Instituto de Física de Cantabria
CSIC-Universidad de Cantabria

y

Universidad de Cantabria
Departamento de Física Moderna

Medida de la sección eficaz de producción de pares top anti-top en el canal dileptonico a $\sqrt{s} = 8\,\text{TeV}$ con el detector CMS

---

por

Javier A. Brochero Cifuentes

Tesis presentada para optar al grado de Doctor en Física

Directores: Dr. Celso Martínez Rivero and Dr. Luca Scodellaro

Julio de 2014

Instituto de Física de Cantabria
CSIC-Universidad de Cantabria

and

Universidad de Cantabria
Departamento de Física Moderna

Top Anti-Top Production Cross Section Measurement at $\sqrt{s} = 8\,\text{TeV}$ in the Dilepton Channel with the CMS Detector

---

by

Javier A. Brochero Cifuentes

A Thesis Submitted in Conformity with the Requirements for the Degree of Doctor of Philosophy

Supervisors: Dr. Celso Martínez Rivero and Dr. Luca Scodellaro

July 2014

"Don't speak to me of anarchy of peace or calm revolt, man, we're in a play of slow decay orchestrated by Boltzmann".

Greg Graffin, Entropy, 1991.

Dedicada a la pandilla de los niños vestidos de colorín.

DECLARACIÓN DE AUTORÍA

Celso Martínez Rivero, Doctor en Ciencias Físicas e Investigador Científico del Consejo Superior de Investigaciones Científicas (CSIC),

y

Luca Scodellaro, Doctor en Ciencias Físicas y Científico Titular del Consejo Superior de Investigaciones Científicas (CSIC),

CERTIFICAN que la presente Tesis titulada

Top Anti-Top Production Cross Section Measurement at $\sqrt{s} = 8\,\mathrm{TeV}$ in the Dilepton Channel with the CMS Detector

ha sido realizada por Javier Andres Brochero Cifuentes bajo nuestra dirección en el Instituto de Física de Cantabria (IFCA), para optar al título de Doctor por la Universidad de Cantabria (UC). Consideramos que este documento contiene aportaciones científicas suficientemente relevantes como para constituir la Tesis Doctoral del interesado.

En Santander, a 31 de Julio de 2014,

Dr. Celso Martínez Rivero   Dr. Luca Scodellaro

ACKNOWLEDGEMENTS

Los agradecimientos son la mejor forma de acabar cualquier trabajo o experiencia, y si son los correspondientes a la tesis de doctorado aun mejor, ya que me dan la oportunidad de mirar atrás y ver todo con una perspectiva diferente. El ir atrás un poco en el tiempo también me permite sentir de nuevo el esfuerzo, dedicación y empeño que se necesitó para llegar hasta este punto. Pero ese arduo trabajo no fue únicamente mío, sino, como todo en la vida, de un grupo de personas que estuvo constantemente apoyándome. Así que en estas líneas espero poder decirles a cada uno: $\lim_{n\to\infty}(\text{GRACIAS})^n$.

Primero que todo, quiero agradecer a mi director Celso Martínez por su ayuda durante estos largos años de trabajo. A Luca Scodellaro, más que mi co-director y compañero de oficina, fue la persona con quién pude contar en mis estancias en el CERN.

En el IFCA conocí a muchas personas que con su ayuda rápidamente se convirtieron en parte importante de mi vida en España. En el inicio de mi doctorado, Rocío Vilar y Gervacio Gómez fueron de gran ayuda tanto a nivel profesional como personal. Desde que empecé a trabajar con rayos cósmicos hasta todos los estudios de la presente tesis, Rocío siempre ha contribuido con ideas, críticas y recomendaciones que han sido relevantes para entender y aclarar lo que se estaba haciendo. Gerva con sus explicaciones, consejos, ayudas y co-direccion, fue de vital importancia en todos los estudios de alineamiento.

distancia que nos separa. Andres Monrok, que siempre lo recuerdo a nivel de GEN y RECO, o RECO y GEN? A Camilo Posada, mentor y amigo que hace los balances de unidades más extraordinarios jamás vistos. A Mele, que nunca cambies, de verdad, nunca "viejita".

A PS3 y a brochigu, amigos de batallas, viajes, estrés y momentos difíciles. Siempre estuvieron para llenar el vacío de la soledad, y aunque el lag nos fastidiara, siempre nos divertimos. A Roponalpa y a Sexino, el equipo Camper, se les extraña mucho.

Por último, a toda mi familia, por su constante ayuda. Maru, ya sabes que no existen palabras para decirte gracias, para decirte lo mucho que te adoro y lo importante que eres para mí. Eres mi compañera de lucha. Ma, saber que siempre cuento contigo es lo máximo, simplemente lo máximo. Cuchis, eres digna de admirar. Cómo te quiero y te extraño. Pa, no pasa un solo día en el que no te admire un poco más, por todo tu cariño, esfuerzo, dedicación y sacrificio. De verdad, gracias. Tampoco pasa un solo día sin que me parezca un poco más a ti. Mona, eres la hermanita más estupenda. Estoy muy orgulloso de ti, y seguro que pronto vendrán tiempos mejores. Por cierto, el favorito soy yo. Chiqui, por traernos tantas pero tantas alegrias y la palabra "FAFA". Finalmente, a la verdadera Familia, Ely, Lida, Totto, el Niño Rata, Juanita, Candy Mora, Galo Luis, Diego "el Minino" y Rolando, gracias.

I switch to English to thanks all the people involved in the CMS experiment. Special thanks to Zoltan and Noemi, for all your help during my alignment years and of course, during my DCS shifts, that was hard. Thanks to all the top quark group, especially, Andreas Meyer, Tae Jong, Martijn Mulders, Roberto Tenchini and Carmen Diez. Your help and support was very important to me.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

The standard model (SM) is a Quantum Field Theory which postulate that all the matter is made of a group of basic point-like structureless constituents called leptons and quarks. The SM also describes, very precisely, the interaction between the particles: strong, weak and electromagnetic forces mediated by gluons, W/Z bosons and photons, respectively.

The SM has been tested during the last 50 years with very accurate measurements and it predicted the existence of new particles that have been later discovered in different accelerators. One of these particles is the top quark, predicted in 1973 by Makoto Kobayashi and Toshihide Maskawa, and discovered in 1995 at the Tevatron collider [4, 5].

The top quark, the haviest fundamental particle with a mass about 173.3 GeV [6], plays a relevant role in the study of the electroweak symmetry breaking (Higgs boson) as well as in the search of physics beyond the standard model (BSM). The production of top anti-top quark pairs is one of the main background in many of the processes related with the SM and BSM, then it is crucial to measure its production cross section with very high precision.

The Large Hadron Collider (LHC) is operating since 2010, producing proton-proton collisions with a center of mass energy of 7 TeV until 2011 and 8 TeV in 2012. On 4th July 2012 ATLAS and CMS collaborations gave a conference at CERN to announces the discovery of a boson with a mass around 125 GeV [7, 8]. Since then, one of the major goals is to provide information about the recent discovered particle and continue the search of physics beyond the SM. For this purpose, it is necessary to improve the understanding of the apparatus and to expand our knowledege of those processes which are the main backgrounds in the new physics searches. The top quark production has a relevant role in both aspects, especially the production of top anti-top quark pairs.

This thesis presents a measurement of the cross section of top anti-top production at a center of mass energy of 8 TeV with the Compact Muon Solenoid (CMS) [9]. The data used have been collected from April to June 2012, and correspond to a recorded integrated luminosity of $5.3\,\text{fb}^{-1}$.

The most recent theoretical prediction for the top anti-top production cross section is $\sigma_{t\bar{t}}^{\text{NNLO+NNLL}}(8\,\text{TeV}) = 245.8 \pm 9.6\,\text{pb}$ [10] for a top quark mass of $m_t = 173.3\,\text{GeV}$. According to the SM, top quark decays to a W boson and a b quark almost 100% of the times. This leads to final states with two W bosons and two jets coming from the b quark fragmentation. When both W bosons decay leptonically, the event contains two high momentum leptons with opposite charge, two undetected neutrinos that are measured as missing energy in the transverse plane of the beam axis ($\not{E}_T$), and at least two jets, where two of them must be originated from b quarks. The present study conciders final states with $\mu^+\mu^-$, e^+e^-, $\mu^\pm e^\mp$ pairs and those events with $\tau^+\tau^-$ pairs when both taus decay leptonically. The top anti-top cross section measurement is performed with a robust cut based analysis method.

Chapter 2 presents a brief introduction of the standard model and the top quark physics, its properties and production. Chapter 3 gives a description of the Large Hadron Collider (LHC) and the Compact Muon Solenoid (CMS), the detector which collected the data analyzed in this thesis. A detailed explanation of the different sub-detectors that compose the CMS apparatus is discussed.

In a particle detector, the hardware component has a leading role in the precision of the measurements. One of the strengths of the CMS detector is the excellent capability to identify and measure the transverse momentum of the muons. Muon detection is

performed by a combination of tracker system and muon chambers, which demands a very precise alignment between the subdetectors. For this purpose, a muon alignment system has been designed and built. It is divided in three subsystems: barrel, endcaps and link alignment. Chapter 4 contains a description of the link alignment system and a study of the different displacements experimented by the detector structure due to the magnetic forces and temperature changes.

After the hardware calibration, the reconstruction of the physics observables is performed. Chapter 5 describes how the CMS experiment reconstructs all the particles produced in a collision with the particle flow (PF) algorithm [11]. The signature studied in this analysis includes muons, electrons, missing transverse energy, and jets coming from b quarks. All these objects are selected with a specific criteria based on algorithms designed to avoid any false observable.

The dilepton final states have several contributions coming from other SM processes: Drell-Yan, single top, boson-boson and W+jets production. These background processes are studied in detail to reduce their contribution maintaining a high efficiency on the signal selection. A summary of the data and Monte Carlo (MC) simulated samples used together with the triggers and the event selection applied over the kinematic observables are described in Chapter 6.

The Drell-Yan and W+Jets MC predictions are no reliable due to the complex LHC running conditions. This makes necessary to implement a so called data-driven method to estimate these backgrounds from data. The data-driven methods used to estimate Drell-Yan and W+jets backgrounds are described in Chapter 6.

The estimate of the top anti-top cross section and a description of the different sources of systematic uncertainties are presented in Chapter 7. The top anti-top cross section has been derived in all the three final states with a cut and count method. A combined result is also obtained and extrapolated to the most recent top quark mass measured value [3]. Finally, conclusions are presented in Chapter 8.

CHAPTER 2

THE STANDARD MODEL AND THE TOP QUARK

2.1 Standard Model

The Standard Model (SM) of particle physics is a theory which combines special relativity and quantum mechanics to explain the constituents of the matter and their interactions. The elementary particles can be divided into two categories according to their spin: fermions with spin 1/2 and bosons with spin 1 (see Figure 2.1).

The fermions, which are the matter components postulated as point-like and structuless constituents, are divided into leptons (ℓ) and quarks (q). The quarks come in six different flavours: up (u), down (d), strange (s), charm (c), bottom (b) and top (t); formally described by assigning flavour quantum numbers. The u, c and t quarks are known as up-type quarks and all of them carry an electric charge of $\frac{2}{3}e$ while d, s and b are known as down-type quarks carrying a charge of $-\frac{1}{3}e$. The six leptons flavours are: electron (e), muon (μ), tau (τ), electron-neutrino (ν_e), muon-neutrino (ν_μ) and tau-neutrino (ν_τ). They carry electron, muon and tau quantum numbers. The three charged leptons have an electric charge of $-1e$ and their masses vary from $\sim 0.5\,\mathrm{MeV}$
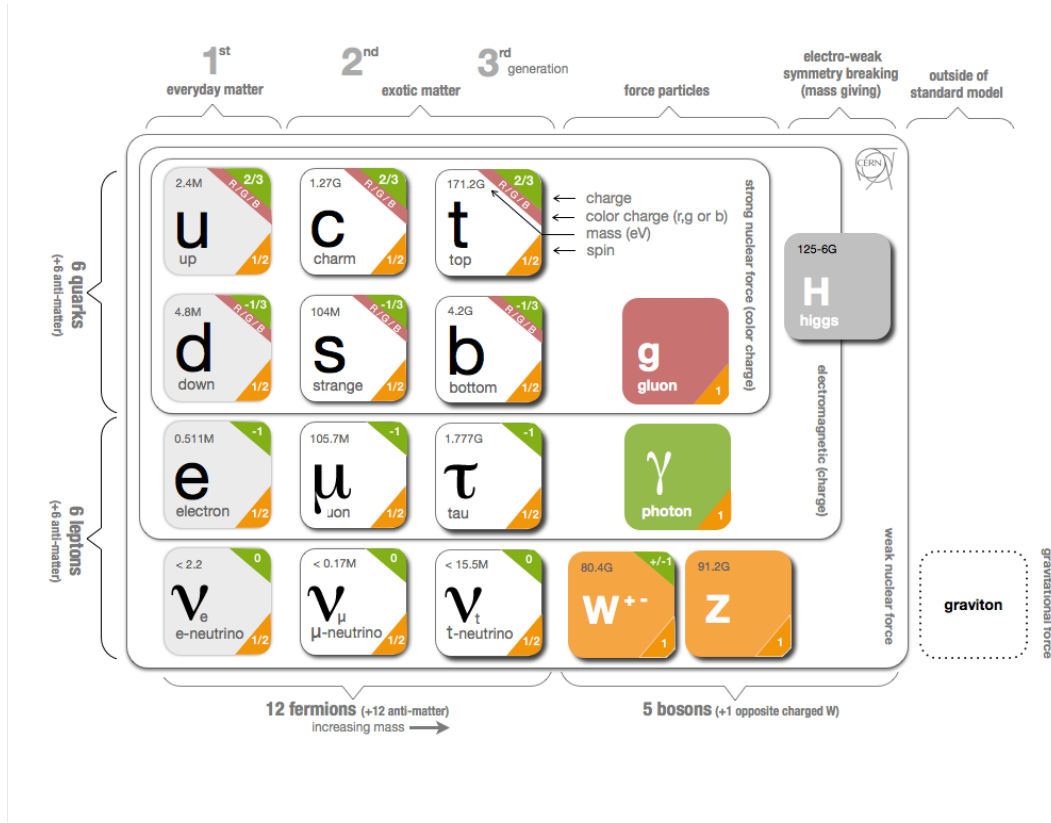
Figure 2.1: Standard model particles.

to $\sim 1\,\text{GeV}$. Neutrinos are chargeless particles and have very small masses ($< 18\,\text{MeV}$) as it has been shown by results from neutrinos oscillation measurements [12]. Quarks and leptons can be grouped into three generations as shown in Table 2.1, which also contains the charges and masses of the particles.

All of the fermions described above have its own anti-particle, which has the same mass but opposite quantum numbers. The anti-quarks are denoted as $\bar{u}$, $\bar{d}$, etc. In the specific case of the electron, its anti-particle is the positron (e^+). However, stable anti-matter has not been detected so far. As a matter of fact, all stable matter observed in the universe is made only of particles from the first generation. The atoms are made of electrons, protons and neutrons, which in turn, are formed of up and down quarks.

The particles of the second and third generations are copies of the first generation particles, except for their masses. The understanding of the mass hierarchy and generation structure is still an open question in the particle physics field.

	Generation					
	First		Second		Third	
Symbol	u	d	c	s	t	b
Mass [GeV]	$(2.3^{+0.7}_{-0.5}) \times 10^{-3}$	$(4.8^{+0.5}_{-0.3}) \times 10^{-3}$	1.28 ± 0.02	$(95 \pm 5) \times 10^{-3}$	173.07 ± 0.89	4.18 ± 0.03
Charge [e]	$+2/3$	$-1/3$	$+2/3$	$-1/3$	$+2/3$	$-1/3$
Symbol	e	ν_e	μ	ν_μ	τ	ν_τ
Mass [MeV]	0.51	$< 3 \times 10^{-6}$	106	< 0.19	1777	< 18.2
Charge [e]	-1	0	-1	0	-1	0

Table 2.1: Charges and massses of the three generations of fermions.

In the SM, the bosons (spin 1) are the quanta of the field responsible for the interactions between particles. The three types of forces experimentally observed in the interaction between fermions are:

- **The electromagnetic (EM) force**, mediated by the photon and responsible for the emission of light from exited atoms. Since the photon is massless, the range corresponding to this interaction is infinite.

- **The weak force**, mediated by the $W^\pm$ and Z^0 bosons. This force causes the nuclear beta decay. The range of the weak interaction is very small ($\sim 10^{-3}$ fm) because of the large mass of the carriers.

- **The strong force**, mediated by the gluons. It keeps the atomic nuclei stable. The quarks are the only fermions that interact via this force.

The three forces are described by so called quantum gauge theories [13, 14]. The fourth fundamental interaction, the gravity, is not included in the SM framework since it dominates only for very large masses, and it is not further considered in the interaction of fundamental particles. Table 2.2 shows the charge and mass of the gauge bosons.

Symbol	Force	Mass[GeV]	Charge[e]	Coupling
γ	Electromagnetic	0	0	10^{-2}
$W^\pm$	Weak	80.4	± 1	10^{-13}
Z	Weak	91.2	0	10^{-13}
g	Strong	0	0	1

Table 2.2: Charges and masses of the gauge bosons.

2.1.1   The Massless SM Lagrangian

The principle of gauge symmetry is essential in the SM construction since the fundamental interactions can be described by local field theories. The electroweak model is based on the gauge group $SU(2)_L \times U(1)_Y$ and QCD is based on the $SU(3)$ symmetry. The most simple Lagrangian for a fermion (spin $1/2$) of mass m is:

$$\mathcal{L} = i\bar{\psi}\gamma^\mu \partial_\mu \psi - m\bar{\psi}\psi \qquad (2.1)$$

which can derive the Dirac equation of motion and it is invariant under phase transformation (U(1) gauge symmetry)

$$\psi \rightarrow e^{i\alpha}\psi \qquad (2.2)$$

where α is a real constant in space and time which has no physical meaning. If the fermion wave function (ψ) changes now under a local phase transformation with rotation parameter $\vec{\epsilon}(x)$ in an internal space represented by the generator $\vec{\tau}$ as

$$\psi' = U\psi = e^{i\vec{\epsilon}(x)\frac{\vec{\tau}}{2}}\psi \qquad (2.3)$$

the Equation 2.1 is not invariant under such transformation, making the Lagrangian a no local gauge invariant. In order to make it invariant under local gauge transformation, it is necessary to introduce the covariant derivative

$$D_\mu = \partial_\mu - ig\frac{\vec{\tau}}{2}\vec{A}_\mu \qquad (2.4)$$

Where $\vec{A}_\mu$ is a new interacting vector field which compensates the local gauge transformation. g is an arbitrary parameter which will determine the universal interaction strength associated to the field. Including the covariant derivative into the Lagrangian, it is obtained

$$\mathcal{L} = i\bar{\psi}\gamma^\mu \partial_\mu \psi - m\bar{\psi}\psi - ig\bar{\psi}\gamma^\mu \frac{\vec{\tau}}{2}\vec{A}_\mu\psi \qquad (2.5)$$

where the last term expresses the coupling between the fermion field and the new vector field $\vec{A}_\mu$. Now, demanding the invariant under phase transformations

$$D'_\mu \psi' = U(D_\mu \psi) \tag{2.6}$$

The relations for the components of the gauge field $\vec{A}_\mu$ are derived to be

$$\frac{\tau^i}{2} A_\mu^{i\,\prime} = -\frac{i}{g}(\partial_\mu U)U^{-1} + U\frac{\tau^i}{2}A_\mu^i U^{-1} \tag{2.7}$$

It is found that the requirement of a theory to be invariant under gauge transformations entails the introduction of associated vector fields called gauge fields. These fields imply the existence of spin 1 particles, the gauge bosons, that couple to fermions.

Electroweak theory

The gauge symmetry group able to give an appropriate description of the observed electroweak phenomena, was determined to be the $SU(2)_L \times U(1)_Y$ group. Requiring the Lagrangian to be gauge invariant towards local phase-space transformations of this group, allowed to unify the weak nuclear force with the electromagnetic force, up to then described by the Quantum Electrodynamics (QED). The electromagnetic force is characterized by unitary $e^{i\epsilon(x)}$ phase transformations in one dimension according to the $U(1)$ symmetry group. The weak force on the other hand is described by $SU(2)$. Consequently, it is convenient to group the fermions into doublets interacting under the weak force:

$$\Psi_{EW} = \begin{pmatrix} u \\ d \end{pmatrix}, \begin{pmatrix} c \\ s \end{pmatrix}, \begin{pmatrix} t \\ b \end{pmatrix}, \begin{pmatrix} \nu_e \\ e \end{pmatrix}, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix} \tag{2.8}$$

Where each doublet corresponds to a field comprising two Dirac spinors. Any $SU(2) \times U(1)$ local phase-space transformation can be defined as:

$$\Psi'_{EM} = e^{i\vec{\epsilon}(x)\cdot\vec{\sigma}}e^{i\theta(x)}\Psi_{EW} \tag{2.9}$$

To make the Lagrangiant invariant under this transformation, the covariant derivative is

$$D_\mu = \partial_\mu - ig\frac{\vec{\sigma}}{2}\vec{W}_\mu + ig'\frac{Y}{2}\vec{B}_\mu \tag{2.10}$$

with $\vec{W}_\mu$ and B_μ the gauge fields associated to the $SU(2)$ and U(1) groups respectively. Consequently, to have a gauge invariant electroweak theory, it is necessary one scalar gauge boson B^0 and three vector gauge bosons W^α. The last bosons can only couple to left-handed fermion doublets; right-handed fermion fields remain unchanged under the $SU(2)$ gauge transformation. Thus, the parity-violating nature of the weak interactions is incorporated into the theory. Adding explicit mass terms to the Lagrangian would break the gauge invariance. In Section 2.1.2 there is a solution to this problem known as spontaneous symmetry breaking. With this electroweak symmetry breaking procedure, the mass terms arise naturally from the Higgs mechanism:

$$\begin{aligned}
W_\mu^\pm &= \sqrt{\frac{1}{2}}\left(W_\mu^1 \mp iW_\mu^2\right) \\
Z_\mu^0 &= W_\mu^3\cos\theta_\omega - B_\mu\sin\theta_\omega \\
A_\mu &= W_\mu^3\sin\theta_\omega + B_\mu\cos\theta_\omega
\end{aligned} \tag{2.11}$$

Where θ_ω is the Weinberg mixing angle.

Quantum chromodinamics

In a similar way to the QED and to the electroweak theory, the quantum chromodinamics (QCD) introduces gluon fields mediating the strong force by requiring gauge invariance. There are three different types of strong charges (colours): red, green and blue. For this interaction the relevant gauge group is the $SU(3)_C$ symmetry group, where the C subscript corresponds to the quark colour triplets. Restoring the gauge invariance of the theory with respect to local $SU(3)$ phase-space transformations invokes the introduction of eight gauge fields, corresponding to the eight $SU(3)$ group generators: the gluons.

The observation of CP violation and processes violating the conservation of strangeness, are allowed in the Standard Model by the assumption that the strong force eigenstates of the quarks slightly differ from their weak force eigenstates. This mismatch of quantum states is given by the Cabibbo-Kobayashi-Maskawa (CKM) matrix:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix}_L = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}_L \tag{2.12}$$

The CKM matrix describes the probability of a transition from one quark q to another q' due to a flavour changing weak interaction, which is proportional to $|V_{qq'}|^2$.

$$|V_{CKM}| = \begin{pmatrix} 0.97427 \pm 0.00015 & 0.22534 \pm 0.00065 & 0.00351^{+0.00015}_{-0.00014} \\ 0.22520 \pm 0.00065 & 0.97344 \pm 0.00016 & 0.0412^{+0.0011}_{-0.0005} \\ 0.00867^{+0.00029}_{-0.00031} & 0.0404^{+0.0011}_{-0.0005} & 0.999146^{+0.000021}_{-0.000046} \end{pmatrix} \tag{2.13}$$

2.1.2 The Higgs Mechanism

In order to guaraantee massive W and Z bosons and consequently weak interactions with a short range (compared to the EM force), the electroweak symmetry has to be broken in a way that conserves the gauge invariance and renormalizability of the theory. The Higgs mechanism provides an explanation for this phenomenon. It is based on the idea of spontaneous symmetry breaking, according to which the vacuum state of a system does not possess the same symmetry as the Lagrangian density [15, 16].

The simplest way to break the $SU(2)_L \times U(1)_y$ gauge symmetry is to add four scalar fields in form of isospin doublet

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tag{2.14}$$

where ϕ^+ and ϕ^0 are complex fields. The Lagrangian density for this field will be

$$\begin{aligned} \mathcal{L} &= |D_\mu \phi|^2 - V(\phi) \\ &= |D_\mu \phi|^2 - \left(\mu^2 + \lambda (\phi^\dagger \phi)^2 \right) \end{aligned} \tag{2.15}$$

where D_μ is the covariant derivative in the $SU(2)_L \times U1_Y$ gauge, μ^2 a mass parameter and λ the strength of the Higgs boson field self interaction.

The parameters μ and λ can be chosen such that the symmetry is spontaneously broken because the minimum of the potential is no longer unique. This minimum is on a continuous ring in th complex plane as shown in Figure 2.2.
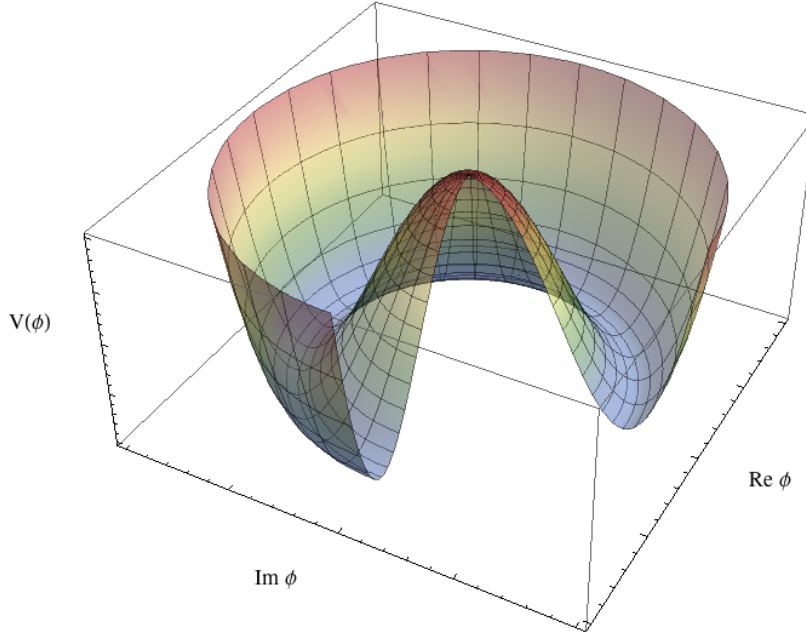
Figure 2.2: The Higgs potential "Sombrero Mexicano".

It is possible to write with an arbitrary choice of the vacuum

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{2.16}$$

where $h(x)$ is the Higgs boson field, expressed as a quantum fluctuation about its vacuum expectation value, and $v = |\mu^2|/\lambda$. When the new parametrization of ϕ is inserted into the Lagrangian, it breaks its symmetry, that is, the Lagrangian is not an even function of the Higgs fields anymore. This mechanism where the ground state does not share the symmetry of the Lagrangian is called spontaneous symmetry breaking. As a result, the Higgs boson field has acquired mass

$$m_H = \sqrt{2\lambda v^2} \tag{2.17}$$

Applying spontaneous symmetry breaking to the Lagrangian and forcing a local gauge invariance, the three electroweak gauge bosons acquire mass

$$m_W = \frac{1}{2}gv, \qquad m_Z = \frac{1}{2}v\sqrt{g^2 + g'^2} \tag{2.18}$$

In contrast to the mass of the bosons, the mass of fermions is not generated by a gauge principle, but by the addition of extra gauge invariant Yukawa couplings in the Lagrangian density, resulting also in an increase of the number of free parameters in the SM. Nevertheless, a gauge invariant way of introducing fermion masses in the theory exists. The coupling strength of fermions to the Higgs boson field is proportional to the mass of the fermion. Hence, due to the extremely high top quark mass, the theoretical prediction of the Higgs boson mass is extremely sensitive to the value of the top quark mass.

2.2 Top Physics

The top quark is considered in the SM as the isospin partner of the bottom quark, and it is required to account for the absence of flavour-changing, charge preserving weak decays of b. As a member of the third generation of quarks, it provides the simplest explanation for CP violation[17] by the weak interaction. The SM does not predict its mass, and experimentally the limit kept changing until its discovery at Tevatron in 1995 by CDF and DØ collaborations [4, 5].

The properties of the top quark differ from those of the other quarks because its mass is much larger. Particularly, it is much heavier than the $W^{\pm}$ boson, allowing its decays by the first-order weak interaction:

$$t \rightarrow q + W^+ \quad \text{with} \quad q = d,\ s,\ b \tag{2.19}$$

Where the only significant decay mode is into b quark due to the small factors $|V_{td}|^2$ and $|V_{ts}|^2$ of the Cabbibo-Kobayashi-Maskawa (CKM) matrix (see Equation 2.13).

The decay width is:

$$\Gamma(t \to bW^+) \approx \frac{G_F}{8\pi\sqrt{2}}m_t^3 \left(1 - 3\frac{m_W^4}{m_t^4} + 2\frac{m_W^6}{m_t^6}\right) \tag{2.20}$$

which has a strong dependence on the top quark mass. With $G_F \approx 10^{-5}\,\text{GeV}^{-2}$, $m_t = 173\,\text{GeV}$ and $m_W = 80\,\text{GeV}$, the decay width obtained is $\Gamma(t \to bW^+) \approx 1\,\text{GeV}$, which corresponds to a t quark lifetime of $\approx 0.6 \times 10^{-24}\,\text{s}$. Because the top quark decays before it can hadronize[1], there are no bound $t\bar{t}$ states and no top-flavored mesons or baryons, different to the situation of the other lighter quarks.

2.2.1   Top Quark Production

The two basic SM production mechanisms of top quarks at the hadron colliders are the dominant top anti-top pair production via strong interaction and the single-top production via electroweak interaction.

$t\bar{t}$ production

$t\bar{t}$ are produced either via quark anti-quark ($q\bar{q}$) annihilation or gluon fusion. Figure 2.3 shows the leading order (LO) Feynman diagrams for the $t\bar{t}$ production. At the LHC accelerator with proton-proton collisions, the dominant process in $t\bar{t}$ production is the gluon fusion with a 83%, follow by the quark anti-quark annihilation with a 17%.
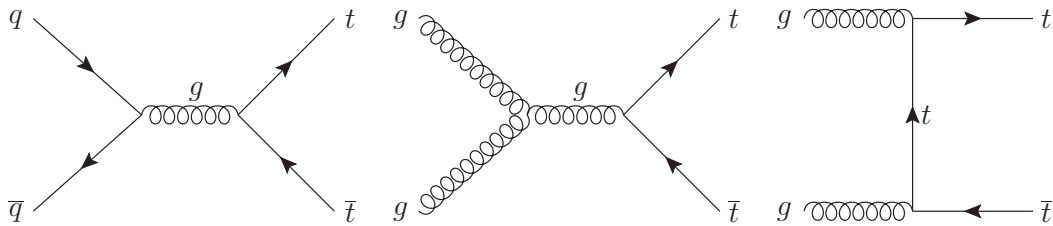
Figure 2.3: Feynman diagrams of the leading order processes for $t\bar{t}$ production: quark anti-quark annihilation ($q\bar{q} \to t\bar{t}$) and gluon fusion ($gg \to t\bar{t}$).

The total $t\bar{t}$ cross section ($\sigma_{t\bar{t}}$) is calculated as a convolution of the parton distribution function (PDF) for the incoming hadrons and the cross section of the partonic processes $q\bar{q}, qq \to t\bar{t}$:

---

[1]The characteristic formation time of hadrons is $\approx 2 \times 10^{-24}\,\text{s}$

$$\sigma\left(s, m_t{}^2\right) = \sum_{i,j} \int_0^1 dx_i \int_0^1 dx_j \, f_i\left(x_i, \mu_F^2\right) f_j\left(x_j, \mu_F^2\right) \hat{\sigma_{ij}}\left(\hat{s}, m_t, \alpha_s(\mu_R^2)\right) \qquad (2.21)$$

Where i and j are the possible combinations of incoming gluons or quark anti-quark pairs. $f(x, \mu_F^2)$ are the PDF evaluated at some factorization scale (μ_F) and for a fraction of the incoming proton energy carried by the parton (x). The partonic subprocess cross section ($\hat{\sigma_{i,j}}$), integrated over the phase space, depends on the centre of mass energy in the collision ($\sqrt{\hat{s}}$), the top quark mass (m_t) and the QCD strong coupling constant α_s evaluated at a renormalization scale μ_R.

The theoretical calculations of the $t\bar{t}$ production cross section with the highest accuracy until now are described in [10]. This estimation corresponds to a precision where it is included two loops corrections in the lowest order diagram (next-to-next-to leading order, NNLO) as well as the logarithmic "Sudakov" correction terms [18] at the same order (next-to-next-to leading logarithm, NNLL).

The SM production cross sections expected at the LHC energies and their incertainties are summarized in Table 2.3. All numbers were computed for a top quark mass of $m_t = 173.3\,\text{GeV}$ and $m_t = 172.5\,\text{GeV}$ using a MSTW2008nnlo68cl PDF set [19] with the program `top++` (`v2.0`) [20]. Scale uncertainty has been determined through independent restricted variation of μ_F and μ_R. Figure 2.4 shows the last theoretical predictions as a function of the center-of-mass energy, compared with the experimental results from the ATLAS and CMS detectors.

$\sqrt{s}$ [TeV]	$\sigma_{t\bar{t}}$ [pb] ($m_t = 172.5\,\text{GeV}$)	$\sigma_{t\bar{t}}$ [pb] ($m_t = 173.3\,\text{GeV}$)	Scale [pb]	PDF [pb]
7	176.3	172.0	+4.4(2.6%) −5.8(3.4%)	+4.7(2.7%) −4.8(2.8%)
8	252.8	245.9	+6.2(2.5%) −8.4(3.4%)	+6.2(2.5%) −6.4(2.6%)
13	824.2	806.4	+19.3(2.4%) −28.5(3.5%)	+13.7(1.7%) −15.3(1.9%)
14	974.8	953.6	+22.7(2.4%) −33.9(3.6%)	+16.2(1.7%) −17.8(1.9%)

Table 2.3: NNLO+NNLL theoretical prediction at the LHC collider energies.
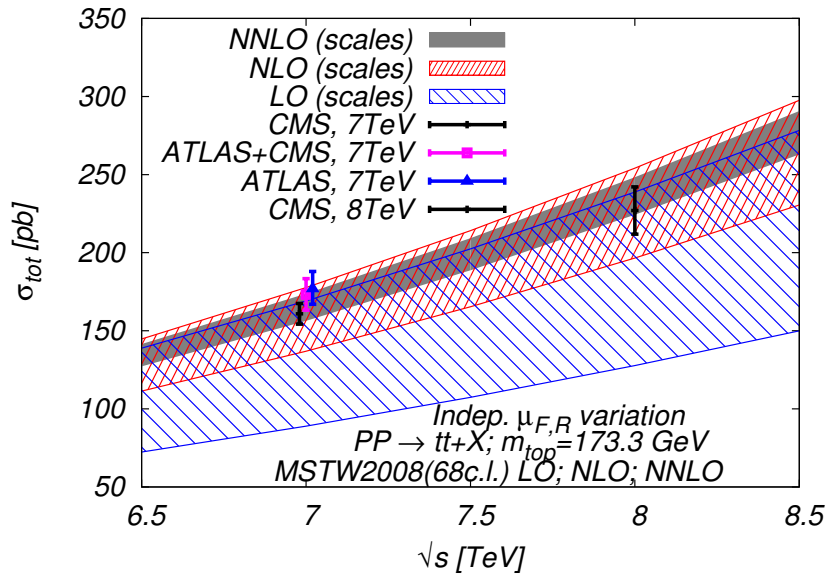
Figure 2.4: Theoretical prediction of $\sigma_{t\bar{t}}$ for the LHC as a function of the collider centre of mass energy ($\sqrt{s}$) compared to the available measurements from the ATLAS and CMS detectors at 7 TeV and 8 TeV.

2.2.2   Single Top Quark Production

Single top quarks can be produced via electroweak interaction involving a *Wtb* vertex. There are three production modes:

1. **t-channel** A virtual *W* strikes a *b* quark (from the sea quark) inside the proton. This mode is also known as *W* gluon fusion, since the b quark originates from a gluon splitting into a $b\bar{b}$ pair. Figure 2.5 (a) and (b) show these processes.

2. **s-channel** This production mode is of Drell-Yan type. A timelike *W* boson is produced by the fusion of two quarks ($q\bar{q}$) belonging to a SU(2) isospin doublet. Figure 2.5 (c) shows this process.

3. **Associated production (*tW*)** The top quark is produced in association with a real *W* boson. The initial *b* quark is a sea quark inside the proton. Figure 2.5 (d) shows this process.

In pp collisions, the single top cross section is dominated by contributions from up and down quarks coupling to the *W* boson on one side of the Feynman diagrams. The production channels involving a *Wtd* or *Wts* vertex are strongly suppressed due

to the CKM matrix elements, $\sim 0.1\%$ and $\sim 1\%$, respectively.
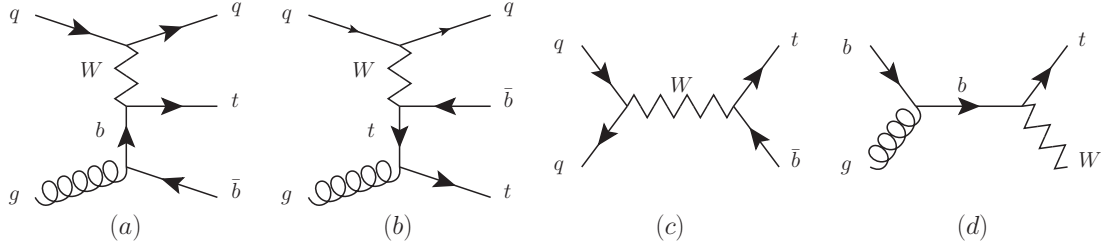
Figure 2.5: Feynman diagrams for the three single top production modes. (a) and (b) show W gluon fusion, (c) s-channel process and (d) associated production. The diagrams for single anti-top quark production can be obtained by interchanging quarks and anti-quarks.

2.2.3 Top Quark Mass

The top quark mass (m_t) is experimentally defined by the position of the peak in the invariant mass distribution of its decay products (W boson and a b-quark jet). This closely corresponds to the pole mass of the top quark, defined as the real part of the pole in the perturbative top quark propagator[21].

The top quark mass has been measured by the CMS collaboration [22, 23, 24] by analysing different decay modes. The combined result obtained is $m_t = 172.2 \pm 0.1(\text{stat.}) \pm 0.7(\text{syst.})$, while the last world average value is $m_t = 173.34 \pm 0.27(\text{stat.}) \pm 0.71(\text{syst.})$ [6]. Figure 2.6 summarizes the top quark mass results obtained by the CMS detector, and the Tevatron and world combinations.

2.2.4 Top Quark Signature in $t\bar{t}$ Events

As it was explained, the top quark decays almost exclusively (99.8%) into a *Wb* pair. A $t\bar{t}$ pair therefore decays to two b quarks (detected as jets in the detector, see Section 5.6) and two W bosons, which will on their turn decay hadronically, full-leptonically or semi-leptonically. Figure 2.7 shows the diagrams for $t\bar{t}$ decay into *Wb* and the subsequent W decay into hadrons, leptons or lepton plus hadrons. Table 2.4 shows the branching ratio of the *W* boson decay.

Experimentally, the decays are distinguished as:
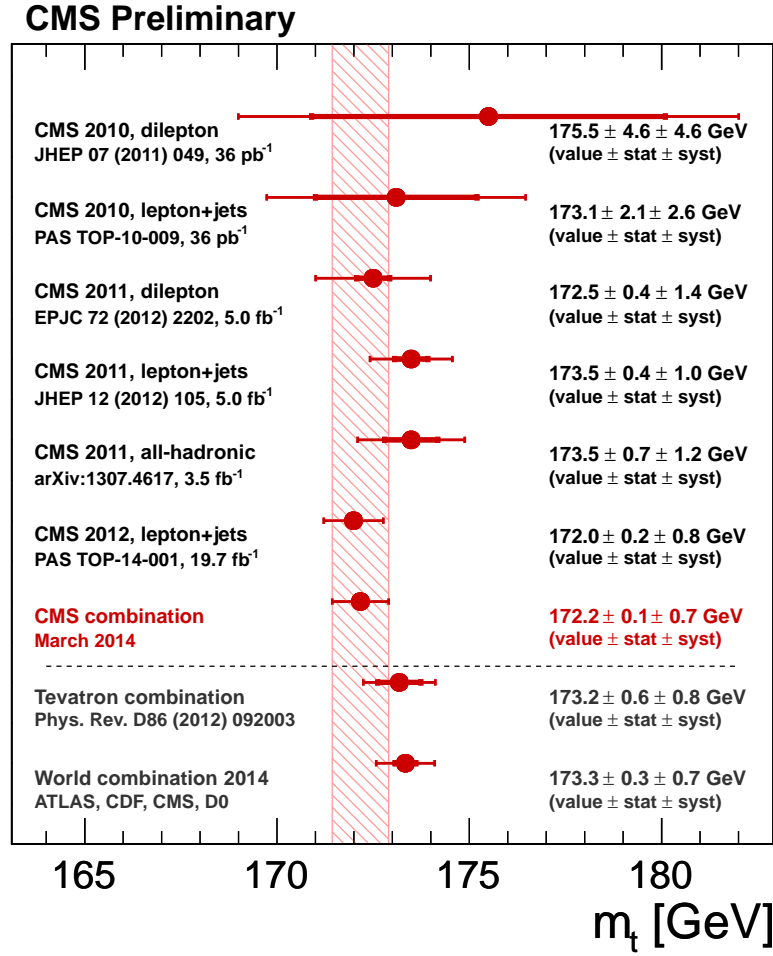
Figure 2.6: Top quark mass measured by the CMS detector, Tevatron[1, 2] combination and the world average value.

- Both *W* bosons decay into electrons and/or muons which can be directly seen in the detector. This decay is called "dileptonic" and it is part of the full leptonic decay.

- One *W* boson decays into an electron or muon and the other *W* decays into quarks. This channel is called semileptonic decay.

- Both *W* bosons decay into quarks. This mode is called full hadronic decay.

- One or both *W* bosons decay into a tau, which can decay leptonically (into electron or muon) or hadronically (into a quark). Even there is specific analysis to study, the $t\bar{t} \rightarrow Wb \rightarrow \tau + X$, in the cases of $W \rightarrow \tau + \nu_\tau \rightarrow e/\mu + \nu_{e/\mu}$

W Decay Modes	
Mode	Br(%)
$e + \nu_e$	10.75 $\pm$ 0.13
$\mu + \nu_\mu$	10.57 $\pm$ 0.15
$\tau + \nu_\tau$	11.25 $\pm$ 0.20
$\ell + \nu_\ell$	10.80 $\pm$ 0.09
Hadrons	67.60 $\pm$ 0.27

Table 2.4: Branching ratio (*Br*) of the W boson decays [3].

signature [25], it is included in the full or semi leptonic channels, depending of the decay of the other *W*.

