope of the transverse mass and the expected number of events above a $M_{T_{min}}$ threshold. The effect of the muon moment scale uncertainty was estimated with two M_{T} templates built smearing the muon momentum curvature[5] by $\kappa' \to \kappa \pm \kappa_b(\eta, \phi)$. The value κ_b depends on the region in pseudo-rapidity of the muon, and correspond to the uncertainties estimated in Table 2.6 for $|\eta| < 2.0$, and the measured values in Figure 2.17 for $|\eta| > 2.0$. For example, the uncertainty in the number of expected events due to the muon momentum scale uncertainty, is 10 (20)% for events with M_{T} above 1 (3) TeV. In order to illustrate this effect, Figure 2.35 shows the nominal transverse mass distribution and the two M_{T} templates obtained from a simulation of SM W generated mass larger than 3 TeV, $M_{\mathrm{W}} > 3$ TeV.

Other sources of systematic uncertainties are related to the theoretical modelling of the simulations which also modify the shape of the M_{T} distribution. The most important ones are associated to the evaluation of the higher order electroweak and QCD corrections of the W* process, as well as the uncertainties associated to the PDFs predictions used in the generation of the samples.

[5] For every modification in the muon momentum assignment, the change was propagated coherently to the missing transverse energy of the event.

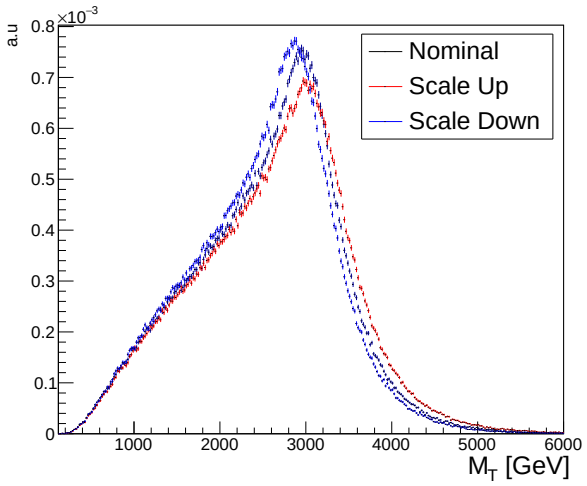

FIGURE 2.35: Nominal transverse mass distribution and two M_{T} templates that corresponds to $\pm 1\sigma$ variation of the scale uncertainty obtained from a simulation of SM with a generated mass larger than 3 TeV, $M_W > 3$ TeV.

Two M_{T} templates were obtained using alternative PDFs replicas following the recommendations from the PDF4LHC group [92]. The effect of the PDF uncertainties strongly depends on the Q^2 that governs the scale of the hard interaction. The higher the mass Q^2, the higher the x of the partons involved in the interaction and so the uncertainty. This uncertainty ranges from 5% at 1 TeV in mass to 30% at 2 TeV.

In the absence at the time of the analysis of a complete NNLO QCD + NLO EW calculation of the W high mass distribution, both corrections were computed separately using different generators. Electroweak NLO corrections were computed with MCSANC [82] while NNLO QCD corrections were derived with FEWZ [36]. A flat 5% uncertainty was considered for the uncertainty of the *k-factors*, that is larger than difference between the derived *k-factors* when they are combined additively (Equation 2.4) or multiplicatively (Equation 2.5). This choice yields an uncertainty which is large enough to cover the differences between both combinations, as well as the individual uncertainty of each of the corrections.

Finally, the uncertainties associated to the expectations from simulation of all the backgrounds other than W plus jets, namely: $\mathrm{t\bar{t}}$, $\mathrm{Z}+\mathrm{jets}$ and dibosons are assumed to be negligible at high mass due to their small contribution at this region. The effect of their uncertainty in the total number of expected background events with mass above 1 TeV, is several orders of magnitude smaller compared to other uncertainties such as: PDF uncertainties, higher order QCD and EW corrections or the muon momentum scale and resolution uncertainties.

2.5 Results and interpretation

Unfortunately, no deviations with respect to the predictions from the SM are present in the 2015 dataset as shown in Figure 2.30. The last step in the analysis is to quantify the "size" of the non-W′ observation into quantities that can be computed directly from the Lagrangian of a new physics model. Particularly this analysis sets an upper limit to the potential $W' \to \mu\nu$ production cross section at 13 TeV as a function of the W′ mass, $\sigma_{W'}\mathcal{B}(W' \to \mu\nu)|_{\text{exc}}$.

In other words, potential new physics models that include the prediction of additional W′ bosons with coupling to leptons in which $\sigma_{W'}\mathcal{B}(W' \to \mu\nu) < \sigma_{W'}\mathcal{B}(W' \to \mu\nu)|_{\text{exc}}$, are not excluded by this search. By excluding theories that are not valid we can further constrain the potential BSM theories that best describe the data. Historically hundreds of searches along different channels and experiments are needed before any discovery.

2.5.1 Statistical model

The transverse mass distribution, M_{T}, presents a completely different behavior for the different backgrounds and a potential W′ signal. This difference is exploited in a dedicated statistical analysis that is designed to quantify the compatibility of the data with the existence of a SSM W′. In the statistical language, this compatibility is given by the probability density function of $P(\text{SM} + W'|\text{Data})$. This probability describes the compatibility of SM plus the existence of a SSM W′ boson, SM + W′, given the measurement. In order to simplify the notation in the formulas along this section, the benchmark SSM W'_{SSM} signal, is denoted simply as W′.

The statistical model is built from the observed and expected counts in the final transverse mass distribution (Figure 2.30), where:

- $i = 1 \ldots N_{bin}$, denotes the i^{th} bin in the M_T distribution.
- n_i number of observed events in a given bin, and $n = \sum_{i=1}^{N_{bin}} n_i$ the total sum of observed events.
- Total number of expected background events, $b_i = \sum_{b=1}^{N_b} \hat{b}_{i,b}$ that contribute to i^{th} bin in M_{T}. Where $\hat{b}_{i,b}$ corresponds to the number of expected events from each of the expected, $b = 1 \ldots N_b$, backgrounds that contribute to the i^{th} bin.
- Number of expected signal events in a given bin, $s_i = L\sigma_{W'}\mathcal{B}(W' \to \mu\nu)A\epsilon_i$. Its value depends on the mass of the W′ and its acceptance times efficiency in that bin, $A\epsilon_i$. $L = 2.3\ \text{fb}^{-1}$ corresponds to the integrated luminosity of the data sample.

The parameter of interest is called the signal strength, μ, and is defined as $\mu = \sigma/\sigma_{W'}$. With this definition, the number of expected events in a given bin is $\lambda_i(\mu) = \mu s_i + b_i$ and $\lambda(\mu) = \sum_{i=1}^{N_{bin}} \lambda_i(\mu)$. For example, $\lambda(1)$, corresponds to the total number of expected events given the nominal signal hypothesis or $\lambda(0)$ to the background-only hypothesis.

With the above definitions in mind, the statistical model is described by the likelihood, $L(n|\lambda(\mu))$, of observing n events given the model $\lambda(\mu)$, and it is built by the product of the Poisson probability evaluated at each of the bins of the M_T distribution.

$$L(n|\lambda(\mu)) = \prod_{i=1}^{N_{bin}} \frac{\lambda_i(\mu)^{n_i} e^{\lambda_i(\mu)}}{n_i!} \tag{2.14}$$

The model presented in Equation 2.14 it is still not complete because it does not include any source of systematic uncertainties in the number of expected events $\lambda(\mu)$. Each source of systematic uncertainty described in Section 2.4.6, $u = 1...N_u$, is defined by a nuisance parameter, θ_u, that could affect the expected background and signal events, $\lambda(\mu, \theta_1...\theta_{N_u})$,

$$\lambda_i(\mu, \theta_1...\theta_{N_u}) = \mu s_i(\theta_1...\theta_{N_u}) + \sum_{b=1}^{N_b} \lambda_{i,b}(\theta_1...\theta_{N_u}) \tag{2.15}$$

Depending of the type of systematic uncertainty, different functional forms are considered for the nuisance parameters that distort the prediction of $\lambda_i(\theta_1...\theta_{N_u})$:

- Multiplicative uncertainties such as theory cross sections, luminosity uncertainty... are parametrized with log normal distributions.
- Shape uncertainties such as muon momentum scale and resolution, trigger and identification efficiencies, PDF uncertainties... are parametrized using as input two alternive M_T shapes that corresponds to ± 1 standard deviation and are parametrized following [93]. For example, in the case of the muon scale uncertainty the two alternative templates were shown in Figure 2.35.
- Statistical uncertainties such as the limited size of the simulated samples... are parametrized with gamma distributions.

This way, the proper statistical model corresponds to the following binned likelihood:

$$L(n|\lambda(\mu, \theta)) = \prod_{i=1}^{N_{bin}} \frac{\lambda_i(\mu, \theta)^{n_i} e^{\lambda_i(\mu,\theta)}}{n_i!} \tag{2.16}$$

Where a new variable $\theta = \theta_1....\theta_{N_u}$ is defined to simplify the notation and encode all the nuisance parameters.

2.5.2 The upper limit

In the statistical model language, the desired probability, $P(\text{SM} + \text{W}'|\text{Data})$, corresponds to the posterior probability $P(\lambda(\mu)|n)$. The latter is computed using Bayes theorem from the likelihood of the statistical model in Equation 2.16, and integrating over all the nuisance parameters θ and with the proper normalization:

$$P(\lambda(\mu)|n) = \frac{\int d\theta' L(n|\lambda(\mu,\theta'))\pi(\mu,\theta')}{\int d\mu' d\theta' L(n|\lambda(\mu',\theta')\pi(\mu',\theta'))} \tag{2.17}$$

Where $\pi(\mu,\theta)$ is called the prior probability density function. In general several options are valid for the choice of the prior function and usually are motivated by other measurements or theoretical arguments. Often in searches, where typically no information about signal cross section is available, the prior probability function is a uniform distribution truncated at $\mu = 0$. With this choice, all positive values for the signal cross section are assumed to be equally likely.

Integrating the posterior probability from zero to μ

$$p(\mu) = \int_0^{\mu} P(n|\lambda(\mu'))d\mu' \tag{2.18}$$

allows to answer to the original question, $P(\text{SM} + \text{W}'|\text{Data})$, or equivalently:

given the measurement, how likely is to be consistent with SM+W′ for the given W′ mass hypothesis and signal strength μ?

The answer to this question is encoded in the integrated $p(\mu)$ and depends on the mass of the W′ as well as on μ. This analysis uses a 95% confidence level (C.L) definition to set an exclusion on μ, this means that if $p(\mu) < 0.05$, that choice of the signal strength for a W′ hypothesis is disfavoured by the observations and is excluded. The value of μ for which $p(\mu) = 0.05$ is what is called the excluded signal strength, μ_{exc} at 95% C.L. Consequently, signal strengths for which $p(\mu) > 0.05$ are not excluded by the measurement. It is important to note that if for some mass hypothesis, $p(\mu) > 0.05$, it does not mean that SM+W′ is true. The latter only means that the considered hypothesis is not inconsistent with the measurement, probably because μ is too weak for the sensitivity of the analysis.

For every W′ mass hypothesis two different μ_{exc} are computed and shown in the results Section 2.5.3, the so-called expected and observed μ_{exc} at 95% C.L:

- The observed μ_{exc} at 95% C.L is computed taking as input in the statistical model the M_{T} measurements from data, n_i.

- Alternatively, the expected μ_{exc} at 95% at C.L is independent of the measurement. It is computed taking as input, n_i, pseudo-data from random experiments following expectations from SM only simulations. Additionally, one and two standard deviations with respect to expected μ_{exc} at 95% C.L are also determined for each mass hypothesis. Often in searches, the expected μ_{exc} at 95% C.L. is computed prior to data taking in order to quantify the potential sensitivity of a search.

For a given W' model and mass hypothesis, if the observed μ_{exc} at 95% C.L. is less than unity, that mass hypothesis is said to be excluded at 95% C.L. in the model. The conversion between a excluded signal strength into a excluded cross section is simple from the definition of signal strength, $\sigma_{exc} = \mu_{exc}\sigma_{W'}$.

In practice, the value of μ_{exc} that gives $\int P(n|\lambda(\mu)) = 0.05$ in Equation 2.18 is computed using the Higgs Combine Tool, a centralized and maintained CMS software package for statistical analysis based on RooStats [94]. The integral in Equation 2.17 is multi-dimensional and not solvable analytically. Due to this, its solution has been computed numerically using a Markov Chain Monte Carlo method [95]. The statistical model was provided to the Higgs Combine Tool using plain text datacard files for each signal mass and the M_T shape via histograms produced with ROOT software package [96].

2.5.3 Constrains to the W′ production at 13 TeV

In the absence of hints of new physics in events with a high momentum muon and large missing transverse energy, the main physics result of the $W' \to \mu\nu$ search are the observed and expected excluded cross sections $\mu_{exc}\sigma_{W'}\mathcal{B}(W' \to \mu\nu)$ at 95% C.L. Both quantities are determined using the statistical analysis described in Sections 2.5.1 and 2.5.2, taking as input the measured transverse mass distribution together with the corresponding systematical uncertainties.

Figure 2.36 shows the observed and expected excluded cross section, $\mu_{exc}\sigma_{W'}\mathcal{B}(W' \to \mu\nu)$, at the 95% C.L. The green and yellow bands correspond to one and two standard deviations from the expected excluded cross section. Taking as reference a potential W′ with mass larger than 2 TeV, new $W' \to \mu\nu$ interactions in which $\sigma_{W'}\mathcal{B}(W' \to \mu\nu) > 3$ fb at 13 TeV are excluded by this measurement at 95% C.L.

Additionally, the plot also shows a black line that corresponds to the theoretical predictions from the benchmark model SSM W', $\sigma_{W'_{SSM}}\mathcal{B}(W' \to \mu\nu)$. Those mass hypothesis in which the observed excluded cross section is below the SSM W′ theory line, are said to be excluded at 95% C.L. in the model. In other words, SSM W′ bosons with masses below 3.9 TeV are excluded with this measurement at 95% C.L.

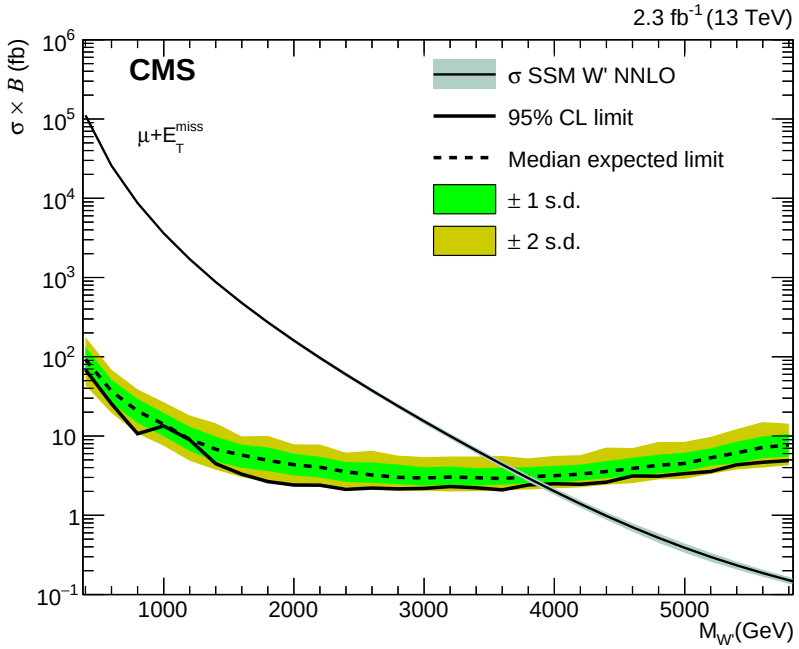

FIGURE 2.36: Expected (dashed line) and observed (solid line) 95% C.L limits to $\sigma_{W'}\mathcal{B}(W' \to \mu\nu)$ for SSM W′ bosons at 13 TeV. The shaded green and yellow bands represent the one- and two-sigma uncertainty bands. Also shown is the QCD NNLO SSM W′ theoretical cross section with the corresponding PDF uncertainty band in grey.

A twin search with similar sensitivity has been analysed independently by other collaborators from CMS in the electron[6] channel, $W' \to e\nu$ [20, 21]. In their study, the same 13 TeV dataset has been analysed using an strategy designed to discover a potential excess in events containing a high momentum electron and large missing transverse energy. Unfortunately, the measurements in the electron channel are also in agreement with the predictions of the SM, and there are no hints of physics beyond the SM.

The statistical interpretation was done using the same formalism as the one described in this manuscript. Figure 2.37 shows the excluded signal cross, $\mu_{exc}\sigma_{W'}\mathcal{B}(W' \to e\nu)$, as a function of the W′ mass. Since the sensitivity of the electron and muon analysis are almost identical, SSM $W' \to e\nu$ interactions are excluded up to 3.8 TeV in mass.

Given the non-observation of new physics in any of the two channels, one could make further assumptions and combine both channels as if they were two independent measurements of the same W′ decay. The combination, done by the $W' \to e\nu$ team, assumes that the potential new W′ presents lepton flavour universality as the SM W. Under this assumption, the excluded signal cross sections is pushed further and SSM W' with masses up to 4.2 TeV are excluded as shown in Figure 2.38.

[6] An introduction to electrons in CMS is provided in the context of the Z + c analysis in Section 3.3.2

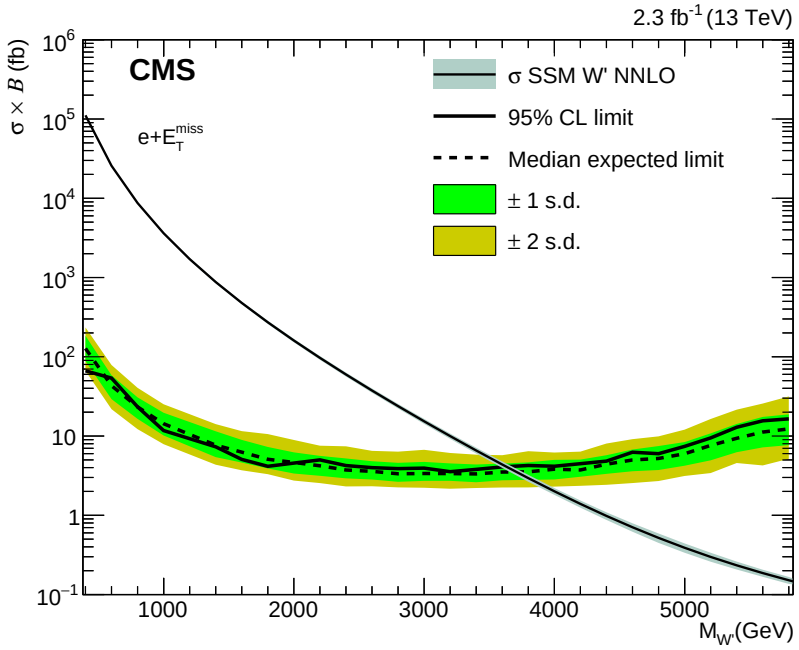

FIGURE 2.37: Expected (dashed line) and observed (solid line) 95% C.L limits to $\sigma_{W'}\mathcal{B}(W' \to e\nu)$ for SSM W′ bosons at 13 TeV. The shaded green and yellow bands represent the one- and two-sigma uncertainty bands. Also shown is the QCD NNLO SSM W′ theoretical cross section with the corresponding PDF uncertainty band in grey. Credit for the study: Collaborators from the $W' \to e\nu$ search analysis team.

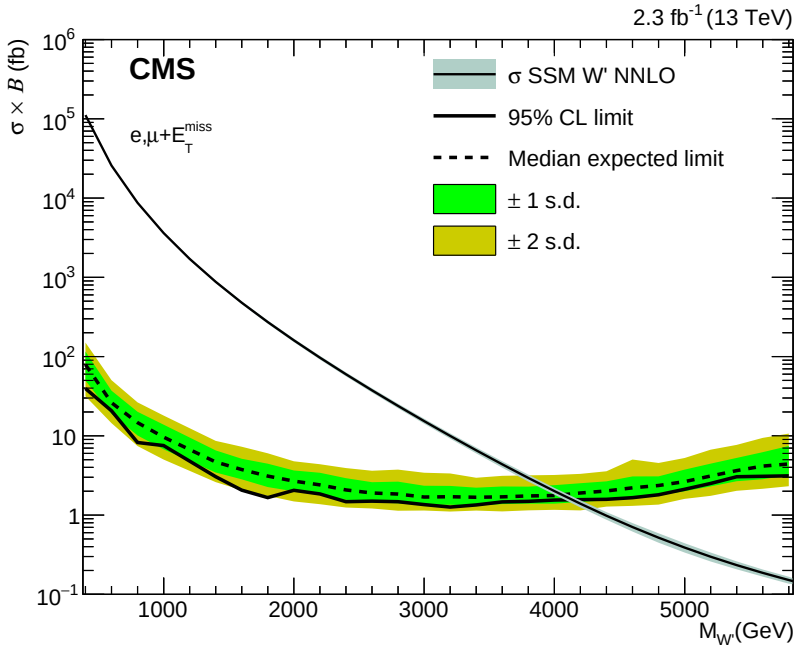

FIGURE 2.38: Expected (dashed line) and observed (solid line) 95% C.L limits to $\sigma_{W'}\mathcal{B}(W' \to \ell\nu)$ for SSM W′ bosons at 13 TeV in the combined channel assuming lepton flavour universality. The shaded green and yellow bands represent the one- and two-sigma uncertainty bands. Also shown is the QCD NNLO SSM W′ theoretical cross section with the corresponding PDF uncertainty band in grey.

How does this new $W' \to \mu\nu$ search at 13 TeV compare to existing constrains at 8 TeV from previous searches? The 13 TeV results presented in this thesis are more stringent than previous searches for the highest W' masses. However for low W' masses, the 8 TeV results are still more stringent due to a much large integrated luminosity. According to simulations, the crossover point is expected to be around a W' with mass 2 TeV where both analysis have a similar sensitivity. The reason why the 13 TeV analysis is more sensitive for the highest masses was discussed in more detail in Section 1.2.4.

Finally the combined 8 TeV + 13 TeV upper limit to the signal strength μ_{exc} at 95% C.L. for SSM $W' \to \mu\nu$ bosons is shown as a black line in Figure 2.39. The separation between the expected 8 TeV result (dashed red line) and the new 13 TeV result (dashed blue line) for a given mass, shows the improvement in sensitivity of the new result. As expected, the largest improvement in sensitivity is for W' bosons with mass larger than 2 TeV. In the context of the SSM W', this new result at 13 TeV extends the current 8 TeV exclusion by about 900 GeV in mass.

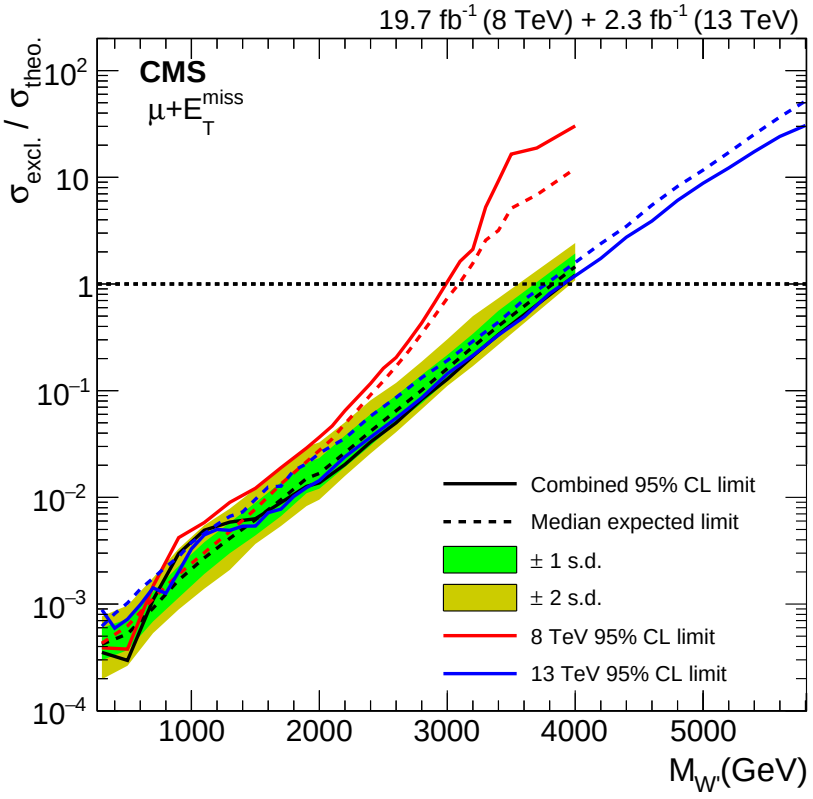

FIGURE 2.39: Expected (dashed line) and observed upper limits (solid line) to the signal strength μ_{exc} at 95% C.L using 19.7 fb^{-1} at 8 TeV of data (red), this work (in blue) and the combination (black) of both datasets 19.7 fb^{-1} (8 TeV) + 2.3 fb^{-1} (13 TeV) for SSM $W' \to \mu\nu$ bosons. The shaded green and yellow bands represent the one- and two-sigma uncertainty bands with respect to the expected combination.

Summary of main physics results of the $W' \to \mu\nu$ search analysis

Result 1: No evidence of excess in the $\mu + E_T^{miss}$ final state

After the analysis of 2.3 fb^{-1} of data recorded at a centre of mass energy of 13 TeV, there are no hints of anomalous production compared to Standard Model expectations in the transverse mass spectrum, M_T, in events with just one isolated high momentum muon and large missing transverse energy.

Result 2: Upper limit to the $W' \to \mu\nu$ production cross section at 13 TeV

This analysis is able to exclude BSM theories that predict new $W' \to \mu\nu$ interactions if their $\sigma_{W'}\mathcal{B}(W' \to \mu\nu) > 3$ fb^{-1} at 13 TeV and their mass is larger than 2 TeV. This result is derived assuming that the hypothetical W' has kinematical properties similar to the W' boson of the SSM benchmark model.

Result 3: SSM $W' \to \mu\nu$ excluded mass

In the context of the W' Sequential Standard benchmark model, this analysis excludes SSM $W' \to \mu\nu$ bosons if their mass is smaller than 3.9 TeV.

Result 4: SSM $W' \to \ell\nu$ excluded mass

Under the assumption of lepton flavour universality, W' bosons are excluded if their $\sigma_{W'_{W'}}\mathcal{B}(W' \to \ell\nu) > 2$ fb^{-1} at 13 TeV and their mass is larger than 2 TeV. In the context of the SSM, this limit translates into an exclusion of W'_{SSM} bosons with masses smaller than 4.2 TeV.

Result 5: Comparison of 13 TeV analysis vs 8 TeV searches

The analysis presented in this manuscript has higher sensitivity for discovery of W' bosons with mass above 2 TeV compared to the previous 8 TeV search. The higher sensitivity of the 13 TeV analysis at the highest masses allows to extend the excluded SSM $W' \to \mu\nu$ mass region by about 900 GeV.

2.6 Search for new W' bosons: conclusions

The Standard Model has demonstrated its capacity to provide a coherent and precise theoretical framework for particles and interactions up to the TeV energy scale in particle accelerators during the past 50 years. The discovery of the Higgs boson in 2012 closes the search for new particles within the SM and provides an experimental evidence of a scalar required to give mass to the particle content in the theory. Despite the undoubted success of this discovery, its light mass brings new conceptual problems in the SM formulation. The hierarchy problem or the vacuum stability of the Higgs potential are two examples of this kind. Additionally, recent observations in cosmology and in neutrino experiments point to modifications in the SM theory to accommodate a dark matter candidate, as well as the mass of the neutrino. Motivated by some of these questions, theoretical physicists try to build extensions of the SM, while keeping the SM as a lower energy limit to avoid tension with current measurements. Among these new theories, extended gauge sectors of the SM predict the existence of new particles and interactions that could be produced in particle physics accelerators.

The goal of this analysis is the search of new interactions/physics, manifested via new charged gauge spin-1 bosons $\mathrm{W}' \to \mu\nu$, produced in the first 2.3 fb^{-1} at 13 TeV of proton proton collisions recorded in CMS. Due to the plethora of extensions of the SM that predict additional W' bosons, historically experimental collaborations used as reference the benchmark model Sequential Standard Model W'. In this model, the predicted W' is a carbon copy of the W boson with the same couplings but higher mass. The W' coupling to vector bosons is assumed to be suppressed. As a consequence, one detectable signature of this new W' boson would be a high momentum muon accompanied with large missing transverse energy.

After more than two years of upgrade from early 2013 to the Summer of 2015, the LHC and the CMS detector have restarted its operation at a new centre of mass energy of 13 TeV. The increase in centre-of-mass energy leads to a higher sensitivity for the discovery of heavy W' bosons compared to previous searches in CMS with already the first fb^{-1} of data at 13 TeV. Due to this, an early calibration and validation of the muon momentum measurement at high p_{T} is a key aspect in the analysis. This calibration was done using cosmic muons recorded prior to the data taking, as well as high momentum muons produced in dimuon events recorded in collision data.

After analysing 2.3 fb^{-1} at 13 TeV accumulated in 2015, there are no signs of new $\mathrm{W}' \to \mu\nu$ like interactions at the new centre-of-mass energy. The statistical analysis of the measurement determined that, in order to make a potential signal compatible with the observations, the new W' bosons should have a production cross section smaller than

$$\sigma_{\mathrm{W}'}\mathcal{B}(\mathrm{W}' \to \mu\nu) < 3 \text{ fb}^{-1} \text{ at 13 TeV if its mass is larger than } M_{\mathrm{W}'} > 2 \text{ TeV}.$$

In the context of the benchmark model SSM, this cross section limit leads to the exclusion of $\mathrm{W}' \to \mu\nu$ bosons with mass lighter than 3.9 TeV, and lighter than 4.2 TeV when combined with the electron channel. This new result pushes the excluded W' masses by about 900 GeV with respect to the 8 TeV analysis. These results were derived with the assumption that the W' has similar kinematical properties to the one described by the SSM and the absence of interference between the W and the W'.

In summary this analysis verifies and extends the validity of the SM at the new centre-of-mass energy, and at the same time, excludes extensions of the SM that predict a $\mathrm{W}' \to \mu\nu$ interaction with production cross section higher than the cross section limits presented in this manuscript.

Chapter 3

$\mathrm{Z} + \mathrm{c}$ cross section measurement

3.1 Standard model against $\mathrm{Z} + \mathrm{c}$ and $\mathrm{Z} + \mathrm{b}$

The Standard Model of Particle Physics has demonstrated a clear success providing theoretical predictions that have been tested experimentally against CMS measurements, as already shown in the first chapter in Figure 1.2.

Z bosons decaying to isolated charged lepton pairs, $\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$ and $\mathrm{Z} \to \mu^+\mu^-$, are one of the experimentally cleanest, theoretically well understood and abundant processes produced at the LHC. The measurement of their production rate in association with additional jets, serves as an ideal benchmark to test the finest theoretical predictions from QCD (Section 1.2.2) and the proton PDFs (1.2.1). The associated production of a vector boson (V=W,Z) with jets is denoted as V + jets in the text, and corresponds to the first two columns in the same Figure 1.2. In this area, CMS has measured the differential cross section at 7 TeV, 8 TeV and 13 TeV for different Z + jets and W + jets processes up to several additional jets. All the measurements show a good agreement compared to the theoretical predictions.

The excellent tracking capabilities (see Section 1.3.1) to identify displaced tracks and reconstruct secondary vertices inside jets from D and B hadron decays, allows CMS also to measure separately the associated production of vector bosons with jets originated in the hadronization of charm or bottom quarks. The latter will be denoted in the text as V + Heavy Flavour (HF). The associated production of bosons with heavy flavour jets is particularly interesting to test the proton PDF for quarks other than up and down. For example, the measurement of the W + c process is sensitive to the modelling of the strange quark PDF [97], or the study of Z + b is sensitive to the bottom quark PDF [98].

Figure 3.1 shows a summary of the measured cross sections for V + jets and separately for V+HF processes compared to theoretical predictions. Measured cross sections for W + c, W + b and Z + b processes are also in agreement with the theoretical predictions within uncertainties.

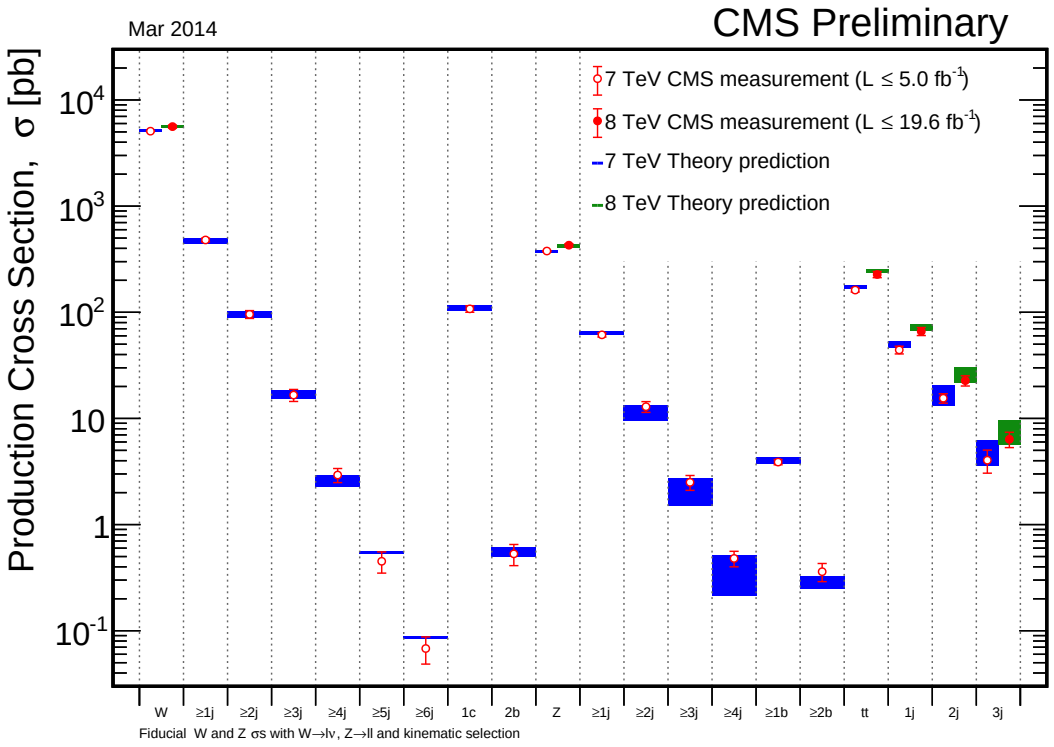

FIGURE 3.1: Measured cross sections (when available) at two different center of mass energies, compared to QCD predictions of the Standard Model. The figure covers processes such as W, Z and $t\bar{t}$ produced in association with multiple jets or heavy flavour jets.

There is an important measurement that is missing in Figure 3.1, the associated production of a Z boson with charm quarks, hereafter denoted as Z + c. The measurement of the Z + c process presented in this thesis fills this hole in the CMS physics program, and sets the first step towards a new era of precision Z + c measurements in CMS. Figure 3.2 shows the leading order Feynman diagram associated to this process. The physical motivation of this analysis is to test the QCD and charm PDF predictions of the Monte Carlo generators in describing the Z + c process, and its relative production to the associated production with bottom quarks, Z + b.