Table 2.5 shows the branching ratios for the three decay modes of the top anti-top quark pairs. In Section 2.2.5 there is a detailed description of the signature studied in this document.

$t\bar{t}$ Decay Modes			
Mode	Br[%]	Mode	Br[%]
Full Leptonic	10.50 $\pm$ 0.12	$ee + \nu_e \nu_e$	1.16 $\pm$ 0.02
		$\mu\mu + \nu_\mu \nu_\mu$	1.12 $\pm$ 0.02
		$\tau\tau + \nu_\tau \nu_\tau$	1.27 $\pm$ 0.03
		$e\mu + \nu_e \nu_\mu$	2.27 $\pm$ 0.04
		$e\tau + \nu_e \nu_\tau$	2.42 $\pm$ 0.05
		$\mu\tau + \nu_\mu \nu_\tau$	2.38 $\pm$ 0.05
Semi Leptonic	43.80 $\pm$ 0.40	$eqq' + \nu_e$	14.53 $\pm$ 0.19
		$\mu qq' + \nu_\mu$	14.29 $\pm$ 0.21
		$\tau qq' + \nu_\tau$	15.21 $\pm$ 0.28
Hadrons	45.70 $\pm$ 0.26	—	

Table 2.5: $t\bar{t}$ branching ratios calculated from the W branching ratios of Table 2.4

2.2.5 $t\bar{t}$ **in the Dilepton Channel**

The dilepton final state includes two leptons (muon or electron) with high p_T, a large imbalance in the total transverse energy (missing transverse energy, $\not{E}_T$) asociated
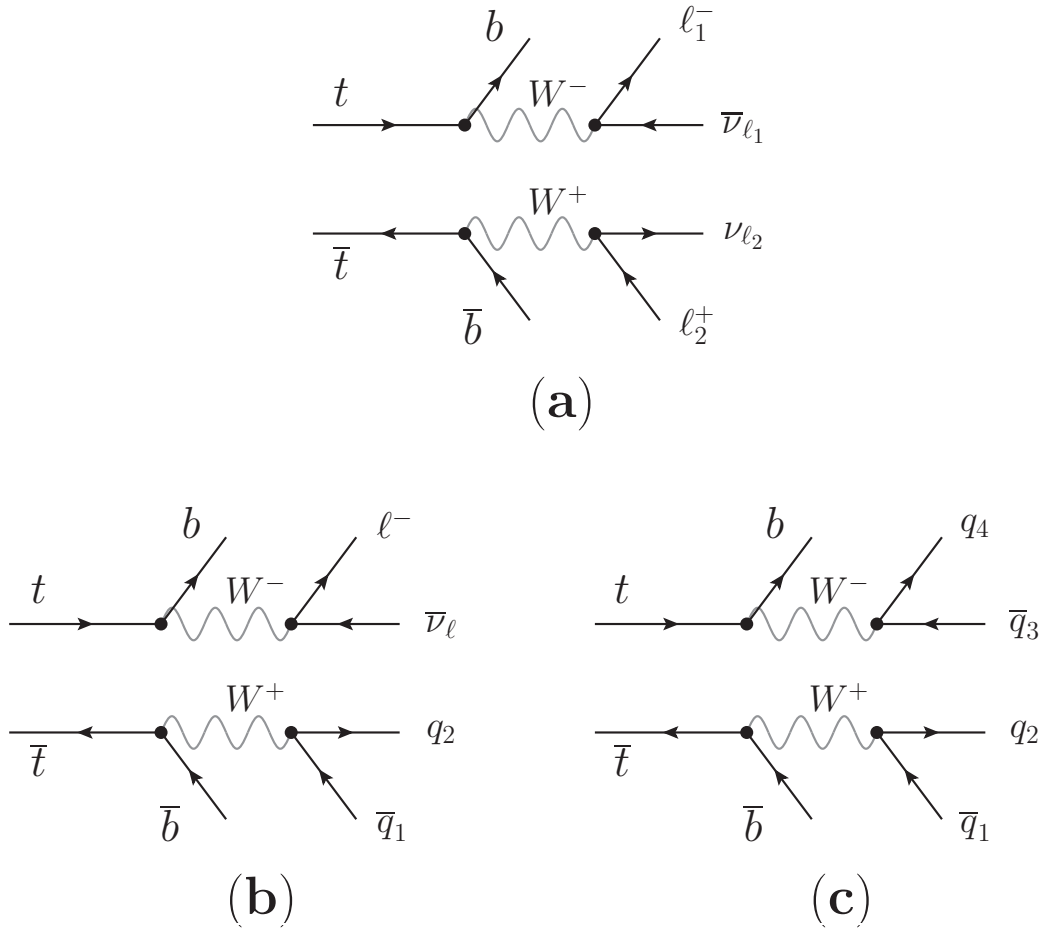
Figure 2.7: Signature events in the top anti-top production: (**a**) dilepton mode; (**b**) lepton plus jets mode; (**c**) hadronic mode.

to the neutrinos and two b quark jets. As it will be shown, one advantage of the dilepton channel is the small background contribution due to other SM processes, especially in the muon-electron channel. However, the downside of this channel is its low branching ratio of about 4.5% as it is shown in Table 2.5. There is a small contribution from tau events to the dilepton channel in the case of a tau decaying into an electron or muon. The $t\bar{t}$ signature can be divided as a function of the lepton flavour into same (electron-electron or muon-muon) and opposite (muon-electron or electron-muon) flavour in order to study the background contributions.

Main backgrounds in the dilepton channel

Some SM processes can have a similar signature to the $t\bar{t}$ dilepton events, especially, those with two leptons in the final state. The leptons in the final state can be clasified as "*prompt*" or "*fake*" leptons. The prompt lepton are electrons or muons coming from the primary interaction vertex through electroweak processes, while fakes include leptons coming from meson decays in jets, cosmic rays, jets misidentified as leptons, etc. The standard model processes considered as backgrounds are:

1. **Drell-Yan ($Z/\gamma \to \ell\ell$)** It is produced via electroweak processes. Its signature is similar to the $t\bar{t}$ one due to the decay of the Z boson into two same flavour leptons. However, in this process there are no neutrinos in the decay chain, thus, there will be no real missing transverse energy in this kind of events. The cross section for the leptonic decay mode is $\sigma_Z^{\ell\ell} \approx 1967$ pb for dilepton invariant mass $m_{\ell\ell} > 20$ GeV at $\sqrt{s} = 8$ TeV [26]. The contribution to the muon-electron channel comes only from the process $Z \to \tau\tau \to \mu e + \nu_e\nu_\mu$, making smaller this background in the opposite flavour than in the same flavour channel. Tau channel also contributes in the same flavour channel when the taus decay in same flavour leptons.

2. **Single Top** Due to the production of a single top quark with a W boson, it is possible to have in the final state two leptons, missing transverse energy and a b quark. This process appears when the W boson from the top decay and the W from the tW vertex (Figure 2.5 (a)) decay into leptons.

3. **W+jets** This background consists of a real W boson produced in association with quarks or gluons. The gluon can split into a pair of heavy flavour quarks producing W plus jets as $Wb\bar{b}$. As a result, in the final state there will be two b-jets, one lepton and $\not{E}_T$, wich is very similar to the $t\bar{t}$ signature (when an additional lepton comes from a meson decay in a jet, cosmic rays, etc).

4. **$t\bar{t}$ semileptonic** This process can contribute to the background in the dilepton channel if there is one fake lepton.

5. **Dibosons** These backgrounds occur when vector bosons are created in pairs as WW, WZ and ZZ. The WW process has two real W bosons where both can decay leptonically or one can decay leptonically and the other one hadronically. Similarly, the WZ process can have the W decaying leptonically, and the Z de-

caying in heavy flavour quarks. Another possibility in WZ, is that the W decays hadronically and the Z decays leptonically, obtaining two leptons in the final state. Finally, the ZZ background can have a similar experimental signature as $t\bar{t}$ if one Z decays into leptons and the other Z decays hadronically. Due to the cross sections of these processes, their contribution is small.

Figure 2.8 shows the signatures for the background processes described before. All of them must be reduced using different cuts in order to have a sample dominated by $t\bar{t}$ production.
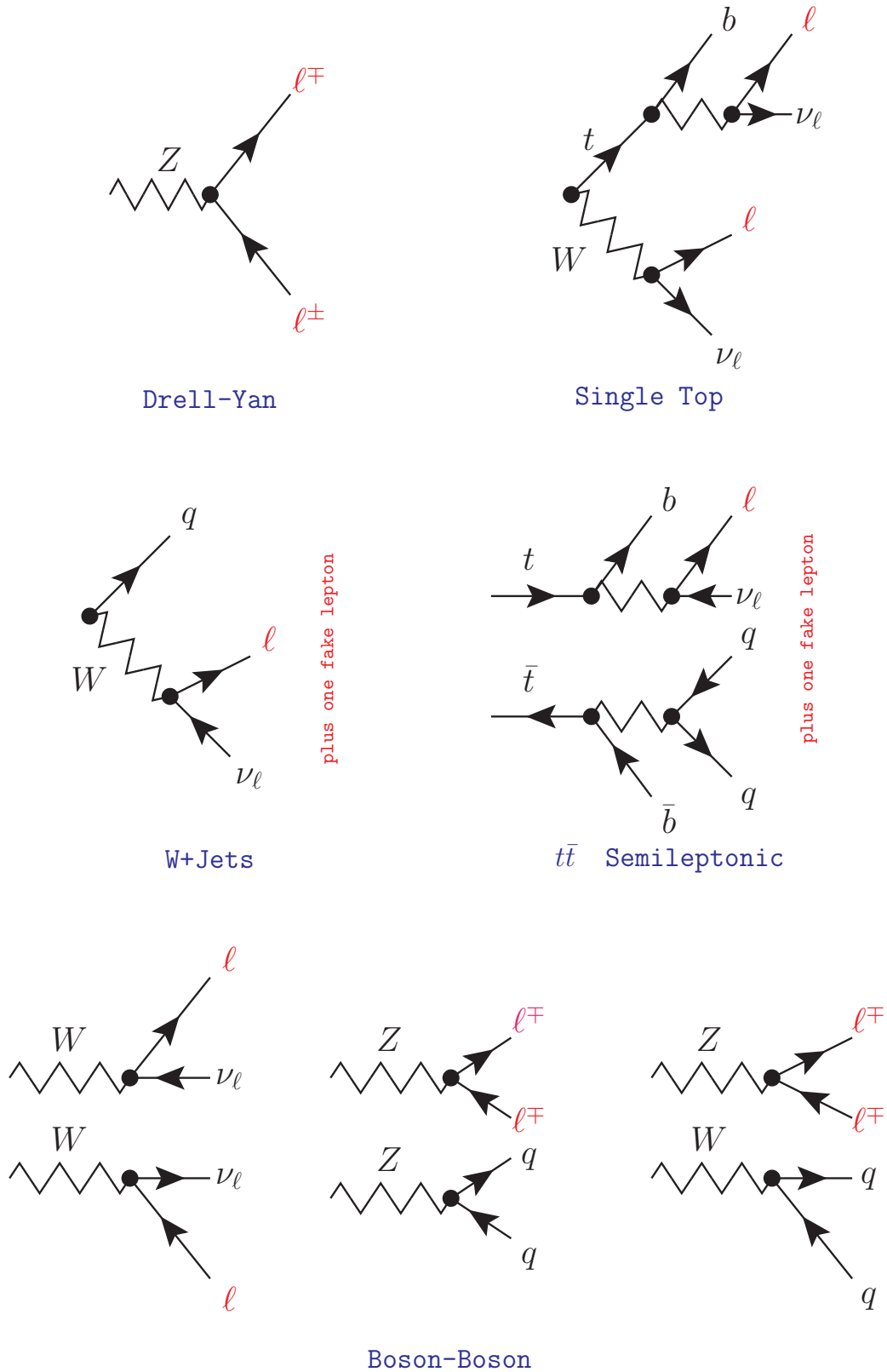
Figure 2.8: Diagrams of the most relevant backgrounds for tt̄ events in the dilepton channel: Drell-Yan, single top, W+jets, tt̄ semileptonic and diboson production. In red color, the same lepton signature (two leptons) of the process studied. Specific cuts must be applied in order to reject these SM processes (see Section 6.2.4).

CHAPTER 3

THE LHC ACCELERATOR AND THE CMS DETECTOR

3.1   Large Hadron Collider (LHC)

The Large Hadron Collider (LHC)[27] is a proton-proton and heavy-ion collider at the European Organization for Nuclear Research, CERN, located in the Swiss-French border. It was constructed in the tunnel of the former Large Electron Positron (LEP) collider which had a circumference of 27 km. It has been designed to accelerate proton beams to a centre of mass energy ($\sqrt{s}$) up to 14 TeV at a maximum luminosity of 10^{34} cm$^{-2}s^{-1}$. However, these nominal values will be achieved only once it becomes fully operational in 2015.

The accelerator complex at CERN is a succession of machines with increasingly higher energies. Each machine injects the beam into the next one, which takes over to bring the beam to an even higher energy and so on. Each of the LHC injectors has its own experimental hall, where the beams can be used for dedicated experiments. The brief description of a proton accelerated through the accelerator complex of CERN is illustrated in Figure 3.1. Protons are obtained by extracting orbiting electrons from

hydrogen atoms. The protons, then, begin their tour in the linear accelerator (LINAC 2), from which they are injected into the Proton Synchrotron (PS) Booster at an energy of 0.12 GeV. The Booster accelerates them to 1.4 GeV. The beam is then transferred to the PS where it is accelerated to 26 GeV. Protons are then sent to the Super Proton Synchrotron (SPS) where they reach 450 GeV, and finally they are transferred to the LHC where they are accelerated to their nominal center of mass energy, 14 TeV (7 TeV in 2011 and 8 TeV in 2012). The beams will counter-rotate for several hours before colliding at the different points where detectors are positioned. Lead ions are produced using a source of vaporized lead before being sent into LINAC 3. They are then accelerated in the Low Energy Ion Ring (LEIR) and take the same route as the protons from the PS accelerator.

The protons are "bunched" together in the LHC rather than to have a continuous beams of particles. A bunch separation of 25 ns (75/50 ns in 2011 and 50 ns in 2012) is maintained with trains of 72 occupied and 12 empty bunches. Out of the 3564 bunch spaces available during each cycle, 2808 are filled.

Different kinds of superconducting cavities (SC) and magnets (dipoles, quadrupoles, sextupoles, decapoles, etc.) are set along the LHC ring in order to accelerate the protons till almost the speed of light and bend their trajectories. The LHC has more than 1200 superconducting magnetic dipoles of 8.3 T operating at a temperature of 1.9 K. The machine parameters relevant for the operation of the CMS detector are listed in Table 3.1, where the nominal values of the parameters designed to reach a center of mass energy of 14 TeV are compared with the values set for the 2010, 2011 and 2012 running periods.

Parameter	2010	2011	2012	Nominal
Energy per beam, $[\text{TeV}]$	3.5	3.5	4	7
Peak Luminosity $[10^{33}\,\text{cm}^{-2}\,\text{s}^{-1}]$	0.2	3.6	7.7	10
Bunch separation $[\text{ns}]$	150	75/50	50	25
Maximum number of bunches, k_B	368	1380	1380	2808
Particles per bunch, N_p $[10^{11}]$	1.2	1.5	1.7	1.15
Beta value at IP, β^* $[m]$	3.5	1.0	0.6	0.55

Table 3.1: LHC parameters for 2010, 2011 and 2012 compared with the nominal values. The point dependent parameter values (β^*, peak luminosity) are taken at point 5, where the CMS detector is located.
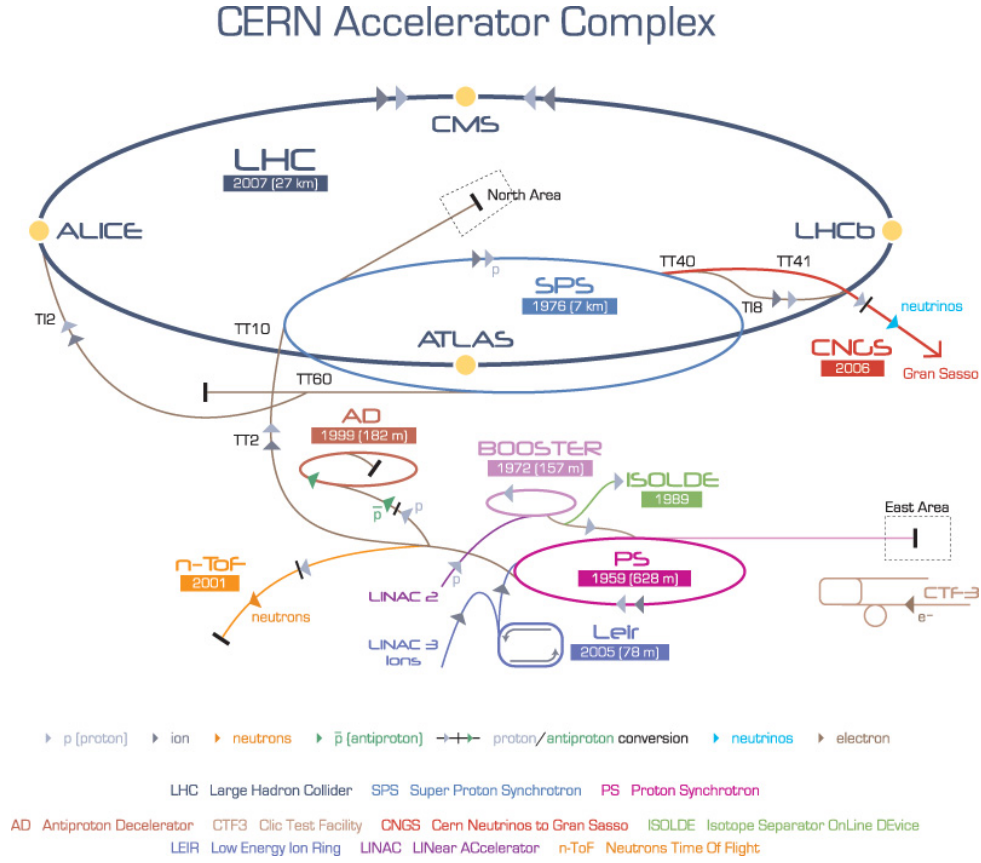
Figure 3.1: Large Hadron Collider scheme.

The luminosity ($\mathcal{L}$) is one of the most important parameter for a collider. It can be written as:

$$\mathcal{L} = \frac{1}{4\pi} \frac{\gamma f k_B N_p^2}{\epsilon_n \beta^*} F \tag{3.1}$$

Where γ is the Lorentz factor, f the revolution frequency, k_B the number of bunches, N_p the number of protons per bunch, ϵ_n is the normalized transverse emittance, β^* the betatron function at the impact point and F the reduction factor due to the crossing angle.

On September 10, 2008, the LHC started up with proton beams successfully circulating in the main ring for the first time. Nevertheless, a faulty electrical connection produced a chain of damages delaying further operations for fourteen months. In November 2009, the machine restarted to work and the first proton-proton collision

at $\sqrt{s} = 900\,\text{GeV}$ was recorded. Later, in March 2010, the LHC produced the world's highest energy collisions at $\sqrt{s} = 2.36\,\text{TeV}$. Successfully, the centre of mass energy was increased up to $7\,\text{TeV}$.

During 2011 the LHC delivered $5.74\,\text{fb}^{-1}$ of proton-proton collisions and the CMS detector recorded $5.21\,\text{fb}^{-1}$ of data, reaching an efficiency of about 91%. The 2011 proton-proton run started in mid March and ended at the end of October, when the heavy ion run started.

The 2012 run started with an increased center of mass energy of $8\,\text{TeV}$. The LHC delivered $23\,\text{fb}^{-1}$ of proton-proton collisions while the CMS detector recorded $21\,\text{fb}^{-1}$ of data. Three technical stops in April, June and September 2012 split the data in four different periods: A, B, C and D. The data studied in this document correspond to the A and B periods which correspond to a total luminosity of $5.3\,\text{fb}^{-1}$.

Along the LHC circumference there are six detectors: ALICE [28], ATLAS [29], CMS [30, 31], LHCb [32], LHCf [33] and TOTEM [34]. The first four detectors are installed in huge underground caverns built at points called 2,1,5 and 8 respectively. CMS and ATLAS are general purpose detectors that share the same physics goals, so a cross-check between their results can be made, although they have different designs. The other experiments are specialized in different topics, such as heavy flavor physics and precise measurements of the properties of the b quark in the case of LHCb, heavy ion studies for ALICE, and forward physics with TOTEM and LHCf located near to CMS and ATLAS detectors, respectively.

3.2 The LHC Detectors

3.2.1 A Large Ion Collider Experiment (ALICE)

The ALICE detector is optimised for high track multiplicities from heavy-ion collisions. Additionally to a silicon pixel detector and a transition radiation detector, a time projection chamber is used that can resolve many tracks in one event but is not designed for the high event rates which are achieved in proton collisions.

3.2.2 The Large Hadron Collider Beauty Experiment (LHCb)

The LHCb experiment is designed for the study of heavy flavour physics, especially CP violation and rare charm and beauty meson decays. At the LHC centre of mass energy charm and beauty mesons are predominantly produced in the forward and backward directions with small angles with respect to the beam line. Contrary to the general purpose experiments, the LHCb detector is a one-arm spectrometer that covers only a small angle in one direction along the beam pipe.

3.2.3 A Toroidal LHC Apparatus (ATLAS)

The ATLAS detector is a multipurpose detector with a length of 44 m and a diameter of 25 m, by its spacial dimensions the largest of the experiments. The detector design is in many aspects complementary to the design of CMS. The magnetic field is generated by a large toroid magnet and only one small solenoid. The most important advantage of ATLAS over CMS is the much better energy resolution of the hadronic calorimeter, while the momentum resolution of the tracking detectors is worse.

3.3 The Compact Muon Solenoid Detector

The Compact Muon Solenoid (CMS) is one of the two general purpose detectors at the LHC accelerator. About 183 institutions of 38 countries, with more than 3600 scientists take part in this collaboration. The Compact Muon Solenoid detector[30, 31] is a 4π multipurpose detector with a length of 28.7 m and a diameter of 15 m. This makes CMS smaller than the other main purpose detector, ATLAS, but much more dense, with a total weight of 14'000 tons.

The general structure of CMS is shown in Figure 3.2. The detector consists of different subdetectors, each with a well defined set of properties to measure within given physics requirements.

CMS adopted a coordinate system where the origin is centred at the nominal collision point of the beams, the y-axis points vertically upward, and the x-axis points radially inward toward the centre of the LHC. The azimuthal angle ϕ is measured from the x-axis in the XY plane, $\phi = arctan(y/x)$, and the polar angle θ is measured from the z-axis, $\theta = arctan(\sqrt{x^2 + y^2}/z)$. A particular reformulation of the polar angle provides
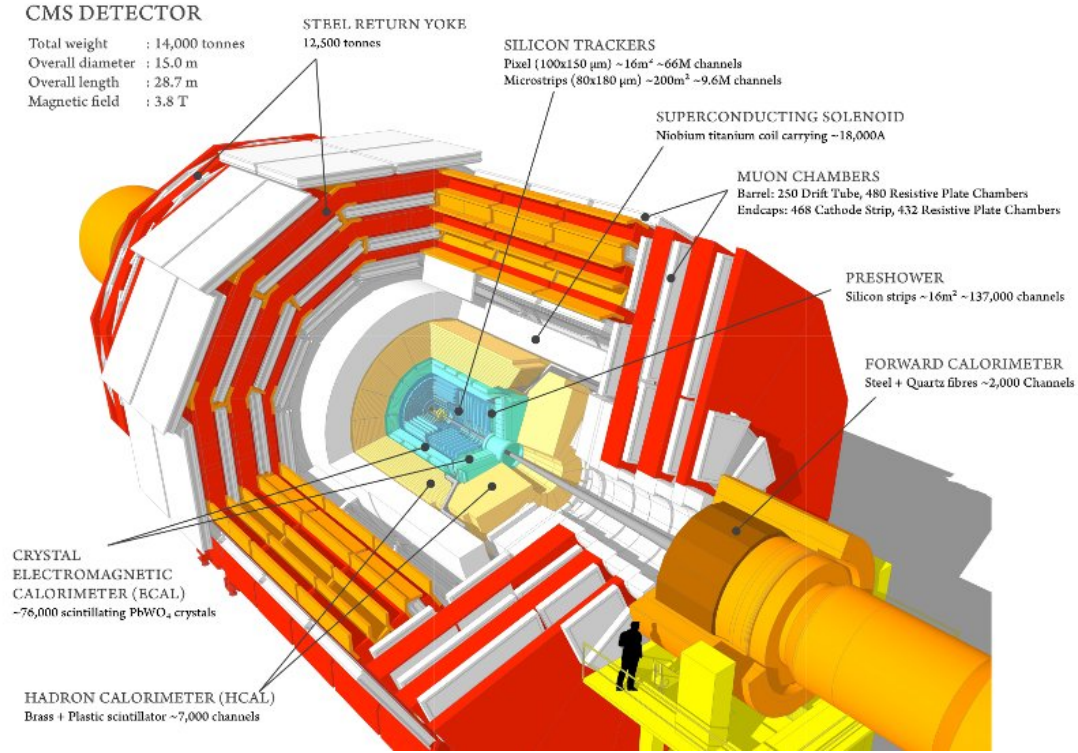
Figure 3.2: Perspective view of the Compact Muon Solenoid detector.

the *pseudo-rapidity η*, defined as:

$$\eta = -\ln \tan \left(\frac{\theta}{2} \right) \tag{3.2}$$

which is also obtained from the rapidity,

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right). \tag{3.3}$$

It is a convenient choice for the LHC as the multiplicity of high energy particles is roughly constant in η. In addition, rapidity intervals are Lorentz-invariant under boosts along the z axis.

The CMS detector has been designed to detect the signatures of new physics by identifying and precisely measuring muons ($\mu^\pm$), electrons ($e^\pm$), photons (γ) and hadrons over a large energy range. The detector requirements for CMS to meet the goals of

the LHC physics program can be summarized as:

- Good muon identification and momentum resolution over a wide range of momentum in the $|\eta| < 2.5$ region, good dimuon mass resolution ($\sim 1\%$ at 100 GeV), and the ability to determine unambiguously the charge of muons with momentum $p < 1\,\mathrm{TeV}$.

- Good charged particle momentum resolution and reconstruction efficiency in the tracker (Section 3.3.2). Precise localization of the primary interaction vertex. Efficient triggering and offline tagging of τ and b-jets, requiring pixel detectors close to the interaction region.

- Good electromagnetic energy resolution, good diphoton and dielectron mass resolution ($\sim 1\%$ at 100 GeV), wide geometric coverage ($|\eta| < 2.5$), measurement of the direction of photons, π_0 rejection and lepton isolation at high luminosities (Section 5.5.1 and 5.5.2).

- Good missing transverse energy and dijet mass resolution, requiring hadron calorimeters with a large hermetic geometric coverage ($|\eta| < 5$) and with fine lateral segmentation ($\Delta\eta$ x $\Delta\phi < 0.1$ x 0.1).

The innermost part of CMS is the silicon tracking detector which measures the momentum of the charged particle in the magnetic field. The tracker is enclosed by the electromagnetic calorimeter, which measures the energy of electrons and photons. Behind the electromagnetic calorimeter, the hadron calorimeter measures the energy of strongly interacting particles. The coil of the superconducting solenoid magnet encloses the previous subdetectors. Around the coil there are 4 stations of muon chambers embedded in the iron yoke of the magnet. Each muon station consists of several layers of aluminum drift tubes (DT) in the barrel region and cathode strip chambers (CSC) in the endcap region, complemented by resistive plate chambers (RPC). The different subdetectors of CMS will be briefly described in the next subsections.

3.3.1 Magnetic Field

The momentum of the charged particles are determined from their curvature in the $3.8\,\mathrm{T}$ (designed for $4\,\mathrm{T}$) magnetic field provided by a $13\,\mathrm{m}$ long superconducting solenoid magnet. It is designed to achieve a momentum resolution of $\Delta p/p \sim 10\%$

at p $= 1$ TeV, in order to have an unambiguous determination of the muon charge for muons with a momentum of ~ 1 TeV and the measurement of narrow states decaying into muons.

The distinctive feature of the 220 tons mass is the four-layer winding made from an established reinforced NbTi conductor. The energy store by the structure is so high that it can cause a large mechanical deformation during energizing. Some of the changes in the detector geometry due to the magnetic field are studied in Chapter 4.

Even if the solenoid was designed to reach 4 T [35], after some tests, and taking into account considerations regarding the lifetime of the system, it was decided that the magnet would operate at 3.8 T. This change does not have a significant impact in the detector performance.

3.3.2 Tracker

The CMS silicon tracker [36] has been designed to reconstruct vertices and the trajectories of the charged particles and the charged tracks within the jets over a large momentum range. Based on silicon semiconductor technology, the central tracker surrounds the interaction point and it has a length of 5.8 m and a diameter of 2.5 m. The CMS solenoid provides a homogeneous magnetic field of 3.8 T over the full volume of the tracker.

The CMS tracker is divided in two different subdetectors characterized by the flux of particles to which they are exposed. At a radius between 4.4 and 10.2 cm, it is located the pixel system, three cylindrical layers of pixel detector modules surrounding the interaction point complemented by two disks on each side. The other one is the silicon strip tracker, which is located at the radial region between 20 and 116 cm. It is composed of four different subsystems:

1. The Tracker Inner Barrel (TIB).

2. The Tracker Inner Disks (TID).

3. The Tracker Outer Barrel(TOB).

4. The Tracker EndCaps (TEC+ and TEC-).

The TIB and TID are composed of 4 barrel layers up to a radius of 55 cm, supplemented by 3 disks at each end. The TOB has an outer radius of 116 cm and consists of 6 barrel layers; it extends until $Z = \pm 118$ cm, surrounding the TIB/TID system. The inner two layers of the TIB and TOB modules are made double-sided with two back-to-back sensors at a relative angle of 100 mrad, providing precise measurements of the hits in two dimensions. Beyond this z range, the TEC+ and TEC- cover the region 124 cm $< |z| <$ 282 cm and 22.5 cm $< |$ radius $| <$ 113.5 cm. Figure 3.3 shows a schematic vision of the CMS tracker with its different subsystems.
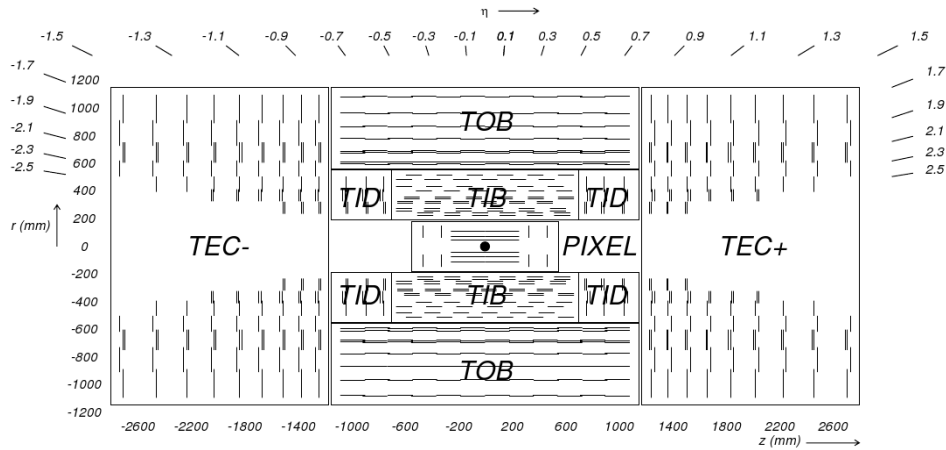
Figure 3.3: CMS tracker. Each line represents a detector module. Double lines indicate double-side modules.

3.3.3 Calorimeter

Electromagnetic calorimeter

The electromagnetic calorimeter (ECAL) [37] is the next subdetector enclosing the tracking system and it is as well located inside the solenoid. It has been designed to measure precisely the energy of photons, electrons and jets as a result of electromagnetic showers. The ECAL has a good energy resolution provided by a homogeneous crystal calorimeter. It can be subdivided into barrel, endcap and preshower.

The barrel part (electromagnetic barrel, EB) covers the pseudorapidity range $|\eta| <$ 1.479, while the endcaps (electromagnetic endcap, EE) the range 1.479 $< |\eta| <$ 3.0. They consist of $\sim$ 68524 lead tungsten ($PbWO_4$) crystals which serve as absorber as well as scintillator. Each of the crystal covers a solid angle of $(\Delta\eta, \Delta\phi) = (0.0174, 0.0174)$.

The aim of the preshower detector is to identify neutral pions in the endcaps at the region $1.653 < |\eta| < 2.6$, where the angle between the two photons from the decay of the neutral pion is small enough to fake the signal of just one photon with larger energy. It also helps the identification of electrons against minimum ionizing particles, and improves the position determination of electrons and photons with high granularity.

Hadronic calorimeter

The hadronic calorimeter (HCAL) [38], which surronds the ECAL calorimeter, measures the energy of the hadrons and their decay products. It is made of copper layers interleaved with scintillator material. The HCAL is divided into the barrel (hadronic barrel, HB), the endcaps (hadronic endcap, HE), the outer barrel (hadronic outer, HO), and the forward (hadronic forward, HF) calorimeters.
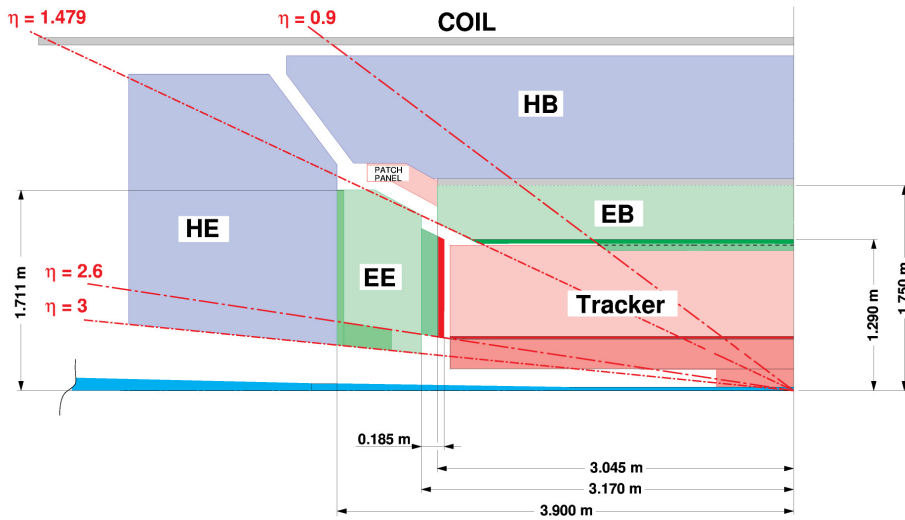
Figure 3.4: Overview of the calorimeter layout with the acceptance of its different components. Outside of the solenoid, the HO and HF calorimeters respectively improve the number of radiation lengths and the acceptance of the calorimeter system.

Figure 3.4 shows a longitudinal cut of a quarter of the CMS plane with the ECAL and the HCAL calorimeters. The HB is radially restricted between the outer part of the electromagnetic calorimeter and the inner part of the magnet coil. This constrains the total amount of material which can be put in to absorb the hadronic shower. The HO calorimeter (not shown in the Figure 3.4) is placed on the central wheel complementing the barrel calorimeter outside the solenoid. The HF calorimeters (not shown in the

Figure 3.4) placed at 11.2 m in z from the interaction point extend the pseudorapidity coverage up to $|\eta| \approx 5.2$ using a Cerenkov-based radiation-hard technology.

3.3.4   Muon System

The muon subdetector [39] is a tracking system placed in the outermost region of the CMS detector. Only muons and non-interacting particles, such as neutrinos, manage to pass through the calorimeters without depositing a large fraction of their energy. The muon system, installed outside the magnet coil, identify this kind of particles and complement the measurement of their momentum (performed by the tracker) using the reflux magnet.

The muon system uses three different types of detectors: drift tubes (DT, in the barrel) covering the pseudorapidity range $|\eta| < 1.2$, cathode strip chambers (CSC, in the endcap) covering the range $0.9 < |\eta| < 2.4$, and resistive plate chambers (RPC) going up to $|\eta| \approx 1.6$.

Drift tube chambers

The barrel muon detector consists of 4 stations forming concentric cylinders around the beam line. Drift tube chambers are installed in and on the wheels of the iron yoke. The three inner cylinders have 60 DT each and the outer cylinder has 70. In each sectors of the yoke there are 4 muon stations per wheel, labeled as MB1, MB2, MB3 and MB4. The first station in each sector is installed on the inner side of the yoke barrel rings, the second and third ones in cavities inside the yoke, and the fourth one on the outside. Figure 3.5 shows the DT arrangement in one of the five wheels of the CMS detector.

The DTs are filled with a mixture of 85% CO_2 and 15% Ar of active gas. The maximum drift distance is 2.1 cm corresponding to a drift time of 380 ns.

Cathode strip chambers

Cathode strip chambers are used to detect muon tracks in the endcaps. They can work reliably in the strong and inhomogeneous magnetic field in this region. The CSCs are arranged to form four disks, called ME1, ME2, ME3, ME4. The disk ME1
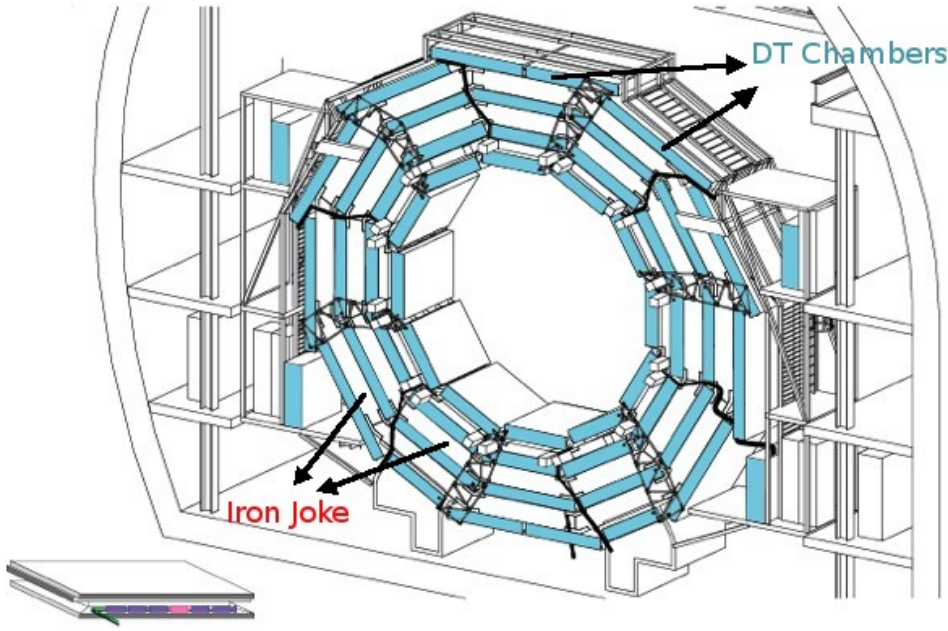
Figure 3.5: Layout of the CMS barrel muon DT chambers in one of the 5 wheels.

is formed by three rings of chambers (ME1/1, ME1/2, ME1/3), while ME2 and ME3 are composed by two rings of chambers (MEn/1 and MEn/2), and finally, ME4 has only one ring (ME4/1).

The CSC chambers are trapezoidal and cover either $10°$ or $20°$ in ϕ. A muon in the pseudorapidity range $1.2 < |\eta| < 2.4$ crosses 3 or 4 CSCs. In the endcap-barrel overlap range, $0.9 < |\eta| < 1.2$, muons are detected both by the barrel DTs and endcap CSCs.

The nominal gas mixture used to fill the CSCs chambers is $40\%Ar$, $50\%CO_2$ and $10\%CF_4$. The CSCs allow precision muon measurement and muon trigger in one device. The chamber position resolution varies from 75 to $200\,\mu$m from the first to the last station to cope with the CMS goal for momentum resolution.

Resistive plate chambers

The resistive plate chambers are dedicated to trigger purpose due to their very fast response and time resolution of the order of 25 ns. This performance is achieved by its structure of gaseous parallel-plate detector. A total of 6 layers of RPCs are embedded in the muon barrel system. In the endcap region, there is a layer of RPCs in each of the first 3 stations, providing an acceptance up to $|\eta| \approx 1.6$. They overlap in ϕ in order

to avoid dead spaces between chambers. Figure 3.6 shows the RPC location as well as the DT and CSC chambers.
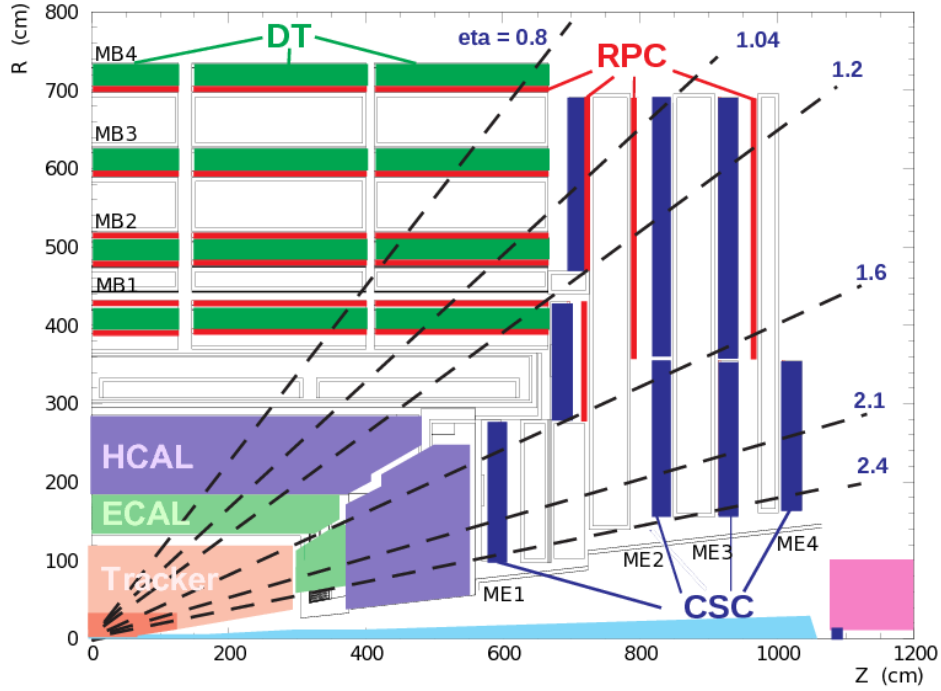
Figure 3.6: The CMS muon system illustrated for a quadrant of the CMS detector.