Recently there is a renewed interest from the theory community to study the charm content of the proton [99]. Current PDF sets at the LHC assume that charm is generated perturbatively through pair production from gluons, $g \to c\bar{c}$. However, it is not experimentally discarded that there could be an additional intrinsic charm component in the proton with non-perturbative origin. Several studies point out to a potential modification of the charm PDF, if there is an additional intrinsic componenent at large

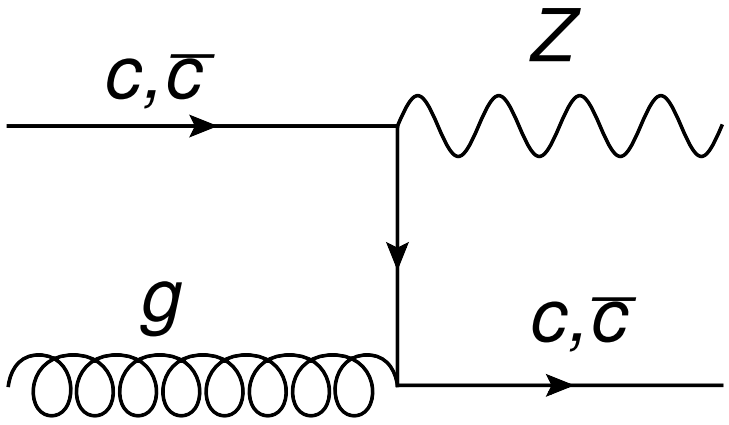

FIGURE 3.2: Leading order production of Z+charm at the LHC

x. This component, if existing, would translate into an enhancement in the differential Z + c production for large transverse momentum of the Z boson, p_{T}^{Z}, or large transverse momentum of the c-jet [100–103]. Charm PDF and potential enhancements due to intrinsic charm are discussed in Section 3.1.1.

Additionally, the associated Z+HF production appears as one of the major backgrounds in many searches in the CMS physics program. For example, the measurement of the SM Higgs boson coupling to bottom quarks, $\mathrm{Z}^{*} \to \mathrm{H}(\mathrm{b\bar{b}})\mathrm{Z}(\ell\ell)$, has a dominant background of Z+HF jets. The theoretically interesting, and at the same time experimentally challenging charm coupling, $\mathrm{Z}^{*} \to \mathrm{H}(\mathrm{c\bar{c}})\mathrm{Z}(\ell\ell)$, would also require a good simulation of the $\mathrm{Z} + c\bar{c}$ background as well as the Z + c/Z + b ratio.

Moving to more exotic searches, Flavour Changing Neutral Currents (FCNC) are highly supressed in the SM. However, the measurement of new physic processes of the kind $t \to \mathrm{Z}+\mathrm{c}$ [104] would require a good knowledge of Z + c production. Diboson resonances of the type $X \to \mathrm{Z}(\ell\ell)\mathrm{V}(\mathrm{qq})$ often use b-tagging techniques to maximize their sensitivity, and deal with Z+HF as dominant background. Another example are searches for third-generation of supersymmetric quarks, $\tilde{t}$, decaying via charm quarks in final states with a c-jet and a large missing transverse energy, $\tilde{t} \to c\tilde{\chi}^{0}$ [105, 106]. In the $\tilde{\mathrm{t}}$ search, $\mathrm{Z}(\nu\nu)+\mathrm{c}$ is one of the dominant backgrounds, and an accurate estimation of its contribution can thus be obtained from the study of the same process where the Z boson decays to charged leptons.

This chapter presents the fiducial cross section measurement of the production of a Z boson associated with at least one charm quark, $\sigma(pp \to \mathrm{Z} + \mathrm{c} + \mathrm{X})$, and its relative production with respect to bottom, $\sigma(pp \to \mathrm{Z} + \mathrm{c} + X)/\sigma(pp \to \mathrm{Z} + \mathrm{b} + X)$. Both measurements are determined inclusively and differentially as a function of the Z transverse momentum, p_{T}^{Z}, and the jet with heavy flavour content, p_{T}^{jet}.

The strategy employed in the analysis selects Z bosons through its decay into electrons or muons pairs, plus an additional jet with heavy flavour content selected in one of the following three independent channels:

- Semileptonic channel: A displaced secondary vertex in which one of the tracks is a well identified muon. This channel is designed to select semileptonic decays of $D \to \mu + X$ and $B \to \mu + X$ hadrons.
- $D^{\pm}$ channel: A displaced secondary vertex made of three tracks in which the invariant mass of the vertex is compatible with a $D^{\pm} \to K^{\mp}\pi^{\pm}\pi^{\mp}$ decay.
- $D^{*\pm}(2010)$ channel: A displaced secondary vertex made of two tracks in which the invariant mass is consistent with a $D^0 \to K^+\pi^-(\bar{D}^0 \to K^-\pi^+)$ decay, together with a primary vertex track called soft pion, π_s, all being consistent with the $D^{*-}(2010) \to D^0\pi_s^-(D^{*+}(2010) \to \bar{D}^0\pi_s^+)$ decay chain.

Finally, there are experimental motivations to develop a Z + c measurement in the context of flavour tagging studies with charm jets. For example, CMS has developed for the first time a dedicated charm tagger for the Run II, and techniques and methods developed in this thesis are currently used to calibrate simulations of the new tagger with 13 TeV data [107].

A preliminary version of this analysis internally reviewed by CMS, was released for the first time during the ICHEP2016 conference. The analysis was originally published in the form of a Physics Analysis Summary (PAS) and documented here [22].

3.1.1 The charm quark PDF

Modern PDFs[1] assume that charm is generated purely perturbatively inside the proton in $g \to c\bar{c}$ processes above a minimum threshold dictated by the charm-quark mass, m_c. Technically this translates in the following constraint in the determination of the charm quark PDF: $f_c(x, Q) = f_{\bar{c}}(x, Q) = 0$ if $Q < m_c$. At higher Q, the charm is generated perturbatively though the DGLAP evolution equation. A direct consequence of this assumption is that charm quark PDFs generated perturbatively are correlated with gluon PDFs, and their predictions are affected by the value and definition of m_c.

However this assumption is not necessarily completely true. Over the years several models have been proposed that predict the existence of an additional non-perturbative or intrinsic charm component. Their existance could lead to a modification of the charm PDF. For example, the BHPS intrinsic charm model [108], predicts a valence-like evolution of the charm PDF that could lead into enhancements at high x. A recent study from the CT14 collaboration [109] studied the compatibility of a intrinsic charm contribution with current measurements, assuming the existence of an additional BHPS charm at high x. They computed an upper limit of $x_{IC} \leq 2.5\%$ at 90% C.L. to the total fraction of momentum carried by a potential intrinsic charm in the proton. The conclusion of their study is that current measurements can be described without a intrinsic charm component, however a small contribution is not completely excluded by data at high x.

As of today some questions still remain open... is there intrinsic charm in the proton?, what is the energy scale (if any) at which the charm PDF is non-zero?, is it m_c? what is the total fraction of the proton's momentum carried by charm? Some of these questions could be answered if the charm PDF was fitted and determined separately with other PDFs. With this spirit, the NNPDF collaboration tried to quantify the charm content of the proton and determine a intrinsic charm component (if any) from available measurements [101]. In their study, the charm PDF is obtained using the NNPDF fit methodology along with the gluon and light PDFs assuming $f_c(x, Q) = f_{\bar{c}}(x, Q)$. Their result suggests that the fitted charm has perturbative origin for $x \leq 0.1$ and vanishes for all x below $Q \approx 1.6$ GeV near the m_c threshold. Additionally the fitted charm PDF, shown in Figure 3.3, presents a mild intrinsic component that peaks around $x \approx 0.5$. The latter is not present when the charm component is only generated perturbatively.

The total fraction of momentum of the proton carried by the fitted charm is measured to be $0.7 \pm 0.3\%$ at 68% C.L. whereas in the case in which the charm is not fitted and produced only perturbatively is found to be $0.239 \pm 0.003\%$. One of the main predictions of the NNPDF fitted charm PDF that could be tested in a CMS measurement,

[1] Prior to read this Section is convenient to read the introduction to PDFs in Section 1.2.1

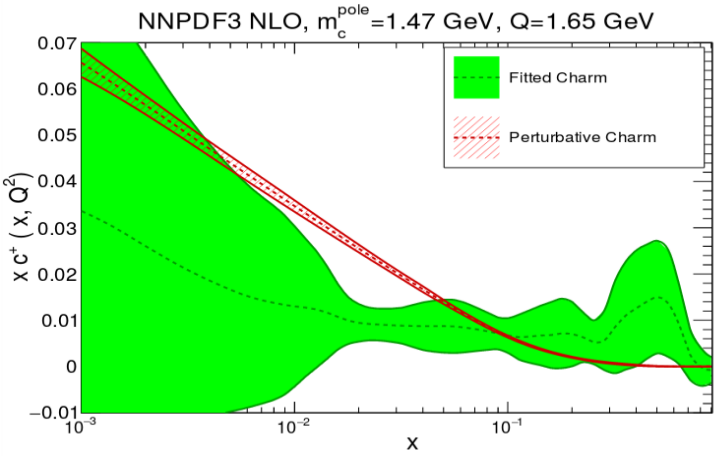

FIGURE 3.3: Charm PDF generated perturbatively compared to fitted charm PDF by NNPDF collaboration. Figure taken from [101]

is a potential enhancement of the Z + c production at large transverse momentum and pseudo-rapidity. A similar Z + c enhancement at high p_T^Z is also present in different intrinsic charm models such as the BHPS studied by the CT14 group. Figure 3.4 shows the size of the potential Z + c enhancement as a function of p_T^Z according to the predictions from the fitted NNPDF charm, as well as the non-excluded predictions from the BHPS intrinsic charm model from the CT14 collaboration.

According to these studies, the size of potential Z + c enhancements could be as large as 20%(100%) at $p_T^Z \approx 100(600)$ GeV in the most optimist predictions [101]. However it must be noted that all these predictions suffer from large theoretical uncertainties, and the size of the enhancement could be much smaller.

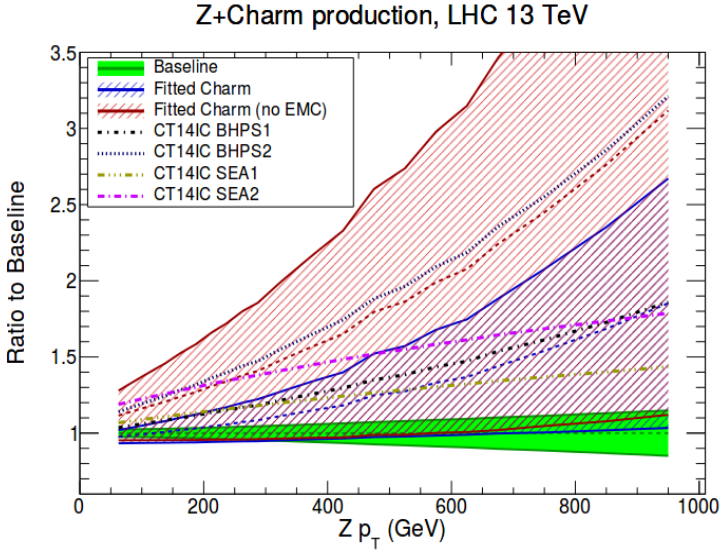

FIGURE 3.4: Z boson transverse mass distribution computed using the fitted charm from the NNPDF group and the different non-excluded predictions from CT14, each of the predictions is normalized to charm generated perturbatively [101].

3.2 Data samples and simulation

3.2.1 2012 8 TeV dataset

The dataset used in the manuscript for the Z + c cross section measurement was recorded at a centre-of-mass energy of 8 TeV between March and December in 2012, and amounts up to 19.7 fb^{-1}. This dataset corresponds to a re-calibrated sample with the latest calibrations derived for the sample, and has been declared with the highest quality for physics analysis by the CMS data certification team.

3.2.2 Monte Carlo simulations

All the samples of simulated events are produced with MC event generators, both for signal and background processes. The signal is generated with a sample of Z boson plus jets events using MADGRAPH [35] event generator and PDF set CTEQ6L [110], interfaced with PYTHIA6 [33] for parton showering and hadronization. The MADGRAPH generator produces parton-level events with a vector boson and up to four partons on the basis of a matrix-element calculation. The matching scale is set to 10 GeV^2, and the factorization and re-normalization scales are set to $Q^2 = M^2_{Z/W} + p^2_{T,Z/W}$.

Z + c, Z + b and Z + light definition and event classification

The MADGRAPH signal sample includes Z-boson events accompanied by generator level jets originating from quarks of all flavours (b, c and light = u, d, s, g). Generator level jets are built with the anti-k_T algorithm with radius 0.5, clustering all stable particles after fragmentation and hadronisation except neutrinos. Events are classified as Z + b, Z + c or Z + light according to the flavour of the generator level jet defined as follows:

A generator level jet is defined to be b-flavoured if there is a b-meson ($500 < |\mathrm{PDG_{id}}| < 600$) or b-baryon ($5000 < |\mathrm{PDG_{id}}| < 6000$) among the particles generated in the event, within a cone of radius $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.5$ of the jet axis. The $\mathrm{PDG_{id}}$ for different mesons or barions is documented here [111]. Similarly, a generator level jet is considered to be c-flavoured if there is a c-meson ($400 < |\mathrm{PDG_{id}}| < 500$) or c-baryon ($4000 < |\mathrm{PDG_{id}}| < 5000$) and no b-hadrons within a cone of $\Delta R = 0.5$ of the jet axis.

A Z + jets event is assigned as a Z + b event if there is a generator level jet with $p_T^{jet} > 15$ GeV identified as a b-flavoured jet. Analogously a Z + jets event is classified as Z + c if there is a c-flavoured generator level jet with $p_T^{jet} > 15$ GeV and no b-flavoured generator level jets. Finally a Z + jets event is classified as Z + light when is not classified as Z + c or Z + b. Signal events refer only to those classified as Z + c or Z + b, whereas Z + light are considered as background.

The main backgrounds other than Z + light are the production of $t\bar{t}$ and dibosons processes (WW, WZ, and ZZ). A sample of $t\bar{t}$ events is generated with POWHEG v1.0 [34], interfaced with PYTHIA6 and CT10 for the PDF [29]. Diboson processes are modelled with samples of events generated with PYTHIA6 using the PDF set CTEQ6L1 [110]. The tau channel, $Z \to \tau^+(\mu^+\nu)\tau^-(\mu^-\nu)$ and $Z \to \tau^+(e^+\nu)\tau^-(e^-\nu)$, is also taken into account, however its contribution is negligible compared to that from electrons and muons. Finally a W plus jets sample is generated with MADGRAPH generator, interfaced with PYTHIA6 and is only used in Section 3.4.3 where complementary studies with an auxiliary W + c sample are presented.

Figure 3.5 shows a sketch of the main backgrounds that could mimic a Z plus heavy flavour jets signal. In the figure, red lines correspond to dilepton pairs (e^+e^- or $\mu^+\mu^-$) and green lines to heavy flavour jets (charm or bottom). Yellow lines correspond to additional particles that contribute to the final state, and black and blue lines correspond to intermediate particles.

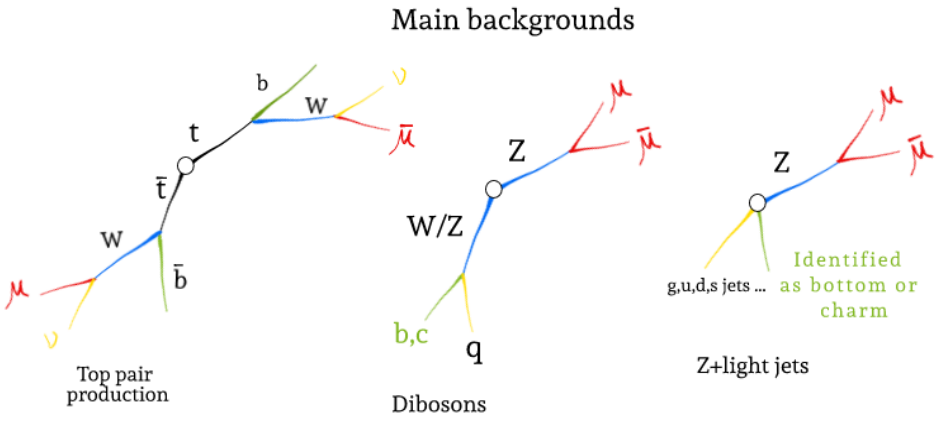

FIGURE 3.5: Sketch of the main processes that could mimic a Z plus heavy flavour signal at the LHC. Details about of each of the processes and the colour code used for the lines in the sketch are provided in Section 2.2.2.

Distributions from the simulated samples along this chapter are normalized to an equivalent integrated luminosity of the recorded data, 19.7 fb^{-1}. For electroweak processes the cross sections are normalized to NNLO calculation computed with FEWZ [36], using the PDF set MSTW2008NNLO [30]. Diboson processes are normalized to NLO predictions from MCFM [37] and the $t\bar{t}$ cross section is taken at NNLL from [112].

Simulated samples are corrected for differences between data and simulation in describing the lepton trigger, reconstruction and identification efficiencies as well as the inclusion of a pileup re-weighting factor. All these corrections were derived in a similar manner as explained in Section 2.3 with the 8 TeV dataset. Correction factors are applied to the simulation as weights on an event-by-event basis as in the W′ search (Equation 2.12).

3.3 Physics objects: leptons and heavy flavour

Lepton pairs with opposite charge, isolated from hadron activity and with the invariant mass compatible with the Z boson mass, provide a clean signature that is used to select Z + jets events. The following Sections 3.3.1 and 3.3.2 describe how lepton candidates from $Z \to \mu^+\mu^-$ and $Z \to e^+e^-$ are selected. Once potential candidates are identified, heavy flavour tagging techniques are used to enrich the Z + jets sample in Z + c and Z + b by removing Z + light events. Section 3.3.3 introduces some of the heavy flavour techniques used in CMS, and particularly in the Z + c cross section analysis.

3.3.1 Muons

Muons were already extensively discussed in the context of the $W' \to \mu\nu$ search in Chapter 2. Due to this and especially to avoid text repetition, this section assumes familiarity with the concepts described already in Section 2.3, and focuses on the main differences with respect to a high p_T single muon analysis.

The trigger strategy employed in the Z + c cross section measurement is slightly different from the one used in the $W' \to \mu\nu$ analysis described in Section 2.3.1. The main difference is the usage of a dimuon trigger instead of a single muon trigger. The choice of a dimuon trigger is important in the Z + c analysis because it allows lower p_T thresholds of the muons, yielding to an increase of the number of selected signal candidates.

Each muon trigger track candidate is built combining muon detector information with silicon tracker detector information. The resulting track must have a transverse momentum larger than 17(8) GeV for the leading (sub-leading) muon trigger candidate. Only reconstructed muons in the pseudo-rapidity range $|\eta| < 2.1$, and with transverse momentum larger than 20 GeV are considered for the analysis. For this selection, more than 90% of the $Z \to \mu^+\mu^-$ candidates in data are selected by the dimuon trigger depending on their muon pseudo-rapidity [113]. The trigger performance in simulations is corrected using data-to-simulation calibration factors derived using the Tag & Probe technique (see Section 2.3.1).

Dimuon pairs in the Z + c analysis use an identification criteria called Tight muon, that is similar to the High-p_T muon identification described in Section 2.3.3. The main difference between both definitions is that there is no requirement in σ_{p_T}/p_T of the track, but there is a requirement in the $\chi^2/n.d.f < 10$ of the track instead. The latter was dropped for High-p_T muons to avoid losses in efficiency for muons that produce an electromagnetic shower in the muon system. Additionally, the isolation definition is also

modified compared to the W′ analysis. An isolation variable, I_{comb}, is built to test the presence of additional activity around the muon in a cone of $\Delta R = 0.4$.

$$I_{comb} = \sum_{\Delta R} p_{\mathrm{T}}(\text{charged hadrons}) + E_{\mathrm{T}}(\text{neutral hadrons}) + E_{\mathrm{T}}(\text{photons}) + \text{Pileup corrections} \tag{3.1}$$

The isolation variable is corrected in order to exclude from the sum in Equation 3.1, particles originated in a different primary vertex. A muon is said to be isolated if $I_{comb}/p_T^{\mu} < 0.20$. The efficiency for a real muon to be identified as Tight muon is close to the one for High-p_{T} muons, as well as for calibration factors that are also close to unity [114].

Finally the muon momentum scale and resolution are mostly dominated by the internal tracker alignment at the p_{T} range relevant for the analysis. Due to this, the muon curvature assignment was calibrated in data and simulation using $\mathrm{Z} \to \mu^+\mu^-$ events, correcting for effects such as potential tracker misalignments as well as mismodellings of the tracker material among other effects.

3.3.2 Electrons

The $\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$ and $\mathrm{Z} \to \mu^+\mu^-$ leptonic channels are treated similarly in the Z + c cross section measurement. Both channels should provide the same measurement and they serve each other as an independent cross check. This section provides a short description about the electron trigger, reconstruction and identification in the Z + c analysis, as well as some of the differences between low and high energy electrons as the ones used in a $\mathrm{W}' \to \mathrm{e}\nu$ search.

The electron trigger used to select $\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$ candidates requires the presence of at least two energy clusters in the ECAL detector, each associated with a track reconstructed in the tracker detector. The total transverse energy of the leading (sub-leading) energy cluster has to be larger than 17(8) GeV. The trigger efficiency for $\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$ candidates is larger than 98 % for potential signal candidates reconstructed with $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.1$ [115].

Electrons are reconstructed using a Gaussian Sum Filter (GSF) algorithm [45]. In this algorithm, candidates are built from energy clusters in the ECAL. These clusters are matched with hits in the silicon pixel detector after recovering energy losses due bremsstrahlung emission along the electron trajectory before reaching the ECAL [116]. This energy spread, mostly along ϕ and small around η due to the magnetic field, defines

a characteristic shape of the energy deposit in the ECAL that is also used to identify electrons efficiently.

A super-cluster (SC) is built from the individual energy clusters consistent with the electron candidate, and its position is determined from a energy weighted average of them. The electron candidate energy is given by the energy of the associated super-cluster, E_{SC}, which needs to be adjusted with calibration and regression methods [45]. Remaining differences in the electron scale and resolution are calibrated using $Z \to e^+e^-$ events.

Different quality criteria are imposed to reconstructed electron candidates in order to ensure that they come from real electrons. This step is called electron identification and refers to a minimum set of quality criteria, shown in Table 3.1, required to an electron in order to be considered as a valid candidate from $Z \to e^+e^-$. The definitions of the variables listed in the table that enter in the identification of electrons are the following:

- $|\Delta\eta|$ and $|\Delta\phi|$: They define the angular separation along η and ϕ between the super-cluster position and the extrapolated tracker track from the primary vertex to the ECAL.
- H/E_{SC}: ratio between the energy as measured in the HCAL in a cone of radius $\Delta R = 0.15$ around the electron position in the ECAL, H, to the energy of the super-cluster, E_{SC}.
- $\sigma_{\eta\eta}$: A measure of the energy spread along η of the energy cluster. Due to the magnetic field pointing towards $+z$, the energy spread for a real electron should be small in η and larger in ϕ.
- $|1/E_{SC} - 1/p|$: A measure of the difference between the super-cluster energy and the measurement of the electron momentum in the tracker. This quantity should be small for real electrons.
- d_{xy} and d_z: transverse and longitudinal impact parameter defined relative to the primary vertex. Real prompt electrons originated at the primary vertex have small impact parameter values.
- Missing hits: number of tracker layers without compatible hits in the reconstructed electron tracker track, this variable is used to reject electrons produced in the conversion of photons.

Variable	ECAL barrel	ECAL endcap
$\|\Delta\eta\|$	< 0.004	< 0.007
$\|\Delta\phi\|$	< 0.06	< 0.03
H/E_{SC}	< 0.12	< 0.10
$\sigma_{\eta\eta}$	< 0.01	< 0.03
$\|1/E_{SC} - 1/p\|$	< 0.05	< 0.05
d_{xy}	< 0.02 cm	< 0.02 cm
d_z	< 0.2 cm	< 0.2 cm
Missing Hits	1	1

TABLE 3.1: Electron identification in the ECAL barrel (second column) and endcaps (third column). More details about each of the variables listed in the first column are provided in the text in Section 3.3.2.

Similarly to muons originated from $Z \to \mu^+\mu^-$ candidates, potential electron candidates from $Z \to e^+e^-$ events are required to be isolated from hadronic activity. The isolation definition for electrons uses the same I_{comb} variable as for the muons (see Equation 3.1) with a reduced radius, $\Delta R = 0.3$. This way, an electron is said to be isolated if $I_{comb}/p_T^e < 0.15$.

The identification efficiency for a real electron, as prescribed with the cuts listed in the Table 3.1, is about 90%(85%) in the barrel (endcaps)[45]. Simulations are corrected to take into account differences between data and simulation in the electron identification plus trigger efficiencies. The size of the calibration factors is between 0.95 and 1.10 [115] and are dominated by corrections to the electron identification efficiency part.

Only well identified electron candidates, reconstructed in the pseudo-rapidity range $|\eta| < 2.1$ and with transverse momentum larger than 20 GeV are considered in the analysis. This particular choice is common with the $Z \to \mu^+\mu^-$ channel, and is adopted to ease the synchronization among both lepton channels.

Finally, the main practical difference between low and high energy electrons such as the ones used in a $W' \to e\nu$ search, is the assignment of the energy measurement. In the case of low momentum electrons, $E_T < 100$ GeV, the best performance is achieved when the electron momentum as measured in the tracker, and the ECAL super-cluster energy are combined together in a single measurement. This combination is done with prior weights based on the individual uncertainty of each measurement. In the case of high energy electrons, the best performance is obtained with the ECAL super-cluster energy alone.

3.3.3 Heavy flavour tagging

This section covers heavy flavour tagging in CMS and is complementary to the Physics Objects described in Section 2.3 since they are also used in the Z + c measurement.

The identification of jets originated in the hadronization of charm or bottom quarks is called heavy flavour tagging. Particularly, the identification of jets coming from charm (bottom) quarks is called c-jet (b-jet) tagging. Heavy flavour tagging benefits from the excellent spatial resolution and tracking capabilities of CMS (see Section 1.3.1 for a description of the CMS Tracker), that allow the reconstruction of primary vertices as well as the identification of displaced tracks with respect to the primary vertex from charm and bottom hadrons (denoted as D and B hadrons) in their decay.

Why is heavy flavour tagging in CMS or in the Z + c measurement needed? The common strategy in most of the analysis in CMS is to maximize the separation between their signal of interest and potential backgrounds. For example, final states that involve isolated leptons such as $Z \to \mu^+\mu^-$ or $W \to \mu\nu$, are excellent signatures to identify potential signals over the overwhelming background of QCD jets[2]. In a similar spirit, heavy flavour tagging is another powerful tool to reduce backgrounds that involve light (u,d,s,g) jets.

Heavy flavour tagging techniques are widely used in CMS analyses. For example, searches of new physics like diboson resonances, Higgs boson searches as well as top or electroweak measurements often use b-tagging techniques to reduce backgrounds of light jets. Currently there are three main algorithms for heavy flavour tagging in CMS. On one hand, the *Combined Secondary Vertex* (CSV) and *Jet Probability* (JP) algorithms, widely used at 8 TeV and 13 TeV, are optimized to maximize the b-tagging against light-tagging. On the other hand, the *charm tagger*, recently released at 13 TeV, is optimized to select charm jets by minimizing the light and bottom tagging power.

The tagging power is quantified by the (b-, c-, light-) tagging efficiencies, defined as the fraction of (b-, c-, light-) jets identified by the algorithm. For a given b-tagging algorithm, there are several working points that correspond to different light-tagging efficiencies. Each analysis has to decide, in a case-by-case basis, what is the ideal balance between b- and light-tagging efficiencies and choose the appropriate working point.

The output of a b-tagging tagging algorithm, for a given jet, is a unique number called the b-jet discriminator. Usually they are defined that the higher the discriminator, more likely the jet to be a true b-jet. Once a working point is chosen, if the b-jet discriminator is larger than the working point, the jet is identified as a b-jet.

[2]This made the $W' \to \mu\nu$ search an experimental "golden final state" to look for new physics

Combined Secondary Vertex

The Combined Secondary Vertex [87] algorithm uses information from the track constituents inside the jet and from possible reconstructed secondary vertices into two likelihood ratio tests. One that discriminates between b- and c-jets and other that discriminates between b- and light-jets. The discriminator is obtained from a combination of both them with prior weights of 0.25 and 0.75 respectively. For example, at 8 TeV a typical working point for the CSV algorithm has a b-tagging efficiency of about $\epsilon_b = 70\%$ with a light-tagging efficiency of $\epsilon_{light} = 1\%$ and for charm $\epsilon_c = 10\%$.

There are many variables that enter in the two CSV likelihood test. For example, some of these variables are the impact parameter for each track in the jet, the mass of a potential secondary vertex, the number of tracks associated to a secondary vertex, the ratio of the energy carried by tracks at the vertex with respect to all tracks in the jet... Among all the variables that enter in the CSV, the one with largest discriminant power between between charm and bottom in jets with a secondary vertex, is the secondary vertex mass. The latter for its simplicity, separation power, and ease to calibrate in simulations was chosen to separate Z + c and Z + b in the semileptonic channel.

Jet Probability

The Jet Probability (JP) [87] uses as b-tag discriminator, P_{jet}, an estimate of the likelihood that all tracks from the jet, N_{track}, are originated at the primary vertex:

$$P_{jet} = \Pi \sum_{i=0}^{N_{track}-1} \frac{(-\ln \Pi)^i}{i!} \quad ; \quad \Pi = \prod_{i=1}^{N} Max(P_i^{IP}, 0.005) \tag{3.2}$$

where P_i^{IP} is the probability density function of the impact parameter for tracks originated at the primary vertex. The latter can be calibrated with tracks from data that have negative impact parameter[3]. Jets in which there are multiple displaced tracks from a secondary decay are highly incompatible with the primary vertex and have large values of the Jet Probability discriminant. In this case, b-jets typically get the highest JP discriminant values because they often contain a larger number displaced tracks with larger impact parameter than c-jets.

In the cases of the reconstructed $D^{\pm}$ and $D^{*\pm}(2010)$ channels, the invariant mass of the secondary vertex mass is resonant at their nominal mass and does not provide any separation power between b- and c- jets. As a consequence, the performance of the Jet Probability is similar to the one provided by Combined Secondary Vertex. Due to the simpler structure of the JP algorithm, the latter was chosen to separate between charm (Z + c) and bottom component (Z + b) in the $D^{\pm}$ and $D^{*\pm}(2010)$ channels.

[3]Primary vertex tracks are equally likely to have positive and negative impact parameter, whereas tracks originated in a secondary vertex mostly have positive impact parameter.

Charm tagging vs Bottom tagging

Charm tagging is challenging and more complicated than bottom tagging in CMS. One of the main reasons for this is the lower mass and lifetime of D hadrons with respect to B hadrons. Charmed hadrons decay closer to the primary vertex and in general complicate a potential secondary vertex reconstruction. In the cases where the secondary vertex is not reconstructed, the kinematical properties of a c-jet resemble a light jet, whereas in the cases where the secondary vertex is reconstructed, it may be misidentified with a b-jet. Probably one of the cleanest way to identify charm jets is by looking at the invariant mass of a reconstructed secondary vertex. The lower mass of charm quarks compared to bottom leads to reconstructed secondary vertices with lower invariant mass. Nonetheless this is not always conclusive. Often b-jets contain $B \to C \to X$ decay chains with a secondary vertex as well as a tertiary vertex. In some cases, only the tertiary with lower mass is reconstructed and can be misidentified with a charm jet.

As an example of the complexity of charm tagging, in the newly released charm tagger in CMS at 13 TeV, for a fixed light and bottom jet efficiency, $\epsilon_{light} = 5\%$ and $\epsilon_b = 20\%$, the charm efficiency is about $\epsilon_c = 25\%$ [107]. In order to increase the charm efficiency to $\epsilon_c = 35\%$ while keeping the light efficiency as low as possible $\epsilon_{light} = 6\%$, the increase in b-tagging efficiency is more than a factor three, $\epsilon_b = 70\%$.

Z+c analysis strategy: reconstruction of exclusive final states

The simulation of the b-tagging discriminants are not perfect and relies in complicated algorithms that need to be calibrated with data. The b-tagging discriminant can be computed for any jet, however the calibration of b-tagging algorithms is usually done in b-jet or light jet enriched topologies. For example, isolated $e\mu$ pairs provide a clean sample that is enriched in b-jets from $t\bar{t}$ decays for b-jet calibration. Unfortunately it is not known experimentally any enriched ($> 85\%$ purity) c-jet topology without any prior selection on the jet content. As a consequence, the calibration of b- and c-tagging algorithms for c-jets is experimentally a challenging task.

The analysis presented in this chapter is a charm precision measurement and unfortunately traditional b-tagging algorithms at 8 TeV, such as CSV and JP, are nor optimized for c-tagging nor calibrated using c-jets. Additionally, due to the similar kinematics of light and c-jets, a strong suppression of the light jet efficiency is needed in order to extract the Z + c in the cleanest possible environment.

In order to overcome some of the above problems, an alternative approach is used in the Z + c analysis in which jets are heavy flavoured tagged in three different final states targeting different jet topologies: semileptonic decays of HF hadrons, and $D^{\pm}$ or $D^{*\pm}(2010)$ secondary vertices. The technical details about how these channels are selected are postponed to Section 3.4.2.

These channels provide a strong suppression of the light-jet efficiency, $\epsilon_{light} < 5 \times 10^{-4}$ leaving a sample with a purity in Z+HF jets of more than 85% in all three cases. The choice of exclusive final states ensures that the largest Z + light background is rejected efficiently, with the additional benefit that the jet topology of such exclusive final states is known. The latter is exploited thanks to W + c and $e\mu$ auxiliary selections that provide a clean samples of charm and b-jets with equivalent jet structure. Their high purity in charm and bottom jets allows to calibrate the differences in jet, track and secondary vertexing performance in simulation with data. The details about how the auxiliary W + c and $e\mu$ samples are selected, and how the simulation is calibrated are provided in Section 3.4.3 and 3.4.4.

Once the Z+HF candidates are selected in one of the three channels, and the simulation calibrated with W+c and $e\mu$ auxiliary samples, the next step is to separate the Z + c and Z + b components using b-tagging algorithms. This is possible thanks to the calibrated discriminant variables of the Jet Probability ($D^{*\pm}(2010)$, $D^{\pm}$) or the secondary vertex mass (semileptonic channel). The details about how the signals are extracted in data are provided in Section 3.4.5.

Finally, the usage of three independent channels allows the extraction of:

$$(\text{semileptonic},\ D^{\pm},\ D^{*\pm}(2010)) \times (Z \to e^{+}e^{-},\ Z \to \mu^{+}\mu^{-}) = 6$$

six independent Z + c cross section measurement that will give an additional level of robustness to the measurement.

3.4 The Z+charm measurement

The experimental signature of the Z + c analysis is defined by two isolated, well identified and oppositely charged leptons (electrons or muons), with an invariant mass compatible with the Z boson mass. Additionally, it is required the presence of at least one reconstructed jet in one of the following three exclusive decay channels: semileptonic and $D^{\pm}$ or $D^{*\pm}(2010)$ vertices.

This section assumes familiarity with the physics objects described in the W′ analysis in Section 2.3, particularly with jets and the missing transverse energy described in Sections 2.3.7 and 2.3.8, as well as with muons in 3.3.1, electrons in 3.3.2 and the heavy flavour tagging techniques in 3.3.3.

The description of the analysis selection is divided in two sequential steps. Firstly, in Section 3.4.1, Z + jets candidates are selected in dilepton plus at least one jet events. In the second step, described in Section 3.4.2, its where this Z + jets sample is enriched in heavy flavour jets and the three exclusive channels explained in detail.

3.4.1 Z+jets selection

The selection of the Z+jets sample follows closely the strategy used in the CMS $Z \to \mu^+\mu^+$ and $Z \to e^+e^-$ cross section analysis at 8 TeV [117]. The leptonic signature is defined by a pair of isolated and well identified[4] muons or electrons with opposite charge. Both leptons have to be reconstructed in the pseudo-rapidity range $|\eta| < 2.1$ with a measured transverse momentum larger than 20 GeV. The invariant mass of the dilepton system, $m_{\ell\ell}$, must lie in the mass range 71 GeV $< m_{\ell\ell} <$ 111 GeV.

In addition to the two leptons, at least one jet with $p_{\mathrm{T}}^{jet} > 25$ GeV has to be reconstructed in the pseudo-rapidity range $|\eta| < 2.5$ defined by the combined tracker and HCAL coverage (see Section 1.3.3 for details about the HCAL). Events in which the angular separation between the jet axis and one of the lepton candidates is less than $\Delta R = 0.5$ are not considered.

Finally the missing transverse energy in the event has to be smaller than $E_{\mathrm{T}}^{miss} <$ 40 GeV. The motivation of this requirement is to reduce the contribution from $t\bar{t}$ background events. The resulting sample with the requirements listed in this section is mostly enriched in Z + light events. As an example, Figure 3.6 shows the dilepton invariant mass in the electron (top), muon (bottom) channel.