3.3.5   Alignment System of the CMS Detector

The extremely precise and complex muon tracking system (silicon tracker and muon spectrometer) needs a very accurate installation of its components in order to take advantage of the intrinsic chamber resolution. Additionally, the high field of the solenoid can produce significant variations of the detector geometry each time that the magnet is turned on. These factors make necessary the implementation of a combination of opto-mechanical and track-based alignment system.

The detector geometry reconstruction using opto-mechanical alignment system starts with precise measurements of the survey and/or photogrammetry [40]. The opto-mechanical system allows the continuous measurement of the position of the chambers during operation. In order to provide precise information about the relative position of the muon chambers, in the barrel and endcaps, and of the various tracker components, the alignment system is organized in three blocks:

- A tracker internal alignment, to measure the positions of the various modules and monitor the eventual internal deformations.

- A muon alignment system, to monitor the relative position among the chambers in the muon barrel and endcaps detectors.

- A link alignment system, that allows to relate the position of the various elements of the muon system (barrel and endcaps) with respect to the tracker and to monitor the eventual relative movements of elements between both subsystems.

The link alignment system is explained in more detail on Chapter 4.

With the accumulation of data collision, a track-based alignment of the tracking system becomes applicable and reduces the alignment uncertainties. The track-based alignment uses tracks from low and high mass resonances (J/Ψ, W and Z) for the alignment of the silicon tracker modules and muon chambers. The precision reached for the pixel detector is of the order of $20\,\mu$m, while for the strip tracker it is $\sim 30\,\mu$m. Additionally the whole muon system is positioned with respect to the inner tracker with a precision of 1 mm and $250\,\mu$rad.

3.3.6   Trigger System

The huge amount of collisions produced by the LHC makes impossible to record all the events by the acquisition system. Besides, there is no reason to store every collision due to the fact that most of them have well known physics (inelastic and/or low energy interactions). The event rate at LHC, $\sim 10^9$ Hz, must be reduced up to $\sim$ $100 - 400$ Hz. It is necessary to design a proper trigger system to reject no interesting events. The trigger must be quick in the reconstruction and identification of the objects and decide if the events are accepted or not due to the very high frequency of the bunch crossing. The trigger system is divided into two sequential levels: the Level 1 Trigger (L1) and the High Level trigger (HLT).

Level 1 trigger

The Level-1 trigger (L1) has been implemented on dedicated hardware and has only access to data from the calorimeters and the muon detectors. This information is used

to identify muons, electrons, photons and jets and permits to make a rough estimation of the p_T of the particles, using only hardware information. The L1 trigger decision has to be taken in 3.2 µs, that is, after 128 bunch crossings. To avoid any data loss, the L1 trigger electronics is housed partly on the detectors and partly in the underground control room.

The Level-1 trigger is divided into three subsystems: the calorimeter trigger, the muon trigger and the global trigger. The muon trigger is further subdivided into three independent systems for the DT, CSC and RPC detectors. Figure 3.7 shows a schematic view of the components of the Level-1 trigger system.

The decision if an event is kept is based on the reconstructed trigger objects and quantities but also on the readiness of the sub-detectors and data acquisition system (DAQ). Once accepted, the event is passed to the High Level Trigger (HLT), which is implemented in software running on a computer farm.
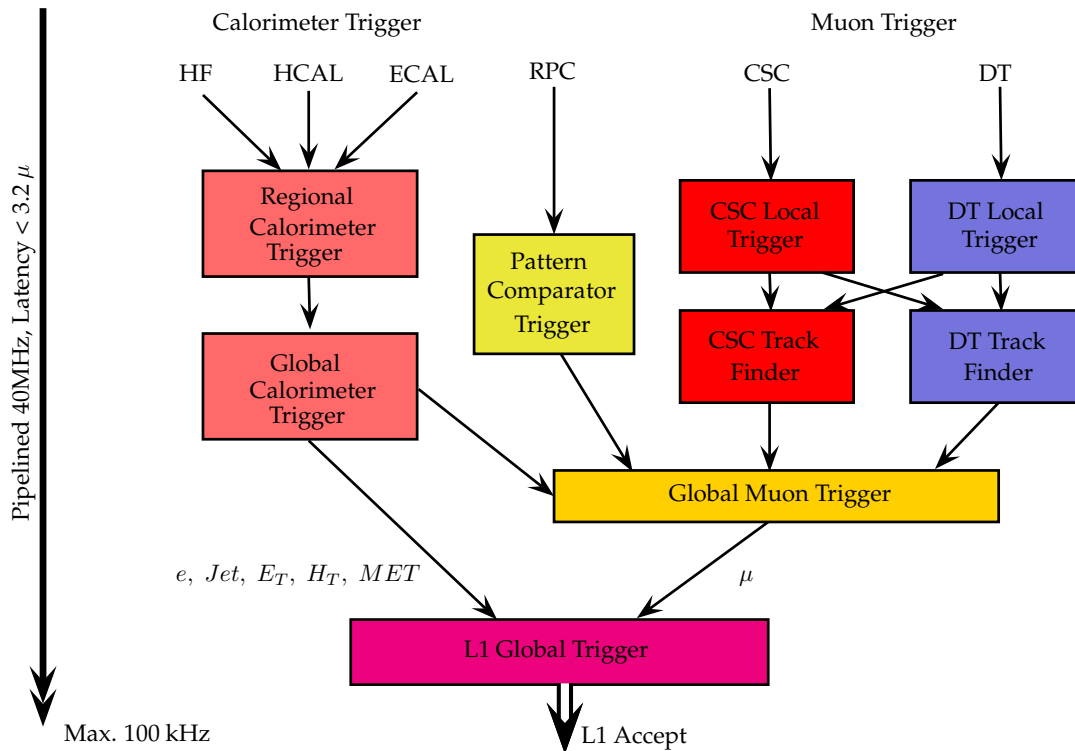
Figure 3.7: Structure of the Level-1 trigger system at CMS.

High level trigger

The rate of events is reduced from 10^9 Hz to ≈ 100 kHz at Level-1. At the high level trigger the rate must reach 100-400 Hz [41], so a factor $\approx 10^3$ reduction must be achieved by using intermediate steps called level two (L2) and level three (L3). There are four systems at the HLT especially designed for the identification of different types of objects: electrons/photons and muons, taus, jets (including b-quark jets) and finally missing transverse energy. The trigger system is organized in trigger paths, which can contain several trigger filters in sequence.

As an example of a trigger, the HLT_Mu17_Mu8_v6 trigger path (which is used later in this thesis) consists of two filters in HLT level. The parameters of the L1 muon candidates are used to seed the reconstruction of a L2 standalone muon, i.e. a muon reconstructed using only information from the muon system. On these L2 candidates, a filter is applied that requires at least two muons with a transverse momentum $p_T > 7$ GeV for the leading muon and $p_T > 0$ GeV for the second leading muon. Having passed the L2 filter, at L3 the muon candidates are reconstructed including also information from the silicon tracker.The L3 filter requires at least two muons, with $p_T > 17$ GeV for the leading muon and $p_T > 8$ GeV for the second leading muon [42].

CHAPTER 4

LINK ALIGNMENT SYSTEM

The purpose of the link alignment system is to measure the relative positions of the muon spectrometer, tracker body, and of the CSC chambers ME1/1 and ME1/2 in a common reference system. The target precision of these measurements is $\sim 150\,\mu$m in position and $\sim 40\,\mu$rad in orientation.

The entire link system is divided into three longitudinal planes $60°$ apart starting at $\phi = 15°$ as is shown in Figure 4.1. Each plane generates four independent quadrants (two in the positive z side and two in the negative one), resulting in twelve laser lines, six in each z side of the CMS detector.

A distributed network of semitransparent Amorphous Silicon Position Detectors (AS-PDs) [43] placed around the muon spectrometer is connected by the laser lines. These 2D sensors with an active area of $28 \times 28\,\text{mm}^2$ and $1\,\text{mm}$ thick, are divided in 64 horizontal and 64 vertical strips. The light spot of the laser is obtained in the form of two ortogonal intensity profiles, which are fitted to gaussian distributions to obtain the position in each sensor coordinate. Figure 4.2 shows a picture of a complete ASPD unit.

The laser-sensor network is complemented by optical and mechanical probes in order
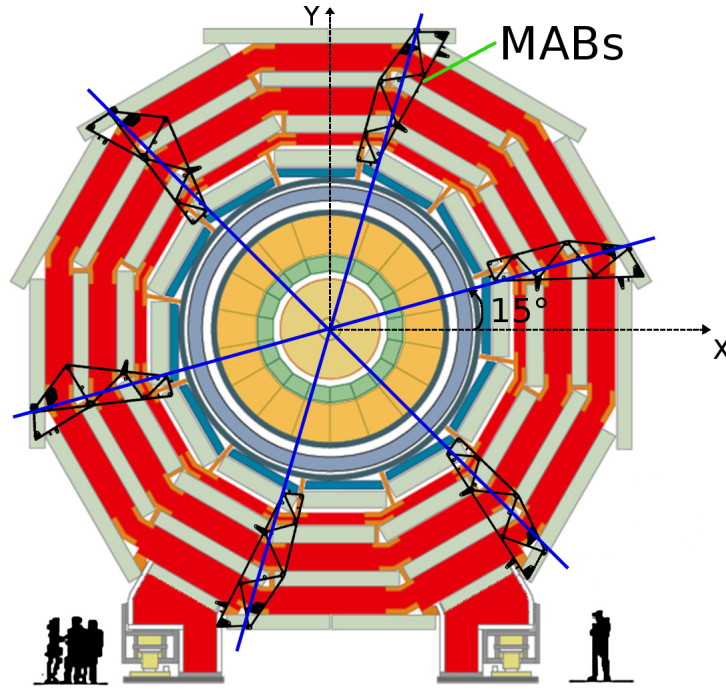
Figure 4.1: Transversal view of the CMS detector showing the optical alignment paths. Each blue path has its corresponding Module for the Alignment of the Barrel (MAB).

to control the possible displacements (in $z - R\phi$ coordinates) of the different structures associated to the link alignment system. In the specific case of the distancemeters, they are conected to two kind of structures: to the radial profiles (RP) and to the longitudinal profiles (LP) monitoring the relative displacements between elements along the radial ($R\phi$) and longitudinal (z) direction, respectively. Temperature probes are used to monitor possible temperature effects on the mechanical components, especially in aluminium structures. Finally, the reconstruction performed with the link alignment system is complemented with measurements from electronic inclinometers to monitor the spacial orientation of the main structures.

Figure 4.3 shows the link alignment components in a quarter of $R - z$ plane. The whole link system is formed by 12 of these sectors. The three lasers showed in the figure are originated in three different components of the system: tracker (Alignment Rings), barrel (Modules for the Alignment of the Barrel) and endcaps (Link Disks).

AR The Alignment Ring has an internal radius of 240 mm and an external radius of 365 mm and it is made of thick carbon fiber material. There are two AR located
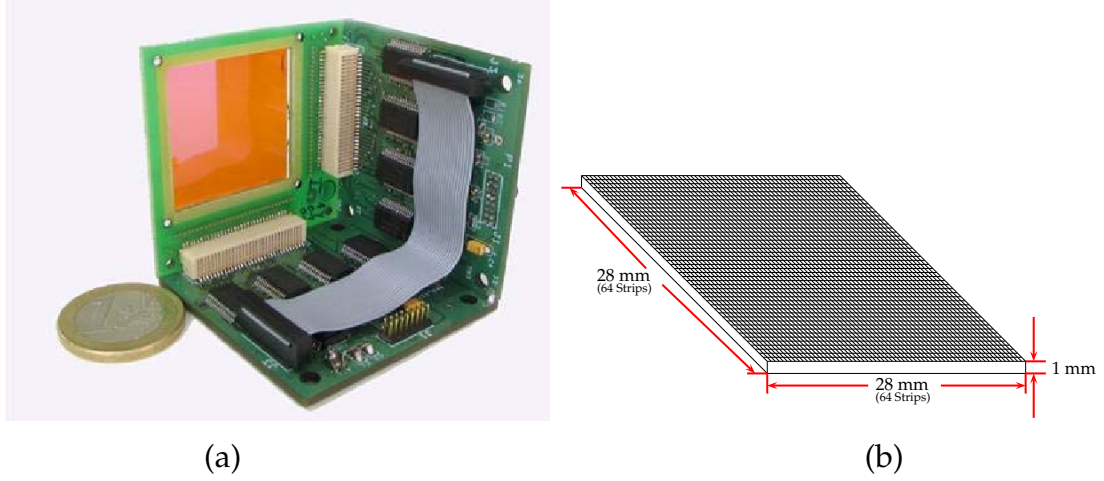
Figure 4.2: (a) Complete ASPD unit and, (b) a sketch of the active sensor area with a 64 × 64 arrange of strips.

in the $\pm z$ end faces of the tracker. Their position and orientation is well known and stable with respect to the tracker. One of the purpose of the AR is to monitor the positions and orientations of the muon chambers with respect to the tracker with a precision of $\sim 200\,\mu$m.

MABs The Modules for the Alignment of the Barrel (MAB) are 36 rigid structures made of carbon fixed to the barrel yoke. The MABs are distributed between the weels of the detector in groups of six, every 60° in azimuthal angle ϕ (in the same planes of the link system, see Figure 4.1). Outer MABs, located in the external weels of the detector, are components of the endcaps and the link alignment system. The link instruments on the outer MABs are two ASPD sensors, one distancemeter and an inclinometer.

LD The Link Disk is a piece of carbon located in the external part of the iron disk of the two endcap muon spectrometers (YN1 structures), supported through 3 radial profiles (RP) connected with their corresponding transfer plates (TP). It has attached three longitudinal profiles (LP) which are connected with the AR by proximity sensors. Additional to these LPs on the LD, there are two inclinometers for angular monitoring and six laser boxes, one for each outer MAB.
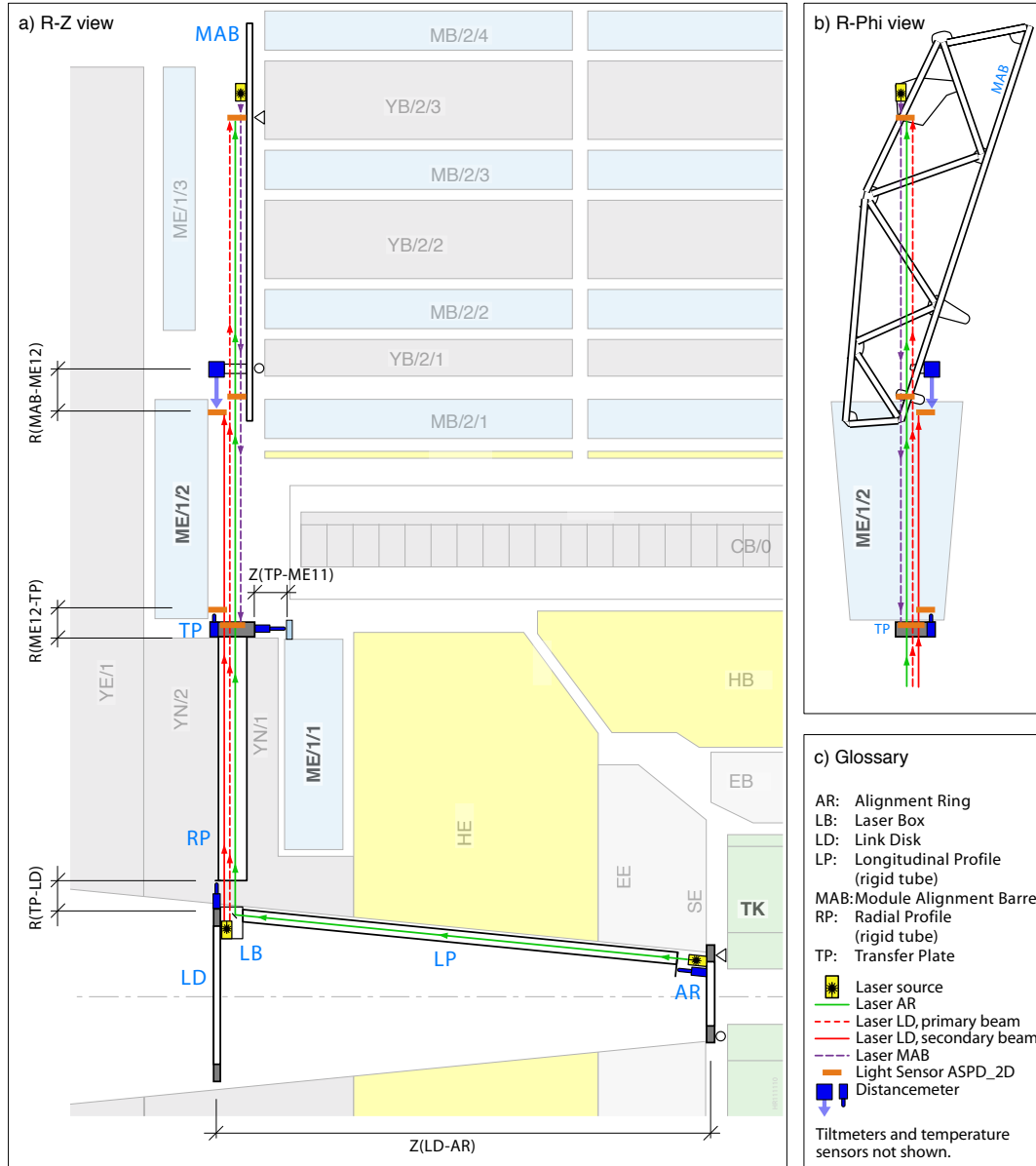
Figure 4.3: Link alignment system in a quarter of $R - z$ plane. Inclinometers and temperature probes are not represented in this figure.

4.1 The CMS Object Oriented Code for Optical Alignment

The CMS Object oriented Code for Optical Alignment (COCOA) [44] is a software based in C++ language designed to reconstruct geometrical positions and angular orientations of an optical-mechanical system. It includes an error propagation of the measurements involved in the geometrical reconstruction. The aim of the software is to construct, in an automatic way, the model of the system through the measurements of the different system elements, the interactions and hierarchical dependences. The derivatives of the coordinates and orientations of the system elements with respect to the measurement values are obtained by a numerical method.

The main optical and mechanical objects used in the reconstruction by COCOA are lasers, lens, mirrors, splitters, ASPD, distancemeters and inclinometers. The user can define additional objects, even more complex than the previous ones.

4.1.1 System Description

To provide to COCOA all the information needed to perform the system reconstruction, it is necessary to create a *System Description File* (SDF) with a special format [45]. This file includes the interconnection elements (as lasers, sensors, etc), the hierarchy (how the elements are attached between them), the known and unknown parameters, and an approximation of the geometry obtained by the nominal positions and orientations or by previous measurements like calibrations and photogrammetry. The SDF structure can be divided as:

1. Global Options: it contains the list of default parameters to be taken into account during the reconstruction.

2. Parameters: they define global values used many times by the SDF to fill the information tree.

3. System Tree Description: it describes the objects contained in the system with their corresponding hierarchy.

4. System Tree Data: it includes the name, position, orientation, and any other information of every object defined in the system tree description.

5. Measurements: it is the COCOA input file with all the measurements from the sensor devices.

4.2 Measurement Strategy

The link alignment reconstruction is divided in three steps. In each step, the position of one main structure (LD, MABs and AR) is adjusted. In the following paragraphs, there is a brief description of each step taking as a reference a quadrant (see Figure 4.3) of the alignment system, also named a *line*. This reconstruction is performed for each of the 12 lines that set up the whole link alignment system.

- **Step 1:** in this step, LD, ME1/1 and ME1/2 are adjusted into the endcaps (Figure 4.4). Two laser beams are originated at the LD collimator and sent radially out. One of this light path, called primary beam, will impact one sensor in the TP as well as the two MAB sensors. The secondary beam, parallel to the primary one at 5 cm, will instead reach the two sensors located in the ME1/2 structures after hitting a sensor in the TP. In addition to the optical system, in this step is used the information provided by the distancemeters located between the LD and the RP and those located between the TP and the ME1/1 and ME1/2 chambers. At the end of this adjustment, the position of the ME1/1 and ME1/2 chambers with respect to the LD is determined.

- **Step 2:** adjustment of the MABs with respect to the LD (see Figure 4.5). Two laser beams are used in this step: The *primary* beam (red path) going from the LD to the MAB, and the *secondary* beam (green) sent from the MAB to the TP. As it was explained in step 1, the primary beam impacts the sensor located in the TP before it reaches the bottom and the top sensor of the MAB. The secondary beam is sent in the same radial direction as the primary, but in opposite sense: it will impact the MAB sensors before reaching the TP sensor. This adjustment uses additional information from the distancementers located between the ME1/2 chambers and the MABs, named *OMRONs*. At the end of this step, the MABs position with respect to the LD is adjusted.

- **Step 3:** adjustment of the LD-MABs system with respect to the AR (Figure 4.6). A laser ray generated at the AR (near to the tracker), follows a path in $|\eta| \sim 3$ direction up to a splitter located in the LD. After the LD deflection, the beam

follows the same path of the *primary* LD radial beam.

Three contact distancemeters placed in the AR at $\phi = 75°$, $195°$ and $315°$ with an angle of $5.7°$, are used to measure the z distance between the LD and the AR. These distancemeters are in contact with a target located at the end of the longitudinal profiles (LP), which in turn are attached to the LD. Because the LP material is aluminium, the effects of changes in temperature should be taken into account at this point (Section 4.3). At the end of this step, the system LD-MABs is adjusted with respect to the tracker, which is considered as fixed.
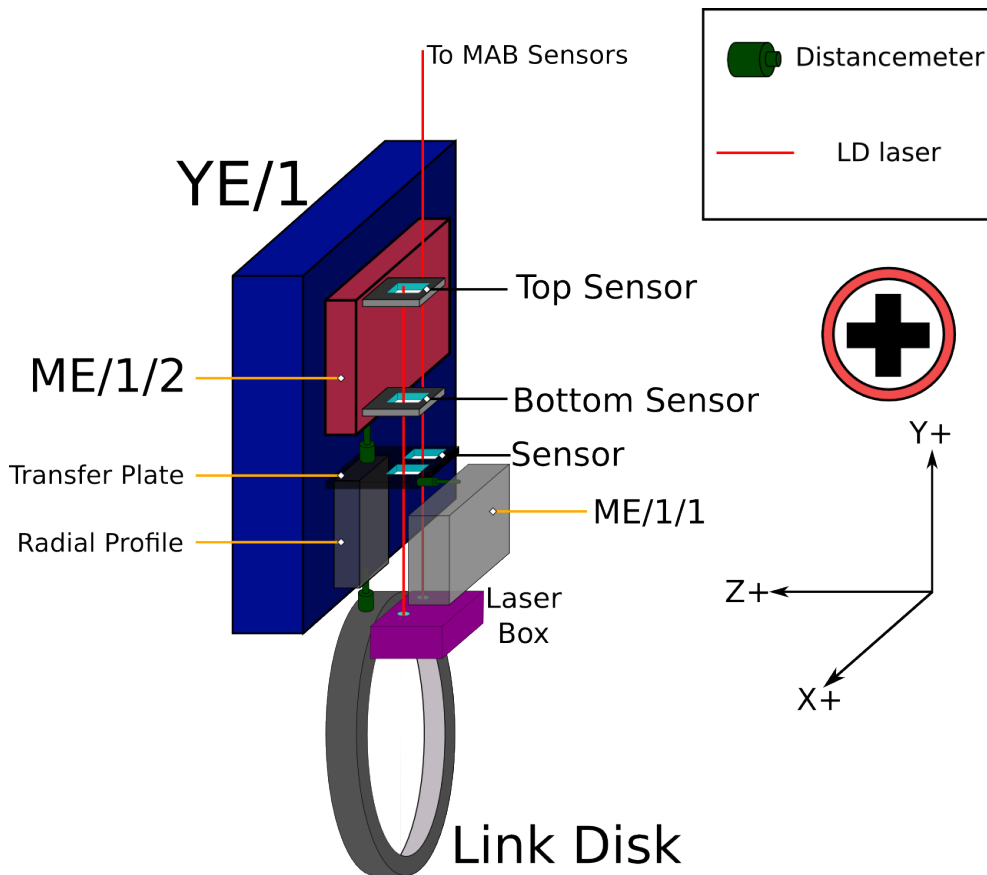
Figure 4.4: In the step 1, the adjustment of the LD, ME1/1 and ME1/2 chambers into the endcaps is performed. The laser beams from the LD run radially to the MAB (primary ray) and to the ME1/1 chamber (secondary ray). The sketch is for a laser path located in the positive Z region of the CMS detector.
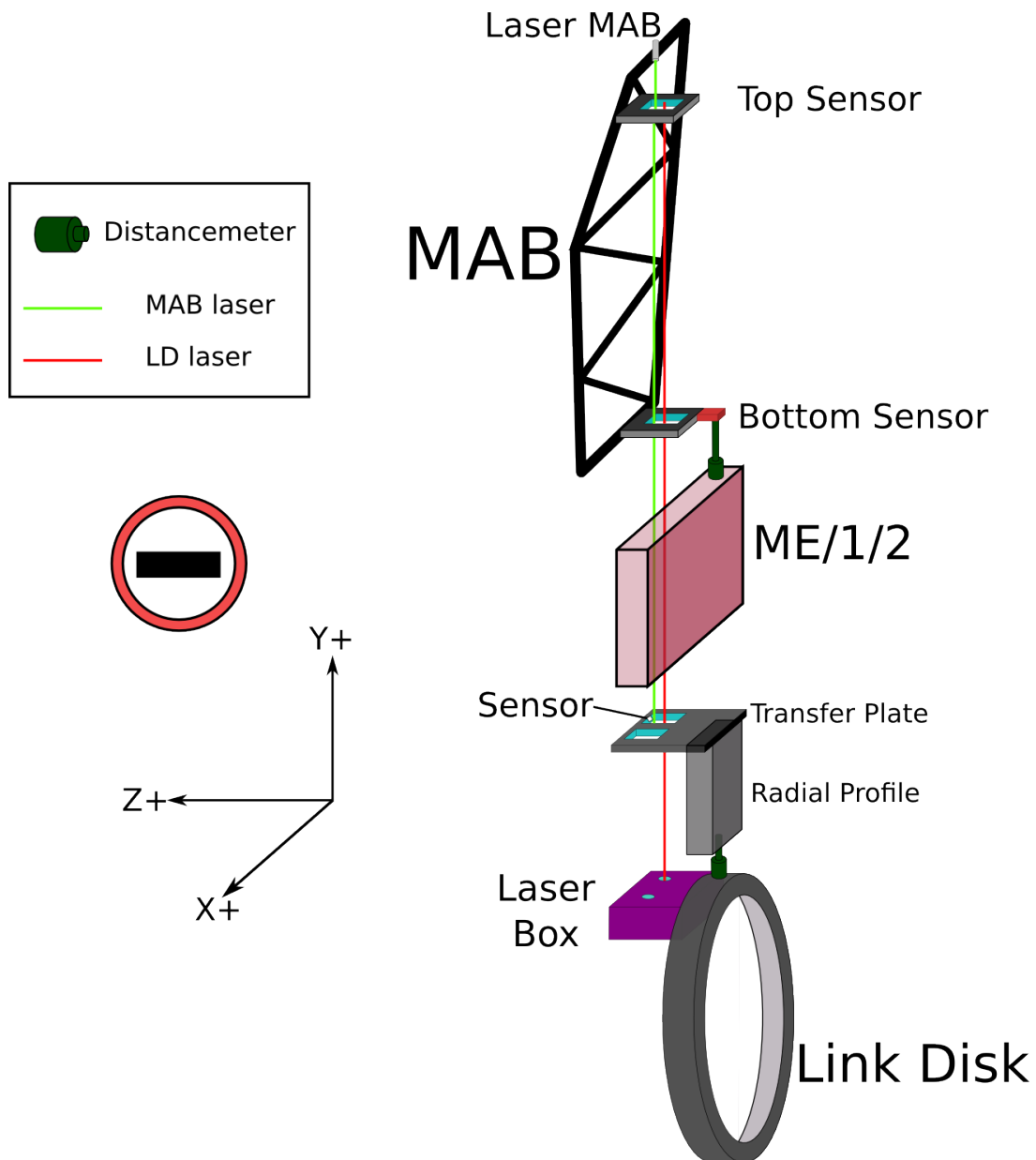
Figure 4.5: In the step 2, the adjustment of the MAB positions and orientations with respect to the LD is performed. The primary ray from the LD to the MAB (red line) and the ray generated at the MAB and running radially to the TP (green line) are shown. The sketch is for a laser path located in the negative Z region of the CMS detector.
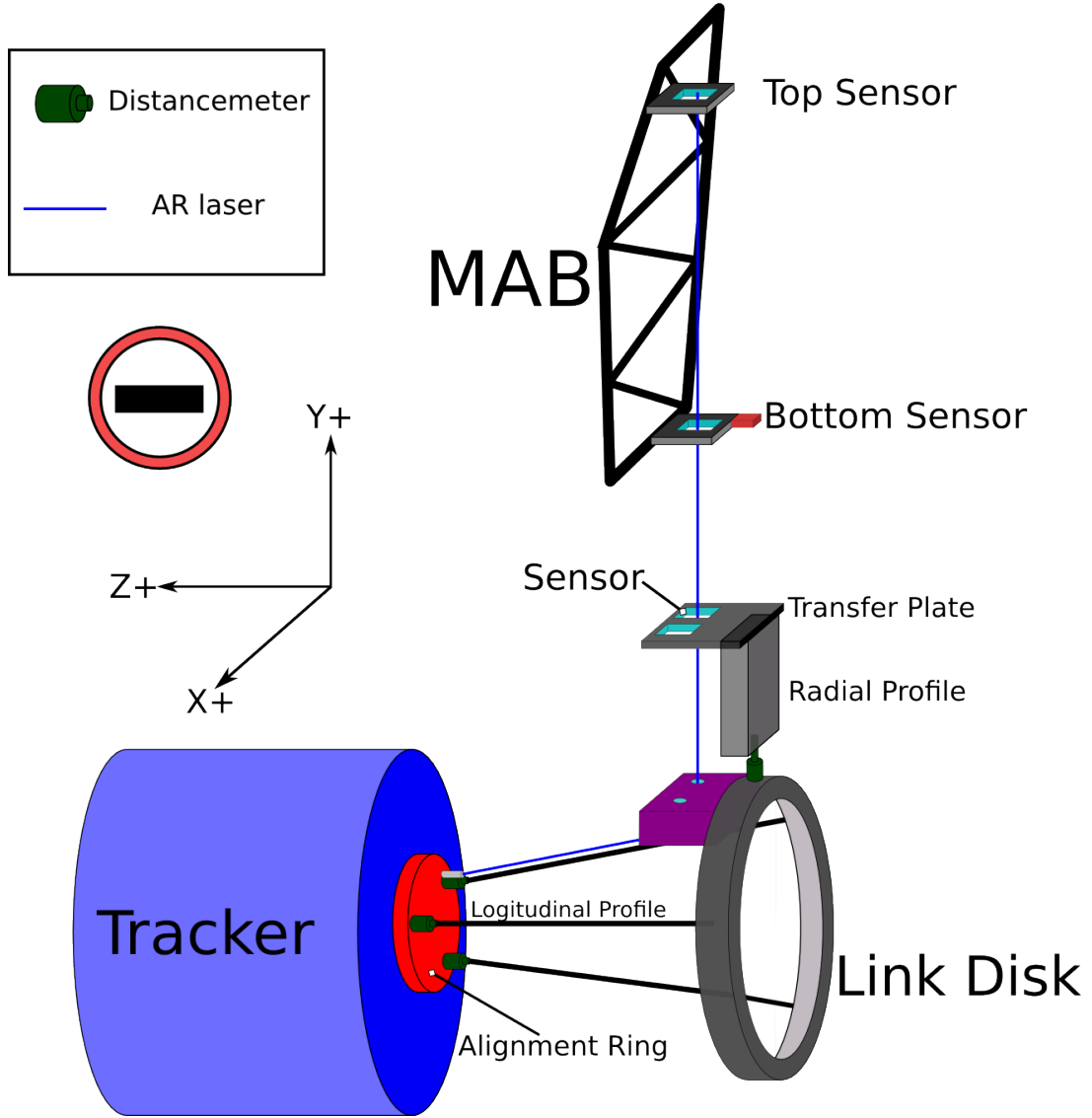
Figure 4.6: In the step 3, the adjustment of the system LD-MABs with respect to the AR (tracker) is performed. A laser ray (blue line) travels from the AR with $|\eta| = 3$ direction to the LD where it is deviated radially to the MAB sensors (blue line). The sketch is for a laser path located in the negative Z region of the CMS detector.

4.3 Temperature Corrections

The high precision of the measurements performed by the link alignment system makes necessary to consider the effects of changes in temperature in the different parts of the system. The most relevant effects can be dilatation or contraction of the structures and changes in the refractive index of the air in which the lasers propagate.

The data of the different sensors impinged by the AR lasers has been taken continuosly during 4 days in order to quantify the temperature dependence of the measurements. The study has been performed only with the AR lasers since the temperature changes near to the tracker are approximately $6 - 7°C$. These changes in the temperature come from the tracker cooling system.

The coordinates in the sensor frame can be parametrized as:

$$x = p_{1x}T + p_{0x}$$
$$y = p_{1y}T + p_{0y} \tag{4.1}$$

As we need just to extrapolate the measurement from one calibration temperature to any other, these equations can be written as:

$$x' = x + p_{1x}\Delta T$$
$$y' = y + p_{1y}\Delta T \tag{4.2}$$

Where x' and y' are the corrected measurements, ΔT is the difference between the calibration temperature and the temperature in which the measurements are performed ($\Delta T = T_{cal} - T$), while x and y are the measured coordinates at the calibration temperature. Figure 4.7 shows the slopes obtained for the top and bottom sensors located in the first MAB (15°) of the positive Z side of the CMS detector. The constants p_{1x} and p_{1y}, obtained from the fits of Figure 4.7, are presented in Table 4.1 for the TP and MAB sensors hit by the AR lasers. The information of some sensors is missing because they were not reached by the lasers in the data taking process.

Table 4.2 shows the effect of the temperature correction in a specific MAB sensor. The dataset used in this example was taken in a small time window, without any variation in the magnetic field but at different temperatures due to the tracker cooling system. Considering that there is no real displacement of the detector structures, it is not expected any change in the probe measurements (distancemeters and sensors). Any
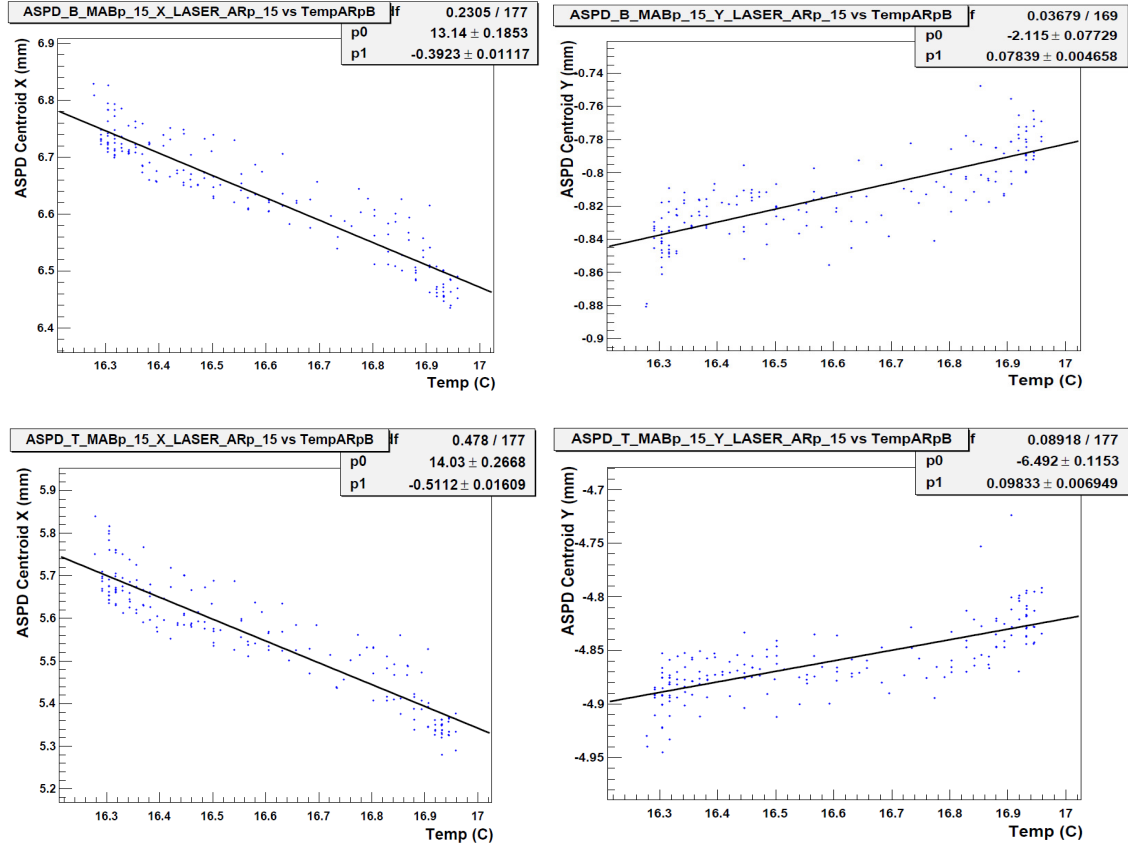
Figure 4.7: Example of the temperature dependency of the measured x (left) and y (right) coordinates in the sensor frame for the bottom (top) and top (bottom) ASPD sensors in the MAB at $15°$ with the corresponding polinomial fit.

Sensor	$p_{1x}[\,\text{mm}/°C]$	$p_{1y}[\,\text{mm}/°C]$
Positive Sector		
TP+195	0.314 ± 0.004	0.147 ± 0.005
TP+255	0.177 ± 0.007	0.106 ± 0.008
TP+315	0.253 ± 0.010	-0.014 ± 0.004
TP+15	0.310 ± 0.008	0.050 ± 0.003
TP+135	0.259 ± 0.004	0.037 ± 0.003
Bottom MAB+195	-0.415 ± 0.006	0.182 ± 0.006
Bottom MAB+315	-0.336 ± 0.015	-0.023 ± 0.011
Bottom MAB+15	-0.392 ± 0.011	0.078 ± 0.005
Top MAB+15	-0.511 ± 0.016	0.098 ± 0.007
Negative Sector		
TP-315	0.094 ± 0.004	0.290 ± 0.007

Table 4.1: p_{1x} and p_{1y} constants ($\,\text{mm}/°C$) for the ASPD sensors reached by the AR lasers. In the negative sector there was only one sensor reached by the laser due to a misalignment of the link disc.

variation is associated to the temperature effect. Results are more compatible with a no real displacement when the temperature correction is applied.

Sensor	coor	Measurement [mm]			Corrected meas. [mm]		
		Date 1	Date 2	Δ	Date 1	Date 2	Δ_{corr}
MAB at 315°	X	-7.554	-6.644	-0.909	-6.849	-6.846	-0.003
	Y	-6.054	-5.836	-0.218	-6.005	-5.850	-0.155

Table 4.2: Example of the temperature corrections for a sensor located on the top of the MAB at 315° on two different dates: **date 1**= May 3rd, 2010 with a $\Delta T = 2.1^o C$ and **date 2**= May 5th, 2010 with a $\Delta T = -0.6^o C$. The difference in the sensor coordinates between this two dates is shown in red. The same difference after applying the temperature correction is shown in green.

4.4   Geometrical Reconstruction Without Magnetic Field

The first link geometry reconstruction is performed when there is no magnetic field in the CMS detector. This represents the starting geometry of the detector which can be

compared with photogrammetry[1] measurements before any motion or deformation in the detector structures due to the magnetic field.

As an input, it is assumed that all YE±1 components (ME1/1, ME1/2 and TPs) are located in the positions measured by photogrammetry, while the LD has an unknown position. From the signals given by the LD lasers over ME1/2 and TP sensors the fit returns the LD position and orientation. Once LD position and orientation are known, its coordinates are fixed, the distancemeters present in $YE \pm 1$ are added into the fit, and all the $YE \pm 1$ components are allowed to vary within their calibrated positions. The result of this iteration is the new fitted position of the structures (TPs, ME1/2, ME1/1 and LD).

The following step is to determine the MAB positions, which are allowed to vary within the uncertainties given by the photogrammetry measurements. All the internal parameters of the MAB are constrained to their calibrated values and their location is obtained in the step 2.

Finally, with the full information on the position of the LD-MABs system, the AR collimator lasers are calibrated.

In situ calibrations

Using the information provided by the photogrammetry measurements and the previous geometry reconstruction, the position of the laser collimators located in the AR is calibrated. This process is necessary because the angular orientation of the lasers can change after their installation [46]. This process is called "in situ calibration" and it is performed without magnetic field. As explained in Section 4.3, the temperature is a relevant parameter for the geometrical reconstruction, so this calibration is performed at different temperatures.

4.5 Geometrical Reconstruction at $B = 3.8$ T

Due to the strong magnetic field produced in the CMS detector, some structures are displaced or rotated with respect to their positions when the magnet is turned off.

---

[1]Photogrammetry allows to determine the relative positions of some structures using photos taken to several physical targets fixed in them.

Consequently, it is necessary to reconstruct the new detector geometry every time the magnetic field is switched on. The starting point of this reconstruction is the geometry obtained at the end of the phase without magnetic field (see Section 4.4). The position and orientation of the AR is assumed to be the same as the one obtained from the fit at $B = 0\,\mathrm{T}$. AR is fixed and used as reference for the rest of the reconstruction. This link alignment reconstruction at $3.8\,\mathrm{T}$ is performed with the three steps presented in Section 4.2.

4.6   Results from the Link Alignment System

The alignment system allows to reconstruct the positions and orientations of $YE \pm 1$, $YB \pm 2$, $ME1/1$ and $ME1/2$ chambers, MABs and the two LDs.

Figure 4.8 shows the CMS and the local coordinate systems used to locate the components of the link system. While the $YE \pm 1$ disks, the $YB \pm 2$ wheels, and the AR are directly described in the CMS coordinate system, the ME1/2 and ME1/1 chambers, the LDs, and the MABs coordinates are related to their parent structures.

Table 4.3 and 4.4 show the reconstructed position and orientation of the main structures of the alignment system in February and August 2010 with a magnetic field of 0 and $3.8\,\mathrm{T}$, respectively.

The major change in the detector geometry due to the magnetic field is that the $YE \pm 1$ disks undergo a displacement of $\sim 14\,\mathrm{mm}$ in the Z axis direction towards the CMS interaction point (IP). This displacement agrees with the prediction reported in [47].

The motion of the endcap disks under the effect of the magnetic forces is quite complex. Besides the displacement in the Z direction, the endcaps experience deformations in the process of detector compression. One of the main results obtained with the link alignment reconstruction is the measurement of these deformations (see Figure 4.9). Some structures designed to prevent that the disks get pushed into the barrel (called Z-stops) together with the magnetic field distribution cause a bend in the endcap disks. The resulting bending angle of the inner ring of the first endcap iron, measured with respect to the transverse plane is found to be $4\,\mathrm{mrad}$ [46].

Figure 4.10 shows the q/p_T resolution as a function of p_T for muons reconstructed with the tracker and the muon system before and after aligned geometry. It can be
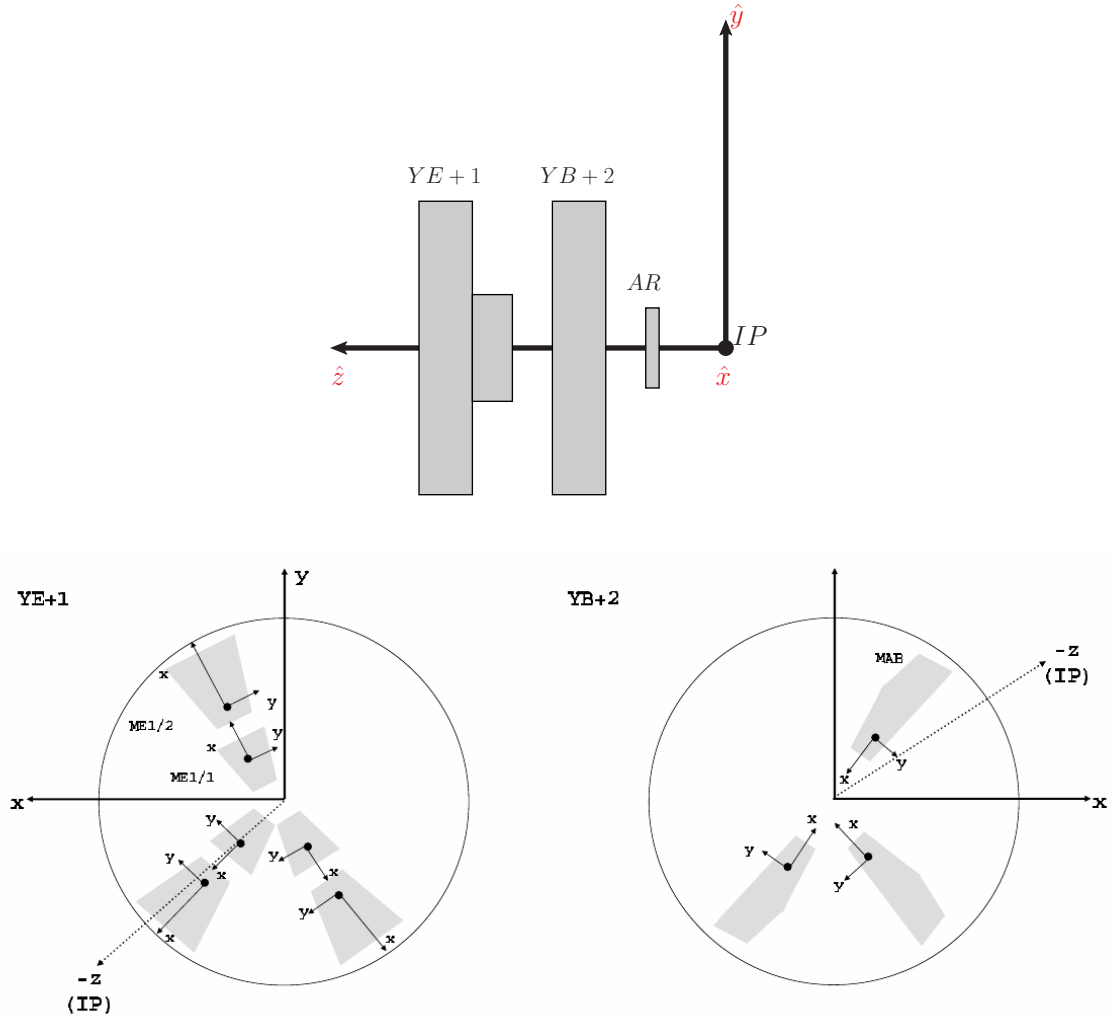
Figure 4.8: Top: the disk YE+1, the wheel YB+2 and the AR with respect to the CMS coordinate system. Bottom: local coordinate systems of the ME1/1 and ME1/2 chambers in YE+1 (left) and of the MAB structures in YB+2 (right).

seen that the alignment of the muon system improves the momentum measurement in the p_T range above 100 GeV, where muons reconstructed using muon chamber hits have a better resolution than muons reconstructed using only the tracker.

Positivo	(mm)			(mrad)		
Estructura	x	y	z	ϕ_x	ϕ_y	ϕ_z
Positive CMS Sector ($+z$)						
AR+	-2.05	-0.92	2932.15	0.67	1.46	-0.53
YE+1	1.50	-0.20	7566.26	-0.47	-0.36	0.41
LD+	4.67	-1.71	6691.16	0.77	-0.28	0.49
ME+1/1	-2666.48	-664.45	6272.49	0.55	0.22	-2879.38
ME+1/2	-3639.74	-643.49	6855.93	0.55	0.20	-2879.43
YB+2	-0.62	0.44	6688.18	-0.34	-0.06	0.34
MAB+195	-4271.47	-1536.34	6693.61	1586.24	0.44	251.79
MAB+255	-796.46	-4486.34	6688.81	1594.86	3.37	1297.58
MAB+315	3485.98	-2930.98	6697.06	1586.58	1.42	2344.31
MAB+15	4276.66	1539.94	6689.61	1586.95	4.07	-2894.67
MAB+75	811.94	4414.60	6690.53	1588.73	-0.19	-1844.63
MAB+135	-3472.21	2922.06	6691.34	1589.57	1.78	-796.55
Negative CMS Sector (z)						
AR-	-1.18	-1.23	-2937.80	-0.25	-0.59	-1.44
YE-1	-1.10	-0.50	-7568.16	0.39	0.03	-0.34
LD-	-0.24	-3.79	-6694.40	-0.08	-0.84	-1.01
ME-1/1	2669.30	655.60	-6275.31	0.38	-0.07	261.46
ME-1/2	-3643.91	-642.27	-6858.59	-0.39	0.22	-2879.87
YB-2	-1.26	-0.33	-6690.01	0.02	0.01	0.01
MAB-195	-4278.58	-1540.95	-6694.87	1576.80	-0.79	251.85
MAB-255	-804.61	-4462.54	-6692.68	1577.32	-4.35	1296.28
MAB-315	3468.96	-2935.24	-6693.06	1583.18	-3.69	2344.74
MAB-15	4275.61	1537.29	-6698.91	1580.50	-3.54	-2892.96
MAB-75	802.96	4466.56	-6693.74	1575.91	-2.95	-1845.81
MAB-135	-3476.83	2925.69	-6694.75	1573.34	-3.24	-797.59

Table 4.3: Position and orientation of the main link alignment structures of the positive and negative z regions of the CMS detector. This reconstruction has been performed without magnetic field ($B = 0$ T).

Positivo	(mm)			(mrad)		
Estructura	x	y	z	ϕ_x	ϕ_y	ϕ_z
Positive CMS Sector ($+z$)						
AR+	-2.05	-0.92	2932.15	0.67	1.46	-0.527
YE+1	1.92	-0.58	7552.01	-0.38	0.20	0.67
LD+	4.31	-1.96	6674.39	1.12	0.22	0.66
ME+1/1	-2666.37	-665.35	6257.34	0.32	-0.29	-2879.12
ME+1/2	-3641.53	-645.22	6847.30	0.324	-3.53	-2879.15
YB+2	-0.76	-0.25	6685.57	-0.52	0.54	0.71
MAB+195	-4271.22	-1538.34	6693.77	1585.02	1.0213	251.73
MAB+255	-794.72	-4487.88	6687.28	1595.16	3.54	1298.05
MAB+315	3487.00	-2930.44	6693.05	1587.27	0.97	2344.94
MAB+15	4276.00	1540.96	6683.96	1586.09	2.27757	-2892.99
MAB+75	810.44	4414.37	6686.93	1590.43	0.35	-1845.08
MAB+135	-3473.63	2920.40	6690.34	1588.46	1.37	-796.25
Negative CMS Sector ($-z$)						
AR-	-1.18	-1.23	-2937.80	-0.25	-0.60	-1.44
YE-1	-0.88	-1.09	-7553.38	0.17	-0.08	-0.38
LD-	-0.73	-4.03	-6678.40	0.61	-3.68	-1.39
ME-1/1	2668.92	655.23	-6258.11	0.15	-0.12	261.42
ME-1/2	-3646.61	-643.34	-6845.12	-0.15	0.25	-2879.89
YB-2	-0.46	-2.60	-6686.10	-0.30	0.04	0.03
MAB-195	-4278.53	-1543.00	-6690.93	1576.50	-0.68	251.87
MAB-255	-804.19	-4465.91	-6686.75	1578.12	-4.68	1296.23
MAB-315	3469.71	-2937.15	-6688.31	1583.48	-4.05	2344.40
MAB-15	4276.38	1535.11	-6695.62	1580.81	-3.65	-2892.94
MAB-75	803.03	4464.42	-6691.44	1576.39	-3.13	-1845.88
MAB-135	-3475.98	2923.43	-6691.49	1573.09	-3.45	-797.57

Table 4.4: Position and orientation of the main link alignment structures of the positive and negative z regions of the CMS detector. This reconstruction has been performed with a magnetic field of $B = 3.8\,\text{T}$.
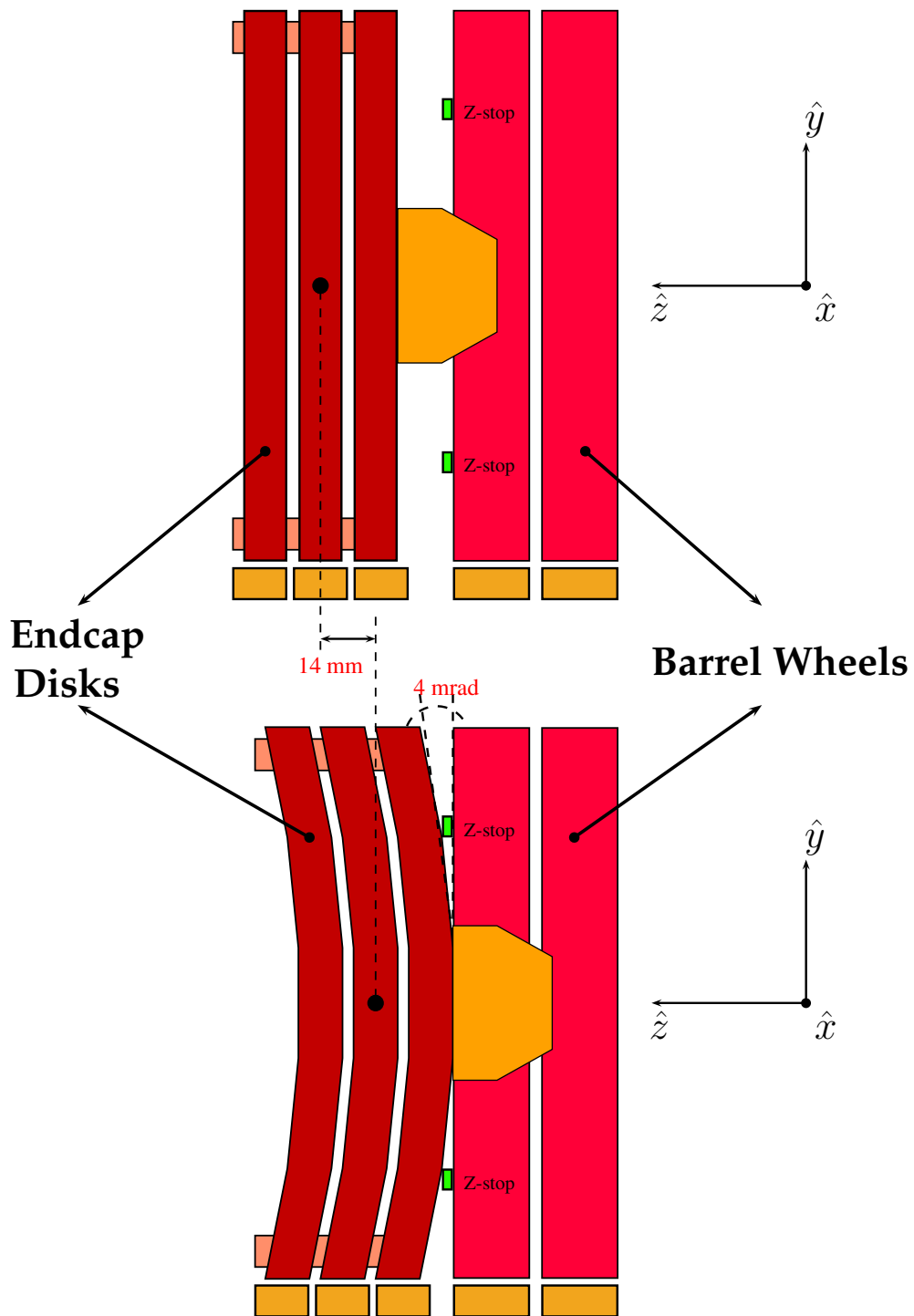
Figure 4.9: Sketch of the displacement and deformation of the endcap iron disks as a result of the detector compression due to the magnetic field.
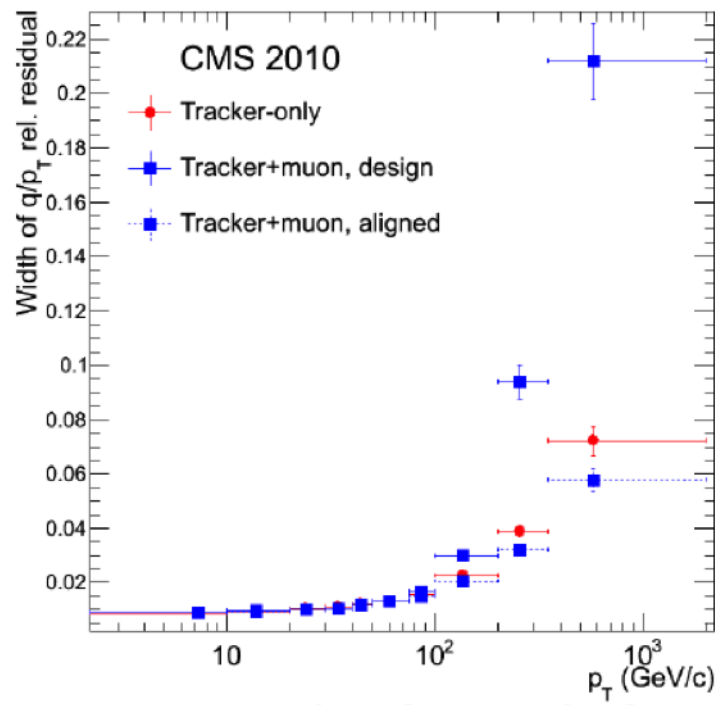
Figure 4.10: q/p_T resolution as a function of p_T for reconstructed muons. The central tracker-only, tracker+muons before alignment and the central tracker plus aligned muon chamber cases are shown.

CHAPTER 5

OBJECT RECONSTRUCTION

This chapter starts with a brief description of the computing model used to process the data collected by the CMS detector. It is followed by a explanation of the measurement of the integrated luminosity. Then, we discuss the algorithm used to process the data, the particle flow algorithm. Finally, we describe the reconstruction procedure to identify the main physics objects used in the analysis: primary vertices, electrons, muons, jets, b-jets and missing transverse energy.

5.1   CMS Computing and Software

The CMS collaboration has had a real challenge developing its computing model in terms of design, calibration, storage, access, reconstruction, and finally, analysis of the huge amount of data. In order to guarantee and distribute the computing services and resources through all the participant institutions, CMS has adopted a GRID computing service. The GRID technology aim at sharing the computing, storage and instrumentation resources through well defined standards [48], creating a software layer between the hardware and the GRID applications. The distributed computing centers available to CMS around the world are configured in a tiered architecture [49].

Each *Tier* level provides different services:

1. **Tier-0**: Based at CERN, it is the stage where the RAW data from the CMS online data acquisition (DAQ) and trigger systems are recorded and distributed to the next stage, the Tier-1 centers (see below). It also archives the RAW data to tape, groups them into data streams and feeds the prompt first-pass reconstruction, producing the full event and in some cases also extracts a first-pass Analysis Object Data (AOD).

2. **Tier-1**: Currently distributed in eight centers, it has the responsability for the safe storage of a copy of CMS real and simulated data, to be fast accessed in reprocessing and skimming activities. Thanks to a large CPU capacity, The Tier-1 performes the reprocessing and analysis of large amounts of data with newly available calibration and alignment information, and allows a fast skimming of the hosted datasets to extract the relevant samples.

3. **Tier-2**: It provides the capacity for user analyses, calibration studies and Monte Carlo production with approximately 49 centres with smaller CPU resources. The main data flows are to import datasets (like AOD subsets) or skimmed RECO data from any of the Tiers-1 hosting the corresponding primary datasets, and to move the produced Monte Carlo simulated samples to a tier-1 for safe custody.

4. **Tier-3**: There are as many tier-3 as the number of institutions and universities take part in the data processing and analysis providing smaller CPU resources.

5.1.1   The CMS Software Design

The CMS SoftWare (CMSSW) is a software collection designed to facilitate the development and deployment of the reconstruction and analysis of CMS data and Monte Carlo simulations. The CMS Event Data Model (EDM) describes every event as a C++ object which contains all the RAW and reconstructed information of the collision. Additionally, every event has relevant information related to the configuration of the software used and the CMS calibration conditions.

The objects in the event are stored in ROOT file format [50]. ROOT is an analysis package written in a C++ object-oriented structure. All the histograms and files with

data objects are produced with user-compiled codes built with ROOT functions.

To reduce the data size storage in each event, information no relevant for a specific stage of the data analysis is removed. Thus, the RAW format is defined by all the signals coming from the detectors, the RECO tier groups all the relevant collections processed from RAW needed for the physic reconstruction plus the collection raised from the reconstruction itself. There is a further data tier designated as Analysis Object Data (AOD), which contains a subset of RECO information, sufficient for most of the physics analysis.

5.2 Luminosity Measurement

For cross section measurements, a precise determination of the integrated luminosity is mandatory since the error of the luminosity is propagated to the final result. In CMS two detectors are exploited for the measurement of the instantaneous luminosity: *the forward hadron calorimeters HF* [51], featuring a dedicated high rate acquisition system independent from the central DAQ and capable of estimating the luminosity per bunch, and *the silicon pixel detector* characterized by very low occupancy and excellent stability over time.

The luminosity measurement based on HF can determine the average luminosity with a 1% statistical accuracy in less than 1s. This measurement, however, is subject to calibration drift as a result of gain changes in the HF. In addition, the detector response has been proved to be non-linear with the 2012 pileup conditions. These two effects together make the usage of HF difficult to measure the luminosity with a high accuracy.

Using the silicon pixel detector, the instantaneous luminosity is evaluated from the number of pixel clusters occurring on average in a zero-bias event as:

$$\mathcal{L} = \frac{\nu \langle n \rangle}{\sigma_{vis}} \tag{5.1}$$

Where ν is the beam revolution frequency (11246 Hz), $\langle n \rangle$ is the average number of pixel clusters *per event* and σ_{vis} is the visible inelastic cross section, defined as the average number of clusters *per inelastic collision* times the total inelastic cross section.

The visible cross section is callibrated via van der Meer scans[52], which consists in moving the beams transversely to each other, measuring the interaction rate (at the same $\langle n \rangle$) as a function of the beam separation.

A convenient minimal time interval to consider for the estimation of the integrated luminosity is the "luminosity section" (LS), defined as 2^{18} LHC orbits and corresponding to $t_{LS} = 23.31$ s. For every luminosity section the average number of clusters per event ($\langle n \rangle$) is measured and the integrated luminosity for that LS is derived multiplying by t_{LS} the instantaneous luminosity $\mathcal{L}$ resulting from Equation 5.1. Finally, The total integrated luminosity for the analysis is then computed summing the integrated luminosity of each LS recorded by CMS during the April to June 2012 data taking period, obtaining $5.3\,\text{fb}^{-1}$.

Figure 5.1 shows the integrated luminosity taken by the CMS detector in the first three years of operation. The relative uncertainty on the luminosity, which is propagated directly into the $t\bar{t}$ cross sections measurement, was found to be 2.6%[53].
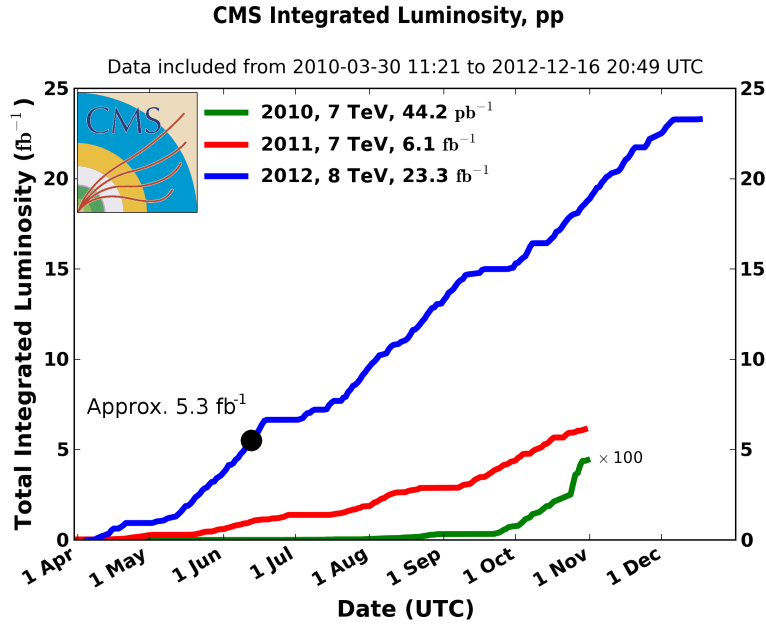
Figure 5.1: Integrated luminosity taken by the CMS detector in proton-proton collisions from 2010 to 2012. The analysis presented in this document uses the data taken from April to June 2012 at $\sqrt{s} = 8\,\text{TeV}$.

5.3 Reconstruction of the Physics Events

All the reconstructed information is based on interpreting the signals from each sub-detector and associating them to decay vertices, trajectories, energy or particle identities. These elements can be grouped to form high level *physics objects* as electrons, muons, photons, jets and missing transverse energy.

5.3.1 The Particle Flow Algorithm

The particle flow (PF) [11] event reconstruction algorithm is a method to reconstruct and identify all stable particles in an event such as electrons, muons, photons and hadrons, using the information of all CMS sub-detectors. All the identified particles are later used to reconstruct jets from quarks hadronization, to determine the missing transverse energy ($\not{E}_T$), to reconstruct the unstable particles, to quantify charged lepton isolation with respect to other particles and to identify jets comming from b quarks. The fundamental "elements" of the PF algorithm to reconstruct events are the charged-particle tracks, the calorimeter clusters and the hits in the muon chambers.

Based on the particle trajectory deviation produced by the magnetic field, the CMS tracker is able to measure the direction and momentum of the charged particle with very high resolution. This attribute, together with the fact that the charged particles carry around two thirds of the jet energy, makes the tracker the most appropriate sub-detector to start the event reconstruction process.

An iterative-tracking strategy [54] based on the need of a reconstruction efficiency close to 100% but a low track fake rate is used to reconstruct the tracks. First, tracks are seeded and reconstructed with very tight criteria, leading to a moderate tracking efficiency. The next steps is to remove hits unequivocally assigned to the tracks found in the previous iteration, and by progressively loosening track seeding criteria. After the first iterations, tracks originated around the beam axis are found with an efficiency of 99.5% for isolated muons in the tracker acceptance, and larger than 90% for charged hadrons in jets. The successive iterations use relaxed constraints on the origin vertex, in order to reconstruct secondary charged particles originating from photon conversions, decay of long-lived particles, etc.

With this iterative technique, charged particles with at least three hits, a $p_T > 150\,\mathrm{MeV}$ and an origin vertex beyond $50\,\mathrm{cm}$ from the beam axis are reconstructed with a fake

rate of the order of 0.01.

The succesive stage in the event reconstruction analyses the information collected in the calorimeters. Its purposes can be divided into:

a) Detect and measure the energy and direction of stable neutral particles such as photons and neutral hadrons.

b) Separate the energy deposits of neutral particles from charged hadron ones.

c) Identify and reconstruct electrons and all accompanying bremsstrahlung photons.

d) Improving the measurement of the energy of the charged hadrons without accurate track reconstruction.

Considering this, a clustering algorithm has been designed with the aim to reach a high detection efficiency in the reconstruction of the energy deposit in the calorimeters, and to separate close energy deposits.

The algorithm is performed separately for each component of the calorimeters in the following steps: first, the local calorimeter cell with the largest amount of energy deposit is identified and called "cluster seed". Second, "topological clusters" are grown from the seed including cells with an energy above a given threshold established as function of the electronic noise in the calorimeter. Finally, a topological cluster gives rise to as many "particle flow clusters" as seeds.

The charged particle tracks, calorimeter clusters and muon tracks are connected to fully reconstruct each particle. The link algorithm is on charge of the connection avoiding double counting of information in different particle candidates. The result of this algorithm are "blocks" of elements linked.

The reconstruction and identification of a set of particles is finally performed by the particle flow algorithm. It is executed for each block, starting by translating each muon reconstructed with the tracker and the muon system (named *global muon*) into a PF-muon if its momentum measured by the tracker is compatible with that determined from the muon chambers. After that, the algorithm starts with the reconstruction and identification of the electrons. Each track of the block is preidentified combining the information of the tracker and pre-shower detector. Preidentified electron tracks are refit with a Gaussian Sum Filter [55] in attempt to follow their trajectories

to the electromagnetic calorimeter. A final identification is performed combining the number of tracking and calorimeter variables in order to give rise to a PF-electron.

For the remaining tracks, a tigther criterium requiring a smaller uncertainty in the p_T track than the relative calorimeter energy resolution expected for charged hadrons is applied. The tracks are directly connected to a number of ECAL and HCAL clusters, giving rise to "PF-charged hadrons". In the case that the energy of the closest ECAL and HCAL cluster linked to the track(s) is significantly larger than the total associated charged-particle momentum, the energy cluster gives rise to a "PF-photon" or "PF-neutral hadrons".

The final list of reconstructed particles constitutes the global description of the event. The jet reconstruction, b-jet identification and the estimation of the missing transverse energy ($\not{E}_T$) are explained in Section 5.6, 5.6.2 and 5.7, respectively. A detailed description of the electron and muon reconstruction is in Section 5.5.2.

5.3.2   PF2PAT

PF2PAT [56] is a particle-flow post-processing step, which starts from the list of particle flow candidates reconstructed by the PF algorithm and creates a set of particle-based RECO physics objects such as PF leptons, with PF isolation, PF jets and PF mising transverse energy.

5.4   Primary Vertex

An accurate reconstruction of the event primary vertex [57] is needed to assign tracks to collisions and determine the event kinematics. After reconstruct and identify the tracks in the event, dedicated algorithms are applied to estimate the primary vertex position and its associated tracks. This vertex reconstruction typically involves two steps:

1. Vertex finding, where clusters of tracks originating from the same vertex are grouped together as vertex candidates. The "prompt" tracks originating from the primary interaction region are selected based on the transverse impact parameter significance with respect to the beam line and some other quality cuts. Then, the tracks are clustered based on their z coordinate at the point of closest

approach to the beam line. This clustering allows for the possibility of multiple primary interactions in the same LHC bunch crossing.

2. Vertex fitting, where from a set of tracks, the most compatible vertex position is computed and used to constrain track parameters at the vertex. Vertex candidates containing at least two tracks are fitted with an adaptive vertex fitter [58] to compute the best estimate of vertex parameters, including the position and covariance matrix, as well as the indicators of the success of the fit, such as the total χ^2/ndf and track weights, which depend on the distance of the track to the vertex.

After obtain the possible primary vertex candidates from the previous reconstruction, the selected origin of the hard interaction is the one with the highest $\sum(p_T^{track})^2$, where the p_T^{track} are the transverse momenta of the tracks associated to the vertex. In order to remove the particles coming from the other vertices (pileup), the tracks of the charged hadrons which do not come from the primary vertex selected are removed. This process is called "charged hadron subtraction (CHS)".

5.5   Leptons

5.5.1   Electrons

An electron for the CMS detector is a single track emerging from the interaction vertex and matched to an electromagnetic supercluster[1]. The electron reconstruction [60] in CMS is hampered by the amount of tracker material discretely distributed in front on the electromagnetic calorimeter (ECAL). Electrons cross the silicon layers of the pixel and inner tracker detectors radiating bremsstrahlung photons. Since the electron direction can change significantly in presence of the 3.8 T magnetic field, the electron reaches the ECAL leaving a spread of energy in the azimuthal direction with respect to the one at the interaction point. This spread has a dependency of the electron p_T. The amount of bremsstrahlung emitted when integrating along the electron trajectory can be very large.

---

[1]"Supercluster" is a group of one or more associated clusters of energy deposits in the ECAL constructed using an algorithm which takes account the clusters narrow width in the η coordinate and their spread in ϕ, produced by the bending in the magnetic field of electrons radiating in the tracker material [59].

Electron reconstruction

Two complementary algorithms are used in the electron reconstruction at the track seeding stage: the *"ECAL driven"* seeding, which is optimized for isolated electron in the p_T range relevant for Z and W decays, and the *"tracker driven"* seeding, more suitable for low p_T electrons and also better performing for electrons inside jets.

The *ECAL driven* seeding starts with the reconstruction of a ECAL "supercluster" of $E_T > 4\,\mathrm{GeV}$. As a filter, the superclusters is matched to tracks seeds (pairs or triplets of hits) in the innermost layers of the tracker, building the electron track from these track seeds.

The *tracker driven* seeding algorithm extrapolates standard tracks reconstructed from the inner tracker to the ECAL, searching for bremsstrahlung clusters. The track is promoted to electron seed if the ratio between the energy E of the cluster and the track momentum p is close to unity.

After the seeding, electron trajectories are reconstructed using a dedicated modeling of the electron energy loss and fitted with a Gaussian Sum Filter (GSF) [55]. The GSF algorithm allows sudden changes in the curvature radius, caused by Bremsstrahlung photon emission, to be properly taken into account in the layer to layer track propagation.

After propagating the electron track candidate, an identification of the potential Bremsstrahlung photon cluster(s) is carried out. A straight line tangent to the direction of the GSF track is extrapolated up to the ECAL entrance. If an ECAL cluster can be linked to the straight line, its energy is assigned to the total electron energy.

The final stage of the electron reconstruction is to apply a loose preselection in order to reject fake electrons. The variables related to the energy and geometrical matching between the track and the ECAL cluster(s) are combined into a multivariate estimator (MVA) [61].

Figure 5.2 shows the invariant mass spectrum for reconstructed electron pairs with opposite charge, using the whole dataset collected during the 2010 data taking.
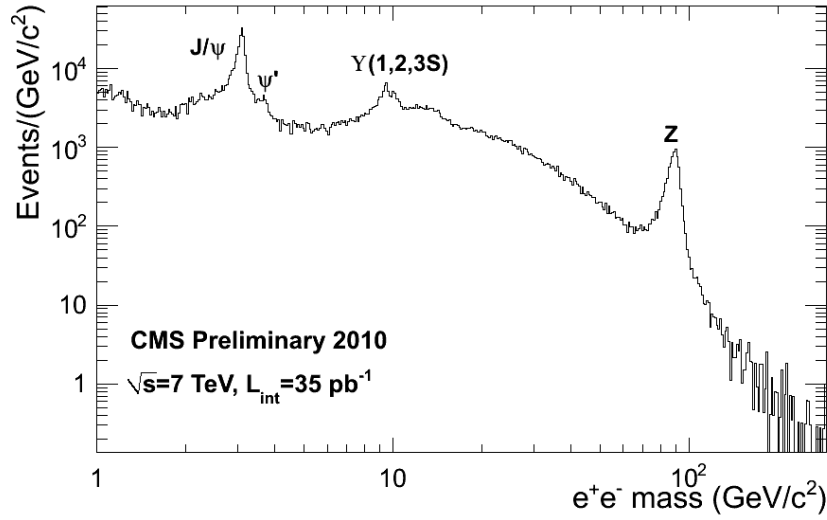
Figure 5.2: Invariant mass spectrum of opposite sign electron pairs with $35\,\mathrm{pb}^{-1}$ of 2010 data.

Electron isolation

The level of isolation of a given particle is quantified by the distribution of neighbouring tracks and calorimeter energy deposits. In the PF algorithm, the isolation is obtained from the distribution of neighbouring particles. Muons and electrons originated from W and Z bosons are expected to be isolated, contrary to leptons coming from QCD processes and b quark decays. The fake leptons can be found within the jet cone, producing a non-zero energy in their nearby region. The isolation algorithms rely on the comparison of the total energy deposited in a cone around the lepton with a predefined threshold.

The isolation cone is represented in Figure 5.3. The geometrical construction of the cone starts with its axis, which is chosen according to the lepton direction with a procedure that is modified on the base of each isolation algorithm. In addition to the axis, the cone is defined by $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, where η is the pseudorapidity and ϕ the azimuthal angle. The lepton contribution to the energy inside the cone (called veto value) is subtracted to improve the discriminating power of the isolation algorithm.

As the analysis described in this thesis uses particle flow candidates, a PF-based isolation variable has been used to reduce the contamination from the non-isolated leptons originating inside the jets. This variable is defined as:
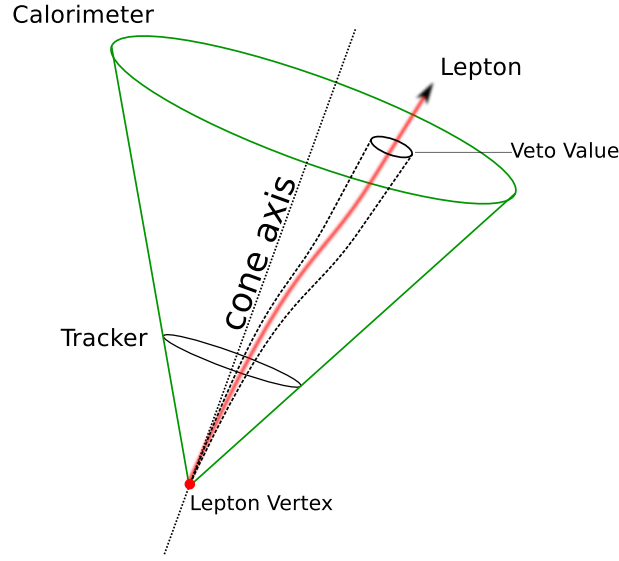
Figure 5.3: Schematic ilustration of the lepton isolation cone.

$$I^\ell = \overset{\text{charged hadrons}}{\sum} \text{p}_\text{T} + \overset{\text{neutral hadrons}}{\sum} \text{p}_\text{T} + \overset{\text{photons}}{\sum} \text{p}_\text{T} \tag{5.2}$$

In order to simplify the isolation cut, it is applied over the relative isolation:

$$I^\ell_R = \frac{I^\ell}{\text{p}_\text{T}^\ell}, \tag{5.3}$$

where p_T^ℓ is the transverse momentum of the lepton. In the specific case of electron isolation, a correction[62] based on the effective area (A_{eff}) of the isolation cone and the average energy expected by neutral particlesfrom pileup (ρ) has been applied. The electron isolation becomes:

$$I^e_R = \frac{\sum^{\text{charged hadrons}} \text{p}_\text{T} + \max\left[0.0, \sum^{\text{photons}} \text{p}_\text{T} + \sum^{\text{neutral hadrons}} \text{p}_\text{T} - \rho \times A_{eff}\right]}{\text{p}_\text{T}^e} \tag{5.4}$$

5.5.2 Muons

The CMS detector has a very robust system of muon detection based on three different subdetectors: RPC, CSC and DT (see Section 3.3.4). Three different approaches has

been developed to reconstruct the muons in CMS, based on stand-alone, tracker and global reconstruction algorithms.

Stand-alone muons

The stand-alone reconstruction uses only the information on the hits in the DT, CSC and RPC subsystems to reconstruct the muon. Hits within each DT and CSC chambers are fitted to form segments. A vector expressing the local track position, momentum and direction is associated to each of these segments. The seed to reconstruct the track of a muon is generated by the group of collected segments. The muon trajectory is fitted using a Kalman Filter technique [63]. The innermost vectors are propagated to the layer surface of the next chamber. The extrapolated vector is compared with the local track segment and the trajectory parameters are computed [64]. The operation is performed until the outermost chamber is reached. The same Kalman Filter is applied in backward direction, working from outside in, to define the track parameters at the innermost muon station.

For each segment added, a cut is applied in the quality of the track fit in order to evaluate the incremental χ^2 of the track fit due the new state and reject possible bad hits due to showering, cosmic rays, etc. Finally, at least two segments (one of which must be the DT or CSC chambers) must be present in the fit in order to reject fake segments.

Tracker muons

Tracker muons are reconstructed by extrapolating tracks from the silicon tracker and looking for compatible segments in the muon spectrometer. For each track with $p > 2.5$ GeV, the algorithm searches for compatible segments in the muon detectors and/or compatible signatures in the calorimeter. An important component of the tracker muon identification is the arbitration, i.e the pattern recognition problem of assigning segments to tracks. The segment arbitration is based on the best ΔX match or the best $\Delta R^2 = \Delta X^2 + \Delta Y^2$ match, where ΔX (ΔY) is the distance in local muon chamber X (Y) axis between the segment and the extrapolated track. Finally, the momentum of the tracker muon is the same as that of the silicon tracker track.

Global muons

Stand-alone muon tracks are matched with tracker tracks to generate global muon. The global muon takes advantage from both the tracker detector and muon spectrometer, to obtain a more accurate description of the muon properties. The first step in reconstructing a global muon is to identify the silicon tracker track to match with compatibles stand-alone muon track. The matching starts defining a rectangular region around the stand-alone track in $\eta - \phi$ space to select a subset of tracker tracks. The determination of the region of interest is based on the track parameters and their corresponding uncertainties of the extrapolated muon trajectory. The second step is to iterate over the subset of tracker tracks, applying more stringent spatial and momentum matching criteria to choose the best tracker track to combine with the stand-alone muon. After the selection of a subset of tracker tracks that match the stand-alone muon track, the global muon track is fitted a using the hits from the tracker track and the stand-alone muon track.

The tracker provides a much higher momentum resolution than the muon system due to its high hit resolution and the greater multiplicity of hits available for the track fit and for the high magnetic field. At low momentum, the best momentum resolution for muons is obtained from the inner silicon tracker. However, at high energies, the reduced bending of the particle limits the resolution of the inner tracker fit. At higher momentum, adding hits from the muon spectrometer significantly improve the curvature measurement providing a better momentum resolution. In Figure 5.4 the momentum resolution for the tracker, stand-alone and global muons is shown. For values below $200\,\mathrm{GeV}$ the measurement of the momentum is dominated by the tracker resolution. For higher values, the intrinsic resolution of the muon plus tracker system starts to become of the same order of magnitude as tracker only, because multiple scattering effects become smaller as $\mathrm{p_T}$ increases.

Figure 5.5 shows the invariant mass of muon pairs with opposite charge for the first $1.1\,\mathrm{fb^{-1}}$ of data collected in 2011. The mass peak for several resonances can be observed, from the low ones, ω and ϕ, moving to higher values of the invariant mass for J/ψ and Upsilon hadrons, to the highest resonance for the Z boson. This figure shows the high muon momentum resolution of the detector for a large kinematic range, from $\mathrm{p_T} \sim 500\,\mathrm{MeV}$ to the TeV momentum.
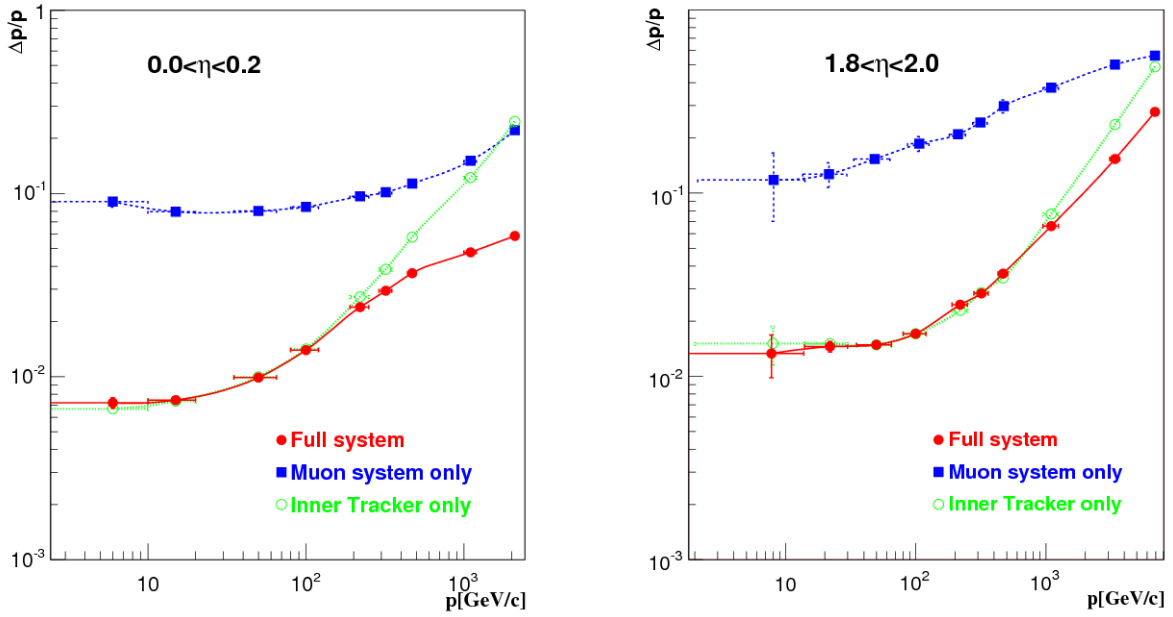
Figure 5.4: Muon momentum resolution as a function of p in Monte Carlo simulation in two regions of muon pseudorapidity: $0.0 < \eta < 2.0$ (left) and $1.8 < \eta < 2.0$. Green curve corresponds to the result of the tracker track fit, blue curve to the stand-alone fit, and red curve the global fit.

Muon isolation

The muon isolation is used to reject the muons coming from W+jets events and QCD processes. A detailed description of the lepton isolation variable can be found on Section 5.5.1.