[4] Muon and electron identification were introduced in Section 2.3.3 and 3.3.2 respectively

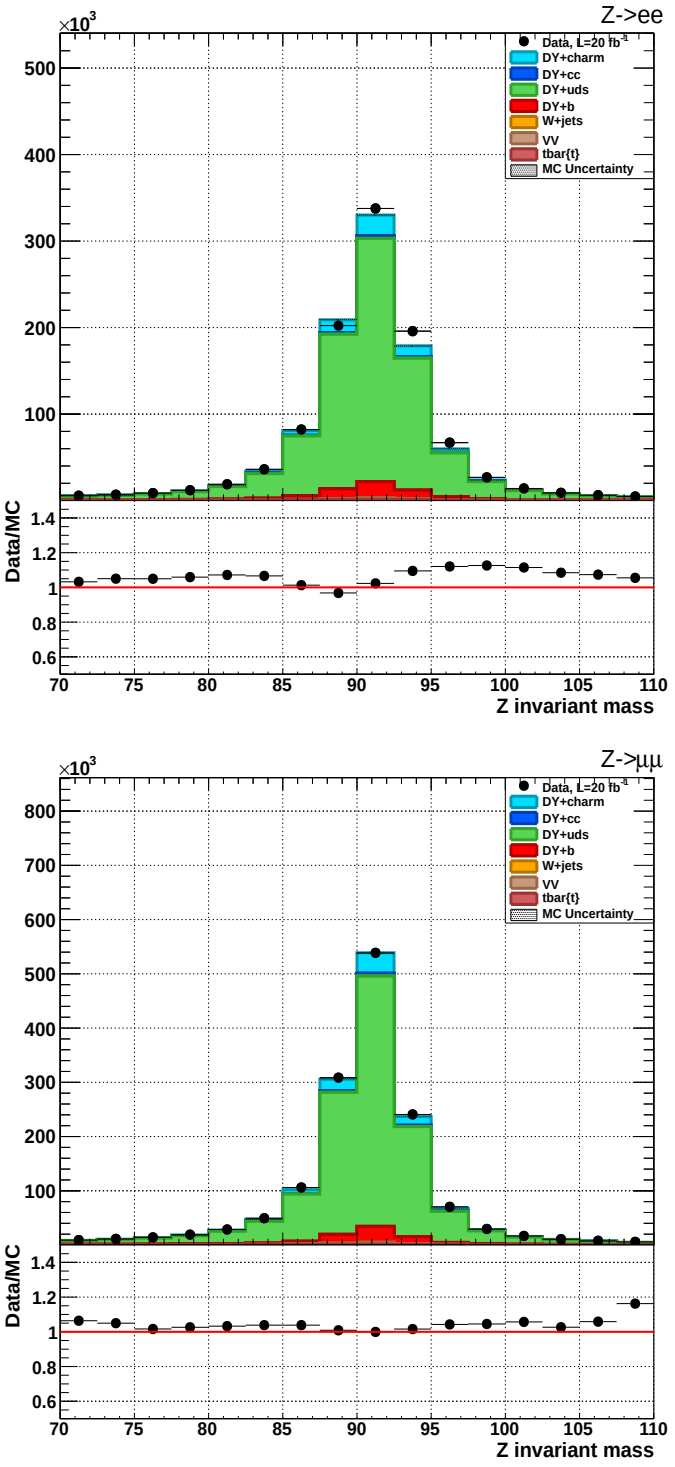

FIGURE 3.6: Dilepton invariant mass in the electron (top) and muon (bottom) channel, after applying the Z + jets selection described in Section 3.4.1. Backgrounds are normalized to the equivalent luminosity of the sample in a similar manner as described in Section 2.4.

3.4.2 Z+heavy flavour selection

This section describes the details of each of the three different channels that are used in this analysis to select heavy flavour jets, as well as the discriminant variables used to separate the Z + c and Z + b signals.

All three channels exploit the fact that bottom and charm hadrons decay weakly with lifetimes of the order of 10^{-12} s and decay lengths at the LHC energies of more than 100 μm. As a consequnce, multiple displaced tracks are originated in a common secondary vertex that may be reconstructed in CMS. Two different algorithms exist to build displaced vertices in CMS: Simple Secondary Vertex (SV) [87] and the Inclusive Vertex Finder Vertex (IVF) [118, 119].

Displaced secondary vertices in the semileptonic and $D^{\pm}$ channels are formed either using a SV or an IVF vertex algorithm; a different technique is employed for the $D^{*\pm}(2010)$ channel. Figure 3.7 shows a sketch of the secondary vertex topology in the three channels. In the figure, the reconstructed primary and secondary vertex are drawn as blue circles, black lines correspond to lines of flights of hadrons, red lines correspond to the minimum reconstructed tracks needed for the vertex category, green tracks are "invisible" tracks (neutrinos) and yellow lines correspond to additional charged tracks associated to the secondary vertex. The particular details about each of the three topologies are provided along this section.

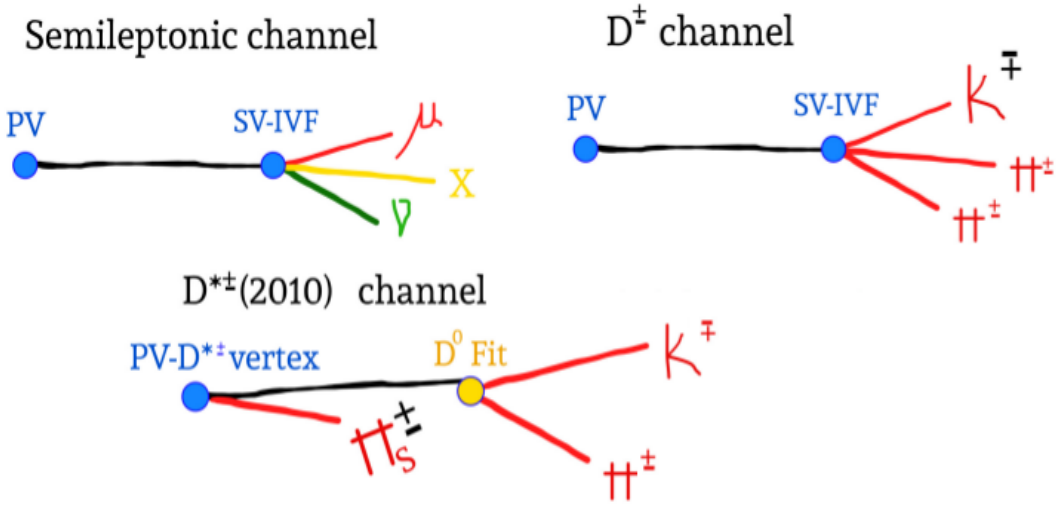

FIGURE 3.7: Sketch of the three channel used to select charm jets: semileptonic (top, left), $D^{\pm}$(top, right), $D^{*\pm}(2010)$(bottom). The description of the color code and the particular details of the channels drawn in the diagram are given in Section 3.4.2.

Semileptonic channel

A Z +HF sample enriched in semileptonic decays of D and B mesons, $D/B \to \mu + X$, is obtained by selecting events that contain a jet with a displaced muons produced in a secondary vertex. A sketch of this channel is shown in Figure 3.7 (top, left). Muons originated in semileptonic decays of hadrons are usually close to the jet axis and non-isolated from hadronic activity. Event candidates in this channel are selected as follows:

- A reconstructed muon, hereafter called *muon-inside-a-jet*, has to be one of the tracks of the selected jet. Only muons with $p_T^{\mu} < 25$ GeV, $p_T^{\mu}/p_T^{jet} < 0.6$ and reconstructed in the pseudo-rapidity region $|\eta| < 2.5$ are considered as candidates.
- The *muon-inside-a-jet* has to be identified with the same quality criteria imposed to the muons from the Z decay except the isolation cut (see Equation 3.1). In this case, the *muon-inside-a-jet* should not be isolated from hadronic activity and the cut is reversed, $I_{comb}/p_T^{\mu} > 0.2$.
- The *muon-inside-a-jet* track is required to be one of the tracks that is used to built a secondary vertex (either SV or IVF). In case of ambiguity, the SV is preferred.

After applying this selection, 4 145 events in the $Z \to e^+e^-$ channel and 5 258 events in the $Z \to \mu^+\mu^-$ channel are used for the measurement in the semileptonic channel. According to simulations, the resulting composition of this channel is 25% in Z + c, 65% in Z + b, 4% in Z + light and 6% of other background such as $t\bar{t}$ and diboson VV. As an example, Figure 3.8 (top, left) shows the *muon-inside-a-jet* transverse momentum distribution.

The discriminant variable used in this channel is the invariant mass distribution of all charged particles that constitute the reconstructed secondary vertex, M_{vertex}. The latter is corrected, M_{vertex}^{corr}, to take into account neutral (or charged) particles that participated in the secondary vertex but were not reconstructed in the CMS tracker [120]:

$$M_{vertex}^{corr} = \sqrt{M_{vertex}^2 + p_{vertex}^2 \sin^2\theta} + p_{vertex} \sin\theta \tag{3.3}$$

In the formula, p_{vertex}, corresponds to the modulus of the momentum-sum of all reconstructed particles that form the secondary vertex. The angle, θ, corresponds to the angular separation between p_{vertex} and the direction of the line between the primary and the secondary vertex. The M_{vertex}^{corr} exploits the larger mass of B hadrons compared to D hadrons to separate the Z + c to the Z + b signal efficiently (see Figure 3.19).

Finally the individual identification of particles in the secondary vertex, such as kaon vs pions separation is not possible (in general) in CMS. Due to this, all tracks associated to the vertex are given the mass hypothesis of the pion, except for the muon itself.

$D^{\pm}$ channel

Event candidates in the $D^{\pm}$ channel are selected requiring a reconstructed secondary vertex (either SV or IVF) made of exactly three charged tracks. Additionally, the invariant mass of the reconstructed secondary vertex has to be consistent with the $D^{\pm}$ mass: 1869.5 ± 0.4 MeV [74]. A sketch of this channel is shown in Figure 3.7 (top, right). Event candidates in this channel are selected as follows:

- The secondary vertex charge, defined as the sum of the individual charges of the three tracks that form the vertex, has to be ±1 and is associated to the charge of the $D^{\pm}$ candidate.
- The kaon track hypothesis is identified as the track with opposite charge to the vertex charge. This assumption is correct more than 99% of the cases since $D^{\pm} \to K^{\pm}\pi^{\pm}\pi^{\mp}$ decays are Cabibbo suppressed.
- Events are selected in the signal region defined in a window of 50 MeV around the difference between the reconstructed vertex mass, $m^{rec}(D^{\pm})$, and the experimentally measured $D^{\pm}$ mass:

  $$\Delta m(D^{\pm}) \equiv |m^{rec}(D^{\pm}) - 1869.5 \text{ MeV}| < 50 \text{ MeV}$$

  Non-resonant backgrounds in the signal region are subtracted using an estimation at the neighboring 50 MeV side-band regions away from the resonance: 50 MeV < $\Delta m(D^{\pm})$ < 100 MeV.

As an example of this channel, the distribution of the invariant mass of the secondary vertex $m^{rec}(D^{\pm})$ is shown in Figure 3.8 (top, right) prior side-band subtraction. After non-resonant backgrounds are subtracted in the signal region, 375 events are selected in the Z → e^+e^- channel and 490 in the Z → $\mu^+\mu^-$ channel. According to simulations, the composition of the resulting sample after background subtraction is 60% in Z + c, 36% in Z + b, less than 1% in Z + light and 3% combining $t\bar{t}$ and diboson VV.

The discriminant variable used in this channel is the Jet Probability algorithm (see Section 3.3.3 for more datails), and is based on the likelihood that all the tracks inside the jet are originated at the primary vertex. In this channel the JP discriminant exploits the closer distance of the secondary $D^{\pm}$ decay tracks to the primary vertex in Z + c than in Z + b events. The explanation for this is that typically $D^{\pm}$ hadron decays in b-jets are produced in a tertiary vertex such as $B \to D^{\pm}(\to K^{\mp}\pi^{\pm}\pi^{\pm})X$, leading to a larger number of tracks with larger impact parameter and incompatibility with the primary vertex. As a consequence, these differences lead to larger values of the JP discriminant in Z + b events than in Z + c events as shown in Figure 3.20.

$D^{*\pm}(2010)$ channel

The first step in the identification of a $D^{*\pm}(2010)$ candidate is done selecting a pair of oppositely charged tracks among the reconstructed tracks inside the jet. A valid two track system is only considered if it forms a common vertex displaced with respect to the primary vertex, and has an invariant mass consistent with a D^0 meson. In a second step, this D^0 candidate is combined with a third track originated at the primary vertex called soft-pion, π_s to form the $D^{*\pm}(2010)$ candidate.

The mass difference between the $D^{*\pm}(2010)$ and D^0 hadrons is experimentally known to be about 145.42 MeV. As a consequence of this small mass splitting, the angular separation between the line of flight of the D^0 candidate and the slow pion is required to be small, $\Delta R(D^0, \pi_s) < 0.1$. A sketch of this decay channel is shown in Figure 3.7 (bottom). The details about how the D^0 candidates are formed as well as its combination with the soft pion to make a $D^{*\pm}(2010)$ candidate, are provided in the following items:

- The charge of the $D^{*\pm}(2010)$ candidate is determined by the charge of the soft-pion track, $\pi_s^{\pm}$.
- The track of the D^0 candidate vertex with opposite charge to the soft-pion is assumed to be a kaon. The remaining track is assumed to be a pion.
- The minimum transverse momentum of the three $D^{*\pm}(2010)$ track candidates is $p_T^{\pi_s} > 0.5$ GeV, $p_T^K > 1.75$ GeV and $p_T^{\pi} > 0.75$ GeV.
- The secondary vertex compatible with the D^0 mass is built using a Kalman Vertex Fitter [85], in which the vertex probability is required to be > 0.05.
- In order to ensure a clean separation between the primary vertex and the secondary vertex, the decay length significance in the transverse plane, $L_{xy}/\sigma(L_{xy})$, has to be larger than three.
- In order to guarantee that the D^0 was produced at the primary vertex, the momentum vector of the D^0 candidate has to be collinear with the line starting from the primary to the secondary vertex. This angular separation is denoted by the angle α_{xy}, and it has to be small, $\cos\alpha_{xy} > 0.99$.
- The reconstructed masses and mass differences of the $D^{*\pm}(2010)$ and D^0 candidates have to agree with their experimental values from the PDG:

$$|m^{rec}(D^0) - 1864.86 \text{ MeV}| < 100 \text{ MeV}$$
$$|m^{rec}(D^{*\pm}) - m^{rec}(D^0)| \subset [145.42 \pm 5] \text{ MeV}$$
$$\Delta m(D^{*\pm}) = |m^{rec}(D^{*\pm}) - 2010.27 \text{ MeV}| < 40 \text{ MeV}$$

- Non-resonant backgrounds in the signal region are subtracted taking as estimate their contribution in the neighboring side-band region away from the $D^{*\pm}(2010)$ resonance: 40 MeV $< \Delta m(D^{*\pm}) <$ 100 MeV.

Figure 3.8 (bottom) shows the distribution of $m^{rec}(D^{*\pm})$ prior the side-band subtraction. The number of selected events after non-resonant backgrounds are subtracted is 234 (308) in the electron (muon) channel. According to simulations the resulting composition is 65% in Z + c, 30% in Z + b, < 1% in Z + light and < 4% of $t\bar{t}$ and diboson, VV.

In this channel the separation power provided by the Jet Probability algorithm between Z + c and Z + b is higher than in the $D^{\pm}$ channel. This is shown in Figure 3.21. The explanation for this is that in addition to potential tertiary tracks in B jets from the D^0 decay, the slow-pion in Z + c events is a primary track whereas in Z + b processes is also another displaced track originated at the decay the B hadron, $B \to D^{*\pm}(2010)X$.

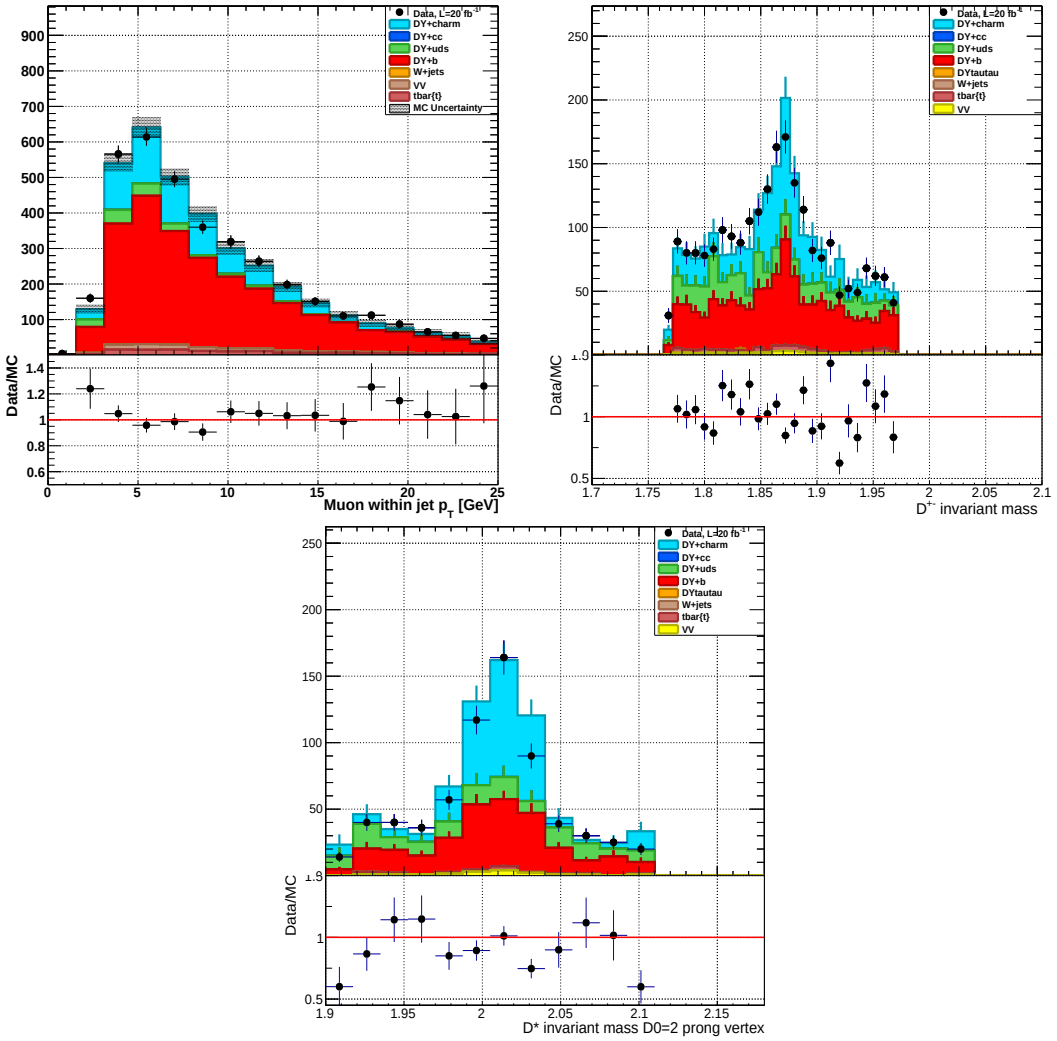

FIGURE 3.8: Muon-inside-a-jet transverse momentum distribution (top, left). Reconstructed secondary vertex mass in $m^{reco}(D^{\pm})$ (top, right) and $m^{reco}(D^{*\pm})$ (bottom).

3.4.3 W+c and $e\mu$ auxiliary samples

In Section 3.4.2, three different channels have been introduced to enrich the Z + jets sample in Z + HF jets, and discriminant variables introduced to distinguish between the Z + c and Z + b components. A summary of this discussion is shown in Table 3.2.

Channel	Discriminant Variable	Z + c	Z + b	Z + light	$t\bar{t}$ + VV
Semileptonic	M_{vertex}^{corr}	25%	65%	4%	6%
$D^{\pm}$	Jet Probability	60%	36%	< 1%	< 3%
$D^{*\pm}(2010)$	Jet Probability	65%	30%	< 1%	< 4%

TABLE 3.2: Summary of the discriminant variables used in each of the three different channels and the composition of the resulting samples according the simulation.

Simulations are not necessarily perfect and need to be calibrated prior the signal extraction. The next step in the analysis is the calibration of the shape of the discriminant variables, as well as the calibration of the secondary vertexing efficiency in simulation for c- and b- jets separately. In this section two auxiliary selections or control samples denoted as W + c and $e\mu$ are introduced to obtain such information. Figure 3.9 shows a summary of the calibration strategy used in the analysis. The different elements of the diagram are going to be presented throughout this section.

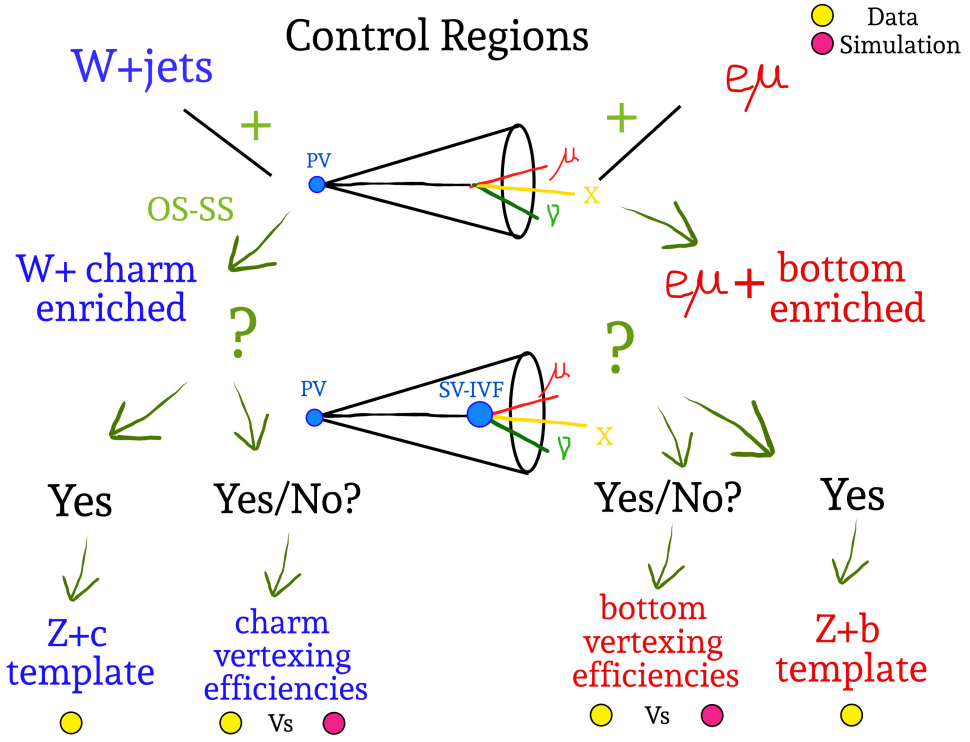

FIGURE 3.9: Sketch of the calibration strategy used to validate the shape and normalization of the discriminant variables in the simulation. All the parts of the diagram will be explained in Section 3.4.3.

Auxiliary sample: W+charm

An auxiliary pure sample in c-jets can be obtained from the associated production of a W-boson and a jet originating from a c-quark. This process will be denoted in the text as W + c and it is used to calibrate the shape of the Z + c templates of the discriminant variables, characterize distributions of c-jets and measure the charm secondary vertexing efficiency.

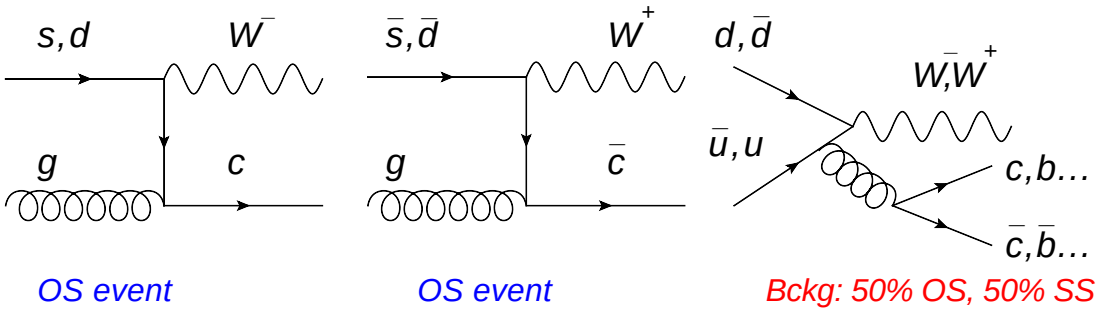

FIGURE 3.10: Left and middle: leading order production of W+c signal with opposite sign charges (OS). Right: production of W+charm final state through the gluon splitting process. In gluon splitting there is an additional charm quark with the same sign as the W boson (SS)

At the LHC, W + c production (Figure 3.10, left and center) is dominated by $\bar{q}g \to W^+ + \bar{c}$ and $qg \to W^- + c$ contributions at the hard scattering level. A key property of a general $qg \to W + c$ process is the presence in the final state of a charm quark and a W boson with opposite sign charges (OS). Gluon splitting processes like $d\bar{u} \to W^- + g \to W^- + q\bar{q}$ (Figure 3.10, right) also give rise to final states with a W and a charm quark, but in this case, there is an additional charm quark with the same sign (SS) as the W boson. Most of the background processes are evenly OS and SS, whereas $qg \to W + c$ is always OS. A pure distribution of c-jets is obtained by subtracting OS-SS events [97]. This subtraction is done by multiplying the event weight by -1 when filling the histograms for SS candidates.

The W + c sample is constructed starting from a W + jets sample[5] separately for electrons and muons as follows:

- Events are recorded by single-lepton triggers that require the presence of an isolated electron or muon.
- The minimum lepton p_T is 30 (25) GeV for electrons (muons). Additionally they have to be in the region $|\eta| < 2.1$ and identified as good leptons (see Section 3.3).
- The transverse mass, defined in Equation 2.13, has to be larger than 55 (50) GeV.
- At least one jet with $p_T^{jet} > 25$ GeV and reconstructed in the region $|\eta| < 2.5$.

[5]Note that already a similar W+jets selection was described at 13 TeV in Section 2.4.1.

The identification of jets originating from c-quarks proceeds exactly in the same three channels described in Section 3.4.2. An advantage of using exclusive channels is that in all cases the charge of the c-quark is unequivocally known. In the semileptonic channel, the charge of the *muon-inside-a-jet*, determines the charge of the c-quark. In the other channels, the charge of the $D^{\pm}$ and $D^{*\pm}(2010)$, determines the charge of the c-quark. OS (SS) events are identified when the isolated lepton from the $W \to l\nu$, has opposite (same) charge to the charge of the *muon-inside-a-jet* (or $D^{\pm}$, $D^{*\pm}(2010)$).

After the OS-SS subtraction the purity in W + c of the resulting selection is more than 90% in the semileptonic channel, and more than 98% in the $D^{\pm}$ and $D^{*\pm}(2010)$ channels. Figure 3.11 shows an example of the high purity in charm of the W+c selection in the three channels: semileptonic (left), $D^{\pm}$ (center), $D^{*\pm}(2010)$ (right). The W + c contribution is shown in light blue in all the figures.

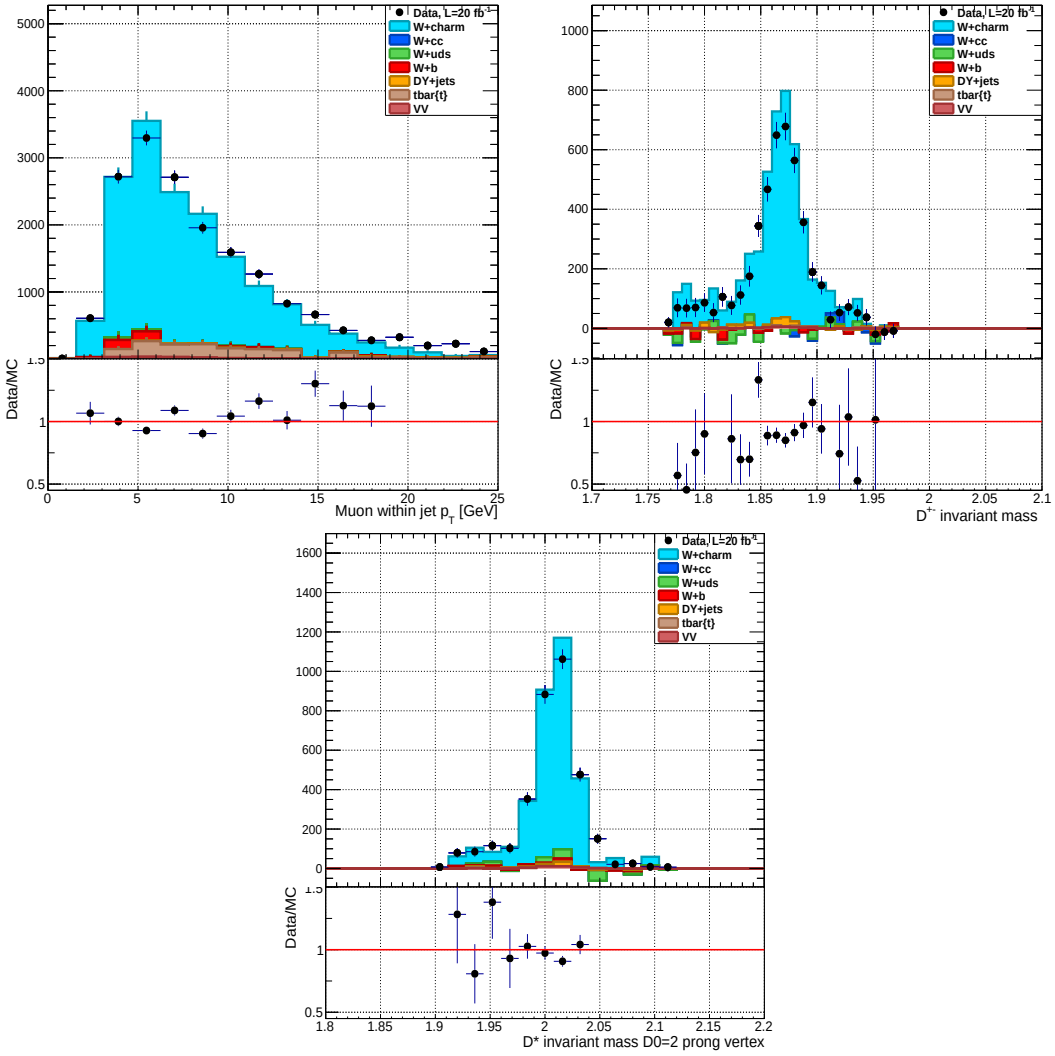

FIGURE 3.11: Data to simulation comparison of the three channels after the W + c selection in the W $\to e\nu$ channel. *Muon-inside-a-jet* transverse momentum distribution (top, left). Reconstructed secondary vertex mass, $m^{reco}(D^{\pm})$ (top, right) and $m^{reco}(D^{*\pm})$ (bottom). All distributions are shown after OS-SS subtraction.

Z+charm template

Due to the high purity in c-jets after the OS-SS subtraction, these W + c distributions are used to validate the Z + c modelling of the discriminant variables used in the analysis. In practice this means that differences between data and simulation seen on the W + c sample, are used to calibrate the Z + c simulation. In other words, if the shape of the discriminant variable is different between data and simulation in the auxiliary and charm enriched W + c sample, then the Z + c simulation is also assumed not to describe accurately the Z + c data. This way, differences between the shape in data and simulation seen in the W + c sample are used to calibrate the shape of the Z + c simulation. This procedure assumes that the shape of any interesting variable obtained with the W + c sample corresponds to the same shape on the Z + c sample. This calibration strategy has been validated in simulation by comparing the discriminant shape between Z + c and W + c.

Main variables of charm jets such as: p_{T}, η, number of charged tracks inside the jet or number of reconstructed IVF vertices inside the jet, are found to be consistent between Z+c and W+charm simulations. As an example, Figure 3.12 shows the distribution of the jet transverse momentum p_{T}^{jet} (left), and number of reconstructed IVF vertices inside the jet (right) in simulated W + c and Z + c compared to W + c data.

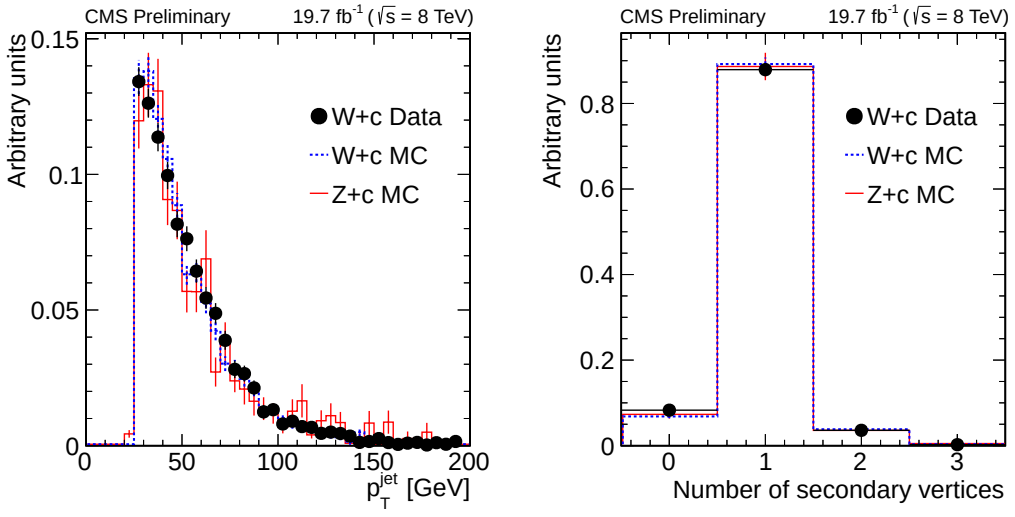

FIGURE 3.12: Transverse momentum distribution of the c-tagged jet (left) and number of reconstructed secondary vertices (right) in simulated W + c and Z + c sample, and in W + c data. W + c distributions are presented after OS-SS subtraction. Remaining backgrounds in W + c data are subtracted according to simulations in both figures.

Figure 3.13 shows the corrected secondary mass, M_{vertex}^{corr} (Equation 3.3), and the Jet Probability discriminant (Equation 3.2), normalized to unity in W + c and Z + c simulation, compared to W + c data for the channels in which they are relevant.