In the case of muons, a different type of correction is applied in order to correct the contribution of the neutral particles from pileup interactions to the isolation. Its name is $\Delta\beta$ correction ($I_{\Delta\beta}$) [65]. Thus, the muon isolation used in this analysis is:

$$I_R^\mu = \frac{\sum^{\text{charged hadrons}} p_T + \max\left[0.0, \sum^{\text{photons}} p_T + \sum^{\text{neutral hadrons}} p_T - \frac{1}{2}I_{\Delta\beta}\right]}{p_T{}^\mu} \quad (5.5)$$

5.6   Jets

Jets of particles from the hadronization of quarks and gluons produce signals in the ECAL and HCAL calorimeters, and their charged components in the pixel and silicon
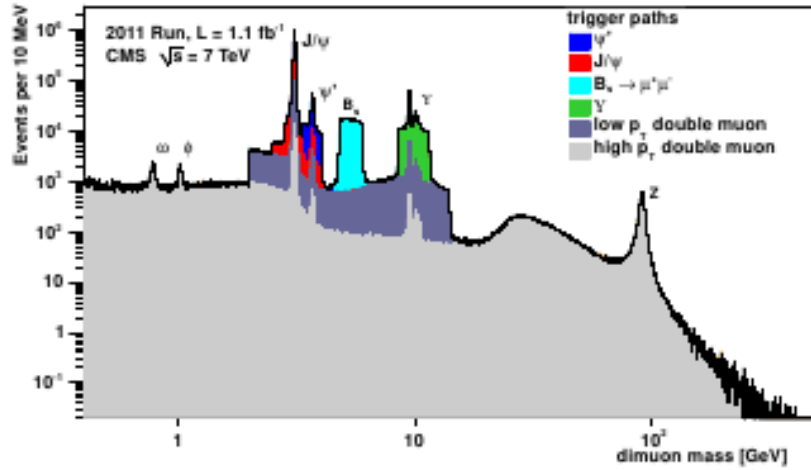
Figure 5.5: Invariant mass spectrum of opposite sign muon pairs in the first $1.1\,\mathrm{fb}^{-1}$ of 2011 data.

strip tracking detector. The energy of the jets is carried in average for 65% by charged particles, 25% by photons and 10% by neutral hadrons. 90% of the jet energy can be reconstructed with good precision by the PF algorithm.

The starting point to reconstruct the jets is given by the PF-particle candidates obtained through the PF algorithm. The particles are clustered by means of the "anti-k_t" clustering algorithm [66]. The algorithm starts with a high-momentum particle as seed to the jet. It is necessary to introduce the distance d_{ij} between the particles i and j and d_{iB} between the particle i and the beam (B). Then, the clustering proceeds by identifiying the smallest of the distances and if it is a d_{ij} recombining entities i and j, while if it is d_{iB} calling i a jet and removing it from the list of particles. The distance is measured as:

$$d_{ij} = \min\left[k_{T_i}^{2p}, k_{T_j}^{2p}\right] \frac{\Delta_{ij}^2}{R^2} \tag{5.6}$$

where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$. k_{T_i}, y_i and ϕ_i are the transverse momentum, rapidity and azimuthal angle of the particle i, respectively. R is a parameter for the jet size. The parameter p, which governs the relative power of energy versus Δ_{ij}, is set to $p = -1$ for the anti-k_t algorithms.

Apart from the anti-k_t, there are other clustering algorithms available, such as the k_t

(where the power p on Equation 5.6 is 1), iterative cone [67] and SIS cone [68].

5.6.1 Jet Energy Corrections

Due to different effects like the not linear calorimeter response to particles, electronic noise and pile-up, the energy of the reconstructed jet does not correspond to the true parton energy. A set of "jet energy corrections" has been developed to correct the measured jet energy to the analysis desired level.

CMS has adopted a factorized solution to the problem of jet energy corrections, where different level of corrections take care of a specific effect. Each correction is essentially a scaling of the jet four momentum with a factor depending on various jet-related quantities, such as p_T, η, flavour, etc. The corrections, which are applied sequentially with a fixed order, are:

- **Level 1 corrections** are applied to remove the energy coming from pile-up events. In principle, this will remove any dataset dependence on luminosity so that the successive corrections are applied upon a luminosity independent sample.

- **level 2 corrections** are applied to make the jet response uniform in pseudorapidity. It is achieved by correcting a jet with arbitrary η relative to a jet in the central region ($|\eta| < 1.3$).

- **Level 3 corrections** are applied to make flat the jet response in the transeverse momentum variable (p_T).

Finally, the jet energy is corrected according to the following formula:

$$E_{\text{corrected}} = (E_{\text{uncorrected}} - E_{\textbf{Level 1}}) \times C_{\textbf{Level 2}}(\eta) \times C_{\textbf{Level 3}}(p_T) \tag{5.7}$$

5.6.2 b-jets

As the $t\bar{t}$ process has almost all the times two b-quarks in its final state, it is very important to identify the jets coming from a bottom quark. These jets can be identified with the b-tagging technique, which exploits the specific properties of the decay of b hadrons [69].

b hadrons have a relatively long life-time ($\tau \sim 1.5$ ps) that produces in the detector a displaced secondary vertex at the point of decay. Additionally, the decay of these hadrons produce final states with a high charged track multiplicity (usually, more than 5). There are two methods to exploit the b hadrons lifetime information: searching tracks with a large impact parameter [70], or looking for displaced secondary vertices [71].

The b-tagging algorithm used in this analysis combines the information on the secondary vertex with the one on the track impact parameters. Its name is combined secondary vertex (CSV). CMS has other algorithms to identify jets coming from a b quark, such as the track counting algorithm, jet probability algorithm and the soft lepton algorithm [72].

The secondary vertices are reconstructed in an inclusive way inside the jet using the Trimmed Kalman Vertex Finder [73]. This algorithm begins by using all tracks in the jet and subsequently rejects outliers, which then are used to reconstruct additional vertices. As not only the presence of a secondary vertex is used in the algorithm, but also topological and kinematical variables related to the vertex, it is desirable to reconstruct as completely as possible the decay vertex in order to increase the discriminating power of these topological and kinematical variables.

The following cuts are applied to the reconstructed vertices to select secondary vertex candidates:

1. The distance between the primary vertex and the secondary vertex in the transverse plane has to exceed $100\,\mu$m and must not exceed 2.5 cm.

2. The distance between the primary vertex and the secondary vertex in the transverse plane divided by its error has to be greater than 3 ($\frac{l_t}{\sigma_{L_t}} > 3$).

3. The invariant mass of charged particles associated to the vertex must not exceed 6.5 GeV.

4. The vertex must not be compatible with a K_S^0 decay. Vertices with two oppositely charged tracks are rejected if their mass is within a window of 50 MeV around the nominal K_S^0 mass ($m_{K_S^0} = 497$ MeV).

The additional topological and kinematic variables used in the b-tagging algorithm

depend on how was reconstructed the secondary vertex [71]. A discriminant is constructed including those variables into a likelihood ratio technique. The CSV algorithm can provide discrimination also in cases when no secondary vertices are found using additional variables and creating "pseudo vertex" to allow the computation of a subset of secondary-vertex-based quantities [72]. Jets coming from the hadronization of b quarks (light quarks) are expected to have a higher (lower) value of the discriminator.

Figure 5.6 shows the probability to tag a jet from a light parton as b-jet (mistag rate) versus the efficiency to identify b-jets for several b-tagging algorithms. The CSV has the higher efficiency (80%) for a mistag rate of 10 % than all the other taggers.
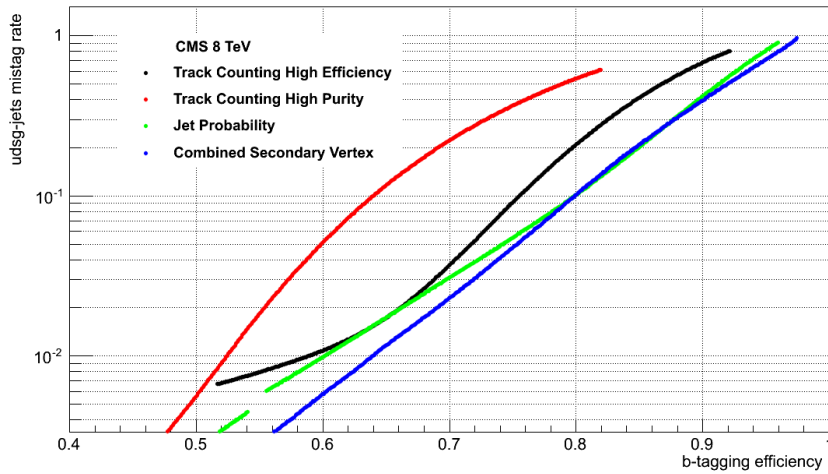
Figure 5.6: Mistag probability versus the efficiency to identify a b-jet for the different b-tagging algorithms. The curves have been derived on a sample of simulated QCD events.

In this analysis, b-jets are identified by requiring the CSV discriminant value to be larger than 0.244 (loose CSV working point, CSVL). This requirement has a identification efficiency in t$\bar{\text{t}}$ samples of about 85%, and a misidentification probability of about 10% [74] for light-flavour jets (u, d, s and gluons).

5.7   Missing Transverse Energy

The missing transverse energy ($\not{E}_T$) refers to energy which is not measured by the detector but is expected by the conservation of momentum. This energy is attributed

to particles which do not interact with the detector, such as neutrinos. However, missing energy can also be caused by mismeasurements of the momentum of the detected particles or by detector resolution effects.

At the LHC, and at any hadron collider, the momentum of the partons along the beam axis is not known. Nevertheless, the momentum along the plane transverse to the beam axis is zero, so the sum of the momentum of all the particles produced in the collisions is constrained to be zero. Figure 5.7 shows a sketch of the $\not{E}_T$ definition.
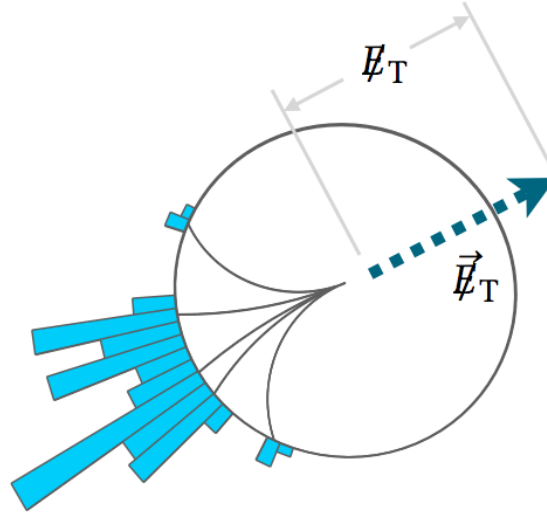
Figure 5.7: Sketch of the missing transverse energy ($\not{E}_T$) definition in the plane transverse to the beam direction.

There are different algorithms to estimate the $\not{E}_T$:

- **Calorimeter $\not{E}_T$ (CaloMET).** It is determined using mostly calorimetric information. In the majority of cases, the muon deposits a little amount of energy in the calorimeters. Hence, precise muon momentum measurement from the central tracker and muon system is used to replace the energy measured along the muon trajectory in the calorimeter.

- **Track corrected $\not{E}_T$ (tcMET).** This algorithm starts from the CaloMET, and apply further corrections using information from the tracker. The transverse momentum of each reconstructed track is included in the $\not{E}_T$ while the calorimetric energy deposit along the extrapolated track trajectory is subtracted. This approach takes advantage of the better resolution of the tracker with respect to the

calorimeters, allowing an overall resolution improvement and a better description of the $\not{E}_T$ distribution in the tails.

- **Particle Flow $\not{E}_T$ (PFMET)**. Since all detector information is included, it is simple to use the PF candidates to estimate the $\not{E}_T$. The PFMET is the magnitude of the momentum imbalance, which is the negative sum of the momenta of all PF-particles in the plane transverse to the beam:

$$\overrightarrow{PFMET} = - \sum_{\text{All PF particles}} \overrightarrow{p_T} \tag{5.8}$$

Figure 5.8 shows a comparison of the $\not{E}_T$ resolution versus the PFMET for the three described algorithms. The $\not{E}_T$ resolution for the tcMET and PFMET algorithms is a factor 2 smaller than the one for the CaloMET algorithm.

Figure 5.9 shows the resolution of the PFMET projection along the x and y axis as a function of the $\sum E_T$ for events with no real $\not{E}_T$ (Drell-Yan processes). The resolution increase slowly between 10 to 30 GeV, for transverse energies from 300 GeV to a few TeV.

Corrections can be applied to the $\not{E}_T$ regardless of the algorithm used. The *Type-I correction* is the most effective $\not{E}_T$ correction in CMS. This correction is a propagation of the jet energy corrections (JEC) to the missing transverse energy. The Type-I correction replaces the vector sum of the transverse momenta of the particles which can be clustered in a jet with the transverse momentum of the jet after the JEC are applied. It can be written as:

$$\overrightarrow{\not{E}_T}^{\text{Type I}} = -\sum_{jet} \overrightarrow{p_T}^{jet}(JEC) - \sum_{\text{unclustered particles}} \overrightarrow{p_T} \tag{5.9}$$
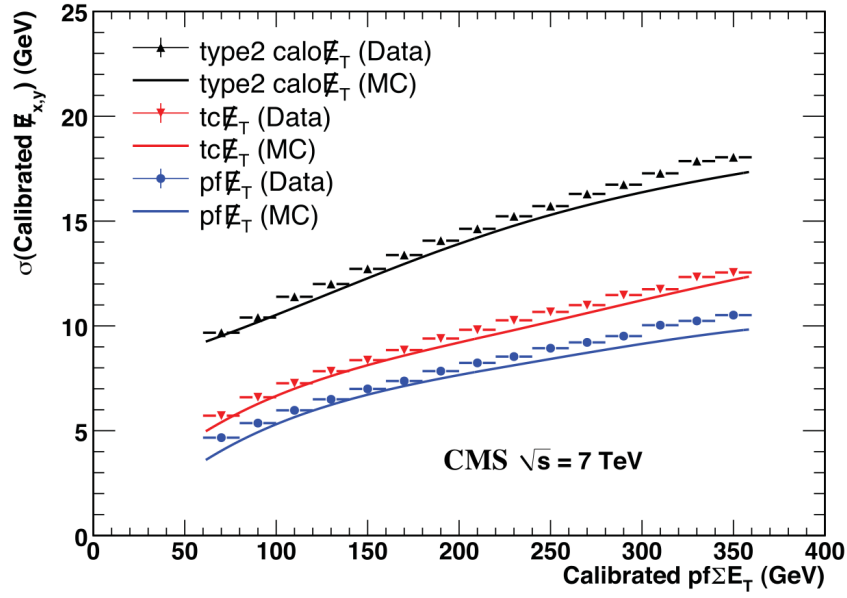
Figure 5.8: $\not{E}_T$ gaussian core resolution versus the PFMET for the CaloMET, tcMET and PFMET algorithms, for events with at least two jets with $p_T > 25\,\text{GeV}$ in 2010 data.

(a)   (b)
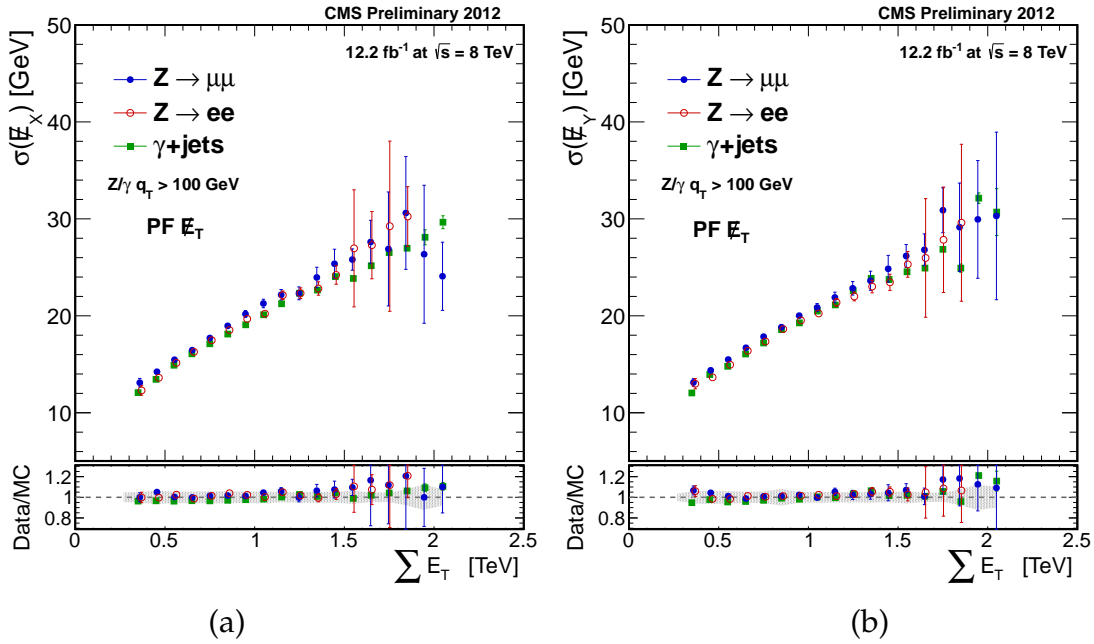
Figure 5.9: Resolution of the PF MET projection along the x-axis (a) and the y-axis (b) as a function of $\sum E_T$ for events with Z to leptons or photons at $\sqrt{s} = 8\,\text{TeV}$. Results are shown for $Z \to \mu\mu$ events (full blue circles), $Z \to$ ee events (open red circles), and photon+jets events (full green squares).

CHAPTER 6

EVENT SELECTION AND BACKGROUND ESTIMATION

This chapter presents the data samples, selection cuts and background estimations used in the measurement of the $t\bar{t}$ cross section in the dilepton channel at $\sqrt{s} = 8\,\text{TeV}$. We first introduce the data and Monte Carlo (MC) samples. Successively, we describe the cuts applied to maximize the signal to background ratio. Finally, the data-driven methods implemented to estimate the Drell-Yan and the Non-W/Z lepton background from data are presented.

6.1   Data and Monte Carlo Samples

The analysis is performed using the CMS event data model and the official software framework of CMS for event generation, simulation and reconstruction (see Chapter 5).

Simulated signal and background events have been centrally produced by CMS with the full simulation of the detector conditions, including miscalibration and misalignment effects described by the start-up conditions. The Physics Analysis Tool (PAT) [75,

76, 77] along with the PF2PAT [56, 78](see also Section 5.3.2) framework are used in the analysis.

The generation of signal (t$\bar{\text{t}}$) and background events (Z+jets, W+jets, WW+jets) is performed using mainly the MADGRAPH generator [79]. Additional backgrounds are produced with PYTHIA [80] (WZ, ZZ production) and POWHEG [81] (single-top events). The single-top quark t-channel and s-channel productions have been studied but their contributions to the events in the signal region are neglible. The theoretical cross sections and the datasets for each MC process are given in Table 6.1. The cross section of t$\bar{\text{t}}$ production is for a top-quark mass of $m_t = 172.5$ GeV.

Simulated Process	Primary Dataset Name	σ [pb]
t$\bar{\text{t}}$	/TTJets_FullLeptMGDecays_8TeV-madgraph-tauola/	24.6
	/TTJets_SemiLeptMGDecays_8TeV-madgraph-tauola/	103
	/TTJets_HadronicMGDecays_8TeV-madgraph/	106.9
$\bar{\text{t}}$W	/Tbar_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola/	11.2
tW	/T_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola/	11.2
WW	/WWJetsTo2L2Nu_TuneZ2star_8TeV-madgraph-tauola/	5.8
WZ	/WZ_TuneZ2star_8TeV_pythia6_tauola/	22.4
ZZ	/ZZ_TuneZ2star_8TeV_pythia6_tauola/	9.0
W + Jets	/WJetsToLNu_TuneZ2Star_8TeV-madgraph-tarball/	37509.0
$Z/\gamma^* \to \ell\ell$: $[10 < m_Z < 50]$	/DYJetsToLL_M-10To50filter_8TeV-madgraph/	860.5
$[50 < m_Z < \infty]$	/DYJetsToLL_M-50_TuneZ2Star_8TeV-madgraph-tarball/	3532.8

Table 6.1: Summary of the Monte Carlo datasets used in the analysis.

The MC samples used to study the systematic uncertainties and the t$\bar{\text{t}}$ cross section as a function of the top quark mass are presented in Table 6.2 and explained in Section 7.1 and 7.4 respectively.

The datasets used are summarized in Table 6.3. The total integrated luminosity corresponds to 5.3 fb^{-1} of proton-proton collisions. Data have been taken in two periods:

- **RunA**: Taken in April 2012. It corresponds to ~ 893 pb^{-1} of data.

- **RunB**: Taken in the May-June period of 2012. It amounts to ~ 4404 pb^{-1} of data.

The analysis is based on the DoubleMuon, DoubleElectron and MuEG Primary Datasets which contain the events collected by dilepton triggers at HLT (see Section 3.3.6).

Dataset Description	Primary Dataset Name	σ (pb)
$t\bar{t}$	/TTJets_MassiveBinDECAY_TuneZ2star_8TeV-madgraph-tauola/ /TTJets_matchingup_TuneZ2star_8TeV-madgraph-tauola /TTJets_matchingdown_TuneZ2star_8TeV-madgraph-tauola /TTJets_scaledown_TuneZ2star_8TeV-madgraph-tauola /TTJets_scaleup_TuneZ2star_8TeV-madgraph-tauola	245.8
$t\bar{t}$	/TTJets_mass161_5_TuneZ2star_8TeV-madgraph-tauola /TTJets_mass163_5_TuneZ2star_8TeV-madgraph-tauola /TTJets_mass166_5_TuneZ2star_8TeV-madgraph-tauola /TTJets_mass169_5_TuneZ2star_8TeV-madgraph-tauola /TTJets_mass175_5_TuneZ2star_8TeV-madgraph-tauola /TTJets_mass178_5_TuneZ2star_8TeV-madgraph-tauola /TTJets_mass181_5_TuneZ2star_8TeV-madgraph-tauola /TTJets_mass184_5_TuneZ2star_8TeV-madgraph-tauola	245.8
$t\bar{t}$	/TT_CT10_TuneZ2star_8TeV-powheg-tauola	245.8
	/TT_8TeV-mcatnlo	245.8
	/TTJets_FullLeptMGDecays_8TeV-madgraph	24.6
	/TT_noCorr_8TeV-mcatnlo	245.8
	/TT_CT10_TuneZ2star_8TeV-powheg-tauola	245.8
	/TTTo2L2Nu2B_8TeV-powheg-pythia6	24.6

Table 6.2: Summary of the Monte Carlo samples used for the systematic studies.

These samples are cleaned by requiring the luminosity sections to be validated by the Data Quality Monitoring (DQM) and the Physics Performance and Dataset (PPD) groups according to specific validation criteria defined for each sub-detector. Besides, low pileup runs and overlap between data samples are removed.

Beam scrapping events are vetoed by selecting events with a significant fraction of high purity tracks with respect to the total number of tracks ($> 25\%$) when the event has at least 10 tracks. Events with anomalous HCAL noise are also rejected. Finally, at least one primary vertex with more than four effective degrees of freedom, a z position within 24 cm of the nominal detector center and a radial position smaller than 2 cm to the beam spot is required.

6.1.1 Comparison Between $t\bar{t}$ Generators

The event generator used to produce the signal MC sample used in the analysis has been selected after comparing different available choises. The criterium established is

Dataset Description	Dataset Name	$\mathcal{L}\,(\,\mathrm{pb}^{-1})$
Run2012A Muon	/DoubleMu/Run2012A-13Jul2012-v1	810
Run2012A Muon	/DoubleMu/Run2012A-recover-06Aug2012-v1	82
Run2012B Muon	/DoubleMu/Run2012B-13Jul2012-v1	4404
Run2012A Electron	/DoubleElectron/Run2012A-13Jul2012-v1	810
Run2012A Electron	/DoubleElectron/Run2012A-recover-06Aug2012-v1	82
Run2012B Electron	/DoubleElectron/Run2012B-13Jul2012-v1	4404
Run2012A MuonElectron	/MuEG/Run2012A-13Jul2012-v1	810
Run2012A MuonElectron	/MuEG/Run2012A-recover-06Aug2012-v1	82
Run2012B MuonElectron	/MuEG/Run2012B-13Jul2012-v1	4404

Table 6.3: Summary of the data samples used in the analysis.

the level of data to MC agreement obtained in some of the most important distributions. The samples studied are listed in Table 6.4.

Figure 6.1 compares the distribution of the difference of the azimuthal angle between the two leptons ($\Delta\phi_{\ell\ell}$) and the dilepton invariant mass ($m_{\ell\ell}$) predicted by different $t\bar{t}$ generators to the ones observed in data. Additional distributions are shown in Appendix A. The $\Delta\phi_{\ell\ell}$ distribution shows that the MADGRAPH $t\bar{t}$ sample with spin correlations (MADGRAPH SC) describes better the data, taking as reference the level of agreement between data and MC in the tails of the distribution. A similar effect can be seen in the low mass region of the dilepton invariant mass distribution, where the other generators such as MC@NLO do not describe data as well as MADGRAPH with SC. Therefore, this generator is used in the analysis in order to estimate the signal acceptance and efficiency of the selection cuts.

Dataset description	Dataset name
$t\bar{t}$ MADGRAPH with SC*	/TTJets_FullLeptMGDecays_8TeV-madgraph-tauola
$t\bar{t}$ MADGRAPH w/o SC*	/TTJets_MassiveBinDECAY_TuneZ2star_8TeV-madgraph-tauola/
$t\bar{t}$ MC@NLO	/TT_8TeV-mcatnlo/
$t\bar{t}$ POWHEG	/TT_CT10_TuneZ2star_8TeV-powheg-tauola/

Table 6.4: $t\bar{t}$ signal samples used to compare and select the generator that has the best agreement with data. (* Spin Correlations (SC))
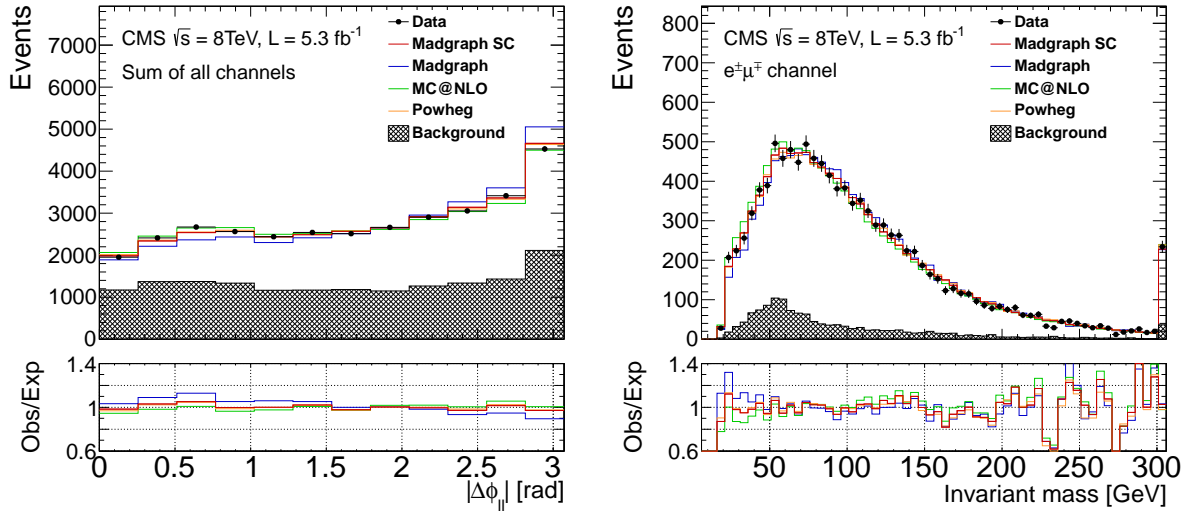
Figure 6.1: Difference of the azimuthal angle between the two selected leptons $\Delta\phi_{\ell\ell}$ (left) and dilepton invariant mass $m_{\ell\ell}$ (right) for the different signal generators and data for events with at least 2 jets.

6.2 Event Selection

6.2.1 Pileup Reweighting Procedure

The simulated events include additional interactions per bunch crossing (pileup), in order to reproduce those observed in data. As the luminosity of the LHC is continuously changing, a MC reweight factor must be applied to reproduce the data taking conditions. The target pileup distribution for data is generated using information on the instantaneous luminosity per bunch crossing for each luminosity section and the total pp inelastic cross section of 69.4 mb. A poissonian smearing is applied to model statistical fluctuations in the actual number of pileup events present in the data. The source pileup distribution for MC simulation is obtained from the true number of pileup events mixed with the particular hard interaction process in each event. Comparisons of the distribution of the number of primary vertices in data and MC after the reweighting procedure are shown in Figure 6.2.
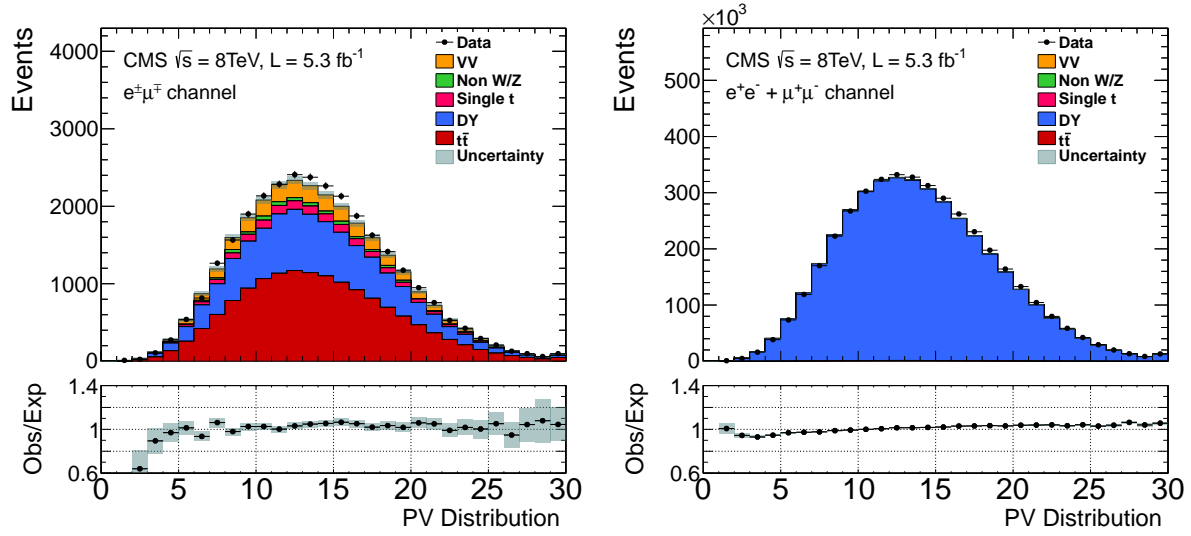
Figure 6.2: Data and MC primary vertex distributions for a luminosity of $5.3\,\text{fb}^{-1}$ in events with two reconstructed leptons in the μe (left) and the sum of ee and $\mu\mu$ (right) channels.

6.2.2 Trigger Selection

As presented in Section 3.3.6, in the CMS design, the online selection of events is achieved in two physical steps, namely the fast Level-1 Trigger and the High-Level Trigger (HLT) operating on longer timescales. The Level-1 trigger is built mostly of hardware level information from the subdetectors, while the HLT selection is implemented as a sequence of reconstruction and filter steps of increasing complexity.

Several High Level inclusive triggers have been considered to select events with two leptons in the final states:

- di-electron (ee):
  HLT_Ele17_CaloIdT_CaloIsoVL_TrkIdVL
  TrkIsoVL_Ele8_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL

- di-muon ($\mu\mu$):
  HLT_Mu17_Mu8 OR
  HLT_Mu17_TkMu8

- electron-muon(μe):
  HLT_Mu17_Ele8_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL OR

HLT_Mu8_Ele17_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL

The trigger selection for MC studies relies on:

- di-electron (ee):
  HLT_Ele17_CaloIdT_CaloIsoVL_TrkIdVL
  TrkIsoVL_Ele8_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL

- di-muon ($\mu\mu$):
  HLT_Mu17_Mu8 OR
  HLT_Mu17_TkMu8

- electron-muon (μe):
  HLT_Mu17_Ele8_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL OR
  HLT_Mu8_Ele17_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL

6.2.3 Selection of Signal and Background Events

The final states of the $t\bar{t}$ signal process are characterized by the presence of two high-p_T isolated leptons coming from W boson decays, with a large missing transverse energy $\not{E}_T$ associated to the neutrinos involved in the W boson decays, and two b-jets (see also Section 2.2.4). In addition to the object selection presented in Chapter 5, a set of requirements are applied in order to increase the purity of the signal in the dataset. This selection is based on the following requirements [82]:

Electron selection

The reconstruction of the electrons is explained in detail in Section 5.5.1. The additional selection criteria for PF electron candidates are:

- $p_T > 20\,\text{GeV}$ and $|\eta| < 2.5$.

- Transverse impact parameter whith respect to the beam spot $< 0.04\,\text{cm}$, applied on the GSF track [1].

- Photon conversion rejection: number of lost hits in the tracker $N_{\text{lost}} < 1$ and minimal distance between the electron and its closest opposite sign track $|\Delta\cos\theta| >$

---

[1]Gaussian Sum Filter Track, see Section 5.5.1

0.02 and distance in the $r - \phi$ plane $d_{r\phi} > 0.02$ [83].

- A discriminator based in a multivariate analysis of various electron reconstruction variables (MVA) [84]: MVA > 0.5 is required as an optimal compromise between efficiency of electron identification and rejection of fake candidate electrons.

- $\Delta R > 0.1$ (where $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$) between the electron and any global muon in the event.

- Relative Electron Isolation (I_R^e) < 0.15 (see Section 5.5.1) based on particle flow candidates and computed in a cone of 0.3 around the electron direction. Charged PF candidates coming from PU events are removed.

Muon selection

The reconstruction of the muons is explained in detail in Section 5.5.2. The additional selection criteria for PF muon candidates are:

- $p_T > 20\,\text{GeV}$ and $|\eta| < 2.4$,

- The muon should be reconstructed with the tracker or with the tracker and the muon espectrometer (see global and tracker muon Section 5.5.2).

- Relative Muon Isolation (I_R^μ) < 0.15 (see Section 5.5.2) based on particle flow candidates and computed in a cone of 0.3 around the muon direction. Charged PF candidates coming from PU events are removed.

Jets selection

As in Section 5.6, the jets are reconstructed using PF candidates with the anti-k_t algorithm and an opening angle of 0.5. Additional selection criteria are applied to jets:

- L1Fastjet corrections compatible with PFnoPU, Level 2 and Level 3 jet energy corrections and L2L3Residual corrections for data (see Section 5.6.1).

- $p_T > 30\,\text{GeV}$ and $|\eta| < 2.5$.

- Loose Jet Identification, requiring a fraction of charged hadronic energy $E_h^q \geq 0$,

a fraction of charged electromagnetic energy $E_e^q \leq 0.99$, a fraction of neutral hadronic energy $E_h^n < 0.99$, and a fraction of neutral electromagnetic energy $E_e^n < 0.99$.

- Jets overlapping with selected leptons (electrons or muons) are excluded if the distance between the jet and the lepton is $\Delta R(jet, \ell) < 0.5$.

Missing transverse energy

The missing transverse energy $\not{E}_T$ is defined as the magnitude of the transverse momentum imbalance, which is the negative sum of the momentum of all reconstructed particles in the transverse plane of the beam (see Section 5.7) . Type I corrections have been applied, as explained in Section 5.7.

b-tagged jets

As the $t\bar{t}$ signal is rich in jets originating from b quarks, a b-tagging requirement is used to improve the background rejection and therefore to clean the signal in the data analysis. Section 5.6.2 describes the b-tag jet identification techniques used in CMS. In this analysis, the combined secondary vertex algorithm with the loose (CSVL) working point is used. This algorithm gives the best b-tagging efficiency ($\sim$80%) for a misidentification rate of $\sim$10% (see Figure 5.6).

Additionally, as the simulations do not reproduce perfectly the b-tagging performance observed in data, data-to-MC scale factors are applied [72, 85] to the MC events.

6.2.4   Event Selection

In order to reduce the contribution of the various background processes while keeping a good signal efficiency, the following sequential cuts are applied:

1. Presence of a pair of selected leptons with opposite charges. Over the muon in the $\mu^\pm e^\mp$ final state and at least one muon in the $\mu^+\mu^-$ final state a $|\eta| < 2.1$ trigger fiducial cut is applied. The requirement of two leptons rejects a large number of W+jets and $t\bar{t}$ semi-leptonic events.

2. Dileptonic invariant mass $m_{\ell\ell} > 20\,\text{GeV}$. This requirement is applied in order

to remove Drell-Yan events with low invariant mass. Additionally, events with $76 < m_{\ell\ell} < 106\,\text{GeV}$ are rejected for the e^+e^- and $\mu^+\mu^-$ channels. This cut rejects around 90% of Z+jets events in those channels.

3. At least two jets with $p_T > 30\,\text{GeV}$. The $t\bar{t}$ final state includes two jets from the b quark hadronization. As Figure 6.3 (b) shows, in events with less than two jets, the background is dominant with a contribution of 98%, versus a 2% of signal. In events with at least two jets, the signal contribution increases up to 58% while the background is reduced to 42%.

4. Missing transeverse energy $\not{E}_T > 40\,\text{GeV}$ for the e^+e^- and $\mu^+\mu^-$ channels. This $\not{E}_T$ cut rejects more than 65% MC Drell-Yan events with a loss of signal efficiency of about 10% [86] giving the best compromise between the signal efficiency and the signal over signal plus background. The $\not{E}_T$ requirement is not applied in the μe channel.

5. As Figure 6.3 (d) shows, to require at least one b-tagged jet is enough to reject most of the remaining Drell-Yan background maintaining a high number of signal events.

Events having more than two selected leptons are classified in the $\mu^+\mu^-$, e^+e^- or $\mu^\pm e^\mp$ channels according to the flavor of the lepton pair of opposite charge maximizing the sum of the transverse momenta.

6.3   Selection Efficiencies from Data

After the pileup reweighting, the Monte Carlo simulations still do not describe properly the performance of the event reconstruction and selection. As a consequence, additional trigger, identification and isolation scale factors are applied to MC in order to increase the agreement with data. They are explained in the following sections. Section 6.3.1 presents the trigger efficiencies, section 6.3.2 describes the lepton identification and isolation efficiencies, and section 6.3.3 gives a summary of the scale factors used in the analysis.
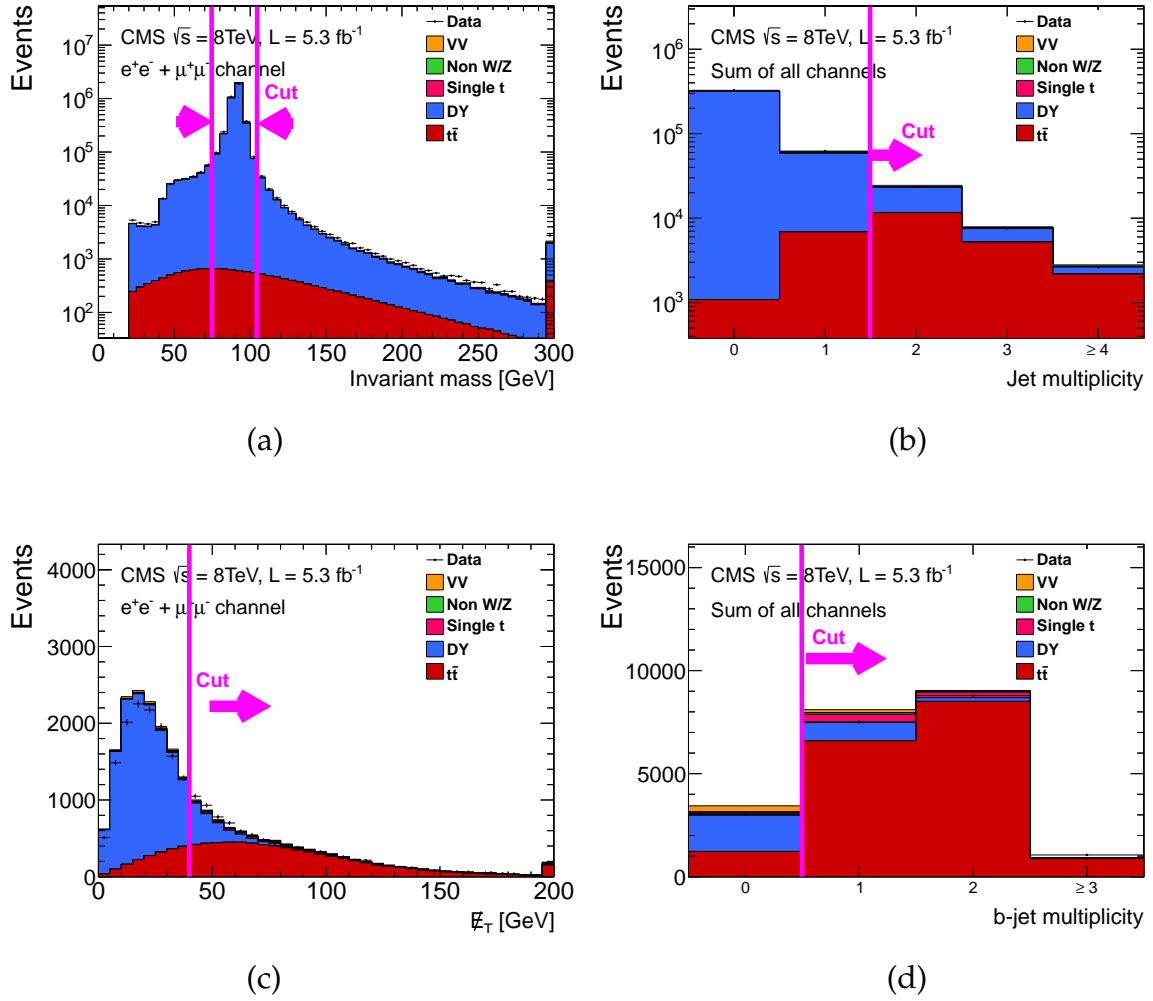
Figure 6.3: Distributions for the dilepton invariant mass (a), jet multiplicity (b), missing transverse energy $\not{E}_T$ (c) and b-jet multiplicity (d) before the corresponding cut. The purple lines show the cut applied on each quatity.

6.3.1 Trigger Efficiencies

The method used to estimate the dilepton trigger efficiencies can be summarized as follow:

- Determine a set of triggers (cross triggers) weakly correlated with the dilepton triggers used in the analysis.

- Count the number of events passing the cross triggers and the $t\bar{t}$ dilepton event selection (N_{Xtrig}).

- Count the number of events which pass the cross trigger selection, the $t\bar{t}$ dilepton event selection and the dilepton trigger selection ($N_{Xtrig+\ell\ell trig}$).

Then, the dilepton trigger efficiency in data is given by the ratio:

$$\varepsilon_{trigg} = \frac{N_{Xtrig+\ell\ell trig}}{N_{Xtrig}} \qquad (6.1)$$

The main difficulty of this method is to find cross triggers weakly correlated with the dilepton triggers which allow at the same time to select enough dilepton events to have a small statistical uncertainty. On the other hand, in some sense the requirement to have weakly correlated cross triggers is equivalent to use $t\bar{t}$ signal events to estimate the dilepton trigger efficiency. Trigger efficiency is then much less biased by kinematic differences between Z and $t\bar{t}$ events as for the tag&probe method.

The missing transverse energy based datasets were selected as the cross triggers, which were found to be weakly correlated with the dilepton ones and to have a large enough number of events to keep the statistical uncertainty below 1%. The chosen datasets are presented in Table 6.5. The measured efficiencies are compared to the efficiencies for $t\bar{t}$ MC events.

Dataset Description	Dataset Name
$\not{E}_T$ ReReco	/MET/Run2012A-13Jul2012-v1/AOD
	/MET/Run2012A-recover-06Aug2012-v1/AOD
	/MET/Run2012B-13Jul2012-v1/AOD

Table 6.5: Datasets used to measure the efficiencies of the dilepton triggers used in the analysis. The total luminosity corresponds to $5.3\,\text{fb}^{-1}$.

In order to keep the statistical uncertainty as low as possible, a set of 50 triggers is se-
lected for the efficiency estimation. The missing transverse energy trigger thresholds
and pre-scales are constant for the data sample considered. Some of the triggers used
are given as reference in Table 6.6.

One source of systematic uncertainty is related to the correlation between $\not{E}_T$ and
dilepton triggers. It is estimated from $t\bar{t}$ MC events by counting the number of events
passing only the $\not{E}_T$ triggers, only the dilepton triggers and passing both of them. In
the case where the dilepton and $\not{E}_T$ triggers are independent, the efficiency to fulfill
both trigger selections can be factorized as the product of the dilepton and $\not{E}_T$ trigger
efficiencies such that:

$$\varepsilon_{\ell\ell trg,METtrg} = \varepsilon_{\ell\ell trg} \times \varepsilon_{METtrg} \tag{6.2}$$

where $\varepsilon_{\ell\ell trg}$ and ε_{METtrg} are the efficiencies to pass the dilepton and the MET trigger
selections, respectively, and $\varepsilon_{\ell\ell trg,METtrg}$ is the efficiency to pass both trigger selections.

The following ratio gives the correlation between the two trigger selections:

$$\alpha = \frac{\varepsilon_{\ell\ell trg}^{MC} \times \varepsilon_{METtrg}^{MC}}{\varepsilon_{\ell\ell trg,METtrg}^{MC}} \tag{6.3}$$

The ratio α is determined from MC, and it was found to be 1.005, 0.999 and 0.995
for the ee, $\mu\mu$ and $e\mu$ channels, respectively, resulting on a systematical uncertainty
of about 0.5%. The total systematic uncertainty on the measured dilepton trigger
efficiencies is conservatively taken to be 1%, to account not only for the correlation
but also for other factors, for instance the possible difference between the behaviour
of triggers in data and simulation or a bias given by the selected $\not{E}_T$ datasets.

The summary of the trigger scale factors, as well as of their uncertainties, is given in
Table 6.7 for the different levels of selection. The scale factors through the different
steps are compatible within the quoted uncertainty.

The scale factors measured for events after the requirement of two leptons is chosen
to correct the simulation yield. The dilepton trigger efficiency measured in data and
MC and the resulting scale factors as a function the transverse momentum p_T and

Cross Triggers
HLT_MET120
HLT_MET200
HLT_PFHT350_PFMET100
HLT_PFHT400_PFMET100
HLT_MET80_Track50_dEdx3p6
HLT_MET80_Track60_dEdx3p7
HLT_MET120_HBHENoiseCleaned
HLT_MET200_HBHENoiseCleaned
HLT_CentralPFJet80_CaloMET50_dPhi1_PFMHT80_HBHENoiseFiltered

Table 6.6: Some of the $\not{E}_T$ trigger used as cross triggers in the estimation of the dilepton trigger efficiencies.

Cut Level	ee	$\mu\mu$	$e\mu$
dilepton	0.974 ± 0.012	0.967 ± 0.010	0.953 ± 0.011
+ Z-Veto	0.964 ± 0.012	0.969 ± 0.010	0.953 ± 0.011
+ $\geq$ 2jets	0.971 ± 0.013	0.968 ± 0.010	0.955 ± 0.011
+ $\not{E}_T$	0.972 ± 0.013	0.968 ± 0.010	0.955 ± 0.011

Table 6.7: Summary of the trigger scale factors at different levels of the event selection. The errors correspond to the sum of the systematic and statistical uncertainties.

pseudorapidity η of the leading lepton are shown in Figure 6.4, 6.5 and 6.6 for the dielectron, dimuon and electron-muon channels, respectively. The scale factors show no significant dependence on lepton kinematics, and their variations are within the total uncertainties.

Table 6.8 summarizes the dilepton trigger efficiencies measured in data (ε_{trg}^{data}) and MC (ε_{trg}^{MC}) and the corresponding scale factors (SF_{trg}) applied to correct the simulation yield.

The scale factors are stable within the uncertainties with respect to the lepton transverse momentum. In order to take into account the small dependence on the lepton pseudorapidity, the scale factors are determined and applied to the simulations in bins of η. Figure 6.7, 6.8 and 6.9 show these scale factors for the dielectron, dimuon, and $e\mu$ channels, respectively.
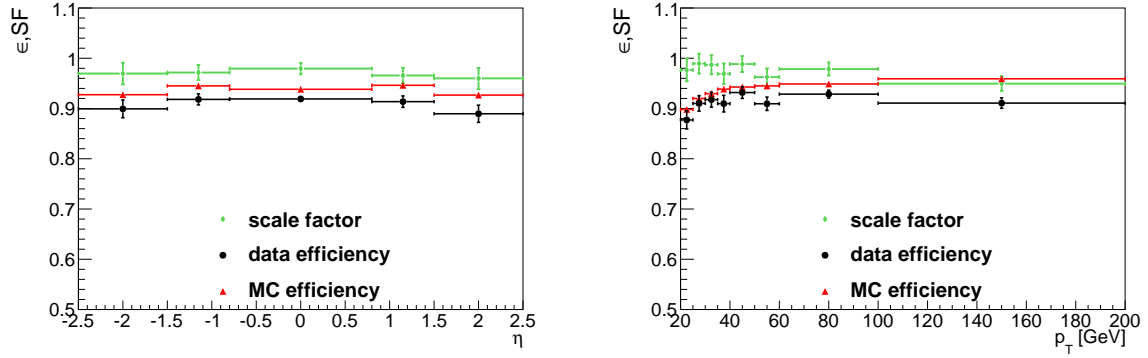
Figure 6.4: Dielectron trigger efficiency as a function of the pseudorapidity η (left) and the transverse momentum p_T (right) of the leading electron, for data (black dots), MC (red triangles) and the respective scale factor (green diamonds). Error bars correspond to the statistical uncertainty.
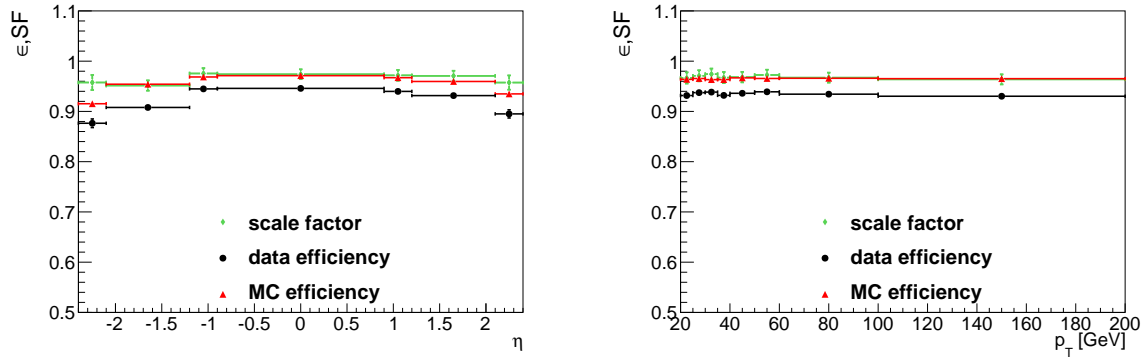
Figure 6.5: Dimuon trigger efficiency as a function of the pseudorapidity η (left) and the transverse momentum p_T (right) of the leading muon, for data (black dots), MC (red triangles) and the respective scale factor (green diamonds).

	ε_{trg}^{data}	ε_{trg}^{MC}	$\alpha - 1$	SF_{trg}
ee	0.914 ± 0.006 (stat.)	0.939 ± 0.001 (stat.)	0.005	0.974 ± 0.007 (stat.) ± 0.010 (syst.)
$\mu\mu$	0.934 ± 0.002 (stat.)	0.965 ± 0.001 (stat.)	-0.001	0.967 ± 0.002 (stat.) ± 0.010 (syst.)
$e\mu$	0.889 ± 0.005 (stat.)	0.933 ± 0.001 (stat.)	-0.005	0.953 ± 0.005 (stat.) ± 0.010 (syst.)

Table 6.8: Dilepton trigger efficiencies for data and MC, and corresponding scale factors measured after the requirement of two leptons. The results correspond to an integrated luminosity of $5.3\,\text{fb}^{-1}$.
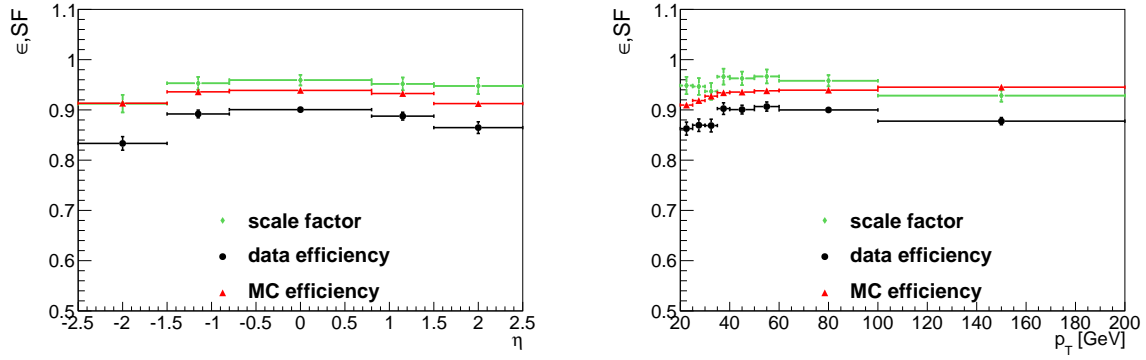
Figure 6.6: Electron-muon trigger efficiency as a function of the pseudorapidity η (left) and the transverse momentum p_T (right) of the leading lepton, for data (black dots), MC (red triangles) and the respective scale factor (green diamonds).
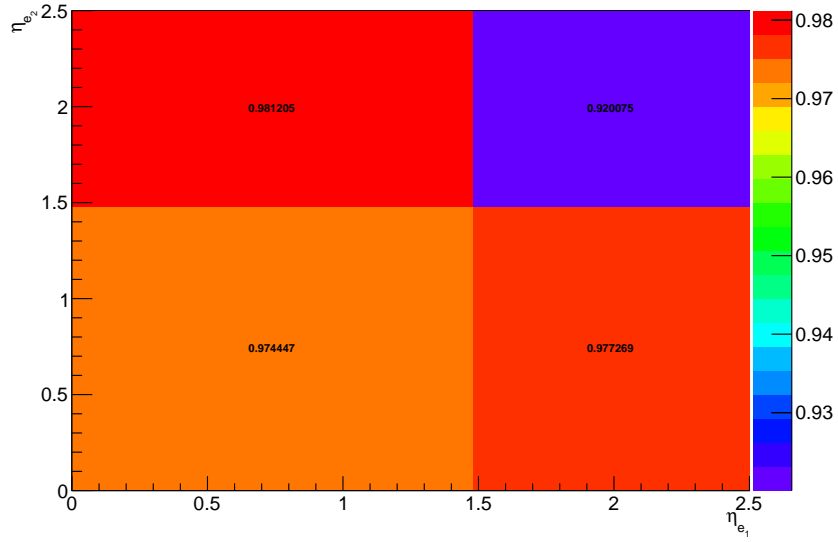
Figure 6.7: Scale factors for the ee trigger as a function of the η of the leading electron (x-axis) and the second leading electron (y-axis).
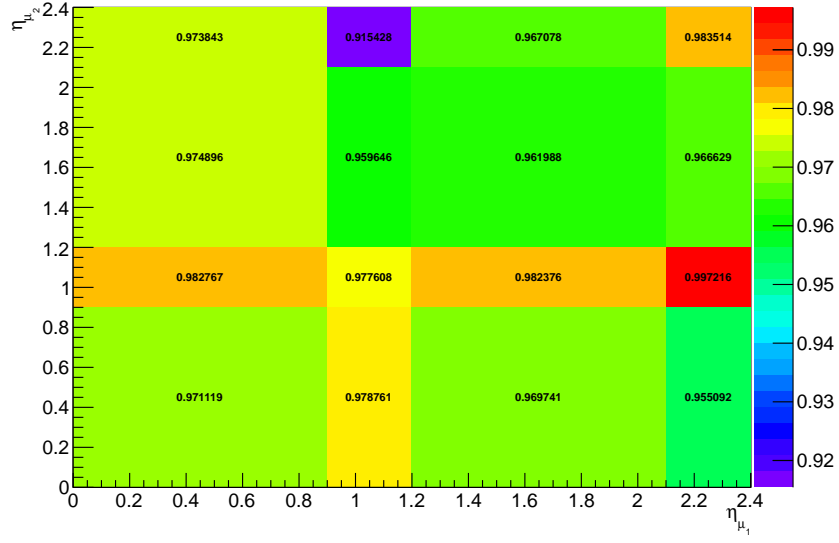
Figure 6.8: Scale factors for the $\mu\mu$ trigger as a function of the η of the leading muon (x-axis) and second leading muon (y-axis).
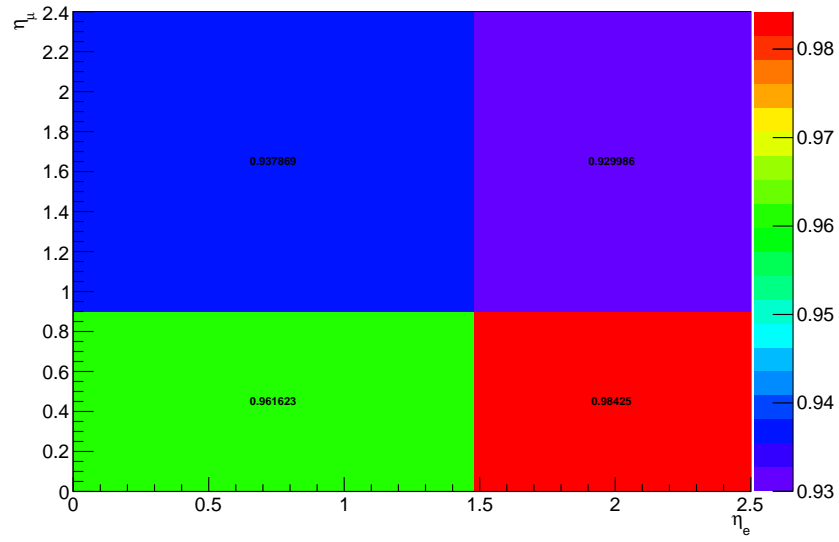
Figure 6.9: Scale factors for the μe trigger as a function of the η of the electron (x-axis) and muon (y-axis).

6.3.2   Lepton Identification and Isolation Efficiencies

The lepton efficiencies are estimated using a tag and probe method, following a simple cut and count approach[2]. In order to estimate the efficiency avoiding bias from the dilepton trigger selection, data samples collected by single lepton triggers are used. The trigger paths used in this section are illustrated in Table 6.9.

Dilepton candidates with invariant mass compatible with the mass of the Z boson are used to estimate the efficiency. The tag and probe leptons are required to have opposite charge and an invariant mass in the range $76\,\text{GeV} < m_{\ell\ell} < 106\,\text{GeV}$. "Tag" electrons corresponds to the complete electron selection, isolation and identification used in the analysis. Tag muons are required to pass tighter criteria than the one used in the analysis, to have a cleaner sample. These criteria involve requirements on the transverse impact parameter, number of hits in the muon chambers and tracker detector, and quality of the track fit. Tag leptons in both channels are selected if they are associated to the lepton from the trigger selection. The trigger bits used are HLT_IsoMu24_eta2p1 and HLT_IsoMu24 for the single muon dataset and HLT_Ele27_WP80 for the single electron one. All of them are unprescaled for the data period considered. The tag lepton is required to have $p_T > 30\,\text{GeV}$ to avoid bias from the p_T threshold of the trigger. An acceptance cut $|\eta| < 2.1$ is applied to tag muons.

The lepton isolation and identification efficiencies are estimated sequentially. The identification efficiency corresponds to the number of probe leptons passing the complete selection criteria but the isolation requirement over the total number of probe leptons. The isolation efficiency is defined as the ratio between leptons passing the selection criteria including the isolation cut and the total number of leptons passing the previous identification requirements.

The measured efficiencies are compared to the one observed in the Drell-Yan Monte Carlo to define scale factors to correct the MC predictions as:

$$SF_{ID,ISO} = \frac{\varepsilon_{ID,ISO}^{data}}{\varepsilon_{ID,ISO}^{MC}}, \tag{6.4}$$

where $\varepsilon_{ID,ISO}^{data}$ and $\varepsilon_{ID,ISO}^{MC}$ are the identification and isolation efficiencies measured in

---

[2]The method of counting the number of events in an established phase space defined by means of a set of cuts is called a "cut and count" approach.

data and MC, respectively.

Data Sample	Dataset Name
SingleMu	/SingleMu/Run2012A-13Jul2012-v1/
	/SingleMu/Run2012A-recover-06Aug2012-v1/
	/SingleMu/Run2012B-13Jul2012-v1
SingleElectron	/SingleElectron/Run2012A-13Jul2012-v1/
	/SingleElectron/Run2012A-recover-06Aug2012-v1/
	/SingleElectron/Run2012B-13Jul2012-v1

Table 6.9: Data samples used for the measurement of the lepton identification and isolation efficiencies. The luminosity used corresponds to $5.3\,\mathrm{fb}^{-1}$.

The identification and isolation efficiencies and scale factors for muons are presented as a function of the transverse momentum p_T and pseudorapidity η in Figure 6.10 and 6.11, respectively, while for electrons are shown in Figure 6.12 and 6.13. Table 6.10 and 6.11 present the average identification and isolation efficiencies and their combined value for data and MC, and the resulting scale factors.
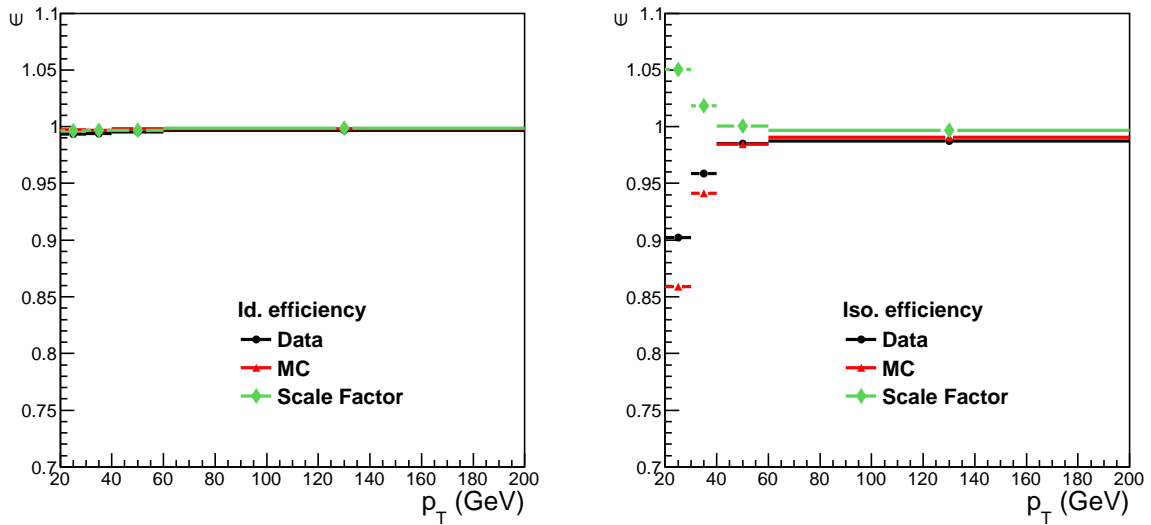
Figure 6.10: Muon identification (left) and isolation (right) efficiencies as a function of the p_T of the probe muon. Black dots correspond to the efficiency in data, red triangles to the efficiency in the Drell-Yan MC, and green squares to the data/MC efficiency ratio.

The systematic uncertainties are estimated by varying the invariant mass window and the tag lepton selection. The largest observed variation of the scale factors with respect
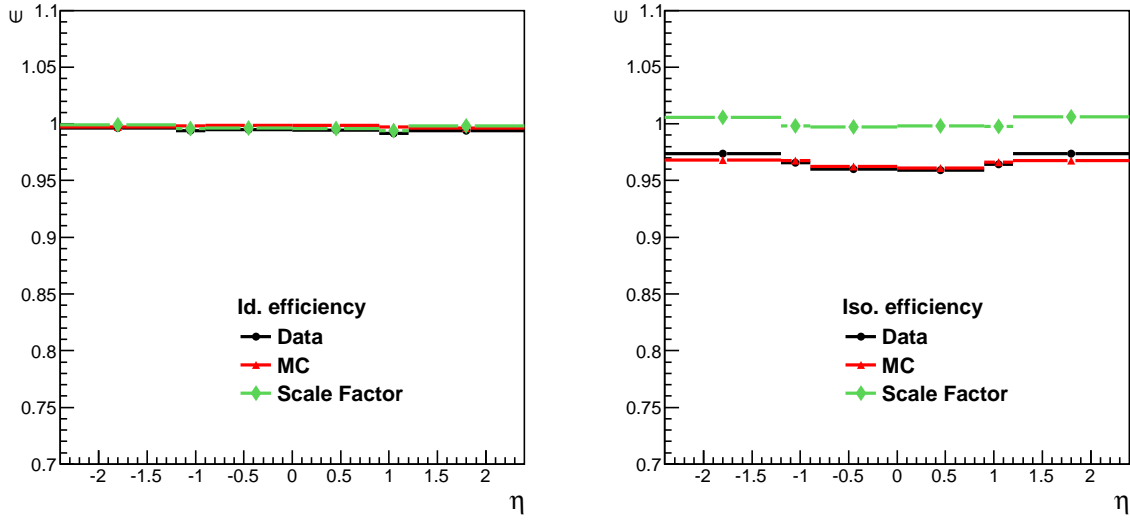
Figure 6.11: Muon identification (left) and isolation (right) efficiencies as a function of the η of the probe muon. Black dots correspond to the efficiency in data, red triangles to the efficiency in Drell-Yan MC, and green squares to the data/MC efficiency ratio.

	ε_{data}	ε_{MC}	SF
Identification (ID)	0.9947±0.0002	0.9978±0.0001	0.9968±0.0002
Isolation (ISO)	0.9669±0.0004	0.9645±0.0002	1.0024±0.0006
Total ID+ISO	0.9617±0.0004	0.9624±0.0002	0.9993±0.0006

Table 6.10: Muon identification and isolation efficiencies and scale factors. The errors shown correspond only to the statistical component.

	ε_{data}	ε_{MC}	SF
Identification (ID)	0.9304±0.0002	0.9361±0.0002	0.9938±0.0004
Isolation (ISO)	0.9590±0.0002	0.9619±0.0002	0.9969±0.0003
Total ID+ISO	0.8922±0.0003	0.9004±0.0002	0.9908±0.0003

Table 6.11: Electron identification and isolation efficiencies and scale factors. The errors shown correspond only to the statistical component.
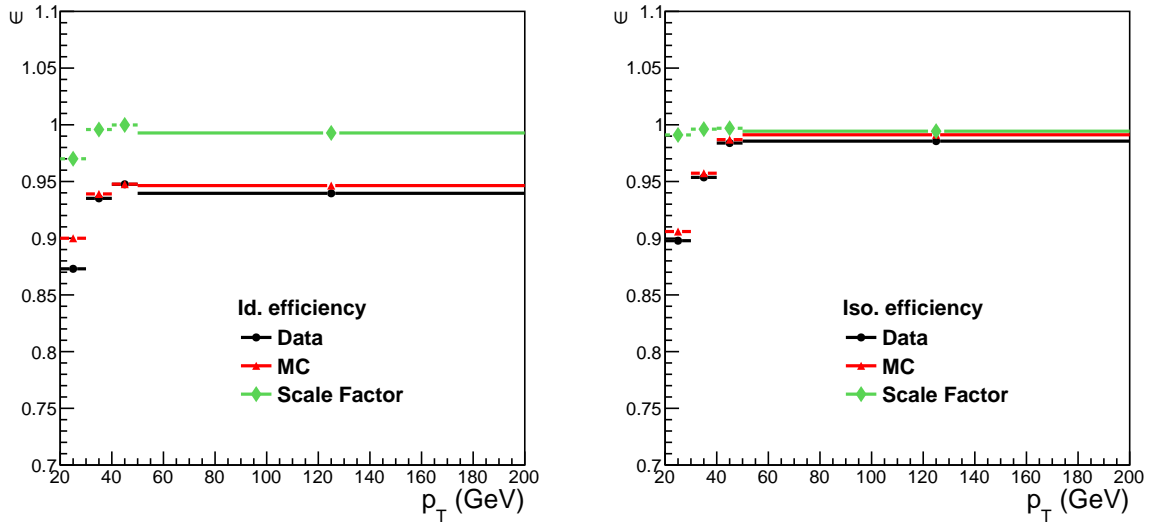
Figure 6.12: Electron identification (left) and isolation (right) efficiencies as a function of the p_T of the probe electron. Black dots correspond to the efficiency in data, red triangles to efficiency in the DY MC, and green squares to the data/MC efficiency ratio.
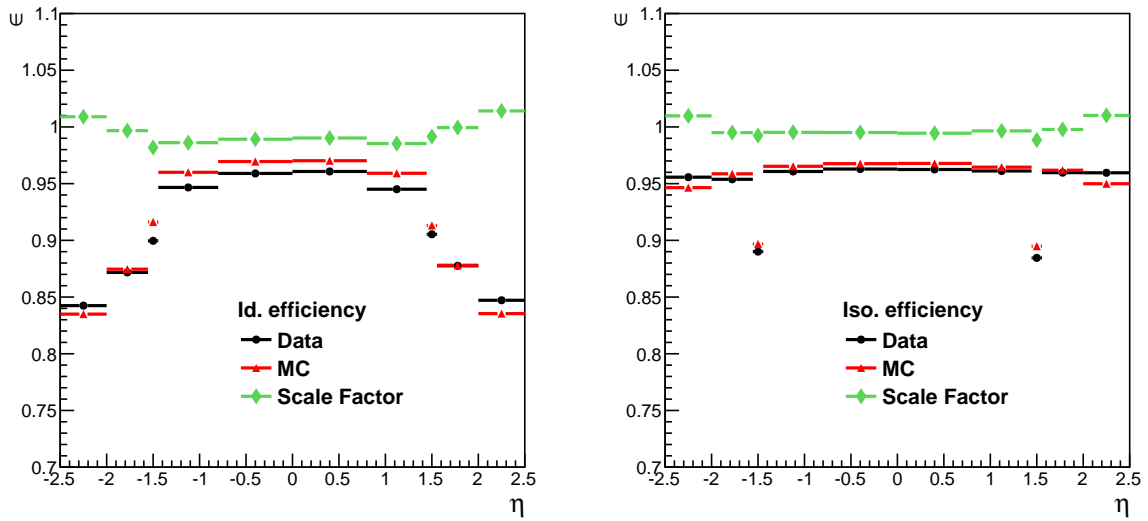
Figure 6.13: Electron identification (left) and isolation (right) efficiencies as a function of the η of the probe electron. Black dots correspond to efficiency in data, red triangles to efficiency from in DY MC, and green squares to the data/MC efficiency ratio.

to their nominal value is about 0.3%. To take into account the different topology of the leptons coming from the top decay and Z decay a conservative systematic uncertainty of 1% is used for the global values.

Figure 6.14 shows the total scale factors as a function of the pseudorapidity η and the transverse momentum p_T of the lepton. These are the scale factors used to correct the simulation in the analysis.
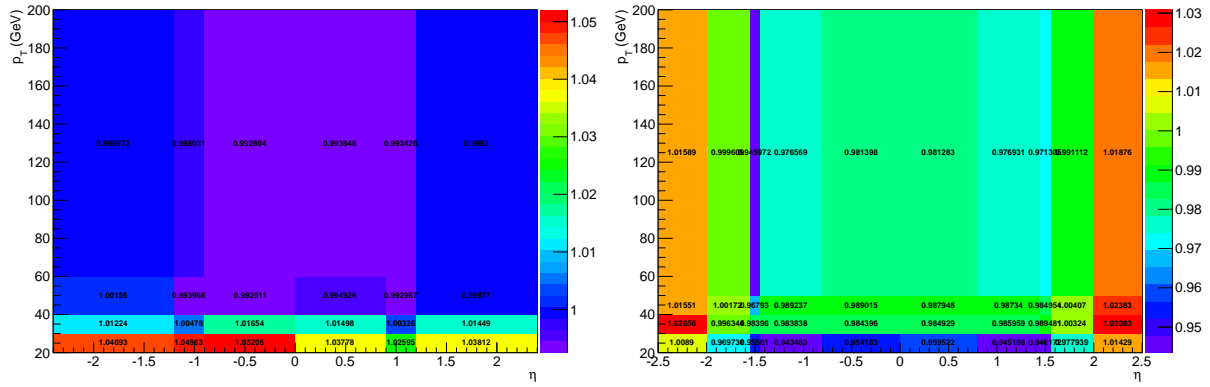
Figure 6.14: Scale factors as a function of the pseudorapidity η and the transverse momentum p_T of the lepton, for muons (left) and electrons (right).

6.3.3   Summary of Trigger and Lepton Efficiencies

The final trigger scale factors (SF_{trg}) used for the analysis are those in bins of lepton pseudorapidity. We list below the average values with their systematic uncertainties. Since the trigger scale factors are stable through the different steps of the selection, the SF_{trg} measured after the dilepton requirement are used in order to reduce the statistical uncertainty, as aforementioned.

- ee **channel :** $SF_{trg}^{ee} = 0.974 \pm 0.012$

- $\mu\mu$ **channel :** $SF_{trg}^{\mu\mu} = 0.967 \pm 0.010$

- eμ **channel :** $SF_{trg}^{\mu e} = 0.953 \pm 0.010$

In the analysis, the identification and isolation scale factors in bins of p_T and η are used. For comparison, average scale factors are presented, including their systematic uncertainties.

- ee **channel :** $SF^{\text{ee}}_{ID,ISO} = 0.982 \pm 0.020$

- μe **channel :** $SF^{\mu\mu}_{ID,ISO} = 0.999 \pm 0.020$

- $\mu\mu$ **channel :** $SF^{\mu\text{e}}_{ID,ISO} = 0.990 \pm 0.014$

The final average scale factors accounting for trigger, isolation and identification efficiencies result to be:

- ee **channel :** $SF^{\text{ee}}_{trg,ID,ISO} = 0.955 \pm 0.023$

- $\mu\mu$ **channel :** $SF^{\mu\mu}_{trg,ID,ISO} = 0.966 \pm 0.022$

- μe **channel :** $SF^{\mu\text{e}}_{trg,ID,ISO} = 0.944 \pm 0.017$

6.4 Background Determination

6.4.1 Determination of the Drell-Yan Background

The Drell-Yan (DY) process is the main background in the e^+e^- and $\mu^+\mu^-$ channels after the full event selection; it also contributes in the $\mu^\pm e^\mp$ channel via tau decays $(Z/\gamma^* \rightarrow \tau\tau \rightarrow e\mu\nu_e\nu_\tau\nu_\mu\nu_\tau)$. In order to estimate the Drell-Yan contribution from data, two different methods has been implemented in the analysis: the $R_{out/in}$ and the template fit methods. Even if the two methods provide compatible results, the $R_{out/in}$ technique is used to estimate the DY contribution in the three channels as explained in Section 6.4.2 and Section 6.4.3. The template fit method is used as a cross check of the results (Section 6.4.4).

6.4.2 $R_{out/in}$ in the $\mu^+\mu^-$ and e^+e^- Channels

To reduce the DY contribuition in the signal region, a veto in the dilepton invariant mass ($76 < m_{\ell\ell} < 106\,\text{GeV}$) is applied. The vetoed events can be used as a control region to estimate the remaining DY background.

The total number of DY events is:

$$N^{Total}_{DY} = N^{in}_{DY} + N^{out}_{DY} \tag{6.5}$$

Where N_{DY}^{in} and N_{DY}^{out} are the number of events inside and outside of the Z mass window, respectively. The ratio $R_{out/in}$ between the number of DY events predicted by the simulation outside and inside the Z mass window is computed. This ratio is used to estimate the DY events outside the veto region in data, starting from the number of events observed in the veto region:

$$N_{DY}^{out} = N_{DY\ Data}^{in} \times \overbrace{\left(\frac{N_{DY\ MC}^{out}}{N_{DY\ MC}^{in}} \right)}^{R_{out/in}} \tag{6.6}$$

This expresion assumes that the Z mass region is dominated by Drell-Yan events.

Non Drell-Yan processes contributing to the Z mass region can be divided in:

- **Peaking backgrounds**: WZ and ZZ processes give a peak in the reconstructed dilepton invariant mass if both leptons come from the Z boson decay.

- **Non peaking backgrounds**: WW, $t\bar{t}$, tW and W+jets give a continuous distribution in the dilepton invariant mass.

In the category of the *peaking backgrounds*, the dominant one is $ZZ \rightarrow \ell^+\ell^- \nu_\ell \overline{\nu_\ell}$. As the contamination come from a real Z boson, the $R_{out/in}$ is expected to be similar to the one from Drell-Yan production. In the case of the *non peaking backgrounds*, the estimation must be obtained from data. This can be achieved by measuring the number of events in the Z mass region for the $e\mu$ channel ($N_{\mu e}^{in}$). This number, scaled by a factor $k_{\ell\ell}$ which takes into account the combinatorics and efficiencies to reconstruct the opposite flavor final state relative to each same flavor final state, is used as an estimate of the non peaking backgrounds in the same flavour channels. Thus, the number of Drell-Yan events outside the veto region can be written as:

$$N_{DY}^{out\ (est)} = R_{out/in} \left(N_{\ell^+\ell^-}^{in} - N_{\mu e}^{in} k_{\ell\ell} \right) \tag{6.7}$$

The constant $k_{\ell\ell}$ is equal to 0.5 for the combinatoric factor between the μe and $\ell\ell$ final states, multiplied by a correction due to the difference in efficiency to reconstruct and select a muon compared to an electron. This correction can be determined from the number of ee and $\mu\mu$ events inside the Z mass window before any $\not{E}_T$ selection. This

is:

$$\frac{n_{\mu\mu}^{obs}}{n_{ee}^{obs}} = \frac{N_{\mu\mu}^{true}\mathcal{A}_{\mu\mu}\mathcal{E}_{\mu}^2}{N_{ee}^{true}\mathcal{A}_{ee}\mathcal{E}_{e}^2} \qquad (6.8)$$

Assuming that the produced number of $q\bar{q} \to e^+e^-$ is equal to the number of $q\bar{q} \to \mu^+\mu^-$, and that the purely geometric acceptances, $\mathcal{A}_{ee}$ and $\mathcal{A}_{\mu\mu}$, are the same:

$$\frac{n_{\mu\mu}^{obs}}{n_{ee}^{obs}} = \frac{\cancel{N_{\mu\mu}^{true}}\cancel{\mathcal{A}_{\mu\mu}}\mathcal{E}_{\mu}^2}{\cancel{N_{ee}^{true}}\cancel{\mathcal{A}_{ee}}\mathcal{E}_{e}^2} \quad \Rightarrow \quad \begin{cases} k_{\mu\mu} = \frac{1}{2}\sqrt{\frac{n_{\mu\mu}^{obs}}{n_{ee}^{obs}}} \\[2ex] k_{ee} = \frac{1}{2}\sqrt{\frac{n_{ee}^{obs}}{n_{\mu\mu}^{obs}}} \end{cases} \qquad (6.9)$$

As the $R_{out/in}$ ratio relies on MC, we study its dependence in data and MC as a function of the number of primary vertices and of the jet multiplicity. In order to guarantee that the sample is dominated by Drell-Yan events, the following cuts are applied:

- number of vertices: two leptons with invariant mass $m_{\ell\ell} > 20\,\text{GeV}$.

- jet multiplicity: two leptons with invariant mass $m_{\ell\ell} > 20\,\text{GeV}$ and $\not{E}_T < 10\,\text{GeV}$.

Comparing the dependences in the number of vertices and jet multiplicity distribution, we derive a correction factor $CF_{out/in} = R_{out/in}^{data}/R_{out/in}^{MC}$ to be applied to the $R_{out/in}$ value. Figure 6.15 and 6.16 show the dependence of the $R_{out/in}$ ratio on the number of primary vertices (N_{PV}) and jet multiplicity (N_{jets}), respectively. The correction factors extracted using a fit to a constant are:

$$\begin{aligned} CF_{out/in}^{ee}(N_{PV}) &= 1.0221 \pm 0.0004 \\ CF_{out/in}^{\mu\mu}(N_{PV}) &= 1.0142 \pm 0.0003 \\ CF_{out/in}^{ee}(N_{jets}) &= 1.163 \pm 0.062 \\ CF_{out/in}^{\mu\mu}(N_{jets}) &= 1.072 \pm 0.051 \end{aligned} \qquad (6.10)$$

The product of both correction factors has been used to scale the final $R_{out/in}$ value. The purity of DY events, also shown in Figure 6.15 and 6.16, is estimated using MC as the ratio between the number of Drell-Yan events over the total number of MC

events (N_{DY}/N_{Total}). This purity shows that the sample used to substract the CF is dominated by Drell-Yan events.
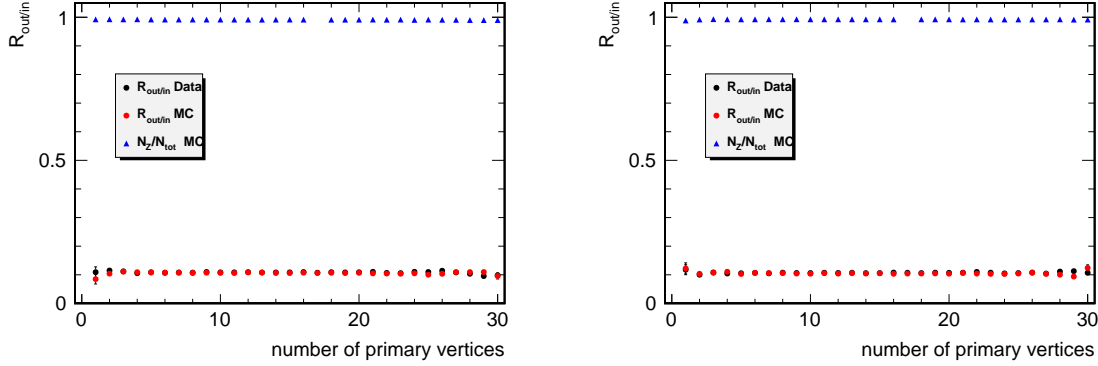
Figure 6.15: $R_{out/in}$ for the e^+e^- (left) and $\mu^+\mu^-$ (right) channels as a function of the number of the reconstructed primary vertices in data (black dots) and in MC (red dots) after the dilepton pair selection. The purity of Drell-Yan events from MC expectations is also shown (blue triangles).
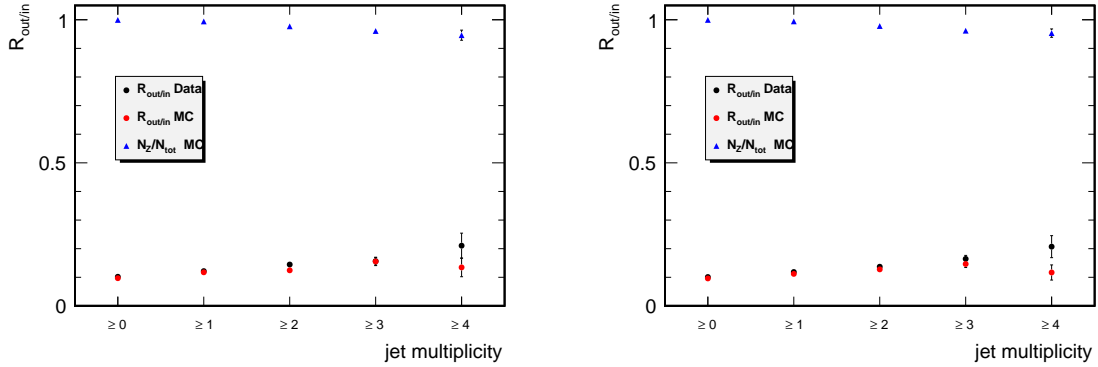
Figure 6.16: $R_{out/in}$ for the e^+e^- (left) and $\mu^+\mu^-$ (right) channels as a function of the jet multiplicity in data (black dots), and in MC (red dots) after the dilepton pair selection. The purity of Drell-Yan events from MC expectations is also shown (blue triangles).

Finally, the systematic uncertainty on the $R_{out/in}$ estimate arises from its dependence on the event selection. The uncertainty is computed as the fractional change in the values of $R_{out/in}$ before and after the $\not{E}_T$ requirement. This uncertainty is found to be of the order of 30% of the final Drell-Yan estimation.

In order to have a better prediction of the Drell-Yan contribution, the $R_{out/in}$ method is applied separately in events with no b-tagged jets and events with at least one b-

tagged jet. Table 6.12 shows the $R_{out/in}$ value, the number of events extracted from MC, the data-driven estimate and the ratio of the MC and data-driven yields (Drell-Yan scale factor, SF_{DY}) in the different channels and at different levels of the event selection.

electron-electron channel				
Cut Level	DY-MC	DY-DD	SF_{DY}^{ee}	$R_{out/in}$
≥ 2 Jets	5189.9 ± 55.7	6079.2 ± 1851.2	1.17 ± 0.35	0.13 ± 0.007
≥ 2 Jets + 0 btag	3398.2 ± 45.0	3714.9 ± 1131.7	1.09 ± 0.33	0.14 ± 0.007
≥ 2 Jets + ≥ 1 btag	1791.7 ± 32.9	2354.3 ± 718.0	1.31 ± 0.39	0.13 ± 0.007
≥ 2 Jets + $\not{E}_T$	669.2 ± 18.7	1082.1 ± 331.2	1.62 ± 0.49	0.18 ± 0.010
≥ 2 Jets + $\not{E}_T$ + 0 btag	448.2 ± 15.3	691.7 ± 212.3	1.54 ± 0.46	0.19 ± 0.012
≥ 2 Jets + $\not{E}_T$ + ≥ 1 btag	221.0 ± 10.8	385.9 ± 119.5	1.75 ± 0.52	0.16 ± 0.012
muon-muon channel				
Cut Level	DY-MC	DY-DD	$SF_{DY}^{\mu\mu}$	$R_{out/in}$
≥ 2 Jets	7218.1 ± 67.8	7778.1 ± 2361.6	1.08 ± 0.32	0.13 ± 0.006
≥ 2 Jets + 0 btag	4716.9 ± 54.7	4687.3 ± 1423.1	0.99 ± 0.30	0.13 ± 0.006
≥ 2 Jets + ≥ 1 btag	2501.1 ± 40.0	3078.9 ± 935.9	1.23 ± 0.37	0.13 ± 0.006
≥ 2 Jets + $\not{E}_T$	884.0 ± 21.9	1267.3 ± 386.4	1.43 ± 0.43	0.17 ± 0.009
≥ 2 Jets + $\not{E}_T$ + 0 btag	575.8 ± 17.7	780.3 ± 238.4	1.36 ± 0.41	0.17 ± 0.009
≥ 2 Jets + $\not{E}_T$ + ≥ 1 btag	308.2 ± 12.9	486.3 ± 149.6	1.58 ± 0.47	0.16 ± 0.010

Table 6.12: Drell-Yan data-driven estimate in the e^+e^- and $\mu^+\mu^-$ channels compared with the expectations from simulation for several steps of the analysis in events with 0 b-tagged jets or at least 1 b-tagged jet.