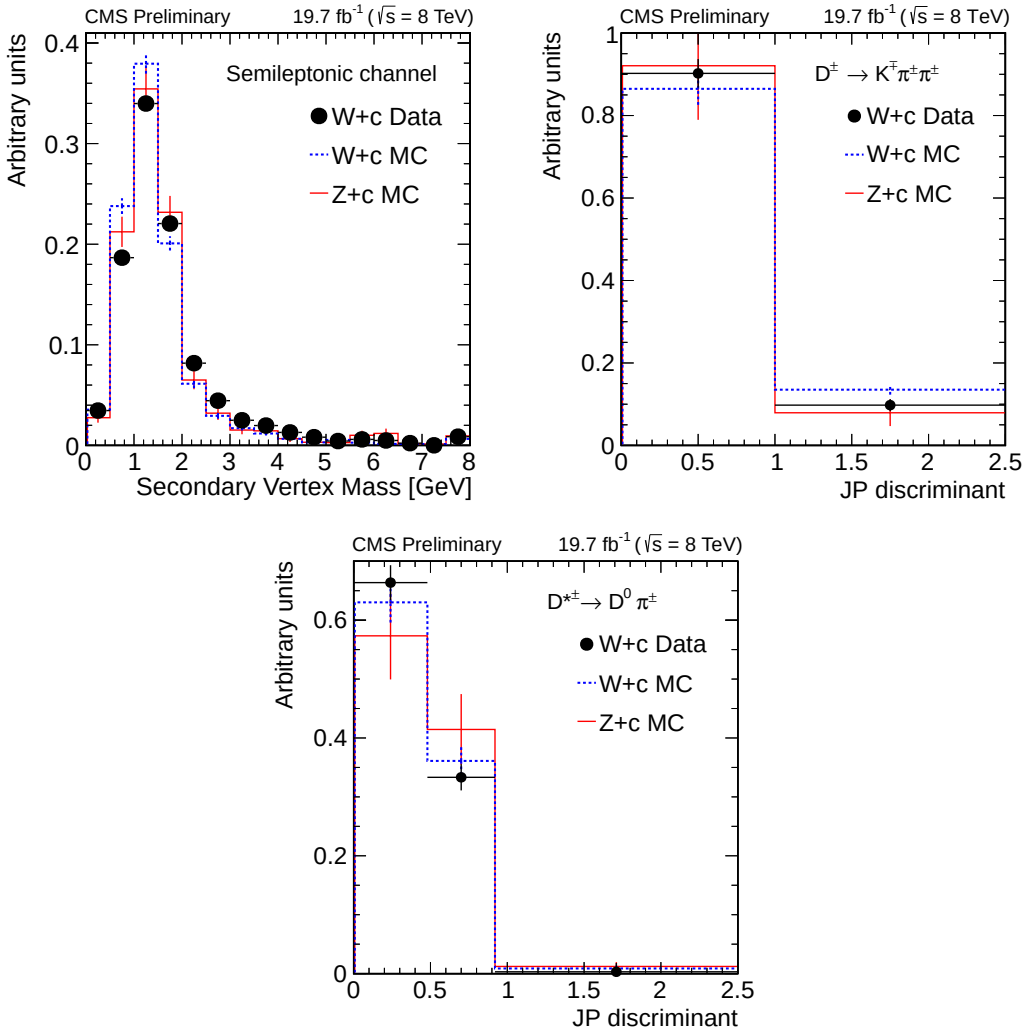

FIGURE 3.13: Distribution of the secondary vertex mass (up, left) and Jet Probability discriminant for $D^{\pm}$ (up, right) and $D^{*\pm}(2010)$ (bottom), in simulated Z + c and W + c compared with W + c data. W + c distributions are presented after OS-SS subtraction and all three figures are normalized to unity. Remaining backgrounds in W + c data are subtracted according to simulations in the figures.

The largest difference between the simulated and measured W + c is seen in the corrected secondary vertex mass distribution (Figure 3.13 top, left). The source of this discrepancy comes from differences in the relative fraction of secondary vertices made of three-track and two-track respectively in data and simulation. This distribution is consistent between Z + c and W + c simulation, however is slightly different when compared with the measured W + c. As a consequence, it is expected that the W + c data describes more accurately the measured secondary vertex mass in Z + c data, given the fact that the simulation do not reproduce the relative W + c vertex track multiplicity.

For the analysis, the shape of M_{vertex}^{corr} in simulation has been re-weighted to match the distribution in W + c data. In the $D^{\pm}$ and $D^{*}(2010)$ channels, the simulations agree with the measured W + c within the statistical uncertainties. However, for consistency

with the semileptonic channel their Jet Probability shapes are also corrected to match the shape of the observed W + c data.

Secondary vertex efficiencies in charm jets

Even if the shapes are well modelled and calibrated in simulations, it is not guaranteed that simulations exactly reproduce the probability of reconstructing a secondary vertex in c-jets in the CMS data. In case differences in the secondary vertex performance are found, they have to be corrected prior the signal extraction.

This calibration is done in the W + c semileptonic channel because it is possible to release the secondary vertex requirement while keeping the OS-SS subtraction. The reason being that in the $D^{\pm}$, $D^{*\pm}(2010)$ channels the charge of the c-quark is determined by the charge of the vertex whereas in the semileptonic channel only the *muon-inside-a-jet* is needed. This particularity of the semileptonic channel in the W + c selection allows to calibrate the charm secondary vertexing efficiency in charm jets. In order to simplify the notation in the calibration, two categories are defined:

- A W+c event is defined to be "pre-tagged", if there is a reconstructed *muon-inside-a-jet* (independently of the presence of a secondary vertex inside the jet).
- A W+c event is defined to be "tagged", if there is a reconstructed secondary vertex and the *muon-inside-a-jet* is one of the tracks constituting the vertex. In case the secondary vertex is of the type SV(IFV), then the event is said to be "SV(IFV)-tagged".

The c-jet secondary vertexing efficiency, ϵ_c, is defined as the fraction of "tagged" W+c events over all the "pre-tagged", both after OS-SS subtraction.

$$\epsilon_c = \frac{N(W+c)^{\text{OS-SS}}_{\text{tagged}}}{N(W+c)^{\text{OS-SS}}_{\text{pre-tagged}}} \tag{3.4}$$

The charm secondary vertex calibration factor is defined as the ratio between the Equation 3.4 evaluated in data and simulation separately, $\epsilon_c^{Data}/\epsilon_c^{MC}$. The efficiencies in data and simulation and calibration factors are computed differentially as a function of the p_{T} of the jet and shown in Figure 3.14. The inclusive calibration factor (yellow band in the same Figure) in an unique p_{T}^{jet} bin above 25 GeV is:

$$0.924 \pm 0.033\ (\text{stat}) \pm 0.020\ (\text{syst})\ \text{for SV vertices in c-jets}$$
$$0.931 \pm 0.027\ (\text{stat}) \pm 0.019\ (\text{syst})\ \text{for IVF vertices in c-jets}$$

The calibration factors in Figure 3.14 are applied to Z + c simulations in the analysis to improve the modelling of the c-jet vertexing efficiency prior the signal extraction.

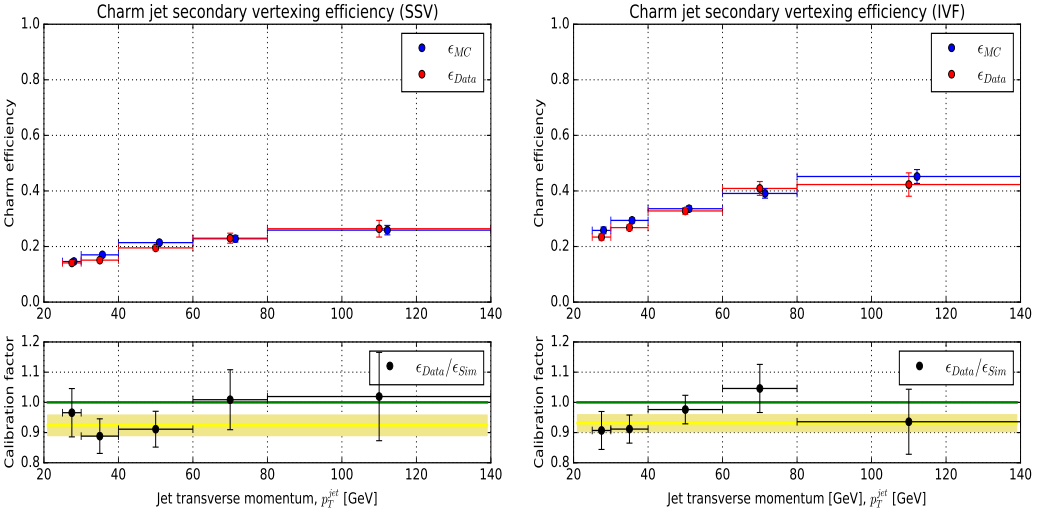

FIGURE 3.14: Secondary vertex efficiencies in c-jets for SV (left) vertices and IVF vertices (right) as a function of the p_{T} of the jet. The upper pad shows the charm secondary vertexing efficiency whereas the lower pad shows the calibration factors. The yellow band shows the inclusive results for only jets with a minimum p_{T} of 25 GeV.

Even if it is not used in the Z + c cross section analysis, the same procedure in Equation 3.4 was used to calibrate the standard b-tagging algorithms described in Section 3.3.3 for the first time using charm jets in CMS. In this case, the "tagged" category is adapted for each studied algorithm and results are shown in Figure 3.15.

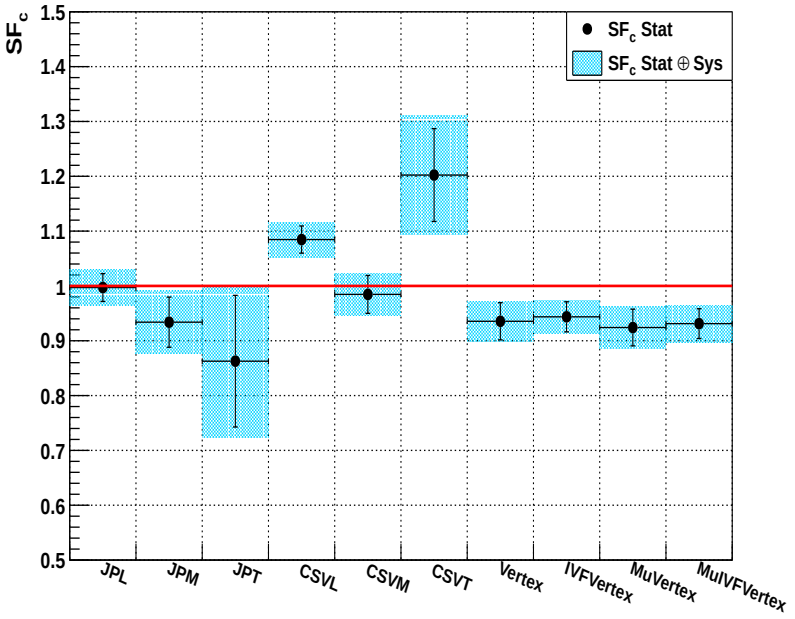

FIGURE 3.15: Calibration factors for the main b-tagging algorithms used in CMS: Combined Secondary Vertex, Jet Probability, Simple Secondary Vertex, Inclusive Vertex Finder for *muon-inside-a-jet* charm jets. More details are provided in AN-14-225.

Currently, the W + c auxiliary sample with the *muon-in-jet selection* is one of the standard tools to calibrate charm and bottom taggers with charm jets in CMS.

Auxiliary sample: $e\mu$ + HF **jets**

Analogously to the W + c selection for charm jets, an auxiliary sample enriched in b-jets is obtained selecting events with two leptons with opposite flavour, $e\mu$, originated in decays of $t\bar{t}$ pairs. This selection is a natural source of b-jets since top quarks decay into $t \to Wb$ more than 99% of the times. The requirement of opposite flavour leptons is applied to remove symmetric backgrounds such as $Z \to e^+e^-$ or $Z \to \mu^+\mu^-$. This way, the $e\mu$ +jets selection is defined as follows:

- Events are recorded using a trigger based on electron-muon pairs.
- Reconstructed muons (electrons) are required to have a p_T larger than 25(30) GeV, and both leptons are reconstructed in the pseudo-rapidity range $|\eta| < 2.1$.
- As in the Z + jets or W + c selections described in this chapter, both leptons have to be well identified and isolated from hadronic activity (see Section 3.4.1).
- At least one jet with $p_T^{jet} > 25$ GeV and reconstructed in the region $|\eta| < 2.5$.

The identification of jets originating from b-quarks proceeds in the same three channels described in this section: semileptonic, $D^{\pm}$, $D^{*\pm}(2010)$. In all three cases the purity of the sample in b-jets is more than $> 97\%$ and are used to calibrate Z + b simulations.

Z+bottom template

Owing to the high purity in b-jets of the $e\mu$+HF sample, the shape of the discriminant variables obtained from Z + b simulations is validated by comparing the shape obtained with both selections.

In the semileptonic channel, the purity in Z + b is expected to be about 65% according to the MC simulation expectations (see Table 3.2). Figure 3.16 shows the shape comparison of the secondary vertex mass in the semileptonic channel, M_{vertex}^{corr}, between a bottom enriched $e\mu$+Semileptonic selection in data and a Z + b simulation. The simulation describes well the shape in data, apart from the region between 3-4 GeV and above 7.5 GeV. These differences are of the order of 13% in the 3-4 GeV region and around 50% above 7.5 GeV, and are used to correct the shape of the Z + b simulation of the M_{vertex}^{corr}. The uncertainty asociated to this correction is propagated as a additional source of systematic uncertainty.

The reduced number of selected events in the $e\mu$+$D^{\pm}$ and $e\mu$+$D^{*\pm}$, combined with the low purity in b-jets in Z + b in those channels does not allow for a meaningful validation of the Jet Probability distribution with data. In both cases the distribution of the Jet Probability discriminant is taken directly from simulation, and the Z + c/Z + b ratio is not measured.

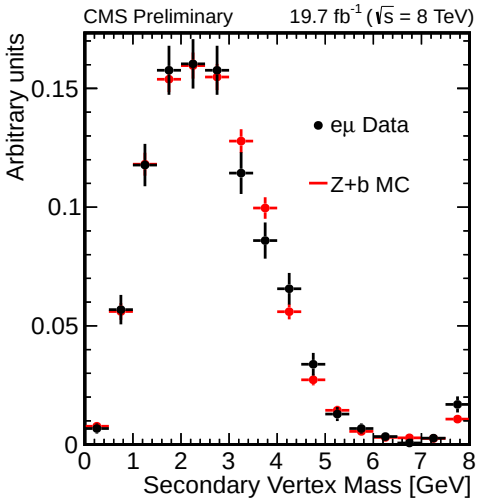

FIGURE 3.16: Secondary vertex mass distribution in $e\mu$ plus *muon-inside-a-jet* events (black) compared with Z + b simulations (red).

Secondary vertexing efficiencies in bottom jets

As in the case of c-jets, it is not expected either from the simulation to fully reproduce the probability of having a secondary vertex in bottom jets. Their calibration can be done also in the semileptonic channel since it is possible to release the vertex requirement while having $> 97\%$ purity in b-jets for calibration. Events are classified into two types in the calibration:

- A $e\mu$ event is defined to be "pre-tagged" if there is an additional *muon-inside-a-jet* (independently of the presence of the secondary vertex).
- A $e\mu$ event is defined to be "tagged" if there is a reconstructed secondary vertex and the *muon-inside-a-jet* is one of the tracks constituting the vertex. In case the secondary vertex is of the type IVF(SV), then is said to be "IVF(SV)-tagged".

$$\epsilon_b = \frac{N(e\mu)_{\text{tagged}}}{N(e\mu)_{\text{pre-tagged}}} \tag{3.5}$$

The b-jet secondary vertex calibration factors are defined as the ratio between Equation 3.5 evaluated in data and simulation, $\epsilon_b^{Data}/\epsilon_b^{MC}$. The efficiencies and calibration factors are computed differentially as a function of the p_{T} of the jet and shown in Figure 3.17. The inclusive calibration factors in an unique p_{T}^{jet} bin above 25 GeV are:

$$0.963 \pm 0.025\ (\text{stat}) \pm 0.007\ (\text{syst})\ \text{for SV vertices in b-jets}$$
$$0.964 \pm 0.024\ (\text{stat}) \pm 0.006\ (\text{syst})\ \text{for IVF vertices in b-jets}$$

In the analysis, Z + b simulations are corrected by the calibration factors in Figure 3.17 to improve the modelling of the b-jet vertexing efficiency prior the signal extraction.

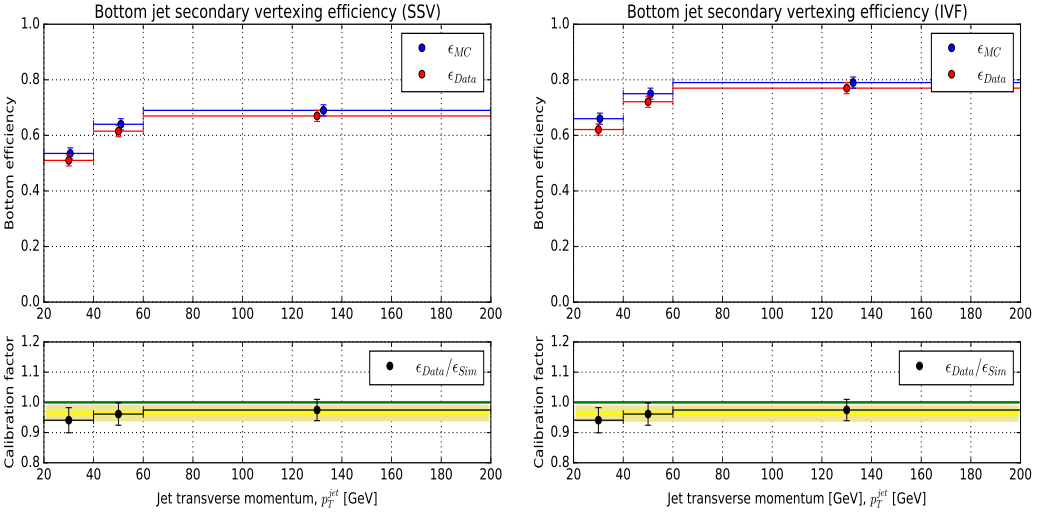

FIGURE 3.17: Secondary vertex efficiencies in b-jets for SV (left) vertices and IVF vertices (right) as a function of the p_{T} of the jet. The upper pad shows the bottom secondary vertexing efficiency, whereas the lower pad shows the calibration factors. The yellow band shows the inclusive results for only jets with a minimum p_{T} of 25 GeV.

3.4.4 Other background: $t\bar{t}$, diboson and Z+light

$t\bar{t}$ pair production

The resulting contribution from $t\bar{t}$ events is about 4% in the semileptonic and less than 2% in the $D^{\pm}$ and $D^{*\pm}(2010)$ channels. In all three cases the expected contribution is evaluated with a POWHEG simulation normalized to a NNLL calculation from [112]. Additionally, the expected contribution in the semileptonic channel is also evaluated with a data-driven technique. This method evaluates $t\bar{t}$ contribution using four orthogonal selections (A,B,C,D) defined using different final states:

- Region A: $\mu\mu(ee)$ + Semileptonic and $E_{\mathrm{T}}^{miss} < 40$ GeV.
- Region B: $e\mu$ + Semileptonic and $E_{\mathrm{T}}^{miss} < 40$ GeV.
- Region C: $\mu\mu(ee)$ + Semileptonic and $E_{\mathrm{T}}^{miss} > 80$ GeV.
- Region D: $e\mu$ + Semileptonic and $E_{\mathrm{T}}^{miss} > 80$ GeV.

In all the regions there is an additional cut on the invariant mass of the leptons 70 GeV $< m_{ll} <$ 111 GeV. The goal of this method is to estimate the shape and normalization of the secondary vertex mass for $t\bar{t}$ events in Region A. For this, the secondary vertex mass shape is obtained in Region B (more than 99% pure in $t\bar{t}$), and the normalization is obtained as the number of observed events in data in Region B times the ratio of the number events in regions C/D. This method assumes that the fraction of events in C/D is the same as in region A/B. The latter has been checked with simulations.

Figure 3.18 shows a summary chart of the 4 regions and how the secondary vertex mass is extracted for Region A. The agreement between the $t\bar{t}$ simulation and the data-driven estimation (baseline in the semileptonic channel) is within 10%. This difference is considered as the uncertainty in the expectation of the $t\bar{t}$ background.

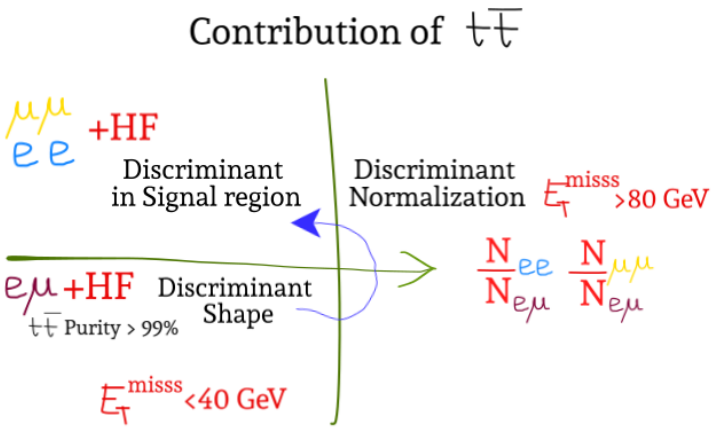

FIGURE 3.18: Sketch of the data-driven estimation of the $t\bar{t}$ contribution in the semileptonic channel. The details of the method are explained in Section 3.4.4.

Diboson: WW, WZ, ZZ

The resulting contribution from diboson processes such as WW, WZ and ZZ is about 2% in the semileptonic channel and less than 1% in the $D^{\pm}$ and $D^{*\pm}(2010)$ channels. The dominant diboson processes that mimics the Z+HF final state are WZ and ZZ production in which the Z boson decays to leptons and the other boson decays hadronically. Their expectation in all three channels is taken directly from a PYTHIA6 simulations normalized to NLO cross sections from MCFM [37] using MSTW2008NLO PDF.

Z+light

One of the motivations to choose an analysis strategy based on three exclusive final states is to minimize the Z + light contribution. The resulting fraction of light jets is about 4% in the semileptonic channel and negligible in the $D^{\pm}$ and $D^{*\pm}(2010)$ channels. Discrepancies in the light-tagging efficiencies between data and simulations are corrected in the simulation using the typical calibration factors derived for the CSV b-tagging algorithm in its medium working point [121]. The uncertainty in the Z + light contribution is taken from the largest variations (10%-30% depending on p_{T}^{jet}) of the light-tagging calibration factors among the different working points of the algorithm.

Finally, the contributions from all backgrounds described in this Section 3.4.4 except for Z + c and Z + b, are subtracted in data according to the predictions prior to the Z + c signal extraction.

3.4.5 Signal extraction

Once contributions from different backgrounds have been evaluated and discriminant distributions calibrated, the next step in the analysis is to extract the total number of observed Z + c and Z + b events in each of the three channels.

Their total contribution is extracted from a χ^2 minimization fit of the Z + c and Z + b templates to the measured data. In the case of the semileptonic channel, the templates correspond to the distribution of the secondary vertex mass, whereas in the cases of the $D^{\pm}$ and $D^{*\pm}(2010)$ channels the templates come from the Jet Probability algorithm. The expression for the χ^2 to minimize is the following:

$$\chi^2 = \sum_i \frac{\left(n_i^{data} - \mu_{\mathrm{Z+c}} N_i^{\mathrm{Z+c}} - \mu_{\mathrm{Z+b}} N_i^{\mathrm{Z+b}}\right)}{(\sigma_i^d)^2 + (\sigma_i^{\mathrm{MC}})^2} \tag{3.6}$$

Where n_i^{data} is the number of entries in bin i of the observed distribution after the subtraction of the Z + light, $t\bar{t}$, and VV contribution in the discriminant variable. $N_i^{\mathrm{Z+c}}$, $N_i^{\mathrm{Z+b}}$ are the number of entries in bin i for the Z + c and Z + b templates according to the predictions. σ_i^d and σ_i^{MC} correspond to the statistical uncertainty of the i bin in data and expectations respectively. Finally the free parameters in the fit are the signal strengths $\mu_{\mathrm{Z+c}}$ and $\mu_{\mathrm{Z+b}}$, that relate the respective contributions predicted by the background expectation with the measured data. Fitted signal strengths close to unity imply a good modelling of the Z + c and Z + b predictions in the reference simulations.

The initial normalization level of all components is set to the predictions obtained with the simulated samples, after all the calibrations described in Section 3.4.3 are included in the Z + c and Z + b templates. The parameters $\mu_{\mathrm{Z+light}}$, $\mu_{t\bar{t}}$ and μ_{VV} associated to the strength of the backgrounds described in Section 3.4.4, are assumed to be equal to one and subtracted prior the fit.

The fitted secondary vertex mass distribution in the semileptonic decay channel is shown in Figure 3.19 for the dilelectron (left) and dimuon (right) channel. Table 3.3 contains the extracted number of Z + c and Z + b in data, denoted as $N_{\mathrm{fitted}}^{\mathrm{Z+c}}$ and $N_{\mathrm{fitted}}^{\mathrm{Z+b}}$ as well as the fitted signal strengths with the corresponding statistical uncertainties.

Similarly Jet Probability fitted distributions for $D^{\pm}$ and $D^{*\pm}(2010)$ channels are shown in Figure 3.20 and 3.21 respectively. Finally Tables 3.4 and 3.5 contain the measured Z + c and fitted signal strengths in the $D^{\pm}$ and $D^{*\pm}(2010)$ channels respectively.

Keeping in mind that the results in the tables only include statistical uncertainties (see Section 3.4.6 for a discussion of the systematic uncertainties), at this point it is interesting to highlight some interesting observations about the first results.

In all three exclusive channels the fitted signal strengths are found to be compatible when comparing the electron and muon channels. The latter is not surprising since it is expected that Z decays to charged leptons are universal, however it serves as an internal check of the consistency of the results.

Additionally it is also an interesting result that the extracted signal strengths are compatible (and not far from unity) among the three exclusive channels. This is important because it gives an additional level of robustness to the measurement since each of the channels has completely different secondary vertex topologies.

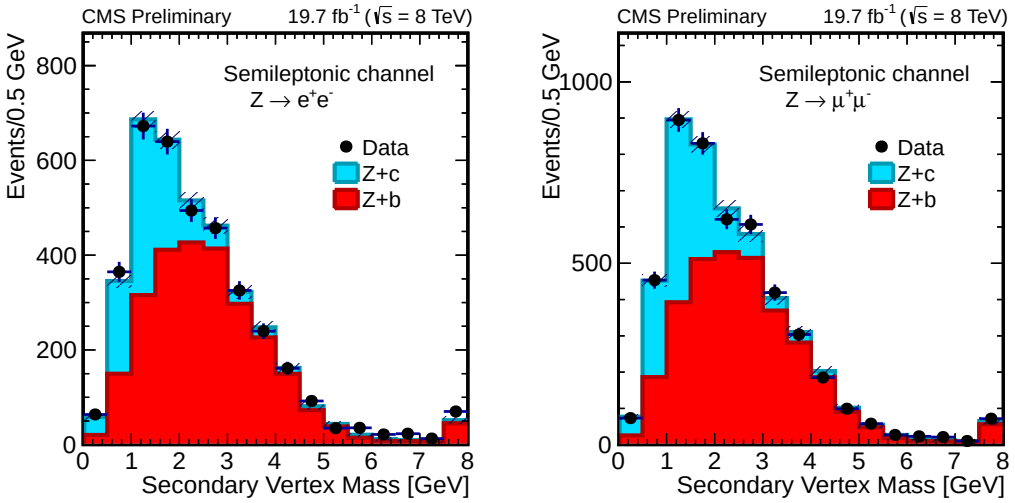

FIGURE 3.19: Corrected vertex mass in the dielectron (left) and dimuon (right) channel. The shape of the Z + c and Z + b contributions is estimated as explained in Section 3.4.3. Their normalization is adjusted to the results of the signal extraction fit in Table 3.3. Sub-dominant backgrounds such as $t\bar{t}$, diboson and Z + light are subtracted in data according to SM expectations as described in Section 3.4.4.

Semileptonic mode					
Channel	N_{sel}	μ_{Z+c}	μ_{Z+b}	N^{Z+c}_{fitted}	N^{Z+b}_{fitted}
$Z \to e^+e^-$	4145	1.05 ± 0.09	1.01 ± 0.04	1066 ± 95	2606 ± 114
$Z \to \mu^+\mu^-$	5258	1.11 ± 0.11	0.92 ± 0.04	1449 ± 143	3240 ± 147

TABLE 3.3: N_{sel} is the total number of selected events in the semileptonic channel prior the background subtraction. μ_{Z+c} and μ_{Z+b} refer to the measured signal strengths to apply to the predicted events by the Z + c and Z + b simulations. The measured Z + c and Z + b events in data are indicated by N^{Z+c}_{fitted} and N^{Z+c}_{fitted} respectively. The errors correspond to the statistical uncertainty from the fit.

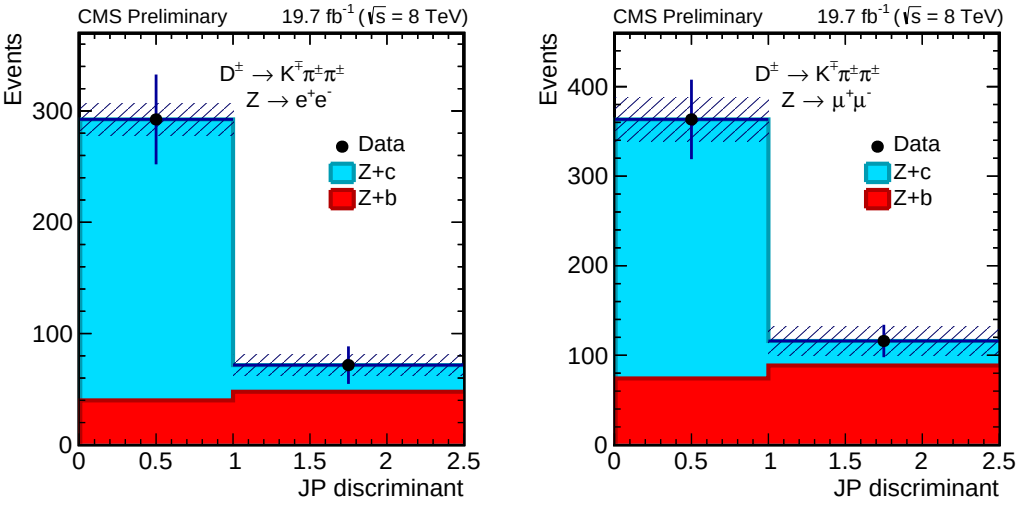

FIGURE 3.20: Distribution of the JP discriminant in the dielectron (left) and dimuon (right) channel for Z + jets events with a $D^{\pm} \to K^{\mp}\pi^{\pm}\pi^{\pm}$ candidate. The shape of the Z + c and Z + b contributions is estimated as explained in Section 3.4.3. Their normalization is adjusted to the results of the signal extraction fit in Table 3.4. Sub-dominant backgrounds such as $t\bar{t}$, diboson and Z + light are subtracted in data according to SM expectations estimated as described in Section 3.4.4.

$D^{\pm} \to K^{\mp}\pi^{\pm}\pi^{\pm}$ mode			
Channel	N_{sel}	μ_{Z+c}	N^{Z+c}_{fitted}
$Z \to e^+e^-$	375	1.35 ± 0.26	276 ± 54
$Z \to \mu^+\mu^-$	490	1.10 ± 0.26	317 ± 74

TABLE 3.4: N_{sel} is the total number of selected events in the $D^{\pm} \to K^{\mp}\pi^{\pm}\pi^{\pm}$ channel prior the background subtraction. μ_{Z+c} refers to the measured signal strength to apply to the Z + c events predicted by the simulation. The measured Z + c in data is indicated by N^{Z+c}_{fitted}. The errors correspond to the statistical uncertainty from the fit.

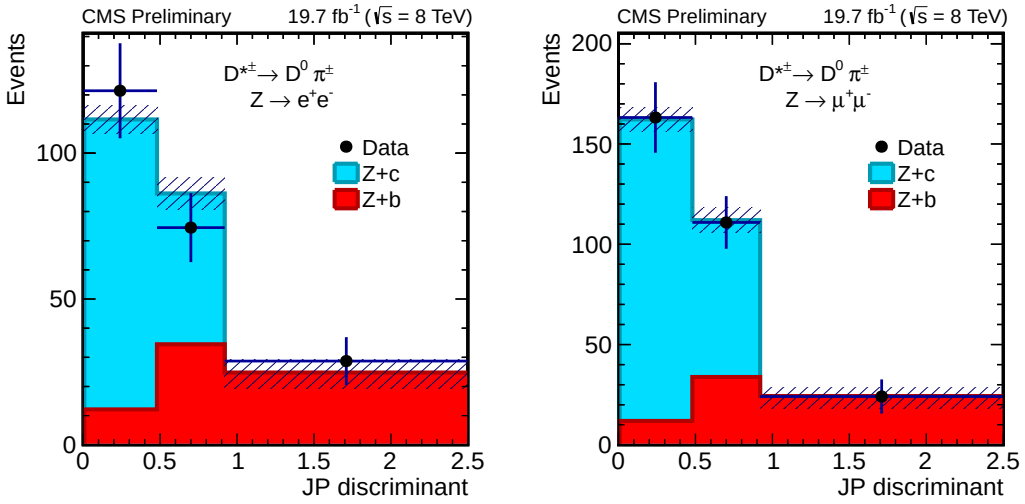

FIGURE 3.21: Distribution of the Jet Probability discriminant in the dielectron (left) and dimuon (right) channel for Z + jets events with a $D^{*\pm}(2010) \to D^0\pi_s^\pm \to K^\mp\pi^\pm\pi_s^\pm$ candidate. The shape of the Z + c and Z + b contributions is estimated as explained in Section 3.4.3. Their normalization is adjusted to the results of the signal extraction fit in Table 3.5. Sub-dominant backgrounds such as $t\bar{t}$, diboson and Z + light are subtracted in data according to the SM expectations estimated as described in Section 3.4.4.

$D^{*+}(2010) \to D^0\pi_s^+$, $D^0 \to K^-\pi^+$ mode			
Channel	N_{sel}	μ_{Z+c}	N^{Z+c}_{fitted}
$Z \to e^+e^-$	234	0.90 ± 0.18	151 ± 31
$Z \to \mu^+\mu^-$	309	1.06 ± 0.13	228 ± 28

TABLE 3.5: N_{sel} is the total number of selected events in the $D^{*+}(2010) \to D^0\pi_s^+$, $D^0 \to K^-\pi^+$ channel prior the background subtraction. μ_{Z+c} refers to the measured signal strength to apply to the Z + c events predicted by the simulation. The measured Z + c in data is indicated by N^{Z+c}_{fitted}. The errors correspond to the statistical uncertainty from the fit.

3.4.6 Systematic uncertainties

There are different sources of systematic uncertainties that could affect the modelling of the discriminant variables in the three channels. Particularly any change in the shape or normalization of the templates prior to the fit will give different fitted signal strengths $\mu_{\mathrm{Z+c}}$, $\mu_{\mathrm{Z+b}}$ and $\mu_{\mathrm{Z+c}}/\mu_{\mathrm{Z+b}}$. For each systematic uncertainty, the discriminant templates are recalculated and fits re-evaluated. Differences with respect to the nominal signal strengths (see Tables 3.3, 3.4, 3.5) are added in quadrature and contribute to the systematical uncertainty associated to the measurement.

In the semileptonic channel, the main contribution to the systematic uncertainty comes from the modelling of the hadronization and decay of charmed hadrons (see Section 1.2.2 for more details about hadronization). The measured inclusive (exclusive) charm semileptonic branching fractions is $\mathcal{B}(c \to \ell) = 0.096 \pm 0.004$ [74], (0.086 ± 0.004) [74, 122] and the average of these two values is $\mathcal{B}(c \to \ell) = 0.091 \pm 0.003$. This value is consistent with the PYTHIA value present in the simulations (9.3%). An additional 5% uncertainty is assigned in order to the cover both central values within one standard deviation. The average of the inclusive bottom semileptonic branching fractions is $\mathcal{B}(b \to \ell) = 0.1069 \pm 0.0022$ [74], that is consistent with the PYTHIA value used in the simulations (10.5%). In the latter case, an additional uncertainty of 2% is considered.

In the $\mathrm{D}^{\pm}$ channel, the decay is fully reconstructed and simulations are corrected to match the experimental values from [122]. In particular, the charm fraction $\mathcal{B}(c \to \mathrm{D}^{\pm})$ in the PYTHIA simulation $(19.44 \pm 0.02\%)$ is lower than the value $(22.7 \pm 0.9 \pm 0.5\%)$ obtained from a combination [122] of published measurements performed at LEP. The branching fraction of the decay $\mathrm{D}^{\pm} \to K^{\mp}\pi^{\pm}\pi^{\pm}$ $(7.96 \pm 0.03\%)$, is also lower than the PDG value $(9.13 \pm 0.19\%)$ [74]. Predicted event rates from the simulation are re-weighted in order to match the experimental values. An additional uncertainty of 5% in the re-weighting procedure is considered.

Similarly, in the $\mathrm{D}^{*\pm}(2010)$ channel the product of the branching fractions $\mathcal{B}(c \to \mathrm{D}^{*+}(2010)) \times \mathcal{B}(\mathrm{D}^{*+}(2010) \to \mathrm{D}^0\pi^+) \times \mathcal{B}(\mathrm{D}^0 \to K^-\pi^+)$ (+c.c.) in the PYTHIA simulation is $(0.741 \pm 0.005)\%$, which is about 15% larger than the measured experimental value, $(0.622 \pm 0.020)\%$ [74, 122]. Expected event rates the from simulation are re-weighted in order to match the experimental values and an uncertainty of 3.2% is considered in this re-weighting procedure. This uncertainty assignment associated to the modelling of the hadronization is similar to the one used in the W + c cross section measurement at 7 TeV from CMS [97].