6.4.3 $R_{out/in}$ in the $\mu^\pm e^\mp$ Channel

Assuming the physics modeling of the Drell-Yan events to be the same for each decay mode (into e^+e^-, $\mu^+\mu^-$ and $\tau^+\tau^-$), the Drell-Yan scale factor can be expressed as:

$$SF_{DY}^{\ell\ell} = SF_{physics} \times SF_\ell^2 \qquad (6.11)$$

which contains the phyisics behavior (the same for the three channels) and the detec-

tor effect (efficiency modeling in the ee and $\mu\mu$ decay modes). It is therefore possible to estimate the scale factor in the $\mu^\pm e^\mp$channel ($SF_{DY}^{e\mu}$) as:

$$SF_{DY}^{\mu\mu} = SF_{physics} \times SF_\mu^2$$
$$SF_{DY}^{ee} = SF_{physics} \times SF_e^2$$
$$SF_{DY}^{\mu e} = SF_{physics} \times SF_e \times SF_\mu$$
$$SF_{DY}^{\mu e} = \sqrt{SF_{DY}^{ee} \times SF_{DY}^{\mu\mu}} \tag{6.12}$$

where the $\mu\mu$ and ee scale factors are estimated without the $\not{E}_T$ cut to be consistent with the eμ selection. The values are extracted in events with 0 b-tagged jet ($SF_{DY}^{\mu e} = 1.04 \pm 0.31$) and at least 1 b-tagged jets ($SF_{DY}^{\mu e} = 1.27 \pm 0.28$).

6.4.4 Template fit: alternative method to estimate the Drell-Yan background

As a cross check, the Drell-Yan contribuiton is estimated through a fit at two components to the dilepton invariant mass distribution, one component reflecting the dilepton mass distribution for Z/γ^* events and the other corresponding mainly to t$\bar{t}$, single-top quark production and diboson events. The templates for these components are extracted from simulations. The results of the fits in the eμ channel are presented in Figure 6.17. Similar fits are performed for the e^+e^- (Figure 6.18) and the $\mu^+\mu^-$ (Figure 6.19) channels after the jet, $\not{E}_T$ and b-tagging selections. The scale factors obtained with the two data-driven methods are compared in Table 6.13.

The SF_{DY} estimated with the template fit have an error of 30% covering the statistical and the systematic uncertainties. The systematic uncertainty comes from the DY and t$\bar{t}$ invariant mass shapes predicted from the MC. The results obtained with the template fit and the $R_{out/in}$ methods are in good agreement.

6.4.5 Non-W/Z Background Estimate

Backgrounds with one jet misidentified as a lepton (W+jets production and t$\bar{t}$ events with semileptonic decay) are estimated in a data-driven way from a sample of events selected using a loose lepton identification and isolation requirement.
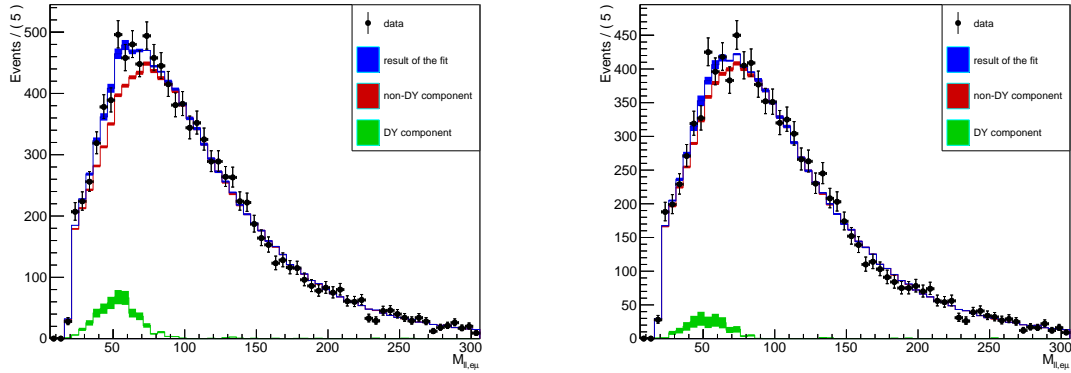
Figure 6.17: Results of the dilepton invariant mass fit, after the jet multiplicity (left) and b-tagging (right) selections in the $\mu^{\pm}e^{\mp}$ channel. Black dots correspond to data, blue band corresponds to the output of the fit, green band to the DY component after the fit, and red band to the contribution from other processes after the fit. The width of the bands corresponds to the statistical uncertainty.

Template Fit			
Channel	≥ 2 jets	≥ 2 jets + $\not{E}_T$	≥ 2 jets + $\not{E}_T$+ ≥ 1 btag
$\mu\mu$	1.04 ± 0.31	1.13 ± 0.34	1.04 ± 0.31
ee	1.13 ± 0.34	1.46 ± 0.44	1.89 ± 0.57
Channel	≥ 2 jets	≥ 2 jets + $\not{E}_T$	≥ 2 jets + ≥ 1 btag
μe	1.08 ± 0.32	$- - -$	1.39 ± 0.42
$R_{out/in}$			
Channel	≥ 2 jets	≥ 2 jets + $\not{E}_T$	≥ 2 jets + $\not{E}_T$+ ≥ 1 btag
$\mu\mu$	1.08 ± 0.32	1.43 ± 0.43	1.58 ± 0.47
ee	1.17 ± 0.35	1.62 ± 0.48	1.75 ± 0.52
Channel	≥ 2 jets	≥ 2 jets + $\not{E}_T$	≥ 2 jets + ≥ 1 btag
μe	1.12 ± 0.34	$- - -$	1.48 ± 0.44

Table 6.13: Data to MC scale factors for Drell-Yan events in the three channels using the template fit (top) and $R_{out/in}$ (bottom) methods.

These *loose* set of lepton quality cuts, showed in Table 6.14 for muons and in Table 6.15 for electrons, defines the *"fakeable object"*. The ratio of "fake" leptons passing the quality requirements used in the analysis (tight lepton selection) over candidate leptons passing the loose criteria is called the fake ratio (f). This ratio is determined in QCD events.
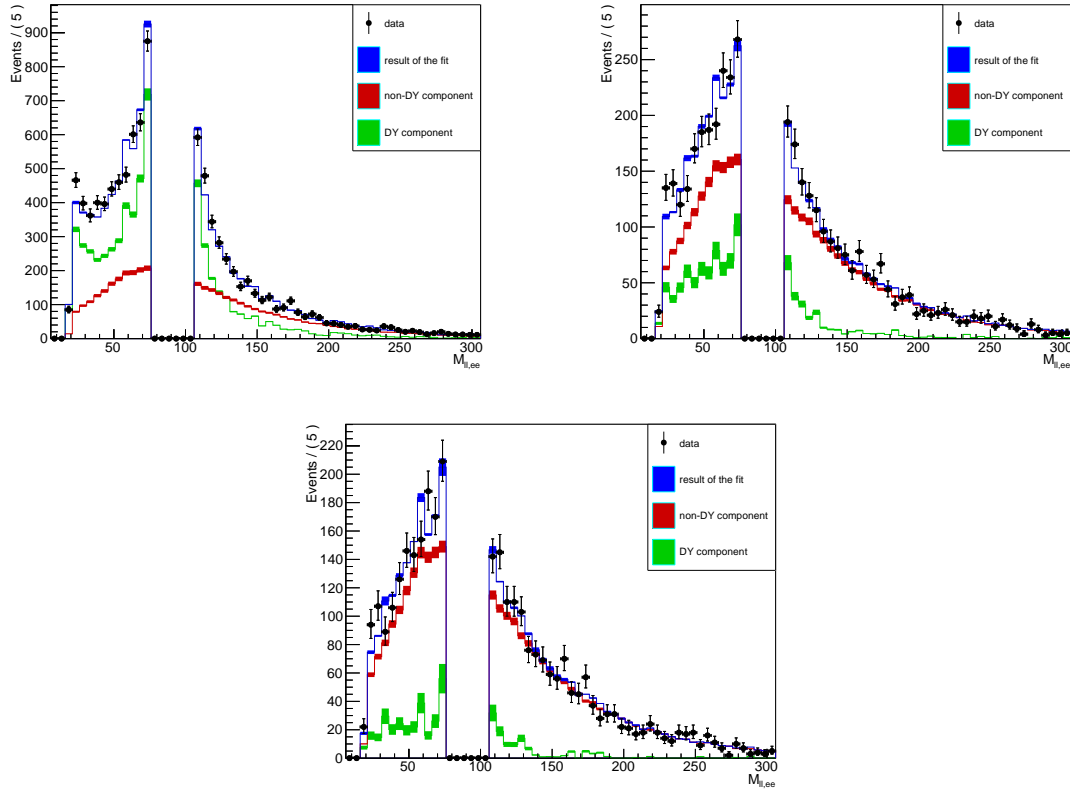
Figure 6.18: Results of the dilepton invariant mass fit after the jet multiplicity (left), $\not{E}_T$ (right) and b-tagging (bottom) selections in the e^+e^- channel. Black dots correspond to data, blue band corresponds to the output of the fit, green band to the DY component after the fit, and red band to the contributions from other processes after the fit. The width of the bands corresponds to the statistical uncertainty.

In contrast, prompt leptons are the leptons characterizing the signal, or coming from background events with isolation and identification properties indistinguishable from the ones of signal leptons. The prompt rate (p) of tight to loose leptons is measured in Z+jets events by a "tag and probe" method.

The method to estimate the Non-W/Z background, which is described in [87] and in [88], can be divided in two steps:

1. The fake and prompt rates are measured on data, in a phase space region enriched with QCD dijet and $Z/\gamma^* \rightarrow \ell\ell$ events, respectively. These rates are parametrized as a function of the p_T and η of the fakeable object.
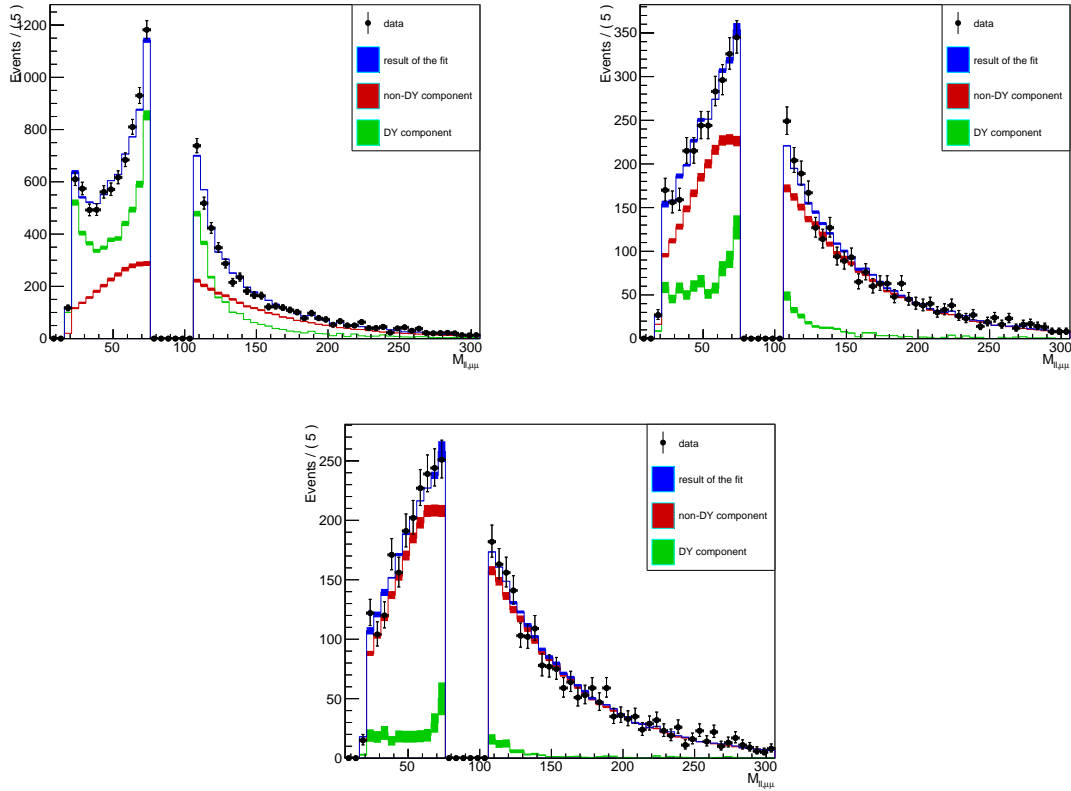
Figure 6.19: Results of the dilepton invariant mass fit after the jet multiplicity (left), $\not{E}_T$ (right), and b-tagging (bottom) selections in the $\mu^+\mu^-$ channel. Black dots correspond to data, blue band corresponds to the output of the fit, green band to the DY component after the fit, and red band to the contributions from other processes after the fit. The width of the bands corresponds to the statistical uncertainty.

2. Data events are required to pass the signal selection cuts with loose lepton requirements. The Non-W/Z event yields are extracted by applying fake and prompt rates to this set of "loose-loose" dilepton events.

Measurement of the rates of prompt and fake leptons

As described in the previous section, the muon and electron prompt rates are obtained with a "tag and probe" technique applied to data. Results are shown in Table 6.16.

Muon and electron fake rates are extracted from data samples dominated by QCD dijet events, selected using the single lepton trigger paths listed in Table 6.17. The cuts defining this control region aim to reduce the contribution from W or Z leptonic

Muon quality cuts		
type	PF muon	
p_T	$p_T > 20\,\text{GeV}$	
η	$\lvert \eta \rvert < 2.4$	
ID	IsGlobalMuon OR IsTrackerMuonArbitrated	
	loose μ	tight μ
isolation	-	PF $I_R^{\mu} < 0.15$ ($\Delta R = 0.3, \Delta\beta$ corr.)

Table 6.14: The set of muon quality cuts used for the tight to loose method.

decays. Events with W decays are rejected by requiring $\not{E}_T < 20\,\text{GeV}$ and, only for the measurement of the muon fake rate, that the W candidate transverse mass is lower than $15\,\text{GeV}$. Events with Z decays are discarded with a Z mass veto: $m_{\mu\mu} \notin [76, 106]\,\text{GeV}$, $m_{ee} \notin [60, 120]\,\text{GeV}$. Events with low-mass resonances are removed by an additional $m_{\ell\ell} > 20\,\text{GeV}$ cut.

The residual bias introduced by leptons with high p_T from W+jets and Z+jets events is suppressed by assuming that the lepton fake rate flattens out above a p_T of $35\,\text{GeV}$.

The energy spectrum of the jets misidentified as leptons can be different from the one of real jets. The relative isolation of a loose lepton is sensitive to the difference in jet energy. In order to properly define the dijet control sample it is therefore useful to compare the relative isolation distribution of loose leptons in that phase-space region and in a sample dominated by W+jets events. This is done by requiring one jet in the dijet sample (the "near-side jet") to be close to a reconstructed lepton, whose relative isolation is studied as a function of the p_T of a jet on the opposite side of the lepton (the "away-side jet") as shown in Figure 6.20. The energies of the two jets are correlated, but the jet p_T requirement is applied on the away-side jet instead of the near-side jet to avoid biases in the isolation distribution.

The loose lepton isolation distributions obtained with different jet p_T requirements are compared with the one extracted from a control sample containing events with exactly two same-sign leptons, one passing the quality criteria used in the analysis, the other one the loose selection. The threshold on the jet p_T is then chosen as the one giving the best agreement.

Electron quality cuts		
type	PF electron	
p_T	$p_T > 20\,\text{GeV}$	
η	$|\eta| < 2.5$	
conversions	pass conversion veto	
hits	$N_{\text{exp.missinghits}} < 1$	
trigger cuts	$\sigma_{i\eta i\eta} < 0.01/0.03$ (barrel/endcaps) $|\Delta\phi_{in}| \quad < \quad 0.15/0.10$ $|\Delta\eta_{in}| \quad < \quad 0.007/0.009$ $\frac{H}{E} \quad < \quad 0.12/0.10$ $\frac{\sum_{trk} E_T}{p_T{}^e} \quad < \quad 0.2$ $\frac{[\sum_{ECAL} E_T]-1}{p_T{}^e} \quad < \quad 0.2$ $\frac{\sum_{HCAL} E_T}{p_T{}^e} \quad < \quad 0.2$	
	loose e	tight e
ID	MVA ID > -0.1	MVA ID > 0.5
e-μ distance	-	ΔR(e - any GlobalMuon) > 0.1
isolation	$I_R^e < 1.00$ $(\Delta R = 0.3, \rho \text{ corr.})$	$I_R^e < 0.15$ $(\Delta R = 0.3, \rho \text{ corr.})$
impact parameter	$IP_{w.r.t.\ PV} < 0.1$	$IP_{w.r.t.\ PV} < 0.04$

Table 6.15: The set of electron quality cuts used for the tight to loose method.

The results are shown in Figure 6.21 for muons and electrons. In the muon case the best jet p_T threshold is around $45\,\text{GeV}$, whereas in the electron case it is not possible to select any particular value. Consequently, a central value of $30\,\text{GeV}$ is chosen for the electron case. A systematic uncertainty is assigned from the dependence of the final results on this choice.

The results obtained for the electron and muon fake rates are listed in Table 6.18.

Extraction of the Non-W/Z background event yield

The signal event selection is applied to data, requiring that both leptons pass the loose lepton quality cuts summarized in Tables 6.14 and 6.15.
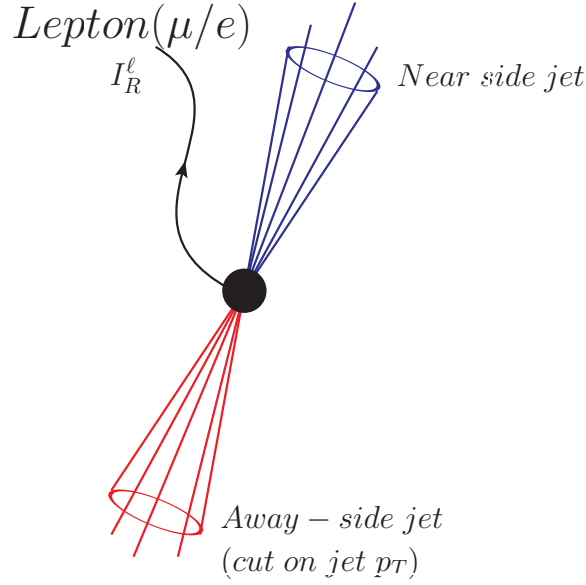
Figure 6.20: Misidentified lepton in a dijet sample. The relative lepton isolation (I_R^ℓ) is studied as a function of the p_T of the away-side jet.
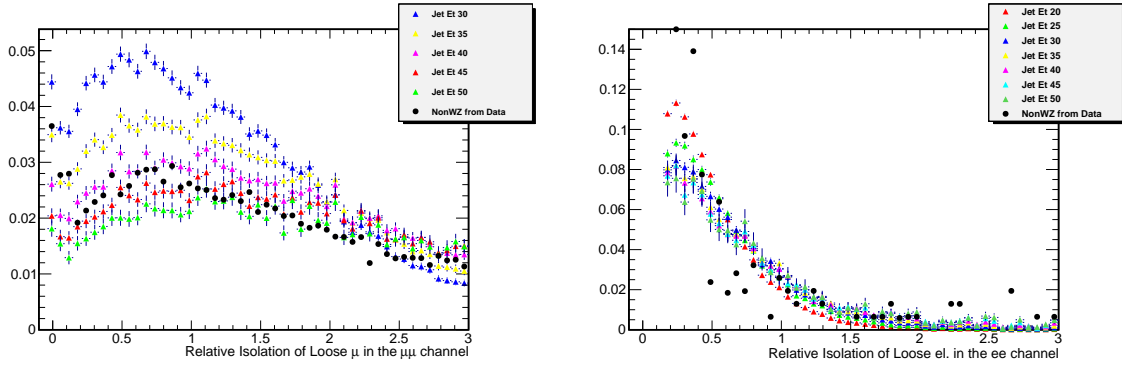
Figure 6.21: Comparison between the relative isolation distributions of the loose leptons in the dijet control sample selected with different jet p_T thresholds (coloured dots), and from a sample enriched in W+jets events (black dots) for muons in the $\mu\mu$ channel (left) and electrons in the ee channel (right).

Electron Prompt Rate			
p_T (GeV)	$0 < \eta \leq 1.4442$	$1.4442 < \eta \leq 1.556$	$1.556 < \eta \leq 2.5$
$10 < p_T \leq 15$	0.418 ± 0.007	0.62 ± 0.05	0.5 ± 0.3
$15 < p_T \leq 20$	0.608 ± 0.004	0.56 ± 0.04	0.589 ± 0.006
$20 < p_T \leq 25$	0.861 ± 0.003	0.78 ± 0.04	0.831 ± 0.003
$25 < p_T \leq 50$	0.9611 ± 0.0012	0.8511 ± 0.0013	0.9428 ± 0.0007
$50 < p_T$	0.9817 ± 0.0003	0.932 ± 0.004	$0.97276 \pm 3E\text{-}05$
Muon Prompt Rate			
p_T (GeV)	$0 < \eta \leq 1.5$		$1.5 < \eta \leq 2.4$
$10 < p_T \leq 15$	0.9946 ± 0.0019		0.9982 ± 0.0012
$15 < p_T \leq 20$	0.9972 ± 0.0007		0.9990 ± 0.0005
$20 < p_T \leq 25$	0.9995 ± 0.0004		$0.99991 \pm 7E\text{-}05$
$25 < p_T \leq 50$	$1 \pm 4E\text{-}07$		$1 \pm 1.2E\text{-}06$
$50 < p_T$	$1 \pm 5E\text{-}06$		$1 \pm 1.4E\text{-}05$

Table 6.16: Measured electron and muon prompt rates in bins of p_T and η of the fakeable object. The error shown corresponds only to the statistical component.

Electron triggers	HLT_Ele17_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL
Muon triggers	HLT_Mu17

Table 6.17: Single lepton trigger paths used for the measurement of the lepton fake rates.

Defining two opportune quantities ε and η as a function of the fake (f) and prompt (p) rates,

$$\varepsilon = \frac{f}{1 - f}$$
$$\eta = \frac{1 - p}{p}, \tag{6.13}$$

it is possible to assign a weight (w) to each event, according to the number of leptons passing or failing the tight requirements. The weights assigned to each event are:

electron fake rate				
p_T (GeV)	$0 < \eta \leq 1.0$	$1.0 < \eta \leq 1.479$	$1.479 < \eta \leq 2.0$	$2.0 < \eta \leq 2.5$
$20 < p_T \leq 25$	0.168 ± 0.005	0.225 ± 0.007	0.152 ± 0.005	0.076 ± 0.004
$25 < p_T \leq 30$	0.200 ± 0.007	0.243 ± 0.009	0.195 ± 0.008	0.092 ± 0.005
$30 < p_T \leq 35$	0.233 ± 0.009	0.250 ± 0.013	0.183 ± 0.010	0.096 ± 0.007
muon fake rate				
p_T (GeV)	$0 < \eta \leq 1.0$	$1.0 < \eta \leq 1.479$	$1.479 < \eta \leq 2.0$	$2.0 < \eta \leq 2.4$
$20 < p_T \leq 25$	0.029 ± 0.004	0.048 ± 0.009	0.112 ± 0.017	0.09 ± 0.04
$25 < p_T \leq 30$	0.037 ± 0.007	0.061 ± 0.015	0.08 ± 0.02	0.12 ± 0.06
$30 < p_T \leq 35$	0.025 ± 0.008	0.044 ± 0.019	0.08 ± 0.03	0.3 ± 0.2

Table 6.18: Measured electron and muon fake rates in bins of p_T and η of the fakeable object. The error shown corresponds only to the statistical component.

$$\text{Pass} - \text{Pass}: \quad w_{P_1 P_2} = -\frac{\varepsilon_1 \eta_1 + \varepsilon_2 \eta_2}{(1 - \varepsilon_1 \eta_1)(1 - \varepsilon_2 \eta_2)} \tag{6.14}$$

$$\text{Fail} - \text{Fail}: \quad w_{F_1 F_2} = -\frac{2\varepsilon_1 \varepsilon_2}{(1 - \varepsilon_1 \eta_1)(1 - \varepsilon_2 \eta_2)} \tag{6.15}$$

$$\text{Pass} - \text{Fail}: \quad w_{P_1 F_2} = \frac{(1 + \varepsilon_1 \eta_1)\varepsilon_2}{(1 - \varepsilon_1 \eta_1)(1 - \varepsilon_2 \eta_2)} \tag{6.16}$$

$$\text{Fail} - \text{Pass}: \quad w_{F_1 P_2} = \frac{(1 + \varepsilon_2 \eta_2)\varepsilon_1}{(1 - \varepsilon_1 \eta_1)(1 - \varepsilon_2 \eta_2)} \tag{6.17}$$

The total yield of Non-W/Z background events is given by the sum of all the event weights in each final state. The two terms in Equation 6.14 and 6.15 give a negative contribution to the total event yield.

Results are shown in Table 6.19 together with the statistical and systematical uncertainties. Systematic uncertainties have been calculated by varying both the muon and the electron fake rates in each bin up and down by their statistical uncertainty and by taking the largest difference between the resulting yields and the central value. Additionally, a variation of the threshold used in the "away-side jet" E_T cut by ± 5 GeV for muons and by ± 10 GeV for electrons is applied.

channel	central value	stat.	syst.	stat. $\oplus$ syst.
$\mu\mu$	114	$\pm32\%$	$\pm26\%$	$\pm41\%$
ee	25	$\pm5\%$	$\pm34\%$	$\pm34\%$
μe	185	$\pm28\%$	$\pm25\%$	$\pm38\%$

Table 6.19: Yields for the Non-W/Z background estimated from data. For each channel, the statistical and systematic uncertainties in % are also given.

6.5 Data and Monte Carlo: Plots and Yields

This section presents the most relevant data to MC distribution comparisons. Figures 6.22 and 6.23 show the p_T for the leading lepton and jet respectively, for all the three channels. Figure 6.24 shows the dilepton invariant mass after the selection on jet multiplicity; Figure 6.25 shows the $\not{E}_T$ distribution. The jet multiplicity and the b-tagged jet multiplicity are shown in Figures 6.26 and 6.27, respectively. Trigger and lepton scale factors discussed in Section 6.3.1 are applied to the simulated signal and backgrounds. In all the plots, the $t\bar{t}$ contribution is normalized to the cross section expected (252.8 pb, see Section 2.3) for a top quark mass of $m_t = 172.5\,\text{GeV}$. Drell-Yan data-driven background estimates from Section 6.4.1 are used at all levels of the selection. Non-W/Z contribution is taken from MC simulation.

The hatched bands on the distributions account for the statistical errors. Additionally, the b-tagging uncertainty is included on the b-tagged jet multiplicity distributions.

Table 6.20 shows the yield for data, signal and background events at each level of selection presented in Section 6.2.4, except the ones corresponding to the final cut level, which are presented in Section 7.2 with the $t\bar{t}$ cross section value obtained.
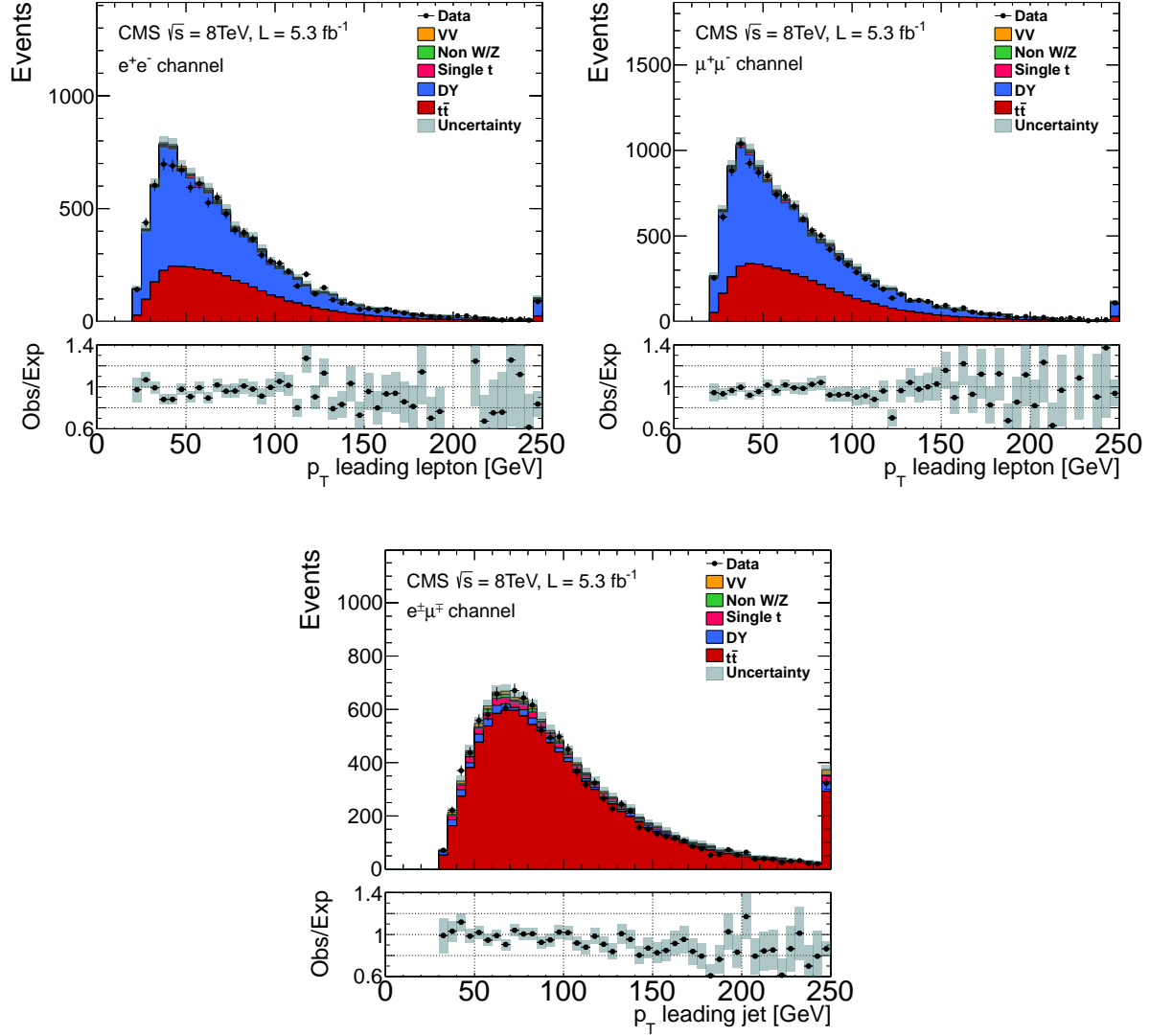
Figure 6.22: The p_T distribution for the leading lepton in the e^+e^- (top left), $\mu^+\mu^-$ (top right) and $\mu^{\pm}e^{\mp}$ (bottom) channels after the selection on jet multiplicity, and the corresponding data-to-simulation ratios. The expected distributions for $t\bar{t}$ signal and background sources are shown by histograms; data are shown by black dots. A $t\bar{t}$ cross section of 252.8 pb is used to normalize the simulated $t\bar{t}$ signal. The statistical uncertainties on the expected events are displayed by the hatched blue bands.
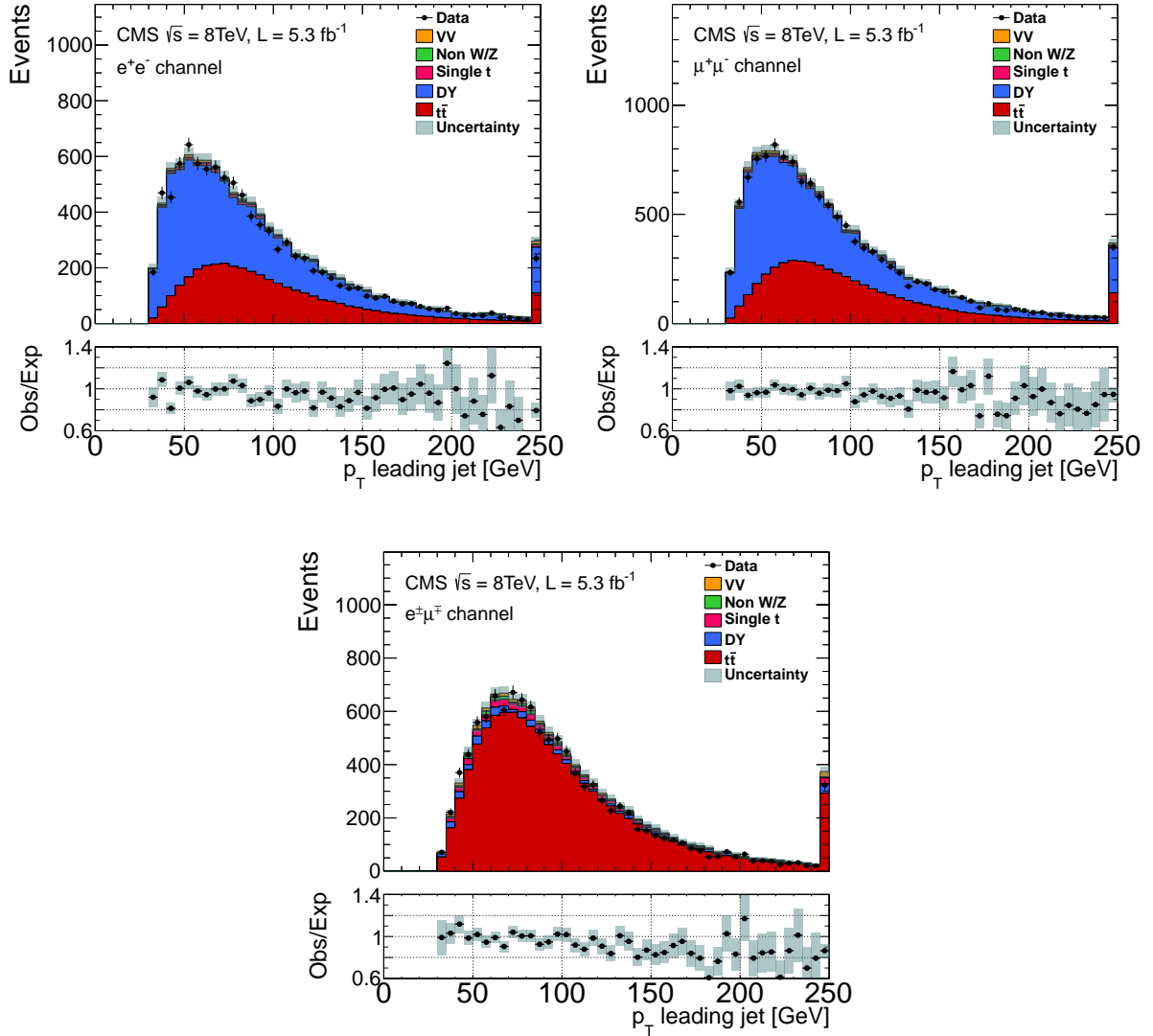
Figure 6.23: The p_T distribution for the leading jet in the e^+e^- (top left), $\mu^+\mu^-$ (top right) and $\mu^\pm e^\mp$ (bottom) channels after the selection on jet multiplicity, and the corresponding data-to-simulation ratios. Details on the distributions are the same as in Figure 6.22.
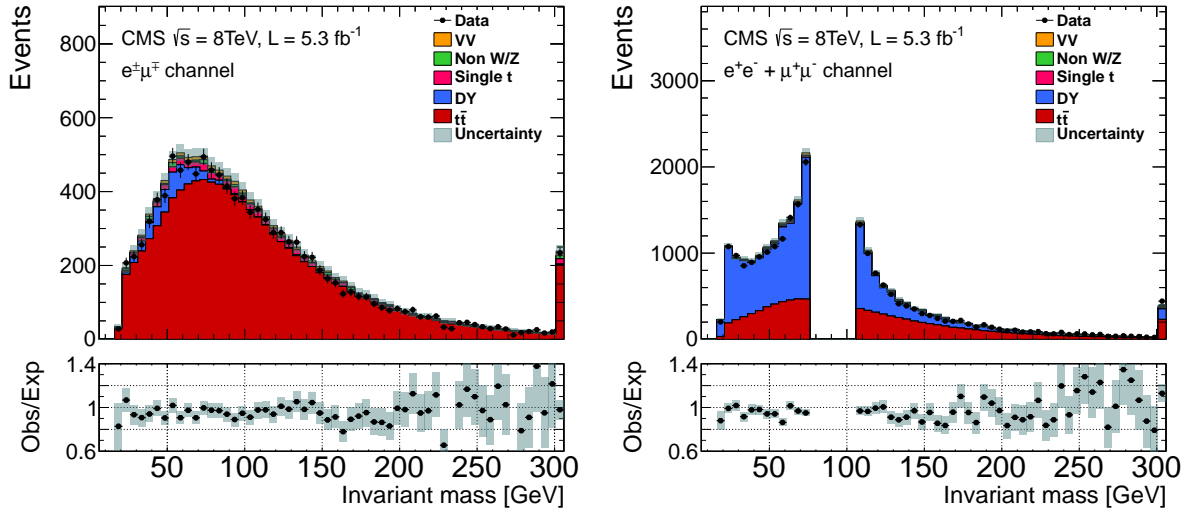
Figure 6.24: The dilepton invariant mass distribution after the selection on jet multiplicity in the $\mu^{\pm}e^{\mp}$ (left) and for the sum of e^+e^- and $\mu^+\mu^-$ (right) channels, and the corresponding data-to-simulation ratios. The expected distributions for $t\bar{t}$ signal and background sources are shown by histograms; data are shown by black dots. The gap in the sum of e^+e^- and $\mu^+\mu^-$ distributions reflects the requirement that removes dileptons from the Z mass window. A $t\bar{t}$ cross section of 252.8 pb is used to normalize the simulated $t\bar{t}$ signal. The statistical uncertainties on the expected events are displayed by the hatched bands.
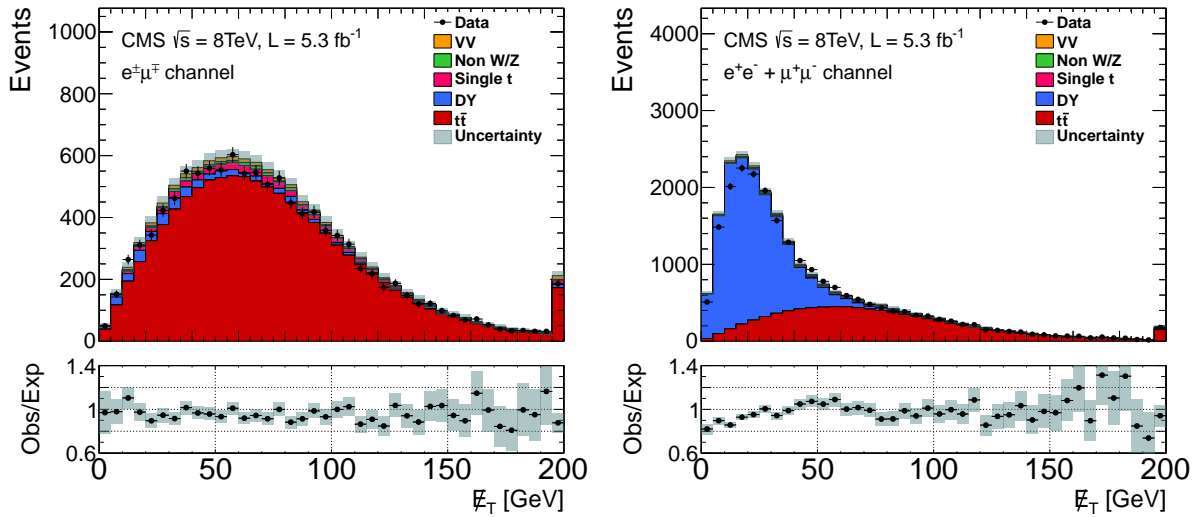
Figure 6.25: The $\not{E}_T$ distribution after the selection on jet multiplicity in the $\mu^{\pm}e^{\mp}$ (left) and for the sum of e^+e^- and $\mu^+\mu^-$ (right) channels, and the corresponding data-to-simulation ratios. Details on the distributions are the same as for Figure 6.24.
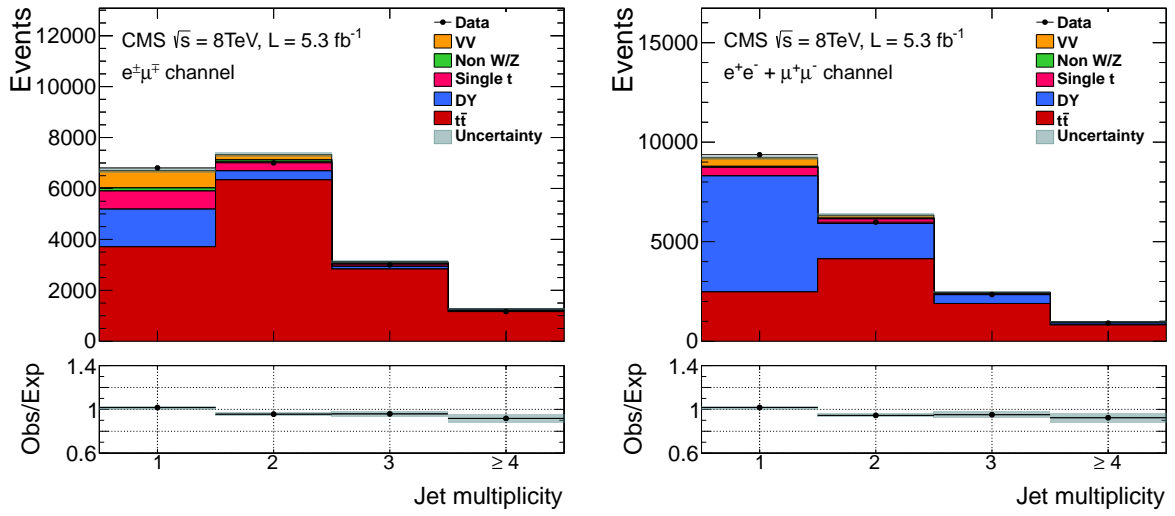
Figure 6.26: The jet multiplicity distribution after the $\not{E}_T$ selection but before jet multiplicity cut in the $\mu^\pm e^\mp$ (left) and in the sum of e^+e^- and $\mu^+\mu^-$ (right) channels, and the corresponding data-to-simulation ratios. The expected distributions for $t\bar{t}$ signal and background sources are shown by histograms; data are shown by black dots. A $t\bar{t}$ cross section of 252.8 pb is used to normalize the simulated $t\bar{t}$ signal. The statistical uncertainties on the expected events are displayed by the hatched bands.
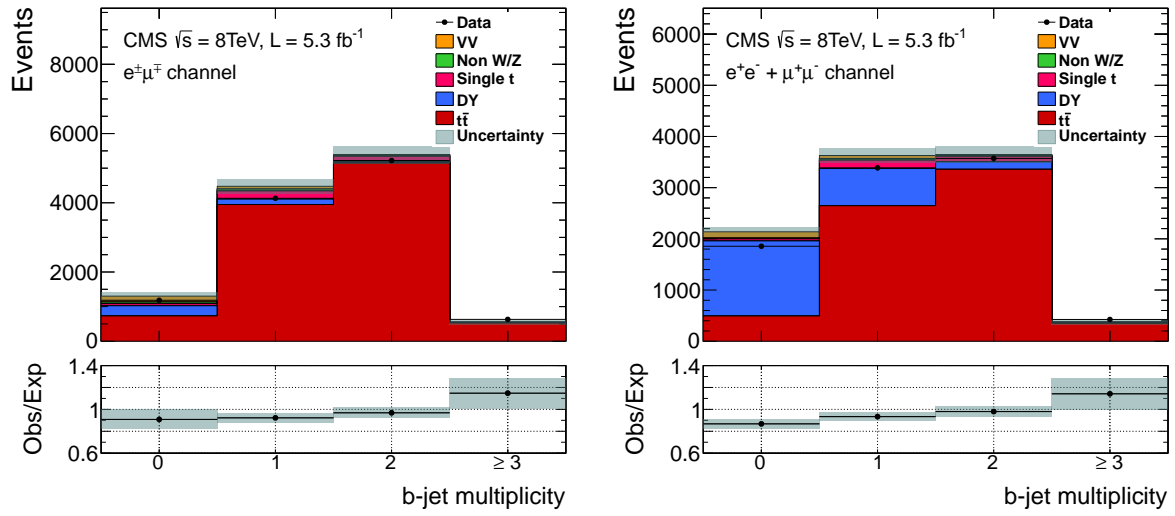
Figure 6.27: The b-jet multiplicity distribution after the $\not{E}_T$ cut in the $\mu^{\pm}e^{\mp}$ (left) and in the sum of e^+e^- and $\mu^+\mu^-$ (right) channels. The expected distributions for t$\bar{\text{t}}$ signal and background sources are shown by histograms; data are shown by black dots. A t$\bar{\text{t}}$ cross section of 252.8 pb is used to normalize the simulated t$\bar{\text{t}}$ signal. The hatched bands show the total statistical and b-jet systematic uncertainties in the event yields for the sum of the t$\bar{\text{t}}$ and background predictions.