The uncertainty associated to the c-jet (b-jet) secondary vertexing calibration factors (see Section 3.4.3 for details) is in the range 3.5-4% (2.5%) in Figure 3.14 and 3.17. This uncertainty directly propagates to the normalization of the templates prior to the fit and so to the fitted signal strengths. In the $D^{*\pm}(2010)$ mode, the candidate building procedure is repeated changing independently by one sigma, in terms of the p_T resolution, the different p_T thresholds imposed and the decay-length significance requirement. The uncertainty is taken as the quadratic sum of the respective differences between simulation and data in the number of selected $D^{*\pm}(2010)$ candidates, 2.8%.

The jet energy scale and resolution (see Section 2.3.7) are corrected in both data and simulation in order to flatten the jet response in p_T and η. These calibrations are derived in γ + jets and Z + jets processes. Energy scale and jet energy resolution uncertainties are assessed by varying the corresponding correction factors within their uncertainties according to the results of CMS dedicated studies at 8 TeV [123]. The variation of them modifies the normalization and shape of the templates and its impact in the fitted signal strength is about 5% in the semileptonic channel, 3% in the $D^{\pm}$ channel and 2% in the $D^{*\pm}(2010)$.

The uncertainty from a mismeasurement of the missing transverse energy of the event is estimated by adding to the reconstructed E_T^{miss} a 10% of the energy of the unclustered objects in the event. The latter refers the energy deposits that are not associated to any reconstructed object. The effect of this change is less than 1% in the extracted signal strength.

The uncertainty from the trigger and identification calibration factors is 2% in the $Z \to \mu\mu$ and 4% in the $Z \to ee$ channels. In the case of the calibration of the identification of non-isolated muons, as the *muon-inside-a-jet* in the semileptonic channel, this uncertainty is enlarged to 3% according to dedicated studies in multi-jet events.

The reference signal sample is generated with MADGRAPH +PYTHIA6 using the PDF CTEQ6L1 and reweighted to NNLO PDF set MSTW2008NNLO (see Section 1.2.1 for more details). The effect of using other PDF sets such as NNPDF2.3 [124] or CT10 is evaluated and the largest difference is found to be less than 2% compared to the predictions with the reference sample.

Collinear emission of partons suffer from large theoretical uncertainties in the calculation. The potential mismodelling of gluon splitting process into collinear $c\bar{c}$ ($b\bar{b}$) pairs in the reference Z + jets simulated sample, is evaluated by increasing by a $\pm 50\%$ the weight of events in which at least two c-hadrons (b-hadrons) at generator level close to the reconstructed jet $\Delta R(jet, c(b)) < 0.5$. The fraction of those events is small, and the effect in the extracted signal strength is smaller than 1%.

The Z + c shapes of the discriminant distributions obtained from the W + c event sample are observed to be stable under jet energy scale variations and changes in the lepton p_{T} threshold in the selection. Thanks to this stability in the templates shape, no additional systematic is included in the extracted shapes from W + c. In the case of Z + b, there are some differences between the shape of the secondary vertex mass estimated from the $e\mu$ (Figure 3.16) sample and directly from the Z + b simulation in the region between 3-5 GeV and above 7.5 GeV. Templates are corrected by this observed difference, and the systematic uncertainty is evaluated varying this correction by their uncertainties. Their effect in the extracted signal strength is less than 2.5%.

Sub-dominant backgrounds (see Section 3.4.4) such as $\mathrm{t\bar{t}}$, Z + light, VV are subtracted in data prior to the signal extraction assuming $\mu_{\mathrm{t\bar{t}}} = \mu_{\mathrm{Z+light}} = \mu_{\mathrm{VV}} = 1$. An additional systematic uncertainty is assigned to account for possible mismodelling of the subtracted backgrounds. In the case of $\mathrm{t\bar{t}}$, an alternative subtraction is performed using directly the expectation from simulation instead of the data-driven estimation described in Figure 3.18. The diboson component, μ_{VV}, is varied by 6% according to the uncertainty in their theoretical cross section. The uncertainty associated to the Z + light contribution is evaluated varying $\mu_{\mathrm{Z+light}}$ prior the subtraction. The size of this uncertainty is 10%-30% depending on the p_{T}^{jet} and corresponds to the largest variation of the light-tagging calibration factors in the CSV algorithm among the different working points. Due to the smallness of these backgrounds in all the cases, the impact of their uncertainties is less than 2% in the extracted signal strengths.

The uncertainty in the determination of the integrated luminosity of the data sample is 2.6% [125], and directly translates to the uncertainty of the fitted signal strengths. Potential mismodellings of the number of proton-proton interactions in simulation are calculated using a modified pileup profile. The latter gives a modified re-weighting function, $C(N_{PU})$ (see Section 2.3.2 for details), obtained varying the minimum bias cross section by its uncertainty, 5%. The effect of this uncertainty is smaller than 1%.

Systematic uncertainties arising from a common source affecting different final states are assumed to be fully correlated across channels e.g: estimation of the integrated luminosity, pileup modelling in simulation, the estimation of E_{T}^{miss}, secondary vertex efficiencies, jet energy scale and resolution... However, those related to lepton trigger, identification and reconstruction efficiencies are assumed to be uncorrelated between the electron and muon channels.

In the extraction of the ratio $\mu_{\mathrm{Z+c}}/\mu_{\mathrm{Z+b}}$, common uncertainties in both processes such as luminosity, lepton efficiencies or PDF uncertainties cancel out, however some other uncertainties that are anti-correlated between the c- and b-jet components are amplified.

Figure 3.22 and Table 3.6 serve as summary of this section and show the systematic uncertainties for each of the channels used in the $\sigma(\mathrm{Z}+\mathrm{c})$ and $\sigma(\mathrm{Z}+\mathrm{c})/\sigma(\mathrm{Z}+\mathrm{b})$ measurements. The systematic uncertainties are about 11% in the semileptonic channel, 8.5% in the $D^{\pm}$ channel and 7% in the $\mathrm{D}^{*\pm}(2010)$ channel. The total size of the systematic uncertainty in the ratio is about 10%.

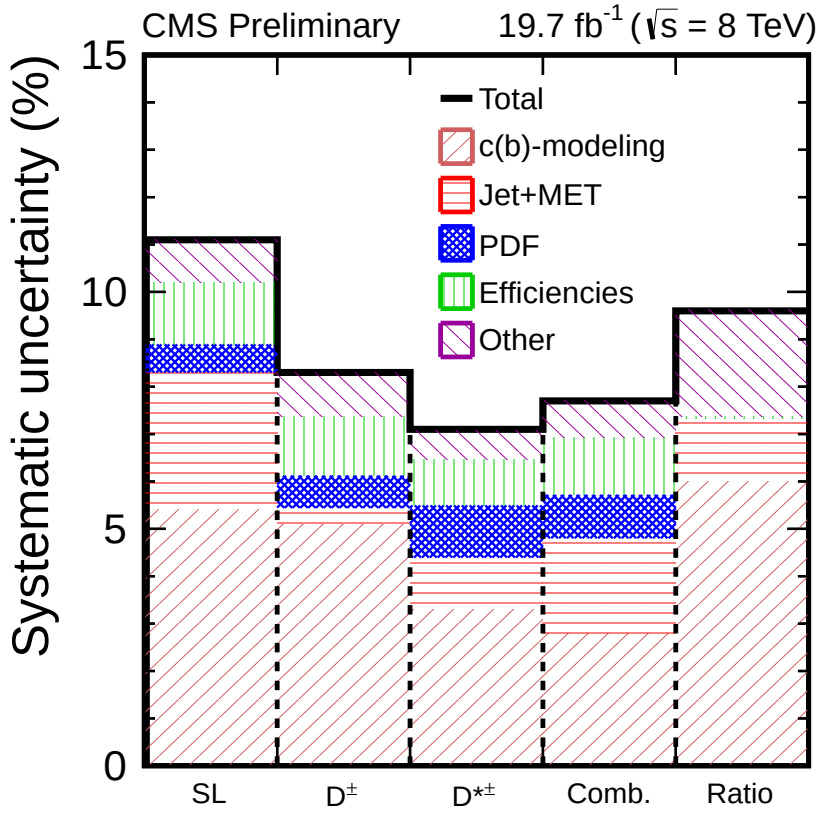

FIGURE 3.22: Contributions to the systematic uncertainty in the measured Z + c cross section and in the Z + c/Z + b cross section ratio. The first three bins in the graphic show the uncertainties in the Z + c cross section in the three decay modes: semileptonic , $\mathrm{D}^{\pm}$, and $\mathrm{D}^{*\pm}(2010)$ calculated from the combination in the dimuon and dielectron Z-boson decay channels. The fourth bin shows the systematic uncertainties in the combined Z + c cross section. The last bin presents the systematic uncertainty in the Z + c/Z + b cross section ratio measured in the semileptonic channel. Contributions are presented in a cumulative manner.

Source of systematic uncertainty	Semileptonic	$D^{\pm}$	$D^{*\pm}(2010)$	Ratio
D and B hadron modelling in simulation	5.5%	5%	3%	5.5%
PDF uncertainties	< 2%	< 2%	< 2%	N/A
Collinear gluon splitting in simulation	< 1%	< 1%	< 1%	< 1%
Calibration of secondary vertexing efficiencies	4%	4%	2.8%	2%
Lepton trigger, reconstruction and identification	< 4%	< 4%	< 4%	N/A
Jet energy scale and resolution	5%	3%	2%	5%
Estimation of E_{T}^{miss}	1%	1%	1%	1%
Calibration of discriminant template shape	2.5%	< 1%	< 1%	6%
Background subtraction	1.5%	1%	1%	1.5%
Pileup modelling in simulation	< 1%	< 2%	< 2%	< 1%
Estimation of integrated luminosity	2.6%	2.6%	2.6%	N/A
Total	11%	8.5%	7%	10%

TABLE 3.6: Breakdown of the individual contributions of each the sources of systematic uncertainties in the measured Z + c cross section and Z + c/Z + b ratio for the three channels. See Section 3.4.6 for a detailed description of each of the sources listed in the first column.

Finally in Section 3.5.2, $\mu_{\mathrm{Z+c}}$ and $\mu_{\mathrm{Z+c}}/\mu_{\mathrm{Z+b}}$ are extracted differentially in the semileptonic channel using a reduced sample binned in p_T^Z and p_T^{jet} separately. In such cases, the uncertainties are larger and of the order of 11%-15%. The main sources of systematic uncertainties in the differential measurement are jet energy scale and resolution as well as charm fragmentation and decay in the reference simulation. Additionally, charm and bottom secondary vertexing calibration factors are evaluated differentially giving rise to an increase in the systematic uncertainty from 3.5%-4% to 7%-8%.

3.5 Results and interpretation

Fiducial cross section σ_{fid}

At the LHC, the cross section (see Equation 1.1) for a given process is a measurement of the rate in which it is produced in the proton-proton collisions. For all the channels under study, the associated production of a Z boson with a charm quark is measured in the fiducial region, $\sigma_{fid}(pp \to \mathrm{Z} + \mathrm{c} + X)$. The latter refers to the process rate in the region of the phase space in which the measurement is done:

- $p_T^{\ell} > 20$ GeV, $|\eta^{\ell}| < 2.1$ where $\ell = e, \mu$ and $p_T^{jet} > 25$ GeV, $|\eta^{jet}| < 2.5$ where jet refers to a charm or bottom jet clustered using the anti-k_T algorithm with radius $R = 0.5$. The latter is also known as AK5 jets in the literature.
- The invariant mass of the lepton pairs, 71 GeV $< m_{\ell\ell} <$ 111 GeV.
- Angular separation between the leptons from the Z and the jet is $\Delta R(jet, \ell) > 0.5$.

The fraction of events produced in the fiducial region with respect to the total number of events is called the acceptance, A, and is estimated using simulated samples. The relation between the fiducial cross section and the total cross section is given by

$$\sigma_{fid}(pp \to \mathrm{Z} + \mathrm{c} + X) = A\sigma(pp \to \mathrm{Z} + \mathrm{c} + X) \tag{3.7}$$

In this analysis only $\sigma_{fid}(pp \to \mathrm{Z} + \mathrm{c} + X)$ and $\sigma_{fid}(pp \to \mathrm{Z} + \mathrm{c} + X)/\sigma_{fid}(pp \to \mathrm{Z} + \mathrm{b} + X)$ are measured, however they are denoted in the text simply as $\sigma(\mathrm{Z} + \mathrm{c})$ and $\sigma(\mathrm{Z} + \mathrm{c})/\sigma(\mathrm{Z} + \mathrm{b})$ to simplify the notation.

In practice, not all the Z + c or Z + b events produced in the fiducial region, $N_{\mathrm{GEN}}^{\mathrm{Z+c}}$ and $N_{\mathrm{GEN}}^{\mathrm{Z+b}}$, are used in the analysis. Only a fraction of them, $N_{\mathrm{RECO}}^{\mathrm{Z+c}}$ and $N_{\mathrm{RECO}}^{\mathrm{Z+b}}$, are reconstructed in CMS and selected in the analysis depending on the channel.

$$N_{\mathrm{RECO}}^{\mathrm{Z+c}} = \epsilon^{\mathrm{Z+c}} N_{\mathrm{GEN}}^{\mathrm{Z+c}} \quad , \quad N_{\mathrm{RECO}}^{\mathrm{Z+b}} = \epsilon^{\mathrm{Z+b}} N_{\mathrm{GEN}}^{\mathrm{Z+b}} \tag{3.8}$$

Usually ϵ is called the reconstruction and selection efficiency and is estimated directly from simulation. Potential mismodellings between the simulation and real performance of the detector affect the efficiencies and need to be corrected (the calibration of the simulation was discussed in Sections 3.3 and 3.4.3). The calibrated reconstruction and selection efficiencies in the simulation are denoted as ϵ_c, and shown in Table 3.7 for all the channels employed in the measurement.

ϵ_c^{Z+c} and ϵ_c^{Z+b}	Semileptonic (%)	$D^{\pm}$ (%)	$D^{*\pm}(2010)$ (%)
Muon channel and Z + c	0.63 ± 0.02	0.13 ± 0.02	0.10 ± 0.01
Electron channel and Z + c	0.81 ± 0.02	0.18 ± 0.02	0.13 ± 0.01
Muon channel and Z + b	2.88 ± 0.08	N/A	N/A
Electron channel and Z + b	3.89 ± 0.10	N/A	N/A

TABLE 3.7: Reconstruction and selection efficiencies in the form of (%) in semileptonic, $D^{*\pm}(2010)$ and $D^{\pm}$ channels for Z + c and Z + b.

With all the ingredients in place, the measured fiducial cross sections, σ_{Z+c} and $\sigma(Z+c)/\sigma(Z+b)$, are computed using Equation 1.1 combined with the definition in Equation 3.8:

$$\sigma(Z+c) = \frac{N_{Z+c}^{fitted}}{\epsilon_c^{Z+c}\mathcal{L}} \quad ; \quad \frac{\sigma(Z+c)}{\sigma(Z+b)} = \frac{N_{Z+c}^{fitted}}{N_{Z+b}^{fitted}} \frac{\epsilon_c^{Z+b}}{\epsilon_c^{Z+c}} \tag{3.9}$$

Where N_{Z+c}^{fitted} (N_{Z+b}^{fitted}) corresponds to the number of reconstructed Z + c (Z + b) in the CMS data and was determined in Section 3.4.5 for all three channels.

Alternatively it is also interesting to re-write the cross section in terms of the signal strengths, μ_{Z+c} and μ_{Z+b}. The latter were defined as the factor that relates the expected number of reconstructed events in data and simulation, $N_{Z+c}^{fitted} = \mu_{Z+c} N_{RECO}^{Z+c}$. This way, the number of fitted events, N_{Z+c}^{fitted}, can be written as

$$N_{Z+c}^{fitted} = \mu_{Z+c} N_{RECO}^{Z+c} = \mu_{Z+c} \epsilon_c^{Z+c} N_{GEN}^{Z+c} \tag{3.10}$$

Combining Equations 3.9 and 3.10, the measured cross section is related with the cross section of the reference simulation, $\sigma(Z+c)_{MC} = N_{GEN}^{Z+c}/\mathcal{L}$ and $\sigma(Z+b)_{MC} = N_{GEN}^{Z+b}/\mathcal{L}$, only by the fitted signal strengths.

$$\sigma(Z+c) = \mu_{Z+c}\sigma(Z+c)_{MC} \quad ; \quad \frac{\sigma(Z+c)}{\sigma(Z+b)} = \frac{\mu_{Z+c}}{\mu_{Z+b}} \frac{\sigma(Z+c)_{MC}}{\sigma(Z+b)_{MC}} \tag{3.11}$$

In other words, the signal strengths are a measurement of the deviation of the measurement with respect to the predictions from the reference simulation obtained with MADGRAPH.

3.5.1 Inclusive Z+c and Z+c/Z+b cross section

The associated Z + c cross section in the fiducial region has been measured in three decay modes and two different Z decay flavours. The extracted signal strengths and number of N_{Z+b}^{fitted} and N_{Z+b}^{fitted} were collected in Tables 3.3, 3.4, 3.5 and the reconstruction and selection efficiencies in Table 3.7. With this information, the measured cross sections are computed equivalently for all the channels using Equation 3.9 or 3.11.

Figure 3.23 (top) shows the measured cross section for each channel. All six measurements are consistent between each other at the level of one standard deviation and they are combined in a unique measurement:

$$\sigma(\mathrm{Z}+\mathrm{c}) = 8.6 \pm 0.5\ (\mathrm{stat}) \pm 0.7\ (\mathrm{syst})\ \mathrm{pb}.$$

The relative Z + c production to Z + b is encoded in the $\sigma(\mathrm{Z}+\mathrm{c})/\sigma(\mathrm{Z}+\mathrm{b})$ ratio that is measured only in the semileptonic channel. Figure 3.24 (top) shows the measured ratio in the electron and muon channels, as well as the combined measurement:

$$\sigma(\mathrm{Z}+\mathrm{c})/\sigma(\mathrm{Z}+\mathrm{b}) = 2.0 \pm 0.2\ (\mathrm{stat}) \pm 0.2\ (\mathrm{syst}).$$

In the following, all these measurements are compared with theoretical predictions, $\sigma(\mathrm{Z}+\mathrm{c})_{\mathrm{MC}}$ and $\sigma(\mathrm{Z}+\mathrm{c})_{\mathrm{MC}}/\sigma(\mathrm{Z}+\mathrm{b})_{\mathrm{MC}}$, from three MC event generators: the reference MADGRAPH sample, MADGRAPH5_AMC@NLO and MCFM. All the theoretical predictions are computed in the same fiducial region as the one employed by the measurement: two leptons with $p_T^{\ell} > 20$ GeV, pseudo-rapidity $|\eta^{\ell}| < 2.1$, and dilepton invariant mass 71 GeV $< m_{\ell\ell} <$ 111 GeV, and a heavy flavour-quark jet with $p_T^{jet} > 25$ GeV and $|\eta^{jet}| < 2.5$. Generated partons are clustered into jets using an anti-k_T algorithm with a radius $R = 0.5$. Generated jets are separated from the leptons originated in the Z-boson decay with an angular requirement $\Delta R(jet, \ell) > 0.5$.

The reference Z + jets sample used in the analysis is generated with MADGRAPH and CTEQ6L1 as PDF. A proper *k-factor* for the normalization is obtained with a NNLO prediction from FEWZ3.1 and MSTW2008NNLO PDF. Generated events are interfaced with PYTHIA6 for parton shower and hadronization. The predicted cross sections in the reference sample are in agreement with the measured values:

$$\sigma(\mathrm{Z}+\mathrm{c})_{\mathrm{MADGRAPH}} = 8.14 \pm 0.03\ (\mathrm{stat}) \pm 0.25\ (\mathrm{PDF})\ \mathrm{pb}$$
$$\sigma(\mathrm{Z}+\mathrm{c})/\sigma(\mathrm{Z}+\mathrm{b})|_{\mathrm{MADGRAPH}} = 1.8 \pm 0.01\ (\mathrm{stat}).$$

The predictions from MADGRAPH5_AMC@NLO and NNPDF3.0 [28] PDF set interfaced with PYTHIA8 [32], are also in agreement with the measurements:

$$\sigma(\mathrm{Z}+\mathrm{c})_{\mathrm{MADGRAPH5\_AMC@NLO}} = 9.47 \pm 0.04\ (\mathrm{stat}) \pm 0.15\ (\mathrm{PDF}) \pm 0.50\ (\mathrm{scales})\ \mathrm{pb}$$
$$\sigma(\mathrm{Z}+\mathrm{c})/\sigma(\mathrm{Z}+\mathrm{b})|_{\mathrm{MADGRAPH5\_AMC@NLO}} = 1.87 \pm 0.02\ (\mathrm{stat}) \pm 0.07\ (\mathrm{scales}).$$

In the above prediccions, the matrix element calculation includes up to 2 partons at NLO and uses the FxFx[126] merging scheme with scale parameter of 20 GeV.

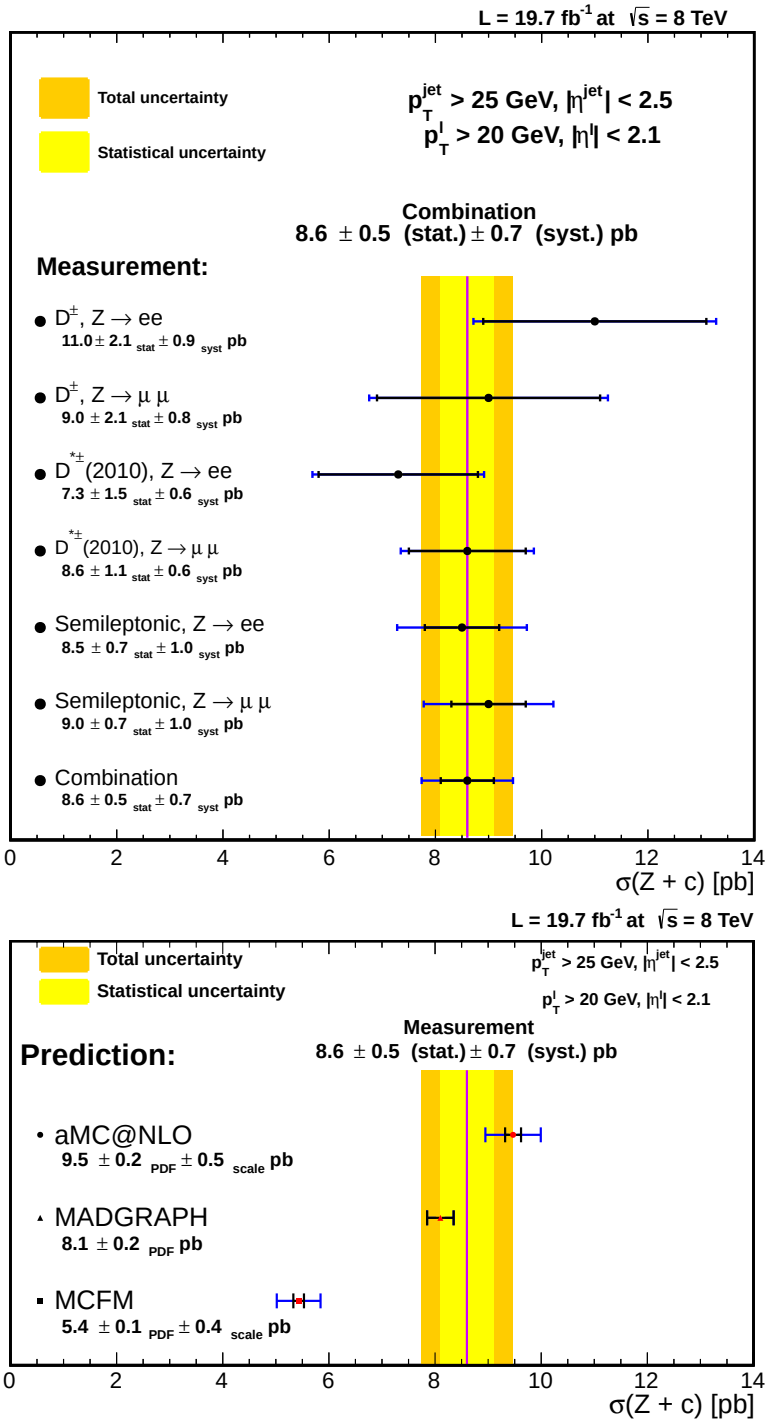

FIGURE 3.23: $\sigma(\mathrm{Z}+\mathrm{c})$ cross section measurement (top) in six different channels: semileptonic decays, $\mathrm{D}^{\pm}$, $\mathrm{D}^{*\pm}(2010)$ each of them in two Z boson decay channels. The error bars in the measurement include statistical uncertainties due to the limited size of the sample in addition to the uncertainty of the fit and systematic uncertainties as described in Section 3.4.6. The combination of all channels is also displayed. The combined measurement is compared with predictions (bottom) from MADGRAPH renormalized to NNLO prediction from FEWZ, MADGRAPH5_AMC@NLO and MCFM.

Finally measurements are compared with predictions from MCFM generator using MSTW2008, CT10 and NNPDF3.0 sets. Predictions are calculated only at parton level, and an additional correction (estimated with the reference MADGRAPH sample) of the order 10% for Z + c are included to take into account non-perturbative effects. Predictions with the three PDF sets agree up to 10% and taking as reference the largest prediction from MSTW08 set,

$$\sigma(\mathrm{Z}+\mathrm{c})_{\mathrm{MCFM}} = 5.32 \pm 0.01\ (\mathrm{stat})^{+0.12}_{-0.06}\ (\mathrm{PDF})^{+0.34}_{-0.38}\ (\mathrm{scales})\ \mathrm{pb}$$
$$\sigma(\mathrm{Z}+\mathrm{c})/\sigma(\mathrm{Z}+\mathrm{b})|_{\mathrm{MCFM}} = 1.58 \pm 0.01\ (\mathrm{stat}) \pm 0.07\ (\mathrm{scales}).$$

The predicted MADGRAPH and MADGRAPH5_AMC@NLO predictions are in agreement with the measurement. On the other hand, MCFM cross sections and the ratio are smaller and disfavored by the measurement. Figures 3.23 and 3.24 (bottom) show a summary of the theoretical predictions compared to the combined measurements.

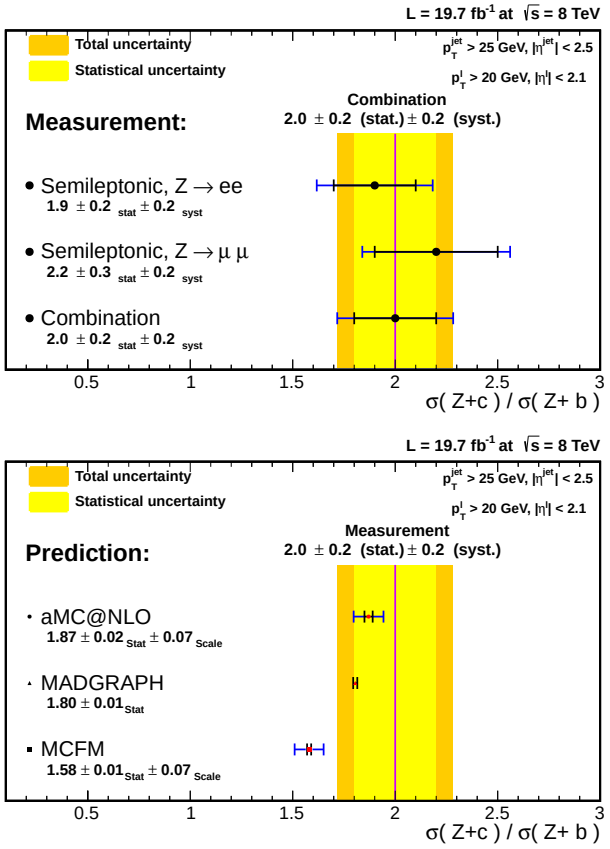

FIGURE 3.24: Measured $\sigma(\mathrm{Z}+\mathrm{c})/\sigma(\mathrm{Z}+\mathrm{b})$ cross section ratio from semileptonic decays in two Z boson decay channel (top), $\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$ and $\mathrm{Z} \to \mu^+\mu^-$. The error bars in the measurement include statistical uncertainties due to the limited size of the sample in addition to the uncertainty of the fit and systematic uncertainties described in Section 3.4.6. The combined measurement is compared (bottom) with predictions from MADGRAPH5_AMC@NLO, MADGRAPH renormalized to a NNLO prediction from FEWZ, and MCFM.

3.5.2 Differential Z+c and Z+c/Z+b cross section

The $\sigma(\mathrm{Z}+\mathrm{c})$ production cross section and the $\sigma(\mathrm{Z}+\mathrm{c})/\sigma(\mathrm{Z}+\mathrm{b})$ ratio are also measured differentially as a function of the transverse momentum of the Z boson and the heavy flavoured jet in the semileptonic mode.

This measurement is important because differential measurements are sensitive to different regions in the phase space and topologies. For example, they are used to isolate the regions in the phase space where the modelling of different generators is not optimal (or best) to describe the data. Additionally, differential distributions in the Z + c measurement are also sensitive to a potential intrinsic component of the charm PDF (more details are given in Section 3.1.1). Several studies suggest that at high transverse momentum of the Z boson, the differential $\sigma(\mathrm{Z}+\mathrm{c})$ measurement could discern among PDF sets that include intrinsic charm component in the proton to others that do not include it.

Technically, the Z + c differential measurement is done when the original sample is further divided into different sub-samples according to the value of the variable of interest, p_{T}^{Z} or p_{T}^{jet}. For each of these sub-samples, the signal extraction fit (Equation 3.10) is performed independently for the $\mathrm{Z} \to \mathrm{e}^{+}\mathrm{e}^{-}$ and $\mathrm{Z} \to \mu^{+}\mu^{-}$ decay modes and the cross section is measured. The size of each of the sub-samples is chosen in such way that the secondary vertex mass distribution is sufficiently populated to perform a robust signal extraction fit (binning is shown in Table 3.10).

Differential measurements are often referred (when relevant) to quantities at particle level prior to the CMS reconstruction. This is needed because allows an easier comparison among different experiments and with theoretical predictions. The transition between an observable after detector reconstruction to the same at particle level is called unfolding. For example, one of the typical reconstruction effects are potential migrations between neighbouring bins in the differential measurement (or inside/outside the acceptance) due to the detector resolution. This effect has been studied in the p_{T}^{jet} differential measurement since migrations are expected due to the experimental jet energy scale and resolution. The largest effect that is observed are potential migrations between the first and second bin ($< 30\%$), and also between the second and last bin ($< 10\%$).

These effects are encoded in a response matrix built from a dedicated study in simulation of bin migrations before and after CMS reconstruction. This unfolding is performed with an analytical inversion of the response matrix and is used to convert fitted signal yields to those at particle level. The uncertainty in the unfolded fitted signal yields is in the range $4\% - 6\%$ for the Z + c and $4.5\% - 7\%$ for the Z + c/Z + b ratio. Statistical and systematic uncertainties discussed in Section 3.4.6 are also propagated through the

unfolding procedure. Table 3.8 shows the measured differential cross section and the ratio, as a function of the p_{T}^{jet} for the electron and muon channels separately.

Channel	$N_{\mathrm{Z+c}}^{\mathrm{signal}}$	$\frac{d\sigma(\mathrm{Z+c})}{dp_{\mathrm{T}}^{jet}}$ (pb)	$N_{\mathrm{Z+b}}^{\mathrm{signal}}$	$\frac{d\sigma(\mathrm{Z+c})}{dp_{\mathrm{T}}^{jet}} / \frac{d\sigma(\mathrm{Z+b})}{dp_{\mathrm{T}}^{jet}}$
$25\ \mathrm{GeV} < p_{\mathrm{T}}^{jet} < 40\ \mathrm{GeV}$				
$\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$	476 ± 58	$0.342 \pm 0.048 \pm 0.041$	1022 ± 67	$2.3 \pm 0.6 \pm 0.2$
$\mathrm{Z} \to \mu^+\mu^-$	583 ± 91	$0.337 \pm 0.059 \pm 0.055$	1393 ± 90	$2.4 \pm 0.5 \pm 0.3$
$40\ \mathrm{GeV} < p_{\mathrm{T}}^{jet} < 60\ \mathrm{GeV}$				
$\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$	289 ± 47	$0.090 \pm 0.027 \pm 0.018$	843 ± 59	$1.3 \pm 0.6 \pm 0.3$
$\mathrm{Z} \to \mu^+\mu^-$	456 ± 66	$0.103 \pm 0.027 \pm 0.014$	1044 ± 75	$1.9 \pm 0.5 \pm 0.3$
$60\ \mathrm{GeV} < p_{\mathrm{T}}^{jet} < 200\ \mathrm{GeV}$				
$\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$	310 ± 56	$0.012 \pm 0.003 \pm 0.008$	686 ± 64	$1.7 \pm 0.5 \pm 0.3$
$\mathrm{Z} \to \mu^+\mu^-$	369 ± 63	$0.013 \pm 0.003 \pm 0.007$	800 ± 75	$1.9 \pm 0.5 \pm 0.3$

TABLE 3.8: Differential cross section $\frac{d\sigma(\mathrm{Z+c})}{dp_{\mathrm{T}}^{jet}}$, and cross section ratio $\frac{d\sigma(\mathrm{Z+c})}{dp_{\mathrm{T}}^{jet}} / \frac{d\sigma(\mathrm{Z+b})}{dp_{\mathrm{T}}^{jet}}$ in the semileptonic mode and in the two Z-boson decay channels. $N_{\mathrm{Z+c}}^{\mathrm{signal}}$ and $N_{\mathrm{Z+b}}^{\mathrm{signal}}$ are the yields of Z + c and Z + b events extracted from the fit. All uncertainties quoted in the table are statistical except for the measured cross sections and cross sections ratio, where the first uncertainty is statistical and the second one is the systematic uncertainty from the sources discussed in Section 3.4.6.

In the case of the differential measurement as a function of p_{T}^{Z}, migration effects are negligible due to the good momentum resolution compared to the size of the bins. In this case, the response matrix used to unfold the fitted signal yields was measured to be almost diagonal. Due to this, no unfolding was applied and an additional uncertainty of 2% was added to the fitted signal yields. Table 3.9 shows the measured differential cross section and the ratio in the electron and muon channel as a function of the p_{T}^{Z}.

Channel	$N_{\mathrm{Z+c}}^{\mathrm{signal}}$	$\frac{d\sigma(\mathrm{Z+c})}{dp_{\mathrm{T}}^{Z}}$ (pb)	$N_{\mathrm{Z+b}}^{\mathrm{signal}}$	$\frac{d\sigma(\mathrm{Z+c})}{dp_{\mathrm{T}}^{Z}} / \frac{d\sigma(\mathrm{Z+b})}{dp_{\mathrm{T}}^{Z}}$
$p_{\mathrm{T}}^{Z} < 30\ \mathrm{GeV}$				
$\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$	212 ± 44	$0.066 \pm 0.014 \pm 0.010$	578 ± 52	$1.5 \pm 0.4 \pm 0.2$
$\mathrm{Z} \to \mu^+\mu^-$	380 ± 61	$0.103 \pm 0.017 \pm 0.018$	693 ± 68	$2.7 \pm 0.6 \pm 0.4$
$30\ \mathrm{GeV} < p_{\mathrm{T}}^{Z} < 60\ \mathrm{GeV}$				
$\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$	501 ± 60	$0.144 \pm 0.017 \pm 0.019$	1035 ± 66	$2.4 \pm 0.4 \pm 0.3$
$\mathrm{Z} \to \mu^+\mu^-$	586 ± 92	$0.123 \pm 0.019 \pm 0.018$	1422 ± 87	$1.9 \pm 0.4 \pm 0.3$
$60\ \mathrm{GeV} < p_{\mathrm{T}}^{Z} < 200\ \mathrm{GeV}$				
$\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$	363 ± 53	$0.017 \pm 0.002 \pm 0.002$	913 ± 67	$1.7 \pm 0.3 \pm 0.2$
$\mathrm{Z} \to \mu^+\mu^-$	474 ± 73	$0.017 \pm 0.003 \pm 0.002$	1056 ± 81	$2.1 \pm 0.4 \pm 0.3$

TABLE 3.9: Differential cross section $\frac{d\sigma(\mathrm{Z+c})}{dp_{\mathrm{T}}^{Z}}$, and cross section ratio $\frac{d\sigma(\mathrm{Z+c})}{dp_{\mathrm{T}}^{Z}} / \frac{d\sigma(\mathrm{Z+b})}{dp_{\mathrm{T}}^{Z}}$ in the semileptonic mode and in the two Z-boson decay channels. $N_{\mathrm{Z+c}}^{\mathrm{signal}}$ and $N_{\mathrm{Z+b}}^{\mathrm{signal}}$ are the yields of Z + c and Z + b events extracted from the fit. All uncertainties quoted in the table are statistical except for the measured cross sections and cross sections ratio, where the first uncertainty is statistical and the second one is the systematic uncertainty from the sources discussed in Section 3.4.6.

Finally in Table 3.10 the unfolded cross sections are shown after the muon and the electron channels are combined together.

$[p_{T_{\min}}^{Z}, p_{T_{\max}}^{Z}]$ [GeV]	$\frac{d\sigma(\mathrm{Z+c})}{dp_T^Z}$ [pb]	$\frac{d\sigma(\mathrm{Z+c})}{dp_T^Z} / \frac{d\sigma(\mathrm{Z+b})}{dp_T^Z}$
$[0, 30]$	$0.075 \pm 0.011 \pm 0.012$	$1.8 \pm 0.3 \pm 0.2$
$[30, 60]$	$0.133 \pm 0.013 \pm 0.018$	$2.1 \pm 0.3 \pm 0.3$
$[60, 200]$	$0.017 \pm 0.002 \pm 0.002$	$1.9 \pm 0.3 \pm 0.2$
$[p_{T_{\min}}^{jet}, p_{T_{\max}}^{jet}]$ [GeV]	$\frac{d\sigma(\mathrm{Z+c})}{dp_T^{jet}}$ [pb]	$\frac{d\sigma(\mathrm{Z+c})}{dp_T^{jet}} / \frac{d\sigma(\mathrm{Z+b})}{dp_T^{jet}}$
$[25, 40]$	$0.341 \pm 0.037 \pm 0.042$	$2.5 \pm 0.4 \pm 0.3$
$[40, 60]$	$0.097 \pm 0.019 \pm 0.012$	$1.5 \pm 0.3 \pm 0.2$
$[60, 200]$	$0.013 \pm 0.002 \pm 0.002$	$1.8 \pm 0.4 \pm 0.3$

TABLE 3.10: Differential $\sigma(\mathrm{Z+c})$ cross section and $\sigma(\mathrm{Z+c})/\sigma(\mathrm{Z+b})$ cross sections ratio. The first block presents the differential measurements as a function of the transverse momentum of the Z boson. The second block shows the cross section and ratio as a function of the transverse momentum of the jet with heavy flavour content. The first uncertainty is the statistical and the second one is the systematic uncertainty arising from the sources discussed in Section 3.4.6.

In the remaining of this section, the differential measurements listed in Table 3.10 are compared with theoretical predictions from the same generators used in Section 3.5.1.

Figure 3.25 shows the differential Z + c cross sections compared to several predictions as a function of the p_{T}^{jet} (left) and p_{T}^{Z} (right). Statistical and theoretical uncertainties in the predictions are added in quadrature and shown as vertical error bars. In both figures, the predictions from MADGRAPH (re-scaled to a NNLO prediction from FEWZ 3.1) and MADGRAPH5_AMC@NLO agree with the measurements at the level of one standard deviation. The largest difference is observed at the lowest jet transverse momentum, where the MADGRAPH5_AMC@NLO gives the best description of the data. In the case of MCFM, the predictions were found to be about 40% lower in the inclusive measurement described in Section 3.5.1. In the differential measurement, this observed disagreement is more pronounced in the region of the phase space where the jet transverse momentum is between 25 GeV and 40 GeV.

The differential measurement with the highest p_{T}^{Z} in Figure 3.25 (right), covers the dilepton momentum between 60 GeV and 200 GeV. This measurement is important because is the one with the highest sensitivity to a potential Z + c enhancement due to a potential contribution from a non-perturbative component of the charm inside the proton (see Section 3.1.1 for more details). According to recent studies, the size of this enhancement could be as high as 20% in the most optimistic scenarios at $p_{\mathrm{T}}^{Z} \approx$ 100 GeV. However it is important to keep in mind that predictions have large theoretical uncertainties and this enhancement could be much smaller.

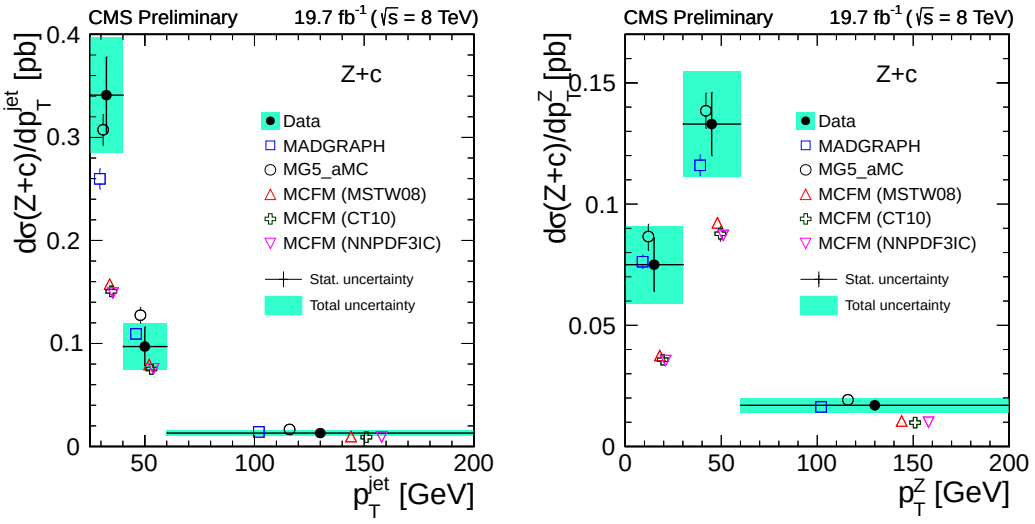

FIGURE 3.25: Differential Z + c cross section as a function of the transverse momentum of the jet (left) and the transverse momentum of the Z boson (right).The combination of the results in the dielectron and dimuon channels are shown (black dots) and compared with different theoretical predictions. Statistical and systematical errors are added in quadrature (green band).

The measurement in the highest p_{T}^{Z} bin agrees within the uncertainties with the prediction from MADGRAPH5_AMC@NLO that uses NNPDF3.0 set that treats charm perturbatively.

The effect of a different treatment of the charm PDF for this 60 GeV $< p_{\mathrm{T}}^{Z} <$ 200 GeV region has been studied using MCFM with different PDFs as input. Predictions of the Z + c differential distributions using NNPDF3.0, MSTW or CT10, all using charm generated perturbatively inside the proton, do not show sizable differences in their predictions. Additionally, the predictions using the traditional NNPDF set were also compared with a modified PDF (NNPDF3IC)[101] prediction in which the charm PDF has been fitted (see Figure 3.3 and 3.4 in Section 3.1.1 for a comparison of both PDFs). As a result of this comparison, both PDF sets do not present sizable differences for the same p_{T}^{Z} regime of the measurement. As expected, differences among both PDF sets start to be sizable when the transverse momentum of the Z boson is greater than 200 GeV.

Unfortunately, the sensitivity of this analysis at high p_{T}^{Z} is limited by the statistics in the data sample and it is not possible to probe the predictions above $p_{\mathrm{T}}^{Z} >$ 200 GeV, where the size of the enhancement is expected to be higher. For example, the size of this enhancement at $p_{\mathrm{T}}^{Z} \approx$ 600 GeV could be as big as a factor two. Due to this, an updated version of this analysis with the future LHC Run II dataset at 13 TeV should be able to extend the differential measurement to higher p_{T}^{Z}, and make a more conclusive statement about potential σ(Z + c) enhancements due to an intrinsic charm component in the proton.

Finally Figure 3.26 shows the combined $\sigma(\mathrm{Z}+\mathrm{c})/\sigma(\mathrm{Z}+\mathrm{b})$ cross section ratio compared to several predictions as a function of the p_{T}^{jet} (left) and p_{T}^{Z} (right). All three generators predict a flat behaviour of the ratio as a function of the observable that is confirmed by the measurement. However the expected central value of the ratio is rather different for the three generators. It is a good new that MADGRAPH and MADGRAPH5_AMC@NLO, baselines generators at 13 TeV in CMS, give an accurate description of the data confirming the inclusive results in Figure 3.24.

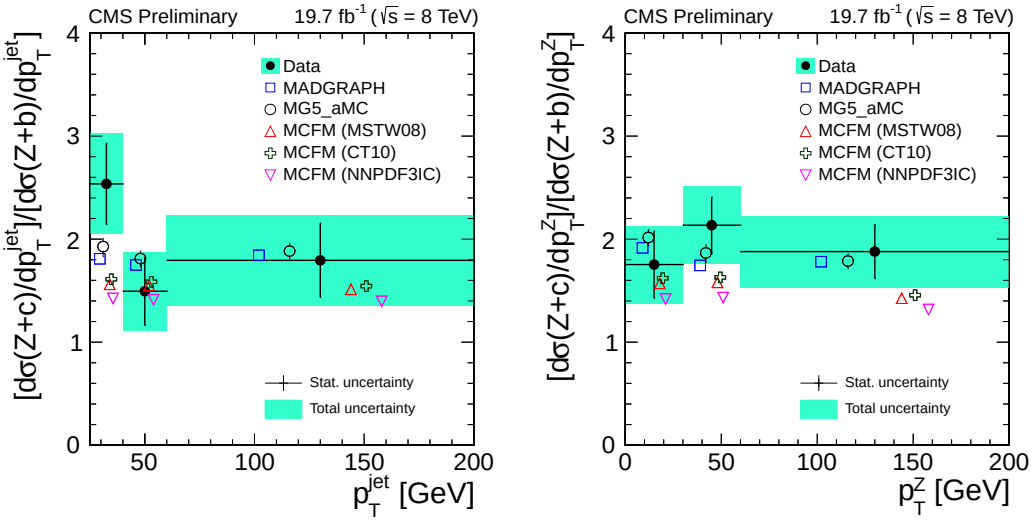

FIGURE 3.26: Differential cross section ratio Z + c/Z + b as a function of the transverse momentum of the jet (left) and the transverse momentum of the Z boson (right). The combination of the results in the dielectron and dimuon channels are shown (black dots) and compared with different theoretical predictions. Statistical and systematical errors are added in quadrature (green band).

The largest difference in the differential $\sigma(\mathrm{Z}+\mathrm{c})/\sigma(\mathrm{Z}+\mathrm{b})$ measurement is concentrated in the lowest p_{T}^{jet} bin between 25 GeV and 40 GeV. In this region, the difference between the measurement and the predictions is of the order of one standard deviation, and the measurement is slightly higher than the predictions of the three generators and the inclusive measurement.

The differential $\sigma(\mathrm{Z}+\mathrm{c})/\sigma(\mathrm{Z}+\mathrm{b})$ measurement presented in this analysis is an interesting quantity because it appears naturally (for example) in direct searches/measurements of the Yukawa coupling of the Higgs boson to charm/bottom quarks. The latter is studied at the LHC with the 13 TeV data in searches of Higgsstrahlung processes of the type: $\mathrm{Z}^{*} \rightarrow \mathrm{H}(\mathrm{c\bar{c}})\mathrm{Z}(\ell\ell)$ and $\mathrm{Z}^{*} \rightarrow \mathrm{H}(\mathrm{b\bar{b}})\mathrm{Z}(\ell\ell)$, in which Z + b and Z + c are dominant background processes. This way, the measurement of the differential cross section ratio presented in this thesis will serve as an important check of the capability of the generators used in CMS to reproduce the relative rate in which the dominant backgrounds in these searches are produced.

Summary of main physics results of the Z + c analysis:

Result 6: Inclusive $\sigma(\mathrm{Z}+\mathrm{c})$ cross section measurement

The fiducial cross section of the associated production of a Z boson with at least one charm quark is measured in six different final states. Measurements in all the channels are compatible between each other given the quoted uncertainties. The combined cross section was determined to be:

$$\sigma_{fid}(\mathrm{Z}+\mathrm{c}) = 8.6 \pm 0.5\ (\mathrm{stat}) \pm 0.7\ (\mathrm{syst})\ \mathrm{pb}$$

in agreement with predictions from MadGraph5_aMC@NLO and Madgraph renormalized to a FEWZ calculation.

Result 7: Inclusive $\sigma(\mathrm{Z}+\mathrm{c})/\sigma(\mathrm{Z}+\mathrm{b})$ ratio

The relative production of charm to bottom in association with a Z boson was measured in the semileptonic channel in the same fiducial region:

$$\sigma_{fid}(\mathrm{Z}+\mathrm{c})/\sigma_{fid}(\mathrm{Z}+\mathrm{b}) = 2.0 \pm 0.2\ (\mathrm{stat}) \pm 0.2\ (\mathrm{syst})$$

in agreement with predictions from MadGraph5_aMC@NLO and Madgraph renormalized to a FEWZ calculation.

Result 8: Differential $\sigma(\mathrm{Z}+\mathrm{c})$ and $\sigma(\mathrm{Z}+\mathrm{c})/\sigma(\mathrm{Z}+\mathrm{b})$ measurement

The differential $\sigma(\mathrm{Z}+\mathrm{c})$ and $\sigma(\mathrm{Z}+\mathrm{c})/\sigma(\mathrm{Z}+\mathrm{b})$ distributions were also measured as a function of the p_{T}^{Z} or p_{T}^{jet} with heavy flavour content.
Measurements in the region 25 GeV $< p_{\mathrm{T}}^{jet} <$ 200 GeV and $p_{\mathrm{T}}^{Z} <$ 200 GeV are collected in Table 3.10, and are in good agreement with predictions from MadGraph5_aMC@NLO and Madgraph.

Result 9: Differential $\sigma(\mathrm{Z}+\mathrm{c})$ at high p_{T}^{Z}

The differential $\sigma(\mathrm{Z}+\mathrm{c})$ cross section measurement at high p_{T}^{Z} could be slightly enhanced due to the existence of a charm component in the proton of non-perturbative origin. This analysis at high p_{T}^{Z} is limited by the statistics of the sample, and the measurement in the most sensitive region, 60 GeV $< p_{\mathrm{T}}^{Z} <$ 200 GeV, can be explained using PDFs that include only charm generated perturbatively. Future measurements with the new 13 TeV dataset at higher $p_{\mathrm{T}}^{Z} >$ 200 GeV, where this enhancement is expected to be larger, should be able to exclude or confirm the existence of a intrinsic charm component inside the proton.

3.6 Z + c cross section measurement: conclusions

Z bosons decaying to electron and muon pairs are one of the most abundant, experimentally clean, and theoretically known process produced at the LHC. Their measurement in association with jets serves as an excellent benchmark to test the SM predictions from different generators against the experimental data.

The excellent capabilities of the tracker system to identify jets originated in the hadronization of heavy flavour quarks allowed CMS to develop a dedicated program of measurement of Z bosons in association with bottom quarks at 7 and 8 TeV. The Z + c measurement described in this chapter aims to extend this program and provides the cross section measurement of associated production of Z bosons with charm quarks for the first time in CMS.

This new measurement serves as an excellent test of the QCD predictions of the SM, as well as the PDF sets used at the LHC. Particularly, there is an ongoing effort to study the charm PDF of the proton. Modern PDF sets assume that charm is produced perturbatively inside the proton, however it is not experimentally discarded that there could be an additional intrinsic component of non-perturbative origin that modifies the charm PDF. For example, according to recent studies from the CT10 and NNPDF groups, its existence could lead to a small enhancement in the differential Z + c cross section as big as 20%(100%) for a transverse momentum of the Z boson of $p_{\mathrm{T}}^{Z} \approx 100$ GeV(600 GeV).

Z plus heavy flavour jets processes are also interesting in measurements or searches for new physics since they are often one of the dominant backgrounds. For example, the *yet-to-be discovered* Higgs boson coupling to bottom/charm in the $\mathrm{Z}^{*} \to \mathrm{H}(\mathrm{b\bar{b}}, \mathrm{c\bar{c}})\mathrm{Z}(\ell\ell)$ channel has a dominant background of Z boson produced in association with heavy flavour jets. Moving to more exotic searches, diboson resonances of the type $X \to \mathrm{Z}(\ell\ell)\mathrm{V}(qq)$ using b-tagging techniques, flavour changing neutral currents of the type $t \to \mathrm{Z} + \mathrm{c}$ or third generation of light squarks decaying via charm and neutralino, $\tilde{\mathrm{t}} \to \mathrm{c}\tilde{\chi}^{0}$, all have in common that they deal with the associated production of $\mathrm{Z} \to \ell\ell$ with heavy flavour jets as background.

The second analysis in this thesis, presents the measurement of the cross section of $\sigma(\mathrm{Z} + \mathrm{c})$ and its relative production to bottom, $\sigma(\mathrm{Z} + \mathrm{c})/\sigma(\mathrm{Z} + \mathrm{b})$, in the fiducial region:

- $p_T^{\ell} > 20$ GeV, $|\eta^{\ell}| < 2.1$, $p_T^{jet} > 25$ GeV, $|\eta^{jet}| < 2.5$, where $\ell = e, \mu$, and jet refers to a charm or bottom anti-k_T with radius $R = 0.5$ (AK5) jet.
- Dilepton invariant mass 71 GeV $< m_{\ell\ell} <$ 111 GeV.
- Angular separation between the leptons from the Z and the jet is $\Delta R(\ell, jet) > 0.5$.

The measurement is done inclusively in six independent channels targeting three different jet topologies. Displaced muons from semileptonic decays of D- and B- hadrons and $D^{\pm}$ or $D^{*\pm}(2010)$ secondary vertices. The Z + c signal is extracted from a χ^2 fit of the secondary vertex mass or the Jet Probability distribution in data, by using those obtained with the simulation after they have been calibrated using W + c and $e\mu$ control regions. The measurements are compatible in all six channels and its combination is:

$$\sigma_{fid}(pp \to \mathrm{Z} + \mathrm{c} + X) = 8.6 \pm 0.5\ (\mathrm{stat}) \pm 0.7\ (\mathrm{syst})\ \mathrm{pb}$$

in agreement with the predictions from MadGraph5_amc@nlo, one of the baseline generator in CMS at 13 TeV for SM precision measurements, and Madgraph renormalized to a NNLO QCD prediction from FEWZ. The prediction from MCFM was lower than the measured value. The relative production of charm to bottom in association with a Z boson was measured in the semileptonic channel in the same fiducial region:

$$\sigma_{fid}(pp \to \mathrm{Z} + \mathrm{c} + X)/\sigma_{fid}(pp \to \mathrm{Z} + \mathrm{b} + X) = 2.0 \pm 0.2\ (\mathrm{stat}) \pm 0.2\ (\mathrm{syst})$$

Again, this measurement agrees with the predictions from MadGraph5_amc@nlo and Madgraph. The MCFM prediction was also lower than the measured value. The data was further analysed and these cross sections were also measured differentially as a function of the p_{T}^{jet} and p_{T}^{Z}. The extracted signals yields were unfolded to particle level and compared with several theory predictions. MadGraph5_amc@nlo generator predictions showed an excellent agreement with the differential measurement. The agreement with Madgraph was found to be also good. The largest discrepancy between the measurement and the predictions was found in comparison with predictions from MCFM. This discrepancy is more pronounced in the region 25 GeV $< p_{\mathrm{T}}^{jet} <$ 40 GeV.

The cross section measurement at the highest momentum, 60 GeV $< p_{\mathrm{T}}^{Z} <$ 200 GeV, is in agreement with predictions from MadGraph5_amc@nlo using the NNPDF set that treats charm perturbatively. According to theoretical predictions in this p_{T}^{Z} range, the size of the intrinsic charm enhancement is expected to be small (and predictions suffer from large theoretical uncertainties). As a consequence, this analysis alone is not able to exclude or confirm the existence of intrinsic charm in the proton. Unfortunately, measurements in the region $p_{\mathrm{T}}^{Z} >$ 200 GeV are limited by the statistics of the sample. An updated measurement with a larger LHC dataset at 13 TeV, should lead to a stronger statement about potential Z + c enhancements at larger momentum $p_{\mathrm{T}}^{Z} >$ 200 GeV.

Finally, this measurement uses novel techniques in CMS, such as W + c, to study the performance of charm jets in the context of heavy flavour tagging that are currently used to calibrate c- and b-tagging algorithms with 13 TeV data.

Summary and outlook

After the discovery of the Higgs boson in 2012, all the particles in the original version of the Standard Model (SM) have been discovered experimentally. In its current form, this model is a powerful theory because it makes precise predictions of phenomena that have been verified to be true in studies using collisions from different particle accelerators during the past 50 years. However, despite its astonishing power to predict phenomena, it leaves un-answered some of the fundamental questions and also cannot explain the origin of some processes observed in nature, e.g the Nature of dark matter or the origin of the neutrino masses.

Nowadays physicists are convinced that the SM is a valid theory up to some unknown energy scale, Λ_{NP}, in which new processes should manifest in terms of new particles/interactions that could give an answer to those problems. The reason why these new processes were not discovered in previous experiments is that their effects at low energy scale are negligible. For this, the new LHC data is an excellent place to seek for new physics at an unique energy scale in the history of accelerator particle physics, 13 TeV.

Our goal as particle physicists is to understand what are the elementary building blocks in Nature, their interactions and in general the universe we live in. There are experimental evidences from cosmology and astrophysics studies that indicate that our universe evolved from a high density and high temperature state to the universe we know today. In order to understand how this evolution happened, it is necessary to know if potential new interactions and particles were relevant at higher energy scales and temperatures.

The physics motivation of the first analysis presented in this thesis is to search for new interactions that manifest via new W' bosons, which could have played a role in the early days of the universe. Their existence could enlarge the symmetry group of the SM and modify its particle content. After analysing the first 2.3 fb^{-1} of proton-proton collision ever recorded at 13 TeV in CMS, there are no signs of potential new W' like interactions at the new centre-of-mass energy. If such states existed in Nature and manifested in a similar way as prescribed by the W' Sequential Standard Model (SSM), their cross section at 13 TeV must be smaller than

$$\sigma_{\mathrm{W}'} Br(\mathrm{W}' \to \mu\nu) < 3 \text{ fb}^{-1} \text{ if their mass is larger than } M_{\mathrm{W}'} > 2 \text{ TeV}.$$

This cross section limit in the context of the W′ predicted in the SSM translates into the exclusion of $\mathrm{W}'_{\mathrm{SSM}} \to \mu\nu$ bosons with masses below 3.9 TeV.

This result, combined with the full searches program of ATLAS/CMS with the early 2015 LHC dataset at 13 TeV, extends the validity of the SM one step higher in energy scale. At this point, a key question for particle physicist is, how far is this new energy scale, Λ_{NP}, in which the new physics is relevant? We do not know the answer, but our goal as particle physicist is to keep and keep searching and testing the validity of the finest predictions of the SM in the laboratory...

Alternatively, the validity of the SM is tested experimentally in precise measurements. For example, its capability to predict the rate at which rare processes occur at the LHC serves as an excellent benchmark of the theory against experimental observations. Precision cross section measurements are an excellent approach to test the accuracy of the latest theoretical predictions, as well as search for enhancements/deficits in this rate due to potential sources of new physics.

With this spirit in mind, the measurement of the Z + c cross section aims to test the latest predictions of the strong sector of the SM, as well as the proton PDF modelling of the Monte Carlo generators used in CMS. For example, several theoretical works point to a potential enhancement in the differential Z + c distribution at large Z transverse momentum due to a possible non-zero contribution of an intrinsic charm component inside the proton. Additionally, the analysis presents the first step towards a program of $\sigma(\mathrm{Z} + \mathrm{c})$ and $\sigma(\mathrm{Z} + \mathrm{c})/\sigma(\mathrm{Z} + \mathrm{b})$ cross section measurements in CMS.

Due to the novelty of the measurement, the designed strategy aims to use a simple and robust analysis. This way, charm jets are selected in different final states targeting three different jet topologies: semileptonic decays of D and B mesons, and reconstruction of $\mathrm{D}^{\pm}$ or $\mathrm{D}^{*\pm}(2010)$ vertices. These three channels are analyzed when they are produced in association with a Z boson in two flavour decay modes, $\mathrm{Z} \to \mathrm{e}^+\mathrm{e}^-$ and $\mathrm{Z} \to \mu^+\mu^-$, thus leading to a total of six different final states enriched in Z + c and Z + b events. Independent final states serve as cross checks of the measurement, giving an additional level of robustness to the presented results. All measurements agree at the level of one standard deviation and the combined results in the fiducial region are determined to be:

$$\sigma_{fid}(pp \to \mathrm{Z} + \mathrm{c} + X) = 8.6 \pm 0.5 \text{ (stat)} \pm 0.7 \text{ (syst) pb}$$
$$\sigma_{fid}(pp \to \mathrm{Z} + \mathrm{c} + X)/\sigma_{fid}(pp \to \mathrm{Z} + \mathrm{b} + X) = 2.0 \pm 0.2 \text{ (stat)} \pm 0.2 \text{ (syst)}$$

both in agreement with predictions from MadGraph5_aMC@NLO and Madgraph. The predictions from MCFM are slightly lower and disfavoured by the measurement.

The cross sections and the ratio were also determined differentially as a function of the transverse momentum of the Z boson up to 200 GeV, and as a function of the the momentum of the jet with heavy flavour content in the range 20 GeV $< p_{\mathrm{T}}^{jet} <$ 200 GeV. The differential measurement agrees with the predictions from MadGraph5_amc@nlo and Madgraph and disfavours the predictions from MCFM, especially at the lowest transverse momentum (25 GeV $< p_{\mathrm{T}}^{jet} <$ 40 GeV).

The differential measurement with the highest sensitivity to potential Z + c enhancements is in the range 60 GeV $< p_{\mathrm{T}}^{Z} <$ 200 GeV. In this interval, the size of the enhancement could be as high as 20% according to the most optimistic predictions. The measurement agrees with predictions made with MadGraph5_amc@nlo and the PDF set from the NNPDF group using charm generated perturbatively. However it is important to remark that the size of this enhancement is expected to grow with the momentum of the Z boson, particularly above 200 GeV. Currently the precision of the measurement is limited by the statistics of the sample. Due to this, it is expected that an updated version of this measurement with the future larger dataset at 13 TeV, to be able to discern between PDF sets that treat charm perturbatively to others that do not do it.

The results presented in this manuscript are a summary of about four years of research carried out at the Centro de Investigaciones Energéticas Mediombientales y Tecnológicas (CIEMAT) as a member of the CMS collaboration. This work was financially supported by the Formación de Personal Investigador (FPI) scholarship BES-2012-057486 awarded by the Ministerio de Economía y Competitividad (MINECO) in December of 2012.

During the four years of research I have been mostly involved in the Z + c and $\mathrm{W}' \to \mu\nu$ analysis, and the results presented in this manuscript correspond to the official CMS results in the topic.

The W' analysis was carried out during the years 2014-2016 in close collaboration with Begoña de la Cruz Martínez and Óscar González López, and it corresponds to the first search of this kind at 13 TeV in CMS. The analysis was carried out after two years of upgrade of the detector and the accelerator (2013-2015). During this time, the calibration of the detector and the readiness of the early searches such as $\mathrm{W}' \to \mu\nu$ using the new 13 TeV data were two of the main priorities for the collaboration. In parallel with the search, I tried to bring new ideas to study the performance and calibration of high p_{T} muons. For example, studies with cosmic muons or the development of the Generalized Endpoint Method, were used as additional methods to calibrate the scale of high p_{T} muons in early 2015 and used by other analysis within the collaboration.

The Z + c measurement at 8 TeV is a completely new analysis in CMS and thus required longer periods of research and brainstorming. The analysis strategy and the methods employed in the analysis were developed during the years 2012-2016 in close collaboration with Isabel Josa Mutuberría, taking as starting point the software framework developed by the W + c at 7 TeV CIEMAT analysis team. The theoretical predictions presented in Section 3.5 and the unfolding of the differential results to particle level in Section 3.5.2 were not part of my research and all the credit must be given to Isabel. In addition to the measurement itself, this work brings new techniques and strategies to handle charm jets for a precision measurement. Particularly it was interesting to recycle traditional ideas like W + c to calibrate charm secondary vertexing and the discriminant variables for a new Z + c measurement. For example studies with the W + c sample developed during the thesis are today one of the standard methods to calibrate the performance of b- and c-taggers in CMS with charm jets at 13 TeV.

Finally I would like to conclude this manuscript with a short remark. The LHC and the CMS experiment are two incredible achievements of human kind and guarantee solid base of particle physics research around CERN. The LHC physics program at 13 TeV has just started and will give us more than 20 years of pioneering research in particle physics. The two analysis presented in this manuscript are at the front of research today, however they were done only with less that 0.1% of the data that the LHC is expected to deliver in the coming years... With this in mind, the only thing I can say is... Fasten your seat-belt, surprises ahead in the coming years!

Alberto Escalante del Valle

Medida de la producción asociada de Z+charm y Búsqueda de bosones W' en el experiemento CMS del LHC

por Alberto Escalante del Valle

¿Sabemos cómo interaccionan las partículas elementales entre ellas?, ¿somos capaces de predecir cual sería el resultado de colisiones entre dichas partículas elementales en el LHC? Después del descubrimiento del bosón de Higgs en 2012, todas las partículas elementales en la versión original del Modelo Estándar de Física de Partículas (ME), modelo teórico más aceptado para explicar dichas interacciones, han sido descubiertas de forma experimental. A día de hoy, el ME es una teoría con una alta capacidad predictiva, ya que es capaz de hacer predicciones teóricas con gran precisión que han sido verificadas de forma experimental en los últimos 50 años. El objetivo científico de esta tesis doctoral es investigar si el ME es capaz de predecir la frecuencia a la cual los procesos $\mathrm{Z} + \mathrm{c}$ y $\mathrm{W}' \to \mu\nu$ son producidos en las colisiones entre protones en el LHC.

La primera mitad de la tesis doctoral se dedica a estudiar la posible existencia de un nuevo proceso, $\mathrm{W}' \to \mu\nu$, mediado por la existencia de un bosón, W', que sería responsable de un nuevo tipo de interacción entre partículas. Este tipo de interacciones no existen el ME, sin embargo aparecen en predicciones teóricas en muchas de sus extensiones. En el caso de que dichas interacciones hubiesen ocurrido en la naturaleza, la existencia del bosón W' podría ser demostrada gracias a las energías alcanzas en las colisiones del LHC. Después de casi dos años de mantenimiento y mejora del LHC y CMS, en verano de 2015 se han reanudado las colisiones y la toma de datos a una energía en el centro de masas récord en la historia de la física de partículas, 13 TeV. Este aumento en la energía permite una sensibilidad mayor desde las primeras colisiones para el descubrimiento de un nuevo bosón W' con masa en la escala del TeV comparado con todas las búsquedas hechas hasta la fecha. La estrategía del análisis se basa en encontrar un exceso de producción en los primeros 2.3 fb^{-1} a 13 TeV en sucesos en los cuales se han formado un muón de alto momento y mucha energía perdida, ambos medidos en el plano transverso del detector.

Para poder certificar un posible exceso en la producción de sucesos con un muón de alto momento en el plano transverso y garantizar el descubrimiento del W' (en caso de haberse producido en las colisiones), es necesaria una rápida calibración y validación de la medida del momento transversal de los muones. Dicha calibración se ha realizado utilizando una muestra de muones cósmicos que fueron tomados antes de las colisiones, y otra muestra de muones de alto momento enriquecida en sucesos $\mathrm{Z} \to \mu^+\mu^-$.

Una vez analizada la muestra, la frecuencia a la cual estos procesos se han detectado en CMS se encuentra en acuerdo con las predicciones del ME, desfavoreciendo la producción de nuevas interacciones del tipo $W' \to \mu\nu$ en los primeros 2.3 fb^{-1} a 13 TeV. Este resultado en el contexto teórico del Modelo Estándar Secuencial (SSM), modelo de referencia usado por colaboraciones experimentales para interpretar sus búsquedas de procesos $W' \to \mu\nu$, excluye la existencia de bosones W'_{SSM} en las cuales el bosón tiene masa menor de 3.9 TeV.

Le segunda parte del manuscrito presenta la primera medida en CMS de la producción de un bosón Z asociada con quarks de tipo charm, $Z + c$. El objetivo del análisis es medir la sección eficaz de producción del proceso $Z + c$, así como medir su producción relativa respecto al quark bottom $Z + b$. Esta nueva medida servirá para poder cuantificar el nivel de precisión las prediciones teóricas del sector fuerte (QCD) del ME y las funciones de distribución de partones (PDF), que explican como se distribuye la energía y el contenido de quarks y gluones dentro del protón. Por otro lado, un posible exceso o déficit en la medida de la sección eficaz en la cual un proceso se produce comparado con predicciones del ME, podría ser una indicación de física nueva o una pista de un nuevo proceso no recogido en el ME.

Para este estudio se ha empleado una muestra de colisiones de protones a una energía en el centro de mases de 8 TeV de 19.7 fb^{-1} de tamaño. La estrategía del análisis experimental se basa en la identificación del bosón Z a través de su desintegración en parejas de electrones o muones con carga opuesta, $Z \to e^+e^-$ o $Z \to \mu^+\mu^-$ y la identificación de quarks de tipo charm en tres estados finales independientes: desintegraciones semileptónicas de hadrones de tipo B y D y la reconstrucción de hadrones $D^{\pm}$ o $D^{*\pm}(2010)$.

Como resultado de este estudio, la sección eficaz del proceso $Z + c$ ha sido medida en la región fiducial $\sigma(Z + c) = 8.6 \pm 0.5\ (stat) \pm 0.7\ (syst)$ pb y su producción relativa respecto al quark bottom $\sigma(Z + c)/\sigma(Z + b) = 2.0 \pm 0.2\ (stat) \pm 0.2\ (syst)$. Además, dichas medidas se han hecho también de forma diferencial en función del momento transversal del bosón Z y del momento transverso del quark de tipo charm. Todas las medidas se han comparado con diferentes predicciones teóricas llevadas a cabo siguiendo las predicciones del ME. El resultado de dicha comparativa muestra que las medidas están en acuerdo con las predicciones de MadGraph5_amc@nlo y Madgraph, generadores Monte Carlo frecuentemente utilizados para producir simulaciones en muchos otros análisis de CMS.

Los resultados presentados en esta tesis ponen de manifiesto, una vez más, la excelente capacidad predictiva del ME utilizando como referencia una nueva medida experimental, el estudio de la producción del proceso $Z + c$. Al mismo tiempo, la búsqueda de nuevas interacciones $W' \to \mu\nu$ a 13 TeV extiende la validez de sus predicciones para explicar las interacciones entre las partículas un paso más alto en energía.

Appendix A

Summary of the Standard Model of Particle Physics

The Standard Model of Particle Physics (SM) is a local gauge theory of quantum fields that provides a theoretical framework which is able to describe almost all the experimental measurements in particle physics with high precision up to date. It is based on the symmetry group $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$, and describes three of the four fundamental interactions in nature: strong, weak and electromagnetism. These interactions are manifested via the exchange of spin-1 gauge fields: gluons, $W^{\pm}$ and Z and the photon respectively. Experimentally we know that the gluons and the photon are massless, whereas the $W^{\pm}$ and Z bosons, responsible of weak interactions, have non-zero mass. The Higgs mechanism provides an explanation to the non-zero mass of the weak gauge bosons, and gives rise to the appearance of a massive scalar spin-0 state, the Higgs boson.

All the visible matter content in the Universe is described in the SM framework by fermionic spin-1/2 constituents, grouped into quarks and leptons. Quarks participate in all the three fundamental interactions, whereas leptons do not participate in strong interactions. As of today, six quarks and six leptons have been discovered experimentally, which are organized into three families or generations.

The first generation is responsible for everyday matter and is made of an up quark (u) with charge +2/3 and a down quark (d) with charge -1/3, the electron (e) and a neutral neutrino (ν_e). The second and third generations appear as heavy replicas of the first generation with the same quantum numbers and a similar particle structure. They are made of one up-type quark (c, t) one down-type quark (s, b) and a negatively charged lepton (μ, τ) with the corresponding neutrino (ν_μ, ν_τ). Additionally each particle in the SM has a carbon copy, called its anti-particle, with opposite electric charge, and in the

case of fermions also opposite flavour quantum numbers. A summary of all the particles in the SM and their properties is shown in Figure A.1.

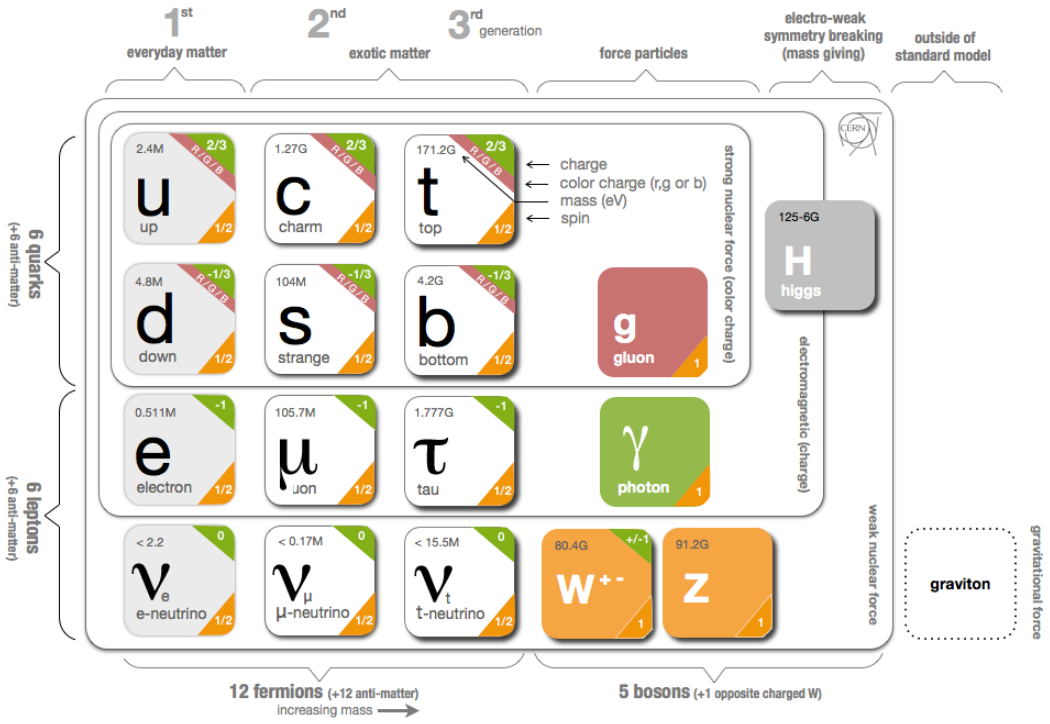

FIGURE A.1: Fundamental particles in the Standard Model with their corresponding experimental masses and properties according to the Particle Data Group

The only known interaction that is missing in the SM framework is the gravitational force. This is because it is challenging to fit in a single theoretical framework general relativity and quantum mechanics. As of today, their theoretical unification is an active front of research in the theory community. The lack of gravity in the SM is not a problem for the LHC phenomenology, and the reason for this is that at the distances and energies probed in particle accelerators, the gravitational force is too weak compared to the other forces and is neglected.

This appendix is not a direct result of my research and it rather contains a short introduction to the Standard Model of Particle Physics. The way the concepts are introduced is heavily inspired by lecture notes from [127], [128] and personal notes from Summer Schools I attended during my formation. The appendix focuses on the presentation of the fundamental interactions in the SM, the problem of the mass in the SM, and its solution through the spontaneous symmetry breaking of the SM symmetries and the Higgs mechanism.

A.1 Local gauge invariance

The Standard Model is a gauge theory of quantum fields and in its mathematical formulation the Lagrangian density, $\mathcal{L}(\phi_i, \partial_\mu \phi_i)$, is a Lorentz-invariant function of the quantum fields, ϕ_i, and its space-time derivatives $\partial_\mu \phi_i$. The fundamental interactions appear in the theory when local gauge invariance is required in the free Lagrangian of the physical system.

This way, $\mathcal{L}(\phi_i, \partial_\mu \phi_i)$, is said to be local gauge invariant when the physical system is invariant under local $\phi\,' \rightarrow U\phi$ transformations of a certain type. In the latter, the transformation matrix, U, belongs to the group $SU(N)$ of $N \times N$ unitary matrices with $\mathrm{Det}(U) = 1$. An important property is that any $SU(N)$ matrix that defines the local gauge transformation, U, can be re-written as:

$$U = e^{iT^a \theta_a(x)} \tag{A.1}$$

Where T^a are $a = 1, ..., N^2-1$ Hermitian traceless matrices that are called the generators of the $SU(N)$ group. Each pair of generators of the group satisfy, $[T^a, T^b] = if^{abc}T^c$, where f^{abc} are real numbers that are called the structure constants the group. The $\theta_a(x)$ are arbitrary parameters of the transformation and depend on the space-time coordinate x. A sum over a repeated upper and lower index is assumed in this appendix.

The complete Standard Model Lagrangian is invariant under local $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ transformations. The fundamental interactions between different particles in the Standard Model Lagrangian, $\mathcal{L}_{\mathcal{SM}}$, are contained into two main pieces. The strong sector (QCD), $\mathcal{L}_{\mathcal{QCD}}$, and the electroweak sector, $\mathcal{L}_{\mathcal{EW}}$, which unifies electromagnetism (QED) and weak interactions.

In the following sections, local gauge $U(1)$, $SU(3)$ and $SU(2)_L \otimes U(1)_Y$ invariance is required separately under a free Lagrangian. These requirements lead to the $\mathcal{L}_{\mathcal{QCD}}$, $\mathcal{L}_{\mathcal{QED}}$ and $\mathcal{L}_{\mathcal{EW}}$ Lagrangians respectively. The aim of this appendix is to show different examples of how the gauge principle is able to generate different interactions via the exchange of spin one bosons.

However it must be noted that the gauge principle alone is not enough to build the SM. This procedure generates interactions mediated only by massless bosons and experimentally it is well known since many years that the W and Z bosons are not massless. The discussion of the mass in the SM requires the inclusion of the Higgs mechanism and the concept of Spontaneous Symmetry Breaking that are presented separately.

A.2 Quantum Electrodynamics: QED

In this section the QED Lagrangian, $\mathcal{L}_{\mathcal{QED}}$, is derived by requiring invariance under a local $U(1)$ transformation to the massive free Dirac fermion Lagrangian:

$$\mathcal{L}_0 = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi \tag{A.2}$$

Under a local $SU(1)$ transformation $\psi \to \psi' = \psi e^{iQ\theta(x)}$, $\mathcal{L}_0$ is not invariant due to the presence of an extra term containing $\partial_\mu\theta$. However there is a possible solution if one adds an extra term to $\mathcal{L}_0$ transforming in such a way that cancels out the extra $\partial_\mu\theta$ term. This is done by introducing an extra spin-1 field to the free Lagrangian, A_μ, transforming as

$$A_\mu \to A'_\mu = A_\mu - \frac{1}{e}\partial_\mu\theta \tag{A.3}$$

Once the new spin-1 field has been included, the new Lagrangian is now invariant under local $U(1)$ transformations:

$$\mathcal{L} = \mathcal{L}_0 - eQ\bar{\psi}\gamma^\mu A_\mu\psi \tag{A.4}$$

Local $U(1)$ gauge invariance required the introduction of new term $-eQ\bar{\psi}\gamma^\mu A_\mu\psi$. This term corresponds to a new interaction between a Dirac fermion and the new spin-1 field, A_μ. A mass term $\sim \frac{1}{2}m^2A^\mu A_\mu$ for the new gauge field is not allowed because violates gauge invariance, so the new spin 1 field has to be massless and is nothing else than a massless photon field of QED. Q is the gauge charge and it depends on the gauge quantum numbers assigned to the fields in the theory. The parameter e is called the electric charge and governs the size of the electromagnetic coupling.

In its current form, $\mathcal{L}$, does not include the propagating properties of the electromagnetic field in the Maxwell equations. Due to this, an additional gauge invariant kinetic term has to be added.

$$\mathcal{L}_{kin} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} \tag{A.5}$$

Where $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$ is called the field strength. Putting everything together, the QED Lagrangian reads:

$$\mathcal{L}_{\mathcal{QED}} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi - eQ\bar{\psi}\gamma^\mu A_\mu\psi \tag{A.6}$$

Finally Equation A.6 can by simplified by defining the covariant derivative of the field: $\gamma^\mu D_\mu\phi = \not{D}\phi = \gamma^\mu[\partial_\mu + ieQA_\mu]\phi$ and the simplified $\mathcal{L}_{\mathcal{QED}}$ reads as

$$\mathcal{L}_{\mathcal{QED}} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + i\bar{\psi}\gamma^\mu D_\mu\psi - m\bar{\psi}\psi \tag{A.7}$$

A.3 Quantum Cromodynamics: QCD

Since the discovery of the quark in 1968 and the gluon in 1979, we know that the proton is not an elementary particle and is made up of quarks and gluons. The posterior discovery of a mysterious new bound state made of three up quarks, Δ^{++}, forced the introduction of a new quantum number, the color quantum number, to avoid the violation of the Pauli principle. The interaction that binds quarks together to form hadrons like the proton is the so-called color or strong interaction, and it is described theoretically by Quantum Cromodynamics (QCD). Each quark exists in three possible different colours (red, green, blue) and forms a color triplet. Gluons are spin-1 bosons that mediate strong interactions and carry color-anticolor charge.

QCD has two important properties whose origin is still not well understood, quark confinement and asymptotic freedom. Confinement refers to the experimental fact that all observed hadrons are colourless, whereas each of its constituent quarks has color. As a consequence, free-quarks have never been observed experimentally and only colourless hadrons made of 2 or 3 quarks, denoted as mesons and barions respectively, exist in Nature. Asymptotic freedom refers to the property of the strength of strong interactions that, in contrast to QED, decreases when the energy scale of the interaction, Q^2, increases. In other words, for large momentum transfer or small distances like at the LHC, the quarks inside the proton behave almost as free particles.

The QCD Lagrangian can be derived requiring local gauge invariance, as in the QED case, but in this case the invariance is under a larger gauge group, $SU(3)$. In contrast to the $U(1)$ group, the generators of $SU(3)$ are eight matrices, λ_a, called Gell-Mann matrices that do not commute with each other. Due to this, QCD is also referred as a non-Abelian theory.

Let's consider the free Lagrangian for a fermionic triplet $q_f = (q_f^r, q_f^g, q_f^b)$, where the superscript indicates the color (red, green, blue) and f is the quark flavour (u,d,c,s,t,b).

$$\mathcal{L}_0 = i \sum_f \bar{\psi}_f (\gamma^\mu \partial_\mu - m_f \bar{\psi}_f) \psi_f \tag{A.8}$$

Under a local $SU(3)$ transformation $\psi \to \psi' = \psi e^{i \frac{\lambda^a}{2} \theta(x)_a}$, $\mathcal{L}_0$ is not invariant similarly to the QED case. The gauge invariance is guaranteed if eight different spin-1 gauge bosons, G_a^μ identified as the gluons, are added to the free Lagrangian with the following transformation properties:

$$G_a^\mu \to G_a^\mu - \frac{1}{g_s} \partial^\mu \theta_a - f^{abc} \theta_b G_c^\mu \tag{A.9}$$

Where g_s is the strength of the strong interactions and f^{abc} in the last term of the transformation are the $SU(3)$ structure constants.

$$[\lambda^a, \lambda^b] = i f^{abc} \lambda^c \tag{A.10}$$

Once the gluon fields with the proper transformation rule are introduced, $\mathcal{L}_0$ becomes gauge invariant,

$$\mathcal{L} = \mathcal{L}_0 - g_s \sum_f \bar{\psi}_f \gamma_\mu G_a^\mu \frac{\lambda^a}{2} \psi_f \tag{A.11}$$

In its current form, the Lagrangian does not include a kinetic term and hence cannot describe a propagating field. Due to this, an additional kinematic term that respects the transformation rules of the gauge field has to be added

$$\mathcal{L}_{kin} = -\frac{1}{4} G_a^{\mu\nu} G^a_{\mu\nu}. \tag{A.12}$$

Where $G_a^{\mu\nu}$ is called the field strength, and has the following gauge invariant form

$$G_a^{\mu\nu} = \partial^\mu G_a^\nu - \partial^\nu G_a^\mu - g_s f^{abc} G_b^\mu G_c^\nu. \tag{A.13}$$

The non-commutativity of the generators λ^a gives rise to additional terms in the kinetic Lagrangian involving the gluon fields themselves. Putting everything together, the final QCD Lagrangian has the following form,

$$\mathcal{L}_{\mathcal{QCD}} = -\frac{1}{4} G_a^{\mu\nu} G^a_{\mu\nu} + i \sum_f \bar{\psi}_f (\gamma^\mu \partial_\mu - m_f \bar{\psi}_f) \psi_f - g_s \sum_f \bar{\psi}_f \gamma_\mu G_a^\mu \frac{\lambda^a}{2} \psi_f \tag{A.14}$$

and if one decomposes the Lagrangian into its different pieces, it is explicitly visible the richer QCD phenomenology compared to QED,

$$\begin{aligned} \mathcal{L}_{\mathcal{QCD}} \quad = \quad & -\frac{1}{4} (\partial^\mu G_a^\nu - \partial^\nu G_a^\mu)(\partial_\mu G^a_\nu - \partial_\nu G^a_\mu) + i \sum_f \bar{\psi}_f (\gamma^\mu \partial_\mu - m_f \bar{\psi}_f) \psi_f \\ & - g_s \sum_f \bar{\psi}_f \gamma_\mu G_a^\mu \frac{\lambda^a}{2} \psi_f \\ & + \frac{g_s}{2} f^{abc} (\partial^\mu G_a^\nu - \partial^\nu G_a^\mu) G_\mu^b G_\nu^c - \frac{g_s^2}{4} f^{abc} f_{ade} G_b^\mu G_c^\nu G_\mu^d G_\nu^e \end{aligned}$$

The non-Abelian nature of the generators of the $SU(3)$ matrices gives rise to cubic and quartic self-interactions of the gluon field encoded in the last line of the expanded QCD Lagrangian. As is the QED case, the $SU(3)$ symmetry forbids a mass term of the form $\frac{1}{2} m_G^2 G_a^\mu G_\mu^a$, and the gluons have to be massless. Finally the presence of self-interactions is a new feature with respect to QED, and one would expect them to explain some of the particular properties of the QCD like asymptotic freedom and confinement.

A.4 Electroweak unification

Weinberg-Salam-Glashow in the 60's proposed a unified description of two of the four known fundamental forces in Nature: electromagnetism and weak force, the latter being responsible for beta decays. Despite the differences between both interactions at low energies, they proposed that above certain energy scale, both interactions should manifest as an unique force, called electroweak interaction (EW). One of the main predictions of the EW theory is the existence of the $W^{\pm}$ and Z bosons. Their discovery in 1983 by the UA1 and UA2 collaborations, confirmed and established the electroweak theory as one of the pillars in the SM.

Prior to the W and Z discovery, it was known experimentally some of the unique properties of weak interactions, not present in QED and QCD, that the unified model had to address. For example, in 1957 the Wu experiment observed a 100% parity violation in the angular distribution of β decays of the type $n \to pe^{-}\nu_e$. Since then, the parity violation in weak decays has been confirmed in many experiments, however, its origin in Nature is still not well understood. Today we know experimentally that only left-handed fermions[1] and right handed anti-fermions participate in weak interactions.

In the electroweak formulation, left handed fermions, L, transform as $SU(2)_L \otimes U(1)_Y$ doublets and right handed fermions, R, as $U(1)_Y$ singlets. This choice is motivated by the fact that we have never observed lepton flavour violating processes in leptonic weak decays. Due to this, a doublet representation that turns left handed electrons into left handed neutrinos of the same $SU(2)$ doublet is appropriate. Additionally, in order to reproduce observed parity violation in weak decays, left and right handed fermions must have different weak isospin $I = \frac{1}{2}$ and $I = 0$ respectively. The third component of the weak isospin, I_3, is conserved in weak interactions and has value of $\sigma_3/2 = \pm 1/2$ or $\sigma_3 = 0$ depending on the fermion.

In local $SU(2)$ gauge transformations, the generators, σ_i, are the Pauli matrices, and in the case of local $U(1)_Y$ transformation, Y, is identified with the weak hypercharge. The physical interpretation of the weak hypercharge and its relation with the weak isospin and the electromagnetic charge in QED, will be postponed to next section. In summary all the needed ingredients with the corresponding transformations are the following:

$$L = \begin{pmatrix} e^- \\ \nu_e \end{pmatrix}_L, \begin{pmatrix} u \\ d' \end{pmatrix}_L \ldots \qquad L \to L' = e^{i(Y\beta(x) + \frac{\sigma_i}{2}\alpha_i(x))}L \qquad \begin{cases} I_3 = \frac{1}{2}\,,\, u\,,\, \nu_e \ldots \\ {}_3 = -\frac{1}{2}\,,\, e^-\,,\, d \ldots \end{cases}$$

$$R = e^-_R, u_R, d_R, \ldots \qquad R \to R' = e^{iY\beta(x)}R \qquad I = I_3 = 0$$

[1] Left and right handed fermions are defined as the projections of Dirac fermions $\psi_{L,R} = \frac{1 \mp \gamma^5}{2}\psi$

In the quark sector, the situation is slightly different since we have observed experimentally mixing between flavours and CP violation in kaon decays. This observation can be encoded in the EW theory if the weak eigenstate d', s', b' are not the same as the mass eigenstates d, s, b. Their relation is given by the V_{CKM} (Cabibbo-Kobayashi-Maskawa) unitary matrix, and its elements have to be measured experimentally:

$$\begin{bmatrix} d' \\ s' \\ c' \end{bmatrix} = \begin{bmatrix} 0.97427 \pm 0.00014 & 0.22536 \pm 0.00061 & 0.00355 \pm 0.00015 \\ 0.22522 \pm 0.00061 & 0.97343 \pm 0.00015 & 0.0414 \pm 0.0012 \\ 0.00886 \pm 0.00033 & 0.0404 \pm 0.0012 & 0.99914 \pm 0.00005 \end{bmatrix} \begin{bmatrix} d \\ s \\ c \end{bmatrix} \tag{A.15}$$

The $\mathcal{L}_{EW}$ is obtained requiring local $SU(2)_L \otimes U(1)_Y$ invariance into the free Dirac Lagrangian, $\mathcal{L}_0$, similarly to the QED and QCD cases. As in the other examples, this requirement turns partial derivatives into covariant derivatives, and introduces as many new fields as generators of the symmetry group:

$$D_\mu L = [\partial_\mu + ig\tilde{W}_\mu + ig'YB_\mu]L \tag{A.16}$$

$$D_\mu R = [\partial_\mu + ig'YB_\mu]R \tag{A.17}$$

where $\tilde{W}_\mu$ is defined as

$$\tilde{W}_\mu = \frac{\sigma^i}{2}W_\mu^i = \frac{1}{2}\begin{bmatrix} W_\mu^3 & W_\mu^1 - iW_\mu^2 \\ W_\mu^1 + iW_\mu^2 & -W_\mu^3 \end{bmatrix} \tag{A.18}$$

The W_μ^i where $i = 1, 2, 3$ are three new spin-1 massless fields that are introduced to guarantee local gauge $SU(2)_L$ invariance. The field B_μ is another massless field introduced to preserve the $U(1)_Y$ local symmetry. The parameters g and g' are the coupling strengths of the new W and B interactions respectively.

As in the QCD and QED example, in order to allow the propagation of the W and B fields, a gauge invariant kinetic term is also needed. This way $\mathcal{L}_{\mathcal{EW}}$ can be written as:

$$\mathcal{L}_{\mathcal{EW}} = -\frac{1}{4}W_a^{\mu\nu}W_{\mu\nu}^a - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} + \sum_{\psi_i = L,R} i\bar{L}_i \not{D} L_i + i\bar{R}_i \not{D} R_i \tag{A.19}$$

Where the field strengths, $B_{\mu\nu}$ and $W_{\mu\nu}^i$, are introduced in a gauge invariant way

$$B_{\mu\nu} = \partial\mu B_\nu - \partial\nu B_\mu \tag{A.20}$$

$$W_{\mu\nu}^i = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a - g\epsilon^{ijk}W_\mu^j W_\nu^k \tag{A.21}$$

Since the Pauli matrices do not commute $[\sigma^a, \sigma^b] = i\epsilon^{abc}\sigma^c$, the kinetic term, $-\frac{1}{4}W_a^{\mu\nu}W_{\mu\nu}^a$, will raise to cubic and quartic self interactions involving the W_μ^i fields themselves.

However it is important to notice that, in contrast to QCD and QED, the gauge symmetry forbids writing a mass term for the fermionic fields. The reason for this is that such a term would mix left handed and right handed chiralities. The latter is not possible in the EW theory since left handed and right handed fermions have different transformation properties. Additionally, a mass term for the gauge fields is also forbidden because such term would also violate local gauge invariance.

Neutral currents and mixing of SU(2) and U(1)

The next step in the electroweak unification is to identify the new gauge fields, W^3 and B, with the known physical Z and the photon γ. Experimentally we know that neutrinos do not couple with the photon field, however the term $i\bar{L}_i \not{D} L_i + i\bar{R}_i \not{D} R_i$ in the $\mathcal{L}_{EW}$ in Equation A.19 presents such coupling. As a consequence, this coupling is problematic and breaks the naïve identification of the B_μ field with the photon field, A_μ. Nonetheless, since both fields are neutral, one can try a linear combination of them:

$$\begin{bmatrix} W^3_\mu \\ B_\mu \end{bmatrix} = \begin{bmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{bmatrix} \begin{bmatrix} Z_\mu \\ A_\mu \end{bmatrix} \tag{A.22}$$

where θ_W is a free parameter in the theory, called the weak mixing angle. Re-writing the relevant kinetic terms of the neutral-current $\mathcal{L}_{EW}$ Lagrangian in terms of the mixed fields Z_μ and A_μ leads to:

$$\mathcal{L}_{NC} = \sum_{\psi_i = L,R} \bar{\psi}_i \gamma^\mu \{A_\mu [g \frac{\sigma_3}{2} \sin\theta_W + g' Y \cos\theta_W] + Z_\mu [g \frac{\sigma_3}{2} \cos\theta_W - g' Y \sin\theta_W]\} \psi_i \tag{A.23}$$

In order to identify A_μ in Equation A.23 as the electromagnetic field from QED in Equation A.6, the following two conditions need to be satisfied simultaneously:

$$g \sin\theta_W = g' \cos\theta_W = eQ \qquad Y = Q - I_3 \tag{A.24}$$

Where now the identification of $I_3 \equiv \frac{\sigma_3}{2}$ becomes manifest and the electric charge, Q, is naturally defined from the weak hypercharge. The relation, $Y = Q - I_3$, is called the Nishijima-Gell-Gann's formula, and allows to compute the weak hypercharge of the different fermions:

$$\begin{aligned} Y(\nu_{e_L}) = Y(e_L) = -\frac{1}{2} \quad Y(\nu_{e_R}) = 0 \quad Y(e_R) = -1 \\ Y(u_L) = Y(d_L) = \frac{1}{6} \quad Y(u_R) = 2/3 \quad Y(d_R) = -1/3 \end{aligned}$$

The θ_W angle is a free parameter in the theory that can be measured with high precision from weak decay rates. Its precise value will have important consequences that can be tested experimentally. For example, its value will fix the relation between the mass of

the physical W and Z bosons in the context of the Higgs mechanism.

$$\sin\theta_W = \frac{g'}{\sqrt{g^2+g'^2}} = 0.215. \tag{A.25}$$

One of the main predictions of the EW theory that has been verified in high energy colliders with high precision, is that the fermionic coupling of the Z_μ shown in Equation A.23, is not universal and is given by $\frac{g}{\cos\theta_W}(I_3 - \sin\theta_W Q)$.

Finally neutrinos in the SM have exactly zero mass and a possible right-handed neutrino, ν_R, is not included in the EW theory because it would not interact with any of the fundamental forces, thus, resulting undetectable.

Charged Currents: W boson in the SM model

The interaction of the W bosons with leptons is encoded in the term $\sum_{\psi_i=L,R} i\bar{L}_i \not{D} L_i$ in Equation A.19. Motivated[2] by Equation A.18, it is possible to define the physical W bosons as $(W^1 + iW^2)/\sqrt{2}$ and write down all the interactions of the fermions with the W field. As an illustrative example and for its simplicity, lets consider the coupling of the W^+ only with the second generation of fermions: $L = \begin{pmatrix}\mu^- \\ \nu_\mu\end{pmatrix}_L$ and $\begin{pmatrix}c \\ s'\end{pmatrix}_L$:

$$\mathcal{L}_{CC}^{2^{nd}Gen} = \frac{ig}{\sqrt{2}}\{\nu_{\mu_L}\not{W}^+\mu_L + \bar{c_L}\not{W}^+ V_{CKM_{21}} d_L + \bar{c_L}\not{W}^+ V_{CKM_{22}} s_L + \bar{c_L}\not{W}^+ V_{CKM_{23}} b_L + h.c\} \tag{A.26}$$

Charged current interactions that involve leptons, $W \to l\nu$, are universal and diagonal with a unique coupling strength g. Table A.1 shows the branching ratio of the W boson as predicted by the theory compared to the measurements.

Parameter	Experiment	Theory
BR($W \to l^+\nu$)	(32.58± 0.27)%	$\approx 3/9 = 0.33$
BR($W \to e^+\nu_e$)	(10.71 ± 0.16)%	$\approx 1/9 = 0.11$
BR($W \to \mu^+\nu_\mu$)	(10.63 ± 0.15)%	$\approx 1/9 = 0.11$
BR($W \to \tau^+\nu_\tau$)	(11.38 ± 0.21)%	$\approx 1/9 = 0.11$
BR($W \to Hadron$)	(67.41 ± 0.27)%	$\approx 2{\cdot}3/9 = 0.66$
BR($W \to cX$)	(33.3 ± 2.3)%	$\approx 3/9 = 0.33$

TABLE A.1: Branching ratio of the SM W boson into different fermions.

In the quark sector, $W \to q_i q_j$, the weak interaction basis is different than the mass basis and breaks the universality. This way flavour changing weak decays are allowed and are not diagonal. The coupling strength depends the CKM matrix element $gV_{CKM_{ij}}$ in Equation A.15.

[2] In next section when discussing the Higgs mechanism, this definition of the W appears naturally when trying to identify the W with the physical boson.

A.5 Spontaneous symmetry breaking and Higgs mechanism

As already described in the previous section, the electroweak SM is invariant under local $SU(2)_L \otimes U(1)_Y$ gauge transformations. It was also noted that, in its original formulation, the $\mathcal{L}_{\mathcal{EW}}$ Lagrangian only contains massless W^1, W^2, W^3, B bosons and fermions. However it is known experimentally since many years that, for example, the physical W and Z bosons have a non-zero mass whereas only the photon is massless [74]:

$$M_W = 80.385 \pm 0.015 \text{ GeV} \qquad M_Z = 91.1876 \pm 0.0021 \text{ GeV} \qquad M_\gamma < 10^{-18} \text{ eV} \quad \text{(A.27)}$$

In order to solve this puzzling situation, a new important ingredient needs to be added to the SM. A complex scalar doublet field, denoted by ϕ,

$$\phi = \begin{bmatrix} \phi^+ \\ \phi^0 \end{bmatrix}. \quad \text{(A.28)}$$

Where ϕ^+ and ϕ^0 are two complex fields and together form a isospin doublet with quantum numbers $I_3 = \pm 1/2$ and $Y = 1/2$ respectively. This particular choice of the quantum numbers is necessary to keep the photon massless.

This new scalar field enters in the $\mathcal{L}_{\mathcal{EW}}$ Lagrangian as a new term, $\mathcal{L}_{\text{Higgs}}$, that is also $SU(2)_L \otimes U(1)_Y$ gauge invariant:

$$\mathcal{L}_{\text{Higgs}} = (D^\mu \phi)^\dagger (D_\mu \phi) - V(\phi) = (D^\mu \phi)^\dagger (D_\mu \phi) - \mu^2 (\phi^\dagger \phi) - \lambda (\phi^\dagger \phi)^2 \quad \text{(A.29)}$$

Additionally, it is also possible to a include Yukawa interaction term, $\mathcal{L}_{\text{Yukawa}}$, between the new scalar field and the fermions with strength Y_{ij},

$$\mathcal{L}_{\text{Yukawa}} = Y_{ij} \left[\bar{L}_i \phi R_j + h.c \right] \quad \text{(A.30)}$$

Firstly it is important to focus on the particular structure of the minimum of the Higgs potential, $V(\phi)$, in Equation A.29. If the self interaction is negative $\mu^2 < 0$ and the quartic coupling $\lambda > 0$ positive, there is a degenerate set of states with non-zero minimum energy[3],

$$\langle 0 | \phi^0 | 0 \rangle = \sqrt{\frac{-\mu^2}{2\lambda}} = \frac{v}{\sqrt{2}}. \quad \text{(A.31)}$$

Where $v = \sqrt{-\mu^2/\lambda} \approx 246$ GeV is denoted as the Higgs vacuum expectation value (VEV) and can be determined from the muon lifetime measurement. This particular

[3] These conditions are imposed in the potential in order to have a non-zero VEV in the Higgs potential. Their particular values cannot be obtained from first principles.

value is a key parameter in the theory because sets the energy scale, also known as electroweak energy scale, in which the Higgs mechanism played a role in the evolution of the early universe.

Keeping this energy scale in mind, it is interesting to distinguish two eras in the evolution of the universe: $T(t) >> v$ and $T(t) << v$, where $T(t)$ refers to the temperature of the universe at some point in time, t. At the very early universe, at high enough energy and temperatures $T >> v$, the SM is invariant under $SU(2)_L \otimes U(1)_Y$ transformations and $V(\phi)$ has a degenerate set of states with minimum energy $\langle\phi\rangle = 0$. However as the universe cools down to energies below the electroweak energy scale, $T << v$, the symmetries in the SM will spontaneously break, and the Higgs potential, $V(\phi)$, will acquire a non-zero vacuum expectation value among the degenerate set of states with minimum energy in Equation A.31. A this point, the minimum of the Higgs potential is not longer around $\langle\phi\rangle = 0$.

With this in mind, and due to the $SU(2)$ invariance of the Lagrangian, it is possible to rewrite the scalar field in terms of a shifted and re-parametrized potential, $H = \phi - v/\sqrt{2}$, that is also known as the Higgs field

$$H = \frac{1}{\sqrt{2}} e^{i\frac{(\sigma_i \theta_i)}{2}} \begin{bmatrix} 0 \\ \phi - v \end{bmatrix} \tag{A.32}$$

This freedom in the choice of the Higgs potential is caused by the presence of three non-physical massless Goldstone bosons, θ_i, that have never been observed in nature. The problem of their existence can be solved if the original $SU(2)_L \otimes U_Y$ symmetry of the Lagrangian, is also broken with the choice of a particular unitary gauge ($\theta_i = 0$). This choice will remove the non-physical massless Goldstone bosons, break the $SU(2)_L \otimes U_Y$ symmetry, and leave the SM Lagrangian invariant under the unbroken $U(1)_{EM}$ symmetry.

As a consequence of the Spontaneous Symmetry Breaking (SSB) of the $SU(2)_L \otimes U_Y$ symmetry and the Higgs mechanism, the W and Z bosons acquire mass and interactions with the Higgs field trough the kinetic term in $\mathcal{L}_{\mathcal{EW}}$ (Equation A.29):

$$D^\mu \phi D_\mu \phi \xrightarrow{SSB} ... + (v + H)^2 \big[\frac{g^2}{4} W_\mu^\dagger W^\mu + \frac{g^2}{8 cos^2 \theta_W} Z_\mu Z^\mu\big] + ... \tag{A.33}$$

Where the W field in the above expression corresponds to $(W^1 + iW^2)/\sqrt{2}$ and is the physical W boson. The masses of the W and Z bosons are identified with the terms $\frac{1}{2} M_W W_\mu^\dagger W^\mu$ and $\frac{1}{2} M_Z Z_\mu Z^\mu$ in Equation A.33 respectively. From their mass relation,

one gets a very important prediction that can be tested in particle accelerators

$$\cos\theta_W = M_W/M_Z \tag{A.34}$$

where θ_W is again the weak mixing angle that appeared in Equation A.25 relating the g and g' couplings. This important relation fixes the mass of the W boson with the mass of the Z boson, and has been confirmed experimentally to be true in measurements in particle accelerators with high precision. This confirmation gave a strong experimental support to the presence of SSB mechanism due to a scalar field to explain the masses of the weak bosons in the SM.

Nevertheless, the final and definitive experimental confirmation of this mechanism, requires the discovery of a new scalar boson in the theory, the Higgs boson. The latter acquires its mass trough his self interactions in the Higgs potential

$$V(\phi) \xrightarrow{SSB} ... - \frac{1}{2}\sqrt{-2\mu^2}H^2 - \frac{1}{2v}\sqrt{-2\mu^2}H^3 - \frac{1}{8v^3}\sqrt{-2\mu^2}H^4 + ... \tag{A.35}$$

Unfortunately for experimental physicist the Higgs boson mass, M_H, is not predicted in the SM as long as λ is a free parameter in the theory

$$M_H = \sqrt{2\lambda}v. \tag{A.36}$$

After decades of searches in previous particle accelerators such as LEP or Tevatron, the Higgs boson was finally discovered in 2012 by the CMS and ATLAS experiments at the LHC. They measured the Higgs boson mass to be:

$$M_H = 125.7 \pm 0.4 \text{ GeV} \tag{A.37}$$

This discovery provides the final experimental evidence of the existence of SSB breaking in the SM, and confirms the Higgs mechanism as our best explanation of the mass of the W and Z bosons.

Finally the Higgs boson also gives mass to the fermions via their Yukawa couplings to the Higgs field. Rewriting Equation A.30 in terms of the Higgs field after SSB, one obtains:

$$\mathcal{L}_{\text{Yukawa}} \xrightarrow{SSB} -\sum_{i,j} Y_{ij}(\frac{v+H}{\sqrt{2}})\bar{f}_i f_j \tag{A.38}$$

Choosing the Yukawa matrix to be diagonal, the masses of the fermions can be written as a function of the Yukawa couplings y_i of the Higgs field with the fermions:

$$m_e = \frac{y_e v}{\sqrt{2}} \qquad m_u = \frac{y_u v}{\sqrt{2}} \qquad ... \tag{A.39}$$

This way the heavier the mass of the fermion, the larger its Yukawa coupling with the Higgs field. The experimental evidence of the coupling of the recently discovered Higgs boson to the heaviest fermions (but lighter than the Higgs itself), $b\bar{b} + \tau\bar{\tau}$, confirmed these Yukawa couplings as our best explanation to the mass of the fermions.

Conclusion

The gauge principle allows, in a simple way, to introduce in the theory the different interactions we know in nature: strong, electromagnetic and weak interactions (excluding gravity) imposing only symmetry arguments. However as a result of this exercise, the built theory only contains interactions mediated by massless spin-1 fields. The latter is in clear contradiction with experimental measurements of the mass of the W and Z bosons.

A simple solution to this problem is to add a scalar field to the theory, and spontaneously break the original symmetry and generate masses for the different particles. This solution was confirmed experimentally in 2012 after the discovery of the Higgs boson. The latter also closes the search for the SM particles, since all of the have been detected experimentally.

In the process of building the SM, it was needed to introduce 'only' 18 different parameters, whose value cannot be predicted by the theory. These parameters are: 3 masses for the leptons (leaving the neutrinos massless...), 6 masses for the quarks, 3-mixing angles and 1-CP violating phase in the CKM matrix, 3 gauge couplings, a vacuum expectation value and the Higgs mass itself.

Appendix B

Publications and technical reports

During the development of this thesis several publications as well as technical reports were written that further expand the content of the thesis. Particularly by publications I refer to CMS reviewed documents (either published or not) that are approved by the CMS Collaboration to be shown in international conferences. Technical reports refer to documents that described the step by step details of an analysis and their content is only available for members of the CMS Collaboration.

Publications:

- "Search for heavy gauge W′ boson in events with an energetic lepton and large missing transverse momentum at $\sqrt{s} = 13$ TeV ". December 2016. arXiv:1612.09274. Submitted to PLB.
- "Measurement of associated Z+charm production in pp collisions at 8 TeV ". August 2016. SMP-15-009.
- "Search for SSM W' production, in the lepton+MET final state at a center-of-mass energy of 13 TeV ". December 2015. EXO-15-005.
- "Performance of b-tagging at $\sqrt{s} = 8$ TeV in multijet, $t\bar{t}$ and boosted topology events". September 2013. PAS-BTV-13-001.

Technical reports:

- "Measurement of the high p_{T} muon momentum scale using the endpoint method". AN-15-228. December 2016.
- "New measurement of c-jet tagging efficiencies in W associated with charm at $\sqrt{s} = 13$ TeV ". AN-2016-162. August 2016.

- "Measurement of associated Z+charm production at 8 TeV ". AN-2014-235. August 2016.
- "Search for new physics in lepton+MET at 13 TeV ". AN-2015-226. December 2015.
- "Preparation for Run II : Search for new physics in lepton+MET at 13 TeV ". AN-2015-094. August 2015.
- "Measurement of c-jet tag efficiencies in $\mathrm{W} + \mathrm{c}$ at $\sqrt{s} = 8$ TeV". AN-2014-225. February 2016.
- "Combination of b-tag efficiency measurements for the 22-Jan-2013 dataset". AN-2013-213. September 2012.
- "B-tagging efficiencies with non-prompt J/ψ". AN-2013-154. September 2012.

Bibliography

[1] J. C. Street and E. C. Stevenson. *"New Evidence for the Existence of a Particle of Mass Intermediate Between the Proton and Electron"*. *Phys. Rev.*, 52:1003–1004, 1937. doi: 10.1103/PhysRev.52.1003.

[2] S. H. Neddermeyer and C. D. Anderson. *"Note on the Nature of Cosmic Ray Particles"*. *Phys. Rev.*, 51:884–886, 1937. doi: 10.1103/PhysRev.51.884.

[3] Hideki Yukawa. *"On the Interaction of Elementary Particles I"*. *Proc. Phys. Math. Soc. Jap.*, 17:48–57, 1935. doi: 10.1143/PTPS.1.1. [Prog. Theor. Phys. Suppl.1,1(1935)].

[4] Steven Weinberg. *"A Model of Leptons"*. *Phys. Rev. Lett.*, 19:1264–1266, 1967. doi: 10.1103/PhysRevLett.19.1264.

[5] S. L. Glashow. *"Partial Symmetries of Weak Interactions"*. *Nucl. Phys.*, 22: 579–588, 1961. doi: 10.1016/0029-5582(61)90469-2.

[6] Murray Gell-Mann. *"A Schematic Model of Baryons and Mesons"*. *Phys. Lett.*, 8:214–215, 1964. doi: 10.1016/S0031-9163(64)92001-3.

[7] Peter W. Higgs. *"Broken Symmetries and the Masses of Gauge Bosons"*. *Phys. Rev. Lett.*, 13:508–509, 1964. doi: 10.1103/PhysRevLett.13.508.

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

[9] Peter W. Higgs. *"Spontaneous Symmetry Breakdown without Massless Bosons"*. *Phys. Rev.*, 145:1156–1163, 1966. doi: 10.1103/PhysRev.145.1156.

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

[11] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. *"Global Conservation Laws and Massless Particles"*. *Phys. Rev. Lett.*, 13:585–587, 1964. doi: 10.1103/PhysRevLett.13.585.

[12] T. W. B. Kibble. *"Symmetry breaking in non-Abelian gauge theories"*. *Phys. Rev.*, 155:1554–1561, 1967. doi: 10.1103/PhysRev.155.1554.

[13] UA1 Collaboration. *"Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s}$ = 540 GeV"*. *Phys. Lett.*, B122:103–116, 1983. doi: 10.1016/0370-2693(83)91177-2. [611(1983)].

[14] UA1 Collaboration. *"Experimental Observation of Lepton Pairs of Invariant Mass Around* 95 *GeV/c^2 at the CERN SPS Collider"*. *Phys. Lett.*, B126:398–410, 1983. doi: 10.1016/0370-2693(83)90188-0.

[15] CDF Colalboration. *"Observation of top quark production in $\bar{p}p$ collisions"*. *Phys. Rev. Lett.*, 74:2626–2631, 1995. doi: 10.1103/PhysRevLett.74.2626.

[16] D0 Collaboration. *"Observation of the top quark"*. *Phys. Rev. Lett.*, 74:2632–2637, 1995. doi: 10.1103/PhysRevLett.74.2632.

[17] DONUT Collaboration. *"Observation of tau neutrino interactions"*. *Phys. Lett.*, B504:218–224, 2001. doi: 10.1016/S0370-2693(01)00307-0. URL http://adsabs.harvard.edu/abs/2001PhLB..504..218D.

[18] CMS Collaboration. *"Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC"*. *Phys. Lett.*, B716:30–61, 2012. doi: 10.1016/j.physletb.2012.08.021. URL http://cds.cern.ch/record/1471016.

[19] ATLAS Collaboration. *"Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC"*. *Phys. Lett.*, B716:1–29, 2012. doi: 10.1016/j.physletb.2012.08.020. URL http://cds.cern.ch/record/1471031.

[20] CMS Collaboration. *"Search for SSM W' production, in the lepton+MET final state at a center-of-mass energy of 13 TeV"*. Technical Report CMS-PAS-EXO-15-006, CERN, Geneva, 2015. URL https://cds.cern.ch/record/2114864.

[21] CMS Collaboration. *"Search for heavy gauge W' bosons in events with an energetic lepton and large missing transverse momentum at $\sqrt{s}$ = 13 TeV"*. Technical Report CERN-EP-2016-281. CMS-EXO-15-006-003, CERN, Geneva, Dec 2016. URL http://cds.cern.ch/record/2239955. Submitted to Phys. Lett. B.

[22] CMS Collaboration. *"Measurement of associated Z + charm production in pp collisions at $\sqrt{s}$= 8 TeV"*. Technical Report CMS-PAS-SMP-15-009, CERN, Geneva, 2016. URL https://cds.cern.ch/record/2202823.

[23] Lyndon R Evans and Philip Bryant. *"LHC Machine"*. *J. Instrum.*, 3:S08001. 164 p, 2008. URL http://cds.cern.ch/record/1129806. This report is an abridged version of the LHC Design Report (CERN-2004-003).

[24] *LHC Programme Coordination web pages*. https://lpc.web.cern.ch/lumi2.html.

[25] CMS Collaboration. *"Search for new physics in final states with a lepton and missing transverse energy in pp collisions at the LHC"*. *Phys. Rev. D*, 87 (arXiv:1302.2812. CMS-EXO-12-010. CERN-PH-EP-2013-010):072005. 27 p, Feb 2013. URL http://cds.cern.ch/record/1514623.

[26] ATLAS Collaboration. *"Search for new particles in events with one lepton and missing transverse momentum in pp collisions at $\sqrt{s}$ = 8 TeV with the ATLAS detector"*. *JHEP*, 09:037, 2014. doi: 10.1007/JHEP09(2014)037. URL http://cds.cern.ch/record/1746306.

[27] Guido Altarelli and G. Parisi. *"Asymptotic Freedom in Parton Language"*. *Nuclear Physics B*, 126(2):298 – 318, 1977. ISSN 0550-3213. doi: http://dx.doi.org/10.1016/0550-3213(77)90384-4. URL http://www.sciencedirect.com/science/article/pii/0550321377903844.

[28] Richard D Ball, Valerio Bertone, Stefano Carrazza, Christopher S Deans, Luigi Del Debbio, Stefano Forte, Alberto Guffanti, Nathan P Hartland, Jose I Latorre, Juan Rojo, and Maria Ubiali. *"Parton distributions for the LHC Run II"*. *J. High Energy Phys.*, 04(arXiv:1410.8849. EDINBURGH 2014-15. IFUM-1034-FT. CERN-PH-TH-2013-253. OUTP-14-11P. CAVENDISH-HEP-14-11):040. 138 p, Oct 2014. URL http://cds.cern.ch/record/1966481.

[29] Jun Gao, Marco Guzzi, Joey Huston, Hung-Liang Lai, Zhao Li, et al. *"The CT10 NNLO Global Analysis of QCD"*. *Phys.Rev.*, D89:033009, 2014. doi: 10.1103/PhysRevD.89.033009. URL http://adsabs.harvard.edu/abs/arXiv:1302.6246.

[30] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt. *"Parton distributions for the LHC"*. *Eur. Phys. J. C*, 63:189, 2009. doi: 10.1140/epjc/s10052-009-1072-5. URL http://adsabs.harvard.edu/abs/2009EPJC...63..189M.

[31] *MSTW PDF*. https://mstwpdf.hepforge.org/.

[32] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. *"A brief introduction to PYTHIA 8.1"*. *Comp. Phys. Comm.*, 178:852, 2008. doi: 10.1016/j.cpc.2008.01.036. URL http://cds.cern.ch/record/1064095.

[33] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Z. Skands. *"PYTHIA 6.4 Physics and Manual"*. *JHEP*, 05:026, 2006. doi: 10.1088/1126-6708/2006/05/026. URL http://adsabs.harvard.edu/abs/2006JHEP...05..026S.

[34] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. *"NLO vector-boson production matched with shower in POWHEG"*. *JHEP*, 0807:060, 2008. doi: 10.1088/1126-6708/2008/07/060. URL http://adsabs.harvard.edu/abs/2008JHEP...07..060A.

[35] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. *"MadGraph 5: Going Beyond"*. *JHEP*, 1106:128, 2011. doi: 10.1007/JHEP06(2011)128. URL http://adsabs.harvard.edu/abs/2011JHEP...06..128A.

[36] Ryan Gavin, Ye Li, Frank Petriello, and Seth Quackenbush. *"W Physics at the LHC with FEWZ 2.1"*. *Comput.Phys.Commun.*, 184:208–214, 2013. doi: 10.1016/j.cpc.2012.09.005. URL http://adsabs.harvard.edu/abs/arXiv:1201.5896.

[37] John M. Campbell and R. K. Ellis. *"MCFM for the Tevatron and the LHC"*. *Nucl. Phys. Proc. Suppl.*, 205-206:10–15, 2010. doi: 10.1016/j.nuclphysbps.2010.08.011. URL http://adsabs.harvard.edu/abs/arXiv:1007.3492.

[38] Törbjorn Sjöstrand. *"The Lund Monte Carlo for Jet Fragmentation"*. *Comput. Phys. Commun.*, 27:243, 1982. doi: 10.1016/0010-4655(82)90175-8. URL http://adsabs.harvard.edu/abs/1982CoPhC..27..243S.

[39] S. Agostinelli et al. *"GEANT4: A Simulation toolkit"*. *Nucl.Instrum.Meth.*, A506:250–303, 2003. doi: 10.1016/S0168-9002(03)01368-8. URL http://geant4.cern.ch/.

[40] CMS Collaboration. *"The CMS experiment at the CERN LHC"*. *JINST*, 3:S08004, 2008. doi: 10.1088/1748-0221/3/08/S08004. URL http://cds.cern.ch/record/1129810.

[41] Tai Sakuma and Thomas McCauley. *"Detector and event visualization with SketchUp at the CMS experiment"*. Technical Report CMS-CR-2013-379. arXiv:1311.4942, CERN, Geneva, Oct 2013. URL http://cds.cern.ch/record/1626816.

[42] CMS Collaboration. *"Description and performance of track and primary-vertex reconstruction with the CMS tracker"*. *J. Instrum.*, 9(arXiv:1405.6569. CERN-PH-EP-2014-070. CMS-TRK-11-001):P10009. 80 p, May 2014. URL https://cds.cern.ch/record/1704291.

[43] CMS Collaboration. *"CMS Tracking Performance Results from Early LHC Operation"*. *Eur. Phys. J. C*, 70(arXiv:1007.1988. CERN-PH-EP-2010-019. CMS-TRK-10-001):1165. 29 p, Jul 2010. URL https://cds.cern.ch/record/1277738.

[44] CMS Collaboration. *"Alignment of the CMS tracker with LHC and cosmic ray data"*. *JINST*, 9:P06009, 2014. doi: 10.1088/1748-0221/9/06/P06009. URL http://cds.cern.ch/record/1667597.

[45] CMS Collaboration. *"Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV"*. *J. Instrum.*, 10(arXiv:1502.02701. CERN-PH-EP-2015-004. CMS-EGM-13-001):P06005. 63 p, Feb 2015. URL http://cds.cern.ch/record/1988091.

[46] CMS Collaboration. *"Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8$ TeV."*. *J. Instrum.*, 10(arXiv:1411.0511. CMS-JME-13-003. CERN-PH-EP-2014-246):P02006, Nov 2014. URL http://cds.cern.ch/record/1966665.

[47] CMS Collaboration. *"Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector"*. Technical Report CMS-PAS-PFT-10-001, 2010. URL http://cds.cern.ch/record/1247373.

[48] CMS Collaboration. *"Performance of CMS Hadron Calorimeter Timing and Synchronization using Test Beam, Cosmic Ray, and LHC Beam Data"*. *JINST*, 5: T03013, 2010. doi: 10.1088/1748-0221/5/03/T03013. URL http://cds.cern.ch/record/1223869.

[49] CMS Collaboration. *"Performance of the CMS Hadron Calorimeter with Cosmic Ray Muons and LHC Beam Data"*. *JINST*, 5:T03012, 2010. doi: 10.1088/1748-0221/5/03/T03012. URL http://cds.cern.ch/record/1223950.

[50] CMS Collaboration. *"Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS"*. *J. Instrum.*, 6(arXiv:1107.4277. CMS-JME-10-011. CERN-PH-EP-2011-102):P11002. 67 p, Jul 2011. URL http://cds.cern.ch/record/1369486.

[51] CMS Collaboration. *"The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s} = 7$ TeV at the LHC"*. *J. Instrum.*, 8(arXiv:1306.6905. CMS-MUO-11-001. CERN-PH-EP-2013-072):P11002, Jun 2013. URL https://cds.cern.ch/record/1558674.

[52] CMS Collaboration. *"Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV"*. *JINST*, 7:P10002, 2012. doi: 10.1088/1748-0221/7/10/P10002. URL http://cds.cern.ch/record/1456510.

[53] CMS and ATLAS Collaboration. *"Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s}$ = 7 and 8 TeV"*. (CMS-PAS-HIG-15-002. ATLAS-CONF-2015-044), 2015. URL http://cds.cern.ch/record/2053103.

[54] CMS Collaboration. *"Measurements of properties of the Higgs boson and search for an additional resonance in the four-lepton final state at $\sqrt{s}$ = 13 TeV"*. Technical Report CMS-PAS-HIG-16-033, CERN, Geneva, 2016. URL https://cds.cern.ch/record/2204926.

[55] CMS Collaboration. *"Updated measurements of Higgs boson production in the diphoton decay channel at $\sqrt{s}$ = 13 TeV in pp collisions at CMS."*. Technical Report CMS-PAS-HIG-16-020, CERN, Geneva, 2016. URL https://cds.cern.ch/record/2205275.

[56] Super-Kamiokande Collaboration. *"Evidence for oscillation of atmospheric neutrinos"*. *Phys. Rev. Lett.*, 81:1562–1567, 1998. doi: 10.1103/PhysRevLett.81.1562.

[57] Rabindra N. Mohapatra and Goran Senjanovic. *"Neutrino Mass and Spontaneous Parity Violation"*. *Phys. Rev. Lett.*, 44:912, 1980. doi: 10.1103/PhysRevLett.44.912.

[58] A. D. Sakharov. *"Violation of CP Invariance, C Asymmetry, and Baryon Asymmetry of the Universe"*. *Pisma Zh. Eksp. Teor. Fiz.*, 5:32–35, 1967. doi: 10.1070/PU1991v034n05ABEH002497. URL http://www.jetpletters.ac.ru/ps/index-v-5_en.shtml. [Usp. Fiz. Nauk161,61(1991)].

[59] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay. *"Evidence for the 2 pi Decay of the k(2)0 Meson"*. *Phys. Rev. Lett.*, 13:138–140, 1964. doi: 10.1103/PhysRevLett.13.138.

[60] Bernard Aubert et al. *"Observation of CP violation in the B^0 meson system"*. *Phys. Rev. Lett.*, 87:091801, 2001. doi: 10.1103/PhysRevLett.87.091801. URL http://adsabs.harvard.edu/abs/2001PhRvL..87i1801A.

[61] Kazuo Abe et al. *"Observation of large CP violation in the neutral B meson system"*. *Phys. Rev. Lett.*, 87:091802, 2001. doi: 10.1103/PhysRevLett.87.091802. URL http://adsabs.harvard.edu/abs/2001PhRvL..87i1802A.

[62] Planck Collaboration. *"Planck 2015 results. XIII. Cosmological parameters"*. 2015. URL http://adsabs.harvard.edu/abs/arXiv:1502.01589.

[63] David B. Kaplan, Howard Georgi, and Savas Dimopoulos. *"Composite Higgs Scalars"*. *Phys. Lett.*, B136:187–190, 1984. doi: 10.1016/0370-2693(84)91178-X. URL http://adsabs.harvard.edu/abs/1984PhLB..136..187K.

[64] Nima Arkani-Hamed, Savas Dimopoulos, and G. R. Dvali. *"The Hierarchy problem and new dimensions at a millimeter"*. *Phys. Lett.*, B429:263–272, 1998. doi: 10.1016/S0370-2693(98)00466-3. URL http://adsabs.harvard.edu/abs/1998PhLB..429..263A.

[65] Marc Sher. *"Electroweak Higgs Potentials and Vacuum Stability"*. *Phys. Rept.*, 179:273–418, 1989. doi: 10.1016/0370-1573(89)90061-6.

[66] Giuseppe Degrassi, Stefano Di Vita, Joan Elias-Miro, Jose R. Espinosa, Gian F. Giudice, Gino Isidori, and Alessandro Strumia. *"Higgs mass and vacuum stability in the Standard Model at NNLO"*. *JHEP*, 08:098, 2012. doi: 10.1007/JHEP08(2012)098.

[67] LIGO Scientific and VIRGO Collaboration. *"Observation of Gravitational Waves from a Binary Black Hole Merger"*. *Phys. Rev. Lett.*, 116(6):061102, 2016. doi: 10.1103/PhysRevLett.116.061102. URL http://adsabs.harvard.edu/abs/2016PhRvL.116f1102A.

[68] Ken Hsieh, Kai Schmitz, Jiang-Hao Yu, and C.-P. Yuan. *"Global analysis of general $SU(2) \times SU(2) \times U(1)$ models with precision data"*. *Phys. Rev. D*, 82:035011, Aug 2010. doi: 10.1103/PhysRevD.82.035011. URL http://link.aps.org/doi/10.1103/PhysRevD.82.035011.

[69] Manuel Masip and Alex Pomarol. *"Effects of Standard Model Kaluza-Klein excitations on electroweak observables"*. *Phys. Rev. D*, 60:096005, Oct 1999. doi: 10.1103/PhysRevD.60.096005. URL http://link.aps.org/doi/10.1103/PhysRevD.60.096005.

[70] Chuan-Ren Chen, Mihoko M. Nojiri, Seong Chan Park, Jing Shu, and Michihisa Takeuchi. *"Dark matter and collider phenomenology of split-UED"*. *Journal of High Energy Physics*, 2009(09):078, 2009. URL http://stacks.iop.org/1126-6708/2009/i=09/a=078.

[71] Kyoungchul Kong, Seong Chan Park, and Thomas G. Rizzo. *"Collider phenomenology with Split-UED"*. *Journal of High Energy Physics*, 2010(4):81, 2010. doi: 10.1007/JHEP04(2010)081. URL http://dx.doi.org/10.1007/JHEP04%282010%29081.

[72] Kingman Cheung and Greg Landsberg. *"Kaluza-Klein states of the Standard Model gauge bosons: Constraints from high energy experiments"*. *Phys. Rev. D*, 65:

076003, Mar 2002. doi: 10.1103/PhysRevD.65.076003. URL http://link.aps.org/doi/10.1103/PhysRevD.65.076003.

[73] G. Altarelli, B. Mele, and M. Ruiz-Altaba. *"Searching for new heavy vector bosons in $p\bar{p}$ colliders"*. *Zeitschrift für Physik C Particles and Fields*, 45(1):109–121, 1989. ISSN 0170-9739. doi: 10.1007/BF01556677. URL http://dx.doi.org/10.1007/BF01556677.

[74] Particle Data Group Collaboration. *"Review of Particle Physics"*. *Chin. Phys.*, C38:090001, 2014. doi: 10.1088/1674-1137/38/9/090001. URL http://pdg.lbl.gov/.

[75] CMS Collaboration. *"Search for new physics in final states with a tau and missing transverse energy using pp collisions at $\sqrt{s}$ = 8 TeV"*. Technical Report CMS-PAS-EXO-12-011, CERN, Geneva, 2015. URL https://cds.cern.ch/record/2002036.

[76] CMS Collaboration. *"Search for Narrow Resonances using the Dijet Mass Spectrum with 19.6 fb-1 of pp Collisions at $\sqrt{s}$ = 8 TeV"*. Technical Report CMS-PAS-EXO-12-059, CERN, Geneva, 2013. URL https://cds.cern.ch/record/1519066.

[77] CMS Collaboration. *"Search for W' → tb decays in the lepton + jets final state in pp collisions at $\sqrt{s}$ = 8 TeV"*. *J. High Energy Phys.*, 05(arXiv:1402.2176. CMS-B2G-12-010. CERN-PH-EP-2014-011):108. 33 p, Feb 2014. URL http://cds.cern.ch/record/1647398. Comments: Replaced with published version. Added journal reference and DOI.

[78] CMS Collaboration. *"Search for heavy resonances in the W/Z-tagged dijet mass spectrum in pp collisions at 8 TeV"*. Technical Report CMS-PAS-EXO-12-024, CERN, Geneva, 2013. URL https://cds.cern.ch/record/1563153.

[79] CMS Collaboration. *"Search for a massive resonance decaying into a Higgs boson and a W or Z boson in hadronic final states in proton-proton collisions at $\sqrt{s}$ = 8 TeV"*. *J. High Energy Phys.*, 02(arXiv:1506.01443. CERN-PH-EP-2015-096. CMS-EXO-14-009):145. 40 p, Jun 2015. URL https://cds.cern.ch/record/2021722.

[80] CMS Collaboration. *"First 48 pb^{-1} of data at 13 TeV: a few selected plots from Exotica non-hadronic analyses"*. Aug 2015. URL https://cds.cern.ch/record/2048094.

[81] *LHC End-of-Year Jamboree 2015*, 2015. http://cms-results.web.cern.ch/cms-results/public-results/preliminary-results/LHC-Jamboree-2015.html.

[82] Sergey G. Bondarenko and Andrey A. Sapronov. *"NLO EW and QCD proton-proton cross section calculations with MCSANC-v1.01"*. *Comput.Phys.Commun.*, 184:2343–2350, 2013. doi: 10.1016/j.cpc.2013.05.010. URL http://adsabs.harvard.edu/abs/2013CoPhC.184.2343B.

[83] Peter Skands, Stefano Carrazza, and Juan Rojo. *"Tuning PYTHIA 8.1: the Monash 2013 Tune"*. *Eur. Phys. J.*, C74:3024, 2014. doi: 10.1140/epjc/s10052-014-3024-y. URL http://cds.cern.ch/record/1695787.

[84] CMS Collaboration. *"Underlying Event Tunes and Double Parton Scattering"*. CMS Physics Analysis Summary CMS-PAS-GEN-14-001, 2014. URL http://cdsweb.cern.ch/record/1697700.

[85] R. Fruhwirth. *"Application of Kalman filtering to track and vertex fitting"*. *Nucl. Instrum. Meth.*, A262:444–450, 1987. doi: 10.1016/0168-9002(87)90887-4. URL https://cds.cern.ch/record/178627.

[86] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. *"The Anti-k(t) jet clustering algorithm"*. *JHEP*, 04:063, 2008. doi: 10.1088/1126-6708/2008/04/063. URL http://adsabs.harvard.edu/abs/2008JHEP...04..063C.

[87] CMS Collaboration. *"Identification of b-quark jets with the CMS experiment"*. *JINST*, 8:P04013, 2013. doi: 10.1088/1748-0221/8/04/P04013. URL http://cds.cern.ch/record/1494669.

[88] CDF and D0 Collaboration. *"Combination of CDF and D0 W-Boson Mass Measurements"*. *Phys. Rev.*, D88(5):052018, 2013. doi: 10.1103/PhysRevD.88.052018. URL http://adsabs.harvard.edu/abs/arXiv:1307.7627.

[89] CMS Collaboration. *"Measurement of inclusive W and Z boson production cross sections in pp collisions at* $\sqrt{s} = 13$ *TeV"*. Technical Report CMS-PAS-SMP-15-004, CERN, Geneva, 2015. URL https://cds.cern.ch/record/2093537.

[90] CMS Collaboration. *"CMS Luminosity Measurement for the 2015 Data Taking Period"*. 2016. URL http://cds.cern.ch/record/2138682.

[91] S. van der Meer. *"Calibration of the effective beam height in the ISR"*. Technical Report CERN-ISR-PO-68-31. ISR-PO-68-31, CERN, Geneva, 1968. URL https://cds.cern.ch/record/296752.

[92] Jon Butterworth et al. *"PDF4LHC recommendations for LHC Run II"*. *J. Phys.*, G43:023001, 2016. doi: 10.1088/0954-3899/43/2/023001. URL http://cds.cern.ch/record/2059563.

[93] J. S. Conway. *"Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra"*. In *Proceedings, PHYSTAT 2011 Workshop on Statistical Issues Related to Discovery Claims in Search Experiments and Unfolding, CERN,Geneva, Switzerland 17-20 January 2011*, pages 115–120, 2011. doi: 10.5170/CERN-2011-006.115. URL http://inspirehep.net/record/891252/files/arXiv:1103.0354.pdf.

[94] Lorenzo Moneta, Kevin Belasco, Kyle S. Cranmer, S. Kreiss, Alfio Lazzaro, Danilo Piparo, Gregory Schott, Wouter Verkerke, and Matthias Wolf. *"The RooStats Project"*. *PoS*, ACAT2010:057, 2010. URL http://cds.cern.ch/record/1289965.

[95] W. K. Hastings. *"Monte Carlo Sampling Methods Using Markov Chains and Their Applications"*. *Biometrika*, 57:97–109, 1970. doi: 10.1093/biomet/57.1.97. URL http://www.jstor.org/stable/2334940.

[96] R. Brun and F. Rademakers. *"ROOT: An object oriented data analysis framework"*. *Nucl. Instrum. Meth.*, A389:81–86, 1997. doi: 10.1016/S0168-9002(97)00048-X. URL https://root.cern.ch/.

[97] CMS Collaboration. *"Measurement of associated W + charm production in pp collisions at $\sqrt{s}$ = 7 TeV"*. *JHEP*, 02:013, 2014. doi: 10.1007/JHEP02(2014)013. URL http://cds.cern.ch/record/1605940.

[98] CMS Collaboration. *"Measurement of the associated production of a Z boson and b quarks in pp collision at 8 TeV"*. Technical Report CMS-PAS-SMP-14-010, CERN, Geneva, 2015. URL https://cds.cern.ch/record/2044919.

[99] S. J. Brodsky, A. Kusina, F. Lyonnet, I. Schienbein, H. Spiesberger, and R. Vogt. *"A review of the intrinsic heavy quark content of the nucleon"*. *Adv. High Energy Phys.*, 2015:231547, 2015. doi: 10.1155/2015/231547. URL http://adsabs.harvard.edu/abs/arXiv:1504.06287.

[100] P-H. Beauchemin, V. A. Bednyakov, G. I. Lykasov, and Yu. Yu. Stepanenko. *"Search for intrinsic charm in vector boson production accompanied by heavy flavor jets"*. *Phys. Rev.*, D92(3):034014, 2015. doi: 10.1103/PhysRevD.92.034014. URL http://adsabs.harvard.edu/abs/arXiv:1410.2616.

[101] Richard D. Ball, Valerio Bertone, Marco Bonvini, Stefano Carrazza, Stefano Forte, Alberto Guffanti, Nathan P. Hartland, Juan Rojo, and Luca Rottoli.

"A Determination of the Charm Content of the Proton". 2016. URL http://cds.cern.ch/record/2155124.

[102] G. Bailas and V. P. Goncalves. *"Phenomenological implications of the intrinsic charm in the Z boson production at the LHC"*. *Eur. Phys. J.*, C76(3):105, 2016. doi: 10.1140/epjc/s10052-016-3941-z. URL http://adsabs.harvard.edu/abs/2016EPJC...76..105B.

[103] A. V. Lipatov, G. I. Lykasov, Yu. Yu. Stepanenko, and V. A. Bednyakov. *"Probing proton intrinsic charm in photon or Z boson production accompanied by heavy jets at LHC"*. 2016.

[104] Victor Mukhamedovich Abazov et al. *"Search for flavor changing neutral currents in decays of top quarks"*. *Phys. Lett.B*, 701:313, 2011. doi: 10.1016/j.physletb.2011.06.014. URL http://adsabs.harvard.edu/abs/2011PhLB..701..313D.

[105] ATLAS Collaboration. *"Search for pair-produced third-generation squarks decaying via charm quarks or in compressed supersymmetric scenarios in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector"*. *Phys. Rev.*, D90(5):052008, 2014. doi: 10.1103/PhysRevD.90.052008. URL http://cds.cern.ch/record/1712975.

[106] ATLAS Collaboration. *"Search for Scalar Charm Quark Pair Production in pp Collisions at $\sqrt{s}$= 8 TeV with the ATLAS Detector"*. *Phys. Rev. Lett.*, 114(16):161801, 2015. doi: 10.1103/PhysRevLett.114.161801. URL http://cds.cern.ch/record/1979931.

[107] CMS Collaboration. *"Identification of c-quark jets at the CMS experiment"*. Technical Report CMS-PAS-BTV-16-001, CERN, Geneva, 2016. URL https://cds.cern.ch/record/2205149.

[108] S. J. Brodsky, P. Hoyer, C. Peterson, and N. Sakai. *"The Intrinsic Charm of the Proton"*. *Phys. Lett.*, B93:451–455, 1980. doi: 10.1016/0370-2693(80)90364-0. URL http://adsabs.harvard.edu/abs/1980PhLB...93..451B.

[109] Sayipjamal Dulat, Tie-Jiun Hou, Jun Gao, Joey Huston, Jon Pumplin, Carl Schmidt, Daniel Stump, and C. P. Yuan. *"Intrinsic Charm Parton Distribution Functions from CTEQ-TEA Global Analysis"*. *Phys. Rev.*, D89(7):073004, 2014. doi: 10.1103/PhysRevD.89.073004. URL http://adsabs.harvard.edu/abs/arXiv:1309.0025.

[110] J. Pumplin, D.R. Stump, J. Huston, H.L. Lai, Pavel M. Nadolsky, et al. *"New generation of parton distributions with uncertainties from global QCD analysis"*. *JHEP*, 0207:012, 2002. doi: 10.1088/1126-6708/2002/07/012. URL http://adsabs.harvard.edu/abs/2002JHEP...07..012P.

[111] *Monte Carlo particle numbering scheme*, 2007. http://pdg.lbl.gov/2007/reviews/montecarlorpp.pdf.

[112] Nikolaos Kidonakis. *"Next-to-next-to-leading soft-gluon corrections for the top quark cross section and transverse momentum distribution"*. *Phys.Rev.*, D82: 114030, 2010. doi: 10.1103/PhysRevD.82.114030. URL http://adsabs.harvard.edu/abs/2003PhRvD..68k4014K.

[113] CMS Collaboration. *"Double Muon Trigger efficiency in 2012 data"*. Oct 2014. URL https://cds.cern.ch/record/1969371.

[114] CMS Collaboration. *"Single Muon efficiencies in 2012 Data"*. Mar 2013. URL https://cds.cern.ch/record/1536406.

[115] CMS Collaboration. *"Measurements of differential and double-differential Drell-Yan cross sections in proton-proton collisions at 8 TeV"*. *Eur. Phys. J. C*, 75(arXiv:1412.1115. CMS-SMP-14-003. CERN-PH-EP-2014-289):147. 39 p, Dec 2014. URL http://cds.cern.ch/record/1973604. Comments: Replaced with published version. Added journal reference and DOI.

[116] S. Baffioni, C. Charlot, F. Ferri, D. Futyan, P. Meridiani, I. Puljak, C. Rovelli, R. Salerno, and Y. Sirois. *"Electron reconstruction in CMS"*. *Eur. Phys. J.*, C49: 1099–1116, 2007. doi: 10.1140/epjc/s10052-006-0175-5. URL http://cds.cern.ch/record/934070.

[117] CMS Collaboration. *"Measurement of the Inclusive W and Z Production Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV"*. *JHEP*, 1110:132, 2011. doi: 10.1007/JHEP10(2011)132. URL http://cds.cern.ch/record/1313495.

[118] CMS Collaboration. *"Measurement of $B\bar{B}$ Angular Correlations based on Secondary Vertex Reconstruction at $\sqrt{s} = 7$ TeV"*. *JHEP*, 03:136, 2011. doi: 10.1007/JHEP03(2011)136. URL http://cds.cern.ch/record/1329400.

[119] CMS Collaboration. *"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"*. *JHEP*, 12:039, 2013. doi: 10.1007/JHEP12(2013)039. URL http://cds.cern.ch/record/1605854.

[120] LHCb Collaboration. *"Identification of beauty and charm quark jets at LHCb"*. *JINST*, 10(06):P06013, 2015. doi: 10.1088/1748-0221/10/06/P06013. URL http://cds.cern.ch/record/2012990.

[121] CMS Collaboration. *"Performance of b-tagging at 8 TeV in multijet, ttbar and boosted topology events"*. 2013. URL http://cds.cern.ch/record/1581306.

[122] Leonid Gladilin. *"Fragmentation fractions of c and b quarks into charmed hadrons at LEP"*. *Eur. Phys. J.*, C75(1):19, 2015. doi: 10.1140/epjc/s10052-014-3250-3. URL http://adsabs.harvard.edu/abs/2015EPJC...75...19G.

[123] CMS Collaboration. *"Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV"*. Technical Report CERN-PH-EP-2015-305. CMS-JME-13-004-003. arXiv:1607.03663, CERN, Geneva, Jul 2016. URL http://cds.cern.ch/record/2198719.

[124] Richard D. Ball et al. *"Parton distributions with LHC data"*. *Nucl. Phys.*, B867: 244–289, 2013. doi: 10.1016/j.nuclphysb.2012.10.003. URL http://cds.cern.ch/record/1460337.

[125] CMS Collaboration. *"Absolute Calibration of the Luminosity Measurement at CMS: Winter 2012 Update"*. Technical Report CMS-PAS-SMP-12-008, CERN, Geneva, 2012. URL https://cds.cern.ch/record/1434360.

[126] Rikkert Frederix and Stefano Frixione. *"Merging meets matching in MC@NLO"*. *JHEP*, 12:061, 2012. doi: 10.1007/JHEP12(2012)061. URL http://cds.cern.ch/record/1481985.

[127] Antonio Pich. *"The Standard Model of Electroweak Interactions"*. (hep-ph/0502010. FTUV-2005-0201. IFIC-2005-13):48 p, Feb 2005. URL https://cds.cern.ch/record/819632.

[128] Yorikiyo Nagashima. *"Beyond the standard model of elementary particle physics"*. Wiley-VCH Verlag, Weinheim, Germany,, 2014. ISBN 9783527411771. URL http://www-spires.fnal.gov/spires/find/books/www?cl=QC793.2.N34::2014.