Number of events with 2 leptons			
Source	e^+e^-	$\mu^+\mu^-$	$\mu^\pm e^\mp$
Drell-Yan	1703596±1000	2355331±1197	9530±75
Non-W/Z leptons	462±20	151±7	557±19
Single top quark	649±9	861±10	1392±13
VV	3754±6	5040±7	2805±6
Total background	1708462±1000	2361383±1197	14284±79
$t\bar{t}$ dilepton signal	6875±9	9171±10	14646±13
Total MC	1715337±1000	2370554±1197	28930±80
Data	1745543	2412381	29731
Number of events with 2 leptons + ≥ 2 jets			
Source	e^+e^-	$\mu^+\mu^-$	$\mu^\pm e^\mp$
Drell-Yan	6067±1820	7759±2328	488±146
Non-W/Z leptons	55±4	44±3	137±6
Single top quark	183±5	227±5	474±7
VV	150±1	197±1	264±2
Total background	6457±1820	8227±2328	1362±147
$t\bar{t}$ dilepton signal	3738±6	4988±8	10356±11
Total MC	10195±1820	13215±2328	11718±147
Data	9718	12767	11165
Number of events with 2 leptons + ≥ 2 jets + $\not{E}_T > 40\,\text{GeV}$			
Source	e^+e^-	$\mu^+\mu^-$	$\mu^\pm e^\mp$
Drell-Yan	1076±323	1267±380	488±146
Non-W/Z leptons	39±3	31±3	137±6
Single top quark	146±4	179±5	474±7
VV	82±1	108±1	264±2
Total background	1344±323	1586±380	1362±147
$t\bar{t}$ dilepton signal	2940±6	3911±7	10356±11
Total MC	4284±323	5497±380	11718±147
Data	4008	5231	11165

Table 6.20: Number of e^+e^-, $\mu^+\mu^-$ and $\mu^\pm e^\mp$ events after applying the different event selection cuts. The results are given for the individual sources of background, $t\bar{t}$ signal with a top-quark mass of 172.5 GeV and $\sigma_{t\bar{t}} = 252.8\,\text{pb}$, and data. The uncertainties correspond to the statistical component.

CHAPTER 7

CROSS SECTION ESTIMATION

This chapter presents the final cross section measurement obtained with $5.3\,\mathrm{fb}^{-1}$ of $8\,\mathrm{TeV}$ data taken by the CMS detector between April and June 2012. A detailed description of the different sources of systematic errors and the methods by which they are estimated is given in the first part of the chapter. The measurement has been performed in the dilepton channel. The results are presented for the three decay channels studied ($\mu^+\mu^-$, e^+e^- and $\mu^\pm e^\mp$) as well as for their combination obtained using the BLUE method[89].

7.1 Systematic Uncertainties

The systematic uncertainties are estimated from the relative change in the number of selected MC events after the full selection. The systematic uncertainties considered can be divided in the following categories:

1. Uncertainties on the detector performance.

2. Effect of extra proton-proton collisions (pileup) in the same bunch crossing.

3. Knowledge of the $t\bar{t}$ signal modeling.

4. Estimation on the yields from background processes in the signal region.

5. Absolute normalization of the sample integrated luminosity.

The uncertainties on detector performance, pile up and signal production are computed for the $t\bar{t}$ signal and the simulated Monte Carlo (MC) backgrounds (single top and VV) and are discussed in Section 7.1.1, 7.1.2 and 7.1.3. The uncertainties on background estimations are those related to the data-driven background determination and are presented in Section 6.4.1 and 6.4.5. The uncertainty on the luminosity, taken as 2.6% [53], is directly propagated to the cross-section measurement (Section 7.2).

7.1.1   Detector Performance

Trigger uncertainty

The trigger efficiencies and their scale factors are treated in Section 6.3.1 and summarized in Section 6.3.3. The uncertainties are propagated to the number of MC events and then to the final cross section value. The trigger uncertainties have been treated as 100% correlated among channels.

Identification (ID) and isolation (ISO) uncertainties

The identification and isolation efficiencies and their scale factors are treated in Section 6.3.2 and summarized in Section 6.3.3. As done on the trigger, the ID and ISO uncertainties are directly propagated to the number of MC events and then to the cross section measurement.

In the specific case of the muon-electron channel, the variations were performed separetly for each lepton, then added in quadrature. The size of the uncertainty is found to be smaller than in the same flavor channels due to the uncorrelation between the uncertainties on the scale factors of muons and electrons. The SFs for the dielectron and dimuon channels are not correlated. The uncertainty on the SFs for the electron-muon channel is correlated to the ones for the dielectron and the dimuon channels. The correlation coefficients, ρ, can be established from the approximate relation:

$$SF^{e\mu}_{Tr,ID,ISO} \approx \sqrt{SF^{ee}_{Tr,ID,ISO} \cdot SF^{\mu\mu}_{Tr,ID,ISO}}, \tag{7.1}$$

as:

$$\rho_{ee/\mu\mu,\mu e} \approx \frac{1}{2} \frac{\delta SF^{ee/\mu\mu}_{Tr,ID,ISO}}{SF^{\mu e}_{Tr,ID,ISO}} \tag{7.2}$$

obtaining that the correlation coefficient of the electron-muon scale factor is 0.64 relative to the dielectron, and 0.55 to the dimuon scale factors.

The rest of the systematic uncertainties discussed in this section are treated 100% correlated among channels.

Lepton energy scale (LES)

The uncertainty from the lepton energy scale has been estimated scaling the muon momentum up and down by 0.2% [90]; a 0.5% is applied for electrons in barrel, and a 1% for electrons in endcap [91]. These changes were propagated to the missing transverse energy adding (subtracting) the correspounding p_T variation.

Jet energy corrections

The jet energy corrections applied correspond to the official prescription adopted in CMS [92]. The uncertainties on the jet energy scale (JES) and jet energy resolution (JER) affect the efficiency of jet selection. The impact of the uncertainty on JES is estimated from the change observed in the number of MC events passing the full selection varying the jet momentum within the JES uncertainties [93]. Similarly, the effect of the JER uncertainty is estimated by changing the jet momentum p_T according to the following formula [94]:

$$p_T^{var} = \max\left[0., p_T^{gen} + c(\eta) \left(p_T - p_T^{gen}\right)\right] \tag{7.3}$$

Where p_T^{var} is the jet momentum after the variation for resolution effects, p_T^{gen} is the jet momentum at generetor level and $c(\eta)$ is the core resolution scaling factor, taken from [94]. Both, JES and JER uncertaities are propagated to the calculation of the missing transverse energy.

b-tagging

The systematic uncertainties associated to the b-jet identification are estimated following the recomendation in [74, 85]. Variations on the scale factors used to correct the MC efficiencies to tag jets from b, c or light partons are applied when estimating both signal and background yields. The scale factors for b-jets are given as a function of jet p_T, from muon-jet data. The scale factors for c-jets, SF_c, are taken to be equal to SF_b with twice the quoted uncertainty, while the scale factors for light jets, SF_l, are provided as a function of jet p_T for several η bins. The average values of the scale factors over the p_T spectrum of the jets from top quark decays are:

$$SF_b = 0.984 \pm 0.016$$
$$SF_c = 0.984 \pm 0.032 \tag{7.4}$$
$$SF_l = 1.080 \pm 0.090$$

7.1.2   Pileup

As discussed in Section 6.2.1, we reweight MC events in order to reproduce the number of additional proton-proton interactions observed in data. To compute the uncertainty on the pileup (PU) reweighting, we recompute the PU distribution expected in data by varying the proton pronton inelastic cross section within its theoretical uncertainty ($\pm 5\%$). This variation covers the uncertainties due to the pileup modeling in MC. The relative change observed in the number of MC events observed after the full selection is quoted as a systematic uncertainty.

7.1.3   t̄t Modeling

The systematic uncertainties on the MC modeling of t̄t production are estimated by using dedicated MADGRAPH samples with different parameter settings (Table 6.2):

1. Factorization and renormalization scales (μ_F and μ_R).

2. Matching of partons from the matrix element with those from parton showers.

The renormalization and factorization scale uncertainty is obtained by varying simultaneously up and down by a factor of two both parameters.

The matching between the matrix elements (ME) and the parton shower (PS) evolution is done by applying the MLM prescription [95]. The systematic uncertainty is estimated changing the thresholds that control the matching of partons from the matrix element with those from PS by factors of 0.5 and 2.0 for one of the parameters (minimum kT measure between partons) and 0.75 and 1.5 for the other (jet matching threshold for the kT -MLM scheme). The relative change observed in the signal acceptance is taken as the systematic uncertainty.

Uncertainties on the t$\bar{\text{t}}$ event description are also evaluated as the relative change observed in the number of MC events when using different generators, MADGRAPH and POWHEG , and different hadronization interface, PYTHIA and HERWIG . The uncertainty from the generator model is estimated as the difference in the signal yields obtained with MADGRAPH and POWHEG samples interfaced with PYTHIA. The effect on the calculated t$\bar{\text{t}}$ cross section is 2.1%. The uncertainty arising from the hadronization model was obtained comparing the number of events having at least two jets in a sample of POWHEG interfaced with PYTHIA with one of POWHEG interfaced with HERWIG. The effect on the calculated t$\bar{\text{t}}$ cross section was 1.4%.

The uncertainties associated to the different generator and hadronization models are not propagated to the cross section since their contributions are already taken into account in other sources of systematic uncertainties:

- The factorization and renormalization scale uncertainty (2.3%) covers the differences due to the generators (2.1%).

- The uncertainty on the JES (2.2%) already contains a contribution from the uncertainty on the hadronization (1.4%).

Single top and diboson events

The uncertainties on the single top (tW) and diboson (VV) backgrounds arise from the same sources affecting the t$\bar{\text{t}}$ signal (as discussed at the beginning of this section), except the errors associated to the theoretical uncertainties (scale and matching). In addition, a 20% uncertainty on the production cross sections is applied on these backgrounds. This uncertainty is conservative with respect to the uncertainties on the inclusive production rate, and is expected to cover the uncertainties on the rate of these backgrounds in the phase space of the event selection used in the analysis.

Finally, Table 7.1 summarizes the systematic uncertainties on the $t\bar{t}$ signal correspond-
ing to the sources discussed above. Table 7.2 shows the corresponding systematic
uncertainties on the single top and VV background estimates.

Source	$t\bar{t}$ signal [%]		
	$\mu^+\mu^-$	e^+e^-	$\mu^\pm e^\mp$
Trigger efficiencies	1.3	1.7	1.5
Lepton efficiencies	2.4	2.4	1.7
Lepton energy scale	0.1	0.3	0.1
Jet energy scale	4.6	4.2	2.2
Jet energy resolution	1.7	1.3	1.3
b-tagging	0.8	0.8	0.7
Pileup	0.7	0.7	0.8
Scale (μ_F and μ_R)	2.3	2.3	2.3
Matching partons to showers	1.6	1.6	1.6
Luminosity	2.6	2.6	2.6

Table 7.1: Summary of the relative systematic uncertainties on the expected number of signal
$t\bar{t}$ events passing the full selection criteria, shown separately for each of the decay channels.
The uncertainties are given in percentage.

Source	tW [%]			VV [%]		
	$\mu^+\mu^-$	e^+e^-	$\mu^\pm e^\mp$	$\mu^+\mu^-$	e^+e^-	$\mu^\pm e^\mp$
Trigger efficiencies	1.3	1.7	1.5	1.3	1.7	1.5
Lepton efficiencies	2.4	2.4	1.7	2.4	2.4	1.7
Lepton energy scale	0.1	0.8	0.4	0.5	1.0	1.0
Jet energy scale	5.4	6.0	3.9	14.3	13.2	6.0
Jet energy resolution	1.6	2.1	1.7	1.6	1.7	1.6
b-tagging	0.9	1.3	1.2	4.9	4.6	4.7
Pileup	1.1	1.2	0.7	0.8	1.3	2.1
Cross-section used to normalize	20.0	20.0	20.0	20.0	20.0	20.0

Table 7.2: Summary of the relative systematic uncertainties on the expected number of single-
top (tW) and diboson (VV) events passing the full selection criteria, shown separately for each
of the decay channels. The uncertainties are given in percentage.

7.2 Estimation of the Top Pair Production Cross Section

The $t\bar{t}$ production cross section $\sigma_{t\bar{t}}$ is extracted from the following expression:

$$\sigma(\text{pp} \rightarrow t\bar{t}) = \frac{N - N_{Bkg}}{\underbrace{(\mathcal{A} \times \mathcal{E} \times \text{Br})}_{\mathcal{A}_{\text{total}}} \int \mathcal{L} \, \mathrm{d}t} \tag{7.5}$$

where N is the total number of events observed in data, N_{Bkg} is the number of estimated background events, $\mathcal{A}_{\text{total}}$ is the productof the mean acceptance ($\mathcal{A}$) by the selection efficiency ($\mathcal{E}$) and by the branching fraction (Br($t\bar{t} \rightarrow \ell\ell\nu\nu bb$)$\sim$ 5%) for $t\bar{t}$ events, and $\mathcal{L}$ is the luminosity. The measurement is performed for events having at least one b-tagged jet.

The number of data and background events are given in Table 7.3. The total acceptance ($\mathcal{A}_{\text{total}}$) and the cross section measured assuming a top quark mass of 172.5 GeV is shown in Table 7.4. Both tables detail the numbers for the three decay channels. The systematic uncertainties discussed in Section 7.1 are included in the event counts, and then propagated in the cross section measurement. Table 7.5 shows the breakdown of the systematic uncertainties in the different channels.

Figure 7.1 shows the theoretical predictions and experimental measurements of the top anti-top cross section for different center-of-mass energies. The numbers correspond to a top quark mass of $m_t = 172.5$ GeV. The value obtained in this thesis is shown by the not filled red circle. It agrees with the expected value at $\sqrt{s} = 8$ TeV.

7.2.1 Data to Monte Carlo Comparison

In this section, we compare several kinematic distributions in data and MC using the measured $t\bar{t}$ cross section. In all the following plots, the $t\bar{t}$ contribution is normalized to the $t\bar{t}$ cross section summarized in Table 7.4. The various data-to-MC scale factors are applied to the simulated signal and background events. Figure 7.2 and Figure 7.3 show the dilepton invariant mass and the $\not{E}_T$ distributions in events with at least two jets, respectively. The jet multiplicity and the b-tagged jet multiplicity are shown in Figure 7.4 and 7.5.

Source	Number of events		
	e^+e^-	$\mu^+\mu^-$	$\mu^\pm e^\mp$
Drell-Yan	386±116	492±148	194±58
Non-W/Z leptons	25±10	114±46	185±72
Single top quark	127±28	157±34	413±88
VV	30±8	39±10	94±21
Total background	569±120	802±159	886±130
$t\bar{t}$ dilepton signal	2728±182	3630±250	9624±504
Total MC	3297±218	4432±296	10510±520
Data	3204	4180	9982

Table 7.3: Observed and expected events passing the event selection after requiring at least one b-tagged jet. The results are given for the individual sources of background, $t\bar{t}$ signal with a top quark mass of 172.5 GeV and $\sigma_{t\bar{t}} = 252.9$ pb, and data. The uncertainties include the statistical and systematic components added in quadrature.

Channel	$\mathcal{A}_{\text{total}}[\%]$	$\sigma_{t\bar{t}} \pm$ stat. $\pm$ syst. $\pm$ Lumi.$[\text{pb}]$
e^+e^-	0.203 ± 0.012	244.3 ± 5.2 ± 18.6 ± 6.4
$\mu^+\mu^-$	0.270 ± 0.017	235.3 ± 4.5 ± 18.6 ± 6.1
$\mu^\pm e^\mp$	0.717 ± 0.033	239.0 ± 2.6 ± 11.4 ± 6.2

Table 7.4: The total acceptance $\mathcal{A}_{\text{total}}$, i.e. the product of event acceptance, selection efficiency and branching fraction for the respective $t\bar{t}$ final states, as estimated from simulation for a top-quark mass of 172.5 GeV, and the measured $t\bar{t}$ production cross sections, where the uncertainties are from statistical, systematic and integrated luminosity components, respectively.

7.3 Combination of $t\bar{t}$ Cross Sections Measured in the $\mu^+\mu^-$, e^+e^- and $\mu^\pm e^\mp$ Channels

A combined measurement of the $t\bar{t}$ cross section on the three decay channels is obtained using the *Best Linear Unbiassed Estimator* (BLUE method, see Appendix B) [89]. The value obtained is:

$$\sigma_{t\bar{t}} = 239 \pm 2\,(\text{stat.}) \pm 11\,(\text{syst.}) \pm 6\,(\text{Lumi.})\,\text{pb} \qquad (7.6)$$

in agreement with the prediction of the standard model, $\sigma_{t\bar{t}}^{Theory} = 252.8 \pm 8.6 \pm 6.6$ pb,

Source	e^+e^-	$\mu^+\mu^-$	$\mu^\pm e^\mp$
	[pb]		
Trigger efficiencies	4.1	3.0	3.6
Lepton efficiencies	5.8	5.6	4.0
Lepton energy scale	0.6	0.3	0.2
Jet energy scale	10.3	10.8	5.2
Jet energy resolution	3.2	4.0	3.0
b-jet tagging	1.9	1.9	1.7
Pileup	1.7	1.5	2.0
Scale (μ_F and μ_R)	5.7	5.5	5.6
Matching partons to showers	3.9	3.8	3.8
Single top quark	2.6	2.4	2.3
VV	0.7	0.7	0.5
Drell-Yan	10.8	10.3	1.5
Non-W/Z leptons	0.9	3.2	1.9
Total systematic	18.6	18.6	11.4
Integrated luminosity	6.4	6.1	6.2
Statistical	5.2	4.5	2.6

Table 7.5: Summary of the individual contributions to the systematic uncertainty on the $\sigma_{t\bar{t}}$ measurement. The uncertainties are given in pb. The statistical uncertainty on the result is given for comparison.

for a top-quark mass of 172.5 GeV.

7.4 Mass Dependency of the Measured tt̄ Cross Section

The measured tt̄ production cross section depends on the top quark mass (m_t) used in the simulated signal sample. The dependency has been studied in signal tt̄ samples simulated using m_t values in the range $160 - 185$ GeV. Figure 7.6 shows the result together with a fit to a second order polynomial:

$$\frac{\sigma_{t\bar{t}}(m_t)}{\sigma_{t\bar{t}}(m_t^o)} = 1.00 - 0.009 \times (m_t - m_t^o) - 0.000168 \times (m_t - m_t^o)^2 \qquad (7.7)$$

Where $m_t^o = 172.5$ GeV is the value used in the cross section estimate. Assuming
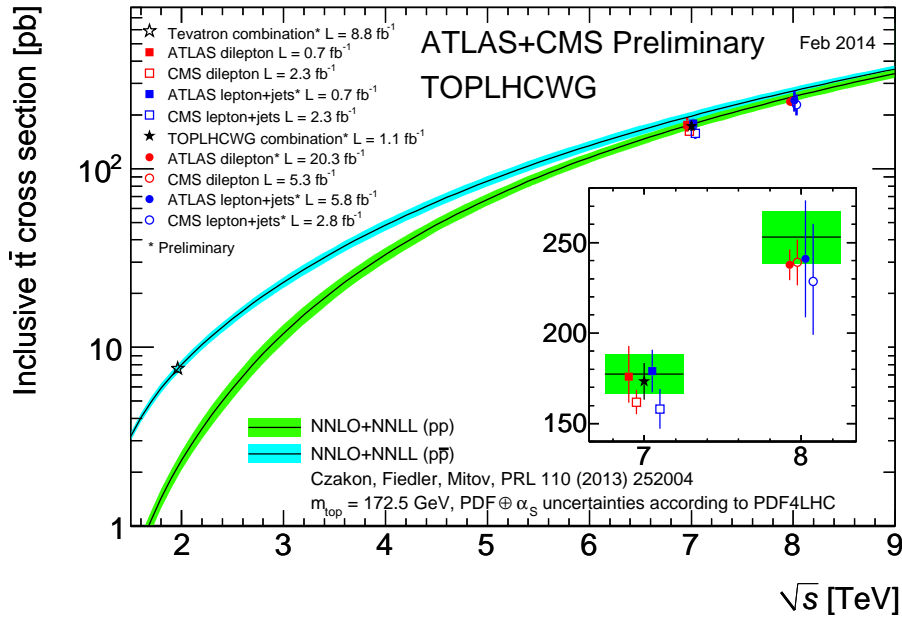
Figure 7.1: Theoretical predictions and experimental mesurements of the t̄t production cross section as a function of the collider center-of-mass energy ($\sqrt{s}$). The available measurements from the Tevatron collider, and from the ATLAS and CMS detectors at 1.96 TeV, 7 TeV and 8 TeV agree with the theoretical predictions.

the last measurement of the top quark mass, 173.2 GeV [3], the cross section value obtained is:

$$\sigma_{t\bar{t}}(m_t = 173.2\,\text{GeV}) = 237.5 \pm 13.1\,\text{pb} \tag{7.8}$$

This value of the t̄t cross section is also in agreement with the theoretical expectation $\sigma_{t\bar{t}}^{Theory} = 245.9 \pm 8.4 \pm 6.4\,\text{pb}$.

7.5   Additional Results

7.5.1   t̄t Cross Section in the $\mu^{\pm}e^{\mp}$ Channel After "At Least Two Jets" Cut

Due to the small background contamination present in the muon-electron channel, it is possible to extract the cross section in events with at least two jets, without requiring
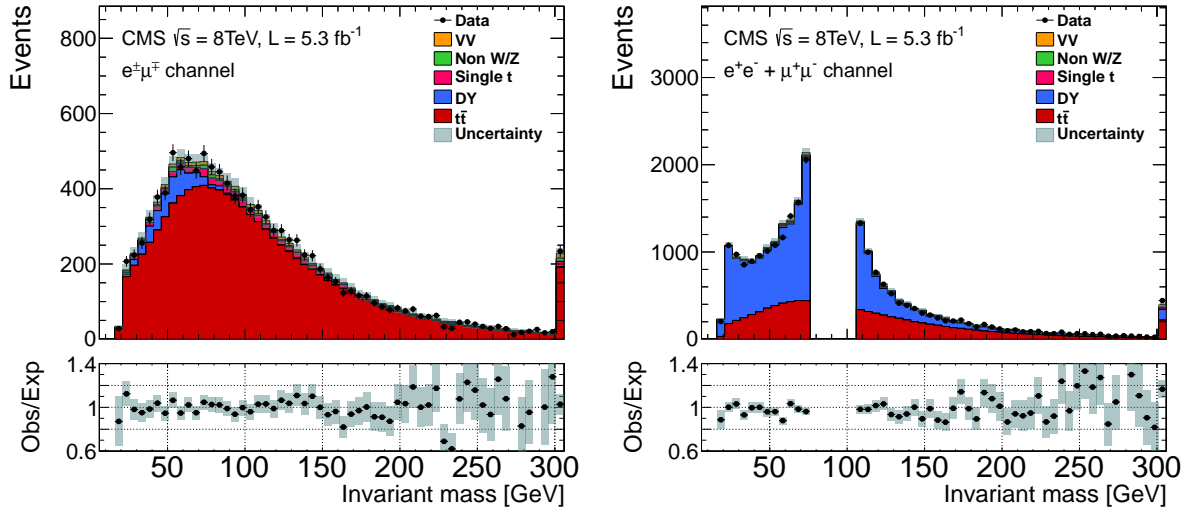
Figure 7.2: Dilepton invariant mass distribution in events with at least two jets with $p_T >$ 30 GeV for the $\mu^{\pm}e^{\mp}$ (left) and for the sum of e^+e^- and $\mu^+\mu^-$ (right) channels, and the corresponding data to simulation ratios. The expected distributions for $t\bar{t}$ signal and background sources are shown by stacked histograms; data are shown by dots. The $t\bar{t}$ cross sections on Table 7.4 are used to normalize the simulated $t\bar{t}$ signal. The systematic uncertainties on the expected events are displayed by the hatched bands.
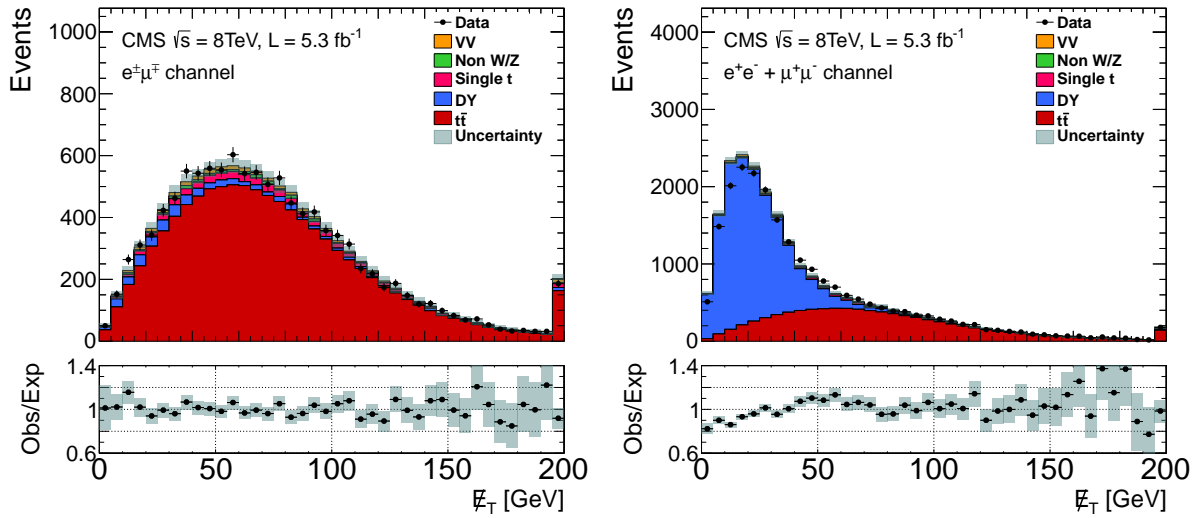
Figure 7.3: Missing transverse energy ($\not{E}_T$) distributions after the selection on at least two jets. Details on the distributions are the same as for Fig. 7.2.
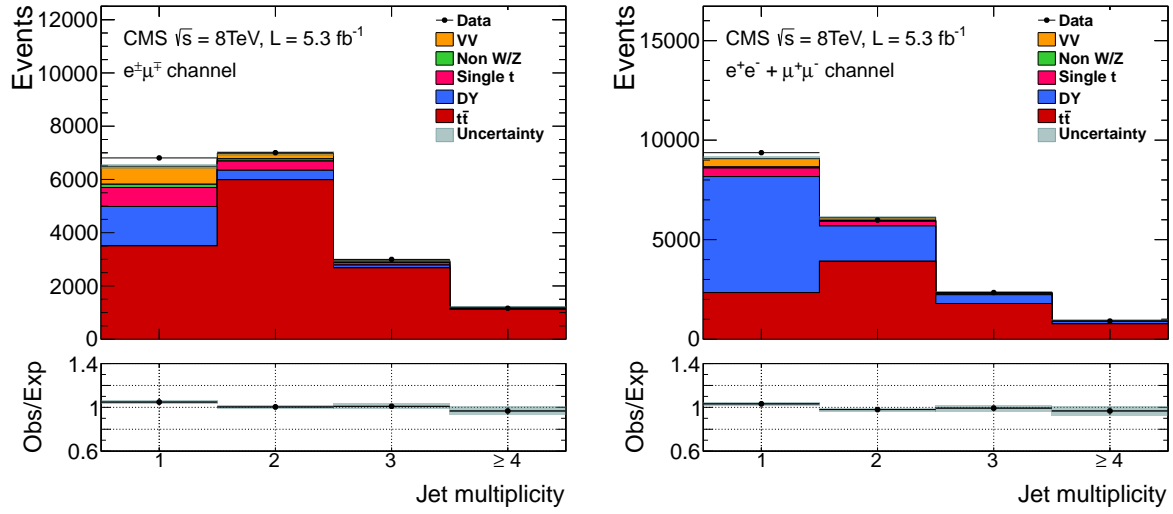
Figure 7.4: Jet multiplicity distribution after the $\not{E}_T$ selection for the $\mu^\pm e^\mp$ (left) and for the sum of e^+e^- and $\mu^+\mu^-$ (right) channels, and the corresponding data to simulation ratios. The expected distributions for t$\bar{\text{t}}$ signal and background sources are shown by stacked histograms; data are shown by black dots. The t$\bar{\text{t}}$ cross sections on Table 7.4 are used to normalize the simulated t$\bar{\text{t}}$ signal. The systematic uncertainties on the expected events are displayed by the hatched bands.

b-tagged jets. As shown in Figure 7.5 and Table 6.20 the background contribution is only about $\sim 12\%$ in the μe channel after this cut.

In this exercise, the Non-W/Z background contribution has been derived from MC while the Drell-Yan background has been estimated using the method described in Section 6.4.1. The t$\bar{\text{t}}$ cross section value obtained is:

$$\sigma_{\text{t}\bar{\text{t}}}^{\mu\text{e}} = 239.4 \pm 2.6(\text{stat.}) \pm 13.1(\text{syst.}) \pm 6.2(\text{Lumi.}) \, \text{pb} \qquad (7.9)$$

This value agrees with the result presented in Table 7.4. The systematic uncertainty is similar to the uncertainty on the measurement on events with at least a b-tagged jet. Even if the b-tag uncertainty is avoided, an increase of the uncertainty associated to the Drell-Yan data-driven estimation is observed.
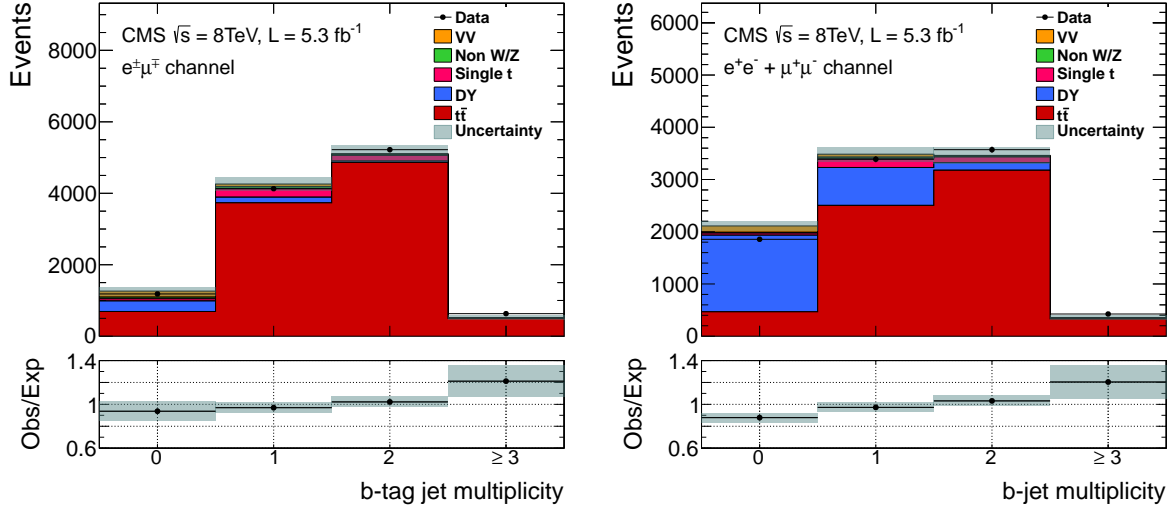
Figure 7.5: b-jet multiplicity distributions after the $\not{E}_T$ cut for the $\mu^{\pm}e^{\mp}$ (left) and for the sum of e^+e^- and $\mu^+\mu^-$ (right) channels. The expected distributions for $t\bar{t}$ signal and background sources are shown by stacked histograms; data are shown by black dots. The $t\bar{t}$ cross sections on Table 7.4 are used to normalize the simulated $t\bar{t}$ signal. The hatched bands show the total statistical and b-jet systematic uncertainties in the event yields for the sum of the $t\bar{t}$ signal and background predictions.

7.5.2 Cross Section in the μe Channel for Different $t\bar{t}$ Samples

In this analysis, the $t\bar{t}$ signal sample is used to estimate the total acceptance $\mathcal{A}_{\text{total}}$. As discussed in Section 6.1, several $t\bar{t}$ signal samples are used to study the effect of different generators and hadronization models in the cross section estimation. Table 7.6 shows the acceptances and the $t\bar{t}$ cross section for these different $t\bar{t}$ samples in the most precise channel. Considering that the obtained relative systematic uncertainty is approximately the same for all the samples, the final results are compatible within the errors.
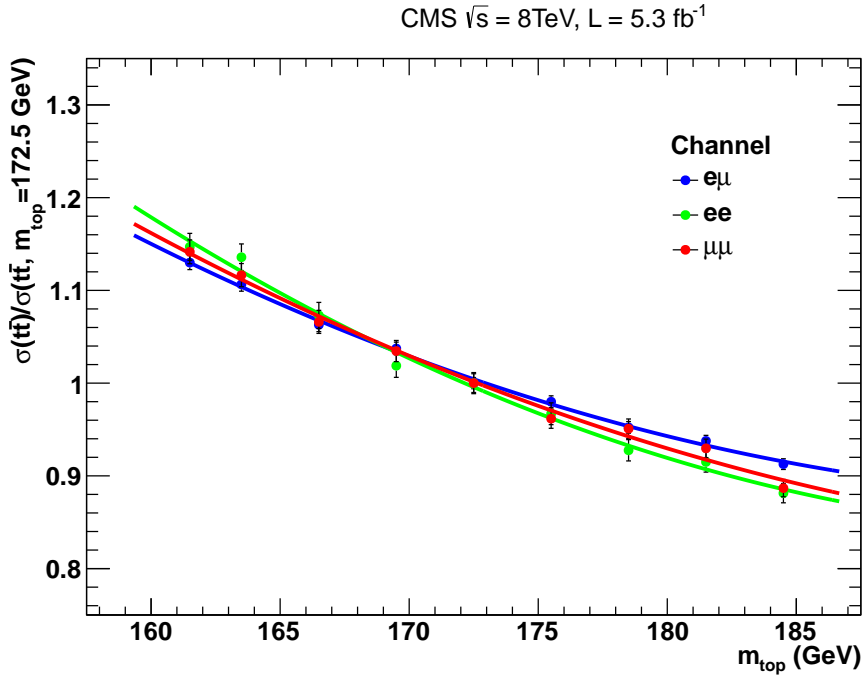
Figure 7.6: Ratio between the measured t$\bar{\text{t}}$ cross section to the one obtained with $m_t^0 = 172.5$ GeV as a function of the top quark mass (m_t) and the corresponding fit for the $\mu^+\mu^-$, e^+e^- and $\mu^\pm e^\mp$ channels.

Sample	$\mathcal{A}_{\text{total}}^{\mu e}$ [%]	$\sigma_{\text{t}\bar{\text{t}}}^{\mu e}$ [pb]
MadGraph w/o spin correlation + Pythia	0.721	237.9
MadGraph with spin correlation + Pythia	0.717	239.0
MC@NLO + Herwig	0.724	236.9
Powheg + Pythia	0.701	244.9
Powheg + Herwig	0.727	235.9

Table 7.6: Total acceptance ($\mathcal{A}_{\text{total}}^{\mu e}$) and t$\bar{\text{t}}$ cross section ($\sigma_{\text{t}\bar{\text{t}}}^{\mu e}$) in the muon-electron channel for signal samples with different generators and interfaced with Pythia or Herwig.

CHAPTER 8

CONCLUSIONS

The Large Hadron Collider (LHC) has produced a huge amount of collision data at a center of mass energies of 7 and 8 TeV between 2010 and 2012. The proton-proton collisions take place in four points of the main ring, where the leading LHC detectors are located: ATLAS, CMS, LHCb and ALICE. The two first have been designed as multipurpose devices, while LHCb and ALICE have specific research topics. This thesis presents results from data collected by the Compact Muon Solenoid (CMS) detector.

The CMS detector is able to perform precise muon identification with a momentum resolution better than 20% for $p_T \approx 1$ TeV in the whole muon system acceptance ($|\eta| < 2.4$). This resolution is achieved when the tracker and muons system are used together in the muon reconstruction. Thus, it is necessary to have a precise knowledge of the relative position of the muon system with respect to the tracker. The link muon alignment system has been designed to perform a continuous monitoring of the muon system components and of the support structures during the detector operation. The link alignment system reconstructs the positions and orientations of the stations ME1, the MABs structures, which are fixed to the barrel, the link disk and alignment rings with respect to the tracker.

The reconstruction process starts using the structure positions determined by photogrametry or the design positions. The first estimation has been performed with data taken when the magnetic field is turned off. Thereafter, several recontructions have been performed in the presence of a magnetic field of 3.8 T.

In addition to monitor the displacements, the performed reconstructions show a temperature dependence in the measurements obtained by the ASPD sensors hitted by the AR lasers. The temperature changes, due to the tracker refrigeration system, affect the laser path in two different ways: generating a temperature gradient in the surrounding air of the laser, and producing deformation in the supporting structures. Consequently, the sensor measurements are corrected by temperature effects using a parametrization derived from data. After the corrections, the reconstructions performed at different temperatures and without changes in the magnetic field are in better agreement.

The studies show that the magnetic field produces a compression towards the CMS detector center of $\approx 14\,\mathrm{mm}$ over the YE$\pm$1 structure. Due to the magnetic field distribution and some structures designed to prevent that the disks get pushed into the barrel, the central part of the YE$\pm$1 is more attracted to the center of the detector than the external part, creating a disk bending of $\approx 4\,\mathrm{mrad}$.

After the hardware calibration, the CMS data are ready to be analyzed. The data taken by the detector allowed to perform precise measurements of the standard model parameters, the discovery of the Higgs boson and searches of new physics. One of the SM particles that contributes to the understanding of the previous topics is the top quark. The properties of the top quark provide relevant information about the electroweak processes. In the study of the properties of the recent discovered Higgs boson, top quark events are one of the main backgrounds in many decay channels (e.g. H→WW). Finally, in the searches for physics beyond standard model, top production is sensitive to new physics signatures as well as one of the main backgrounds.

This thesis presents the measurement of the top anti-top production cross section in the dilepton channel with proton-proton collisions at $\sqrt{s} = 8\,\mathrm{TeV}$. The data used to perform this analysis have been taken by the Compact Muon Solenoid (CMS) detector from April to June 2012. The total amount of data analyzed is $5.3\,\mathrm{fb}^{-1}$.

The data samples have been collected by dilepton triggers based on the selection of

two high p_T lepton ($p_T^{\ell_1} > 17\,\text{GeV}$ and $p_T^{\ell_2} > 8\,\text{GeV}$). Each event has been reconstructed using the particle flow algorithm (PF), which allows to identify all the particles produced in the collision combining the information from all CMS subdetectors.

To identify the signature of the $t\bar{t}$ events in the dilepton channel, the first requirement is that the event must have at least two high-p_T and isolated leptons well identified with opposite electric charge. In the case that the event has more than two leptons, the pair maximising the sum of the transverse momentum is kept. The next cut is over the jets reconstructed using the anti-k_t algoritm. At least 2 jets are required. In order to reject the main background contribution, Drell-Yan, a set of cuts in the $\mu\mu$ and ee channels are applied. The dilepton invariant mass must be outside of the Z mass window ($\notin [76, 106]$) and the missing transverse energy ($\not{E}_T$) must be larger than $40\,\text{GeV}$. Finally, the event must have one jet coming from the hadronization of a b quark. The b-jet has been selected using the *Combined Secondary Vertex (CSV)* algorithm with an operating point corresponding to an identification efficiency of about 85% and a misidentification probability of 10% for jets coming from gluons and light quarks.

The total signal acceptance ($\mathcal{A}_{\text{total}}$), which is the fraction of $t\bar{t}$ events that satisfies the selection cuts, includes trigger, identification and isolation efficiencies, as well the kinematic selection and the b-tagging requirement. $\mathcal{A}_{\text{total}}$ has been measured using a MADGRAPH $t\bar{t}$ Monte Carlo sample generated with a top quark mass of $m_t = 172.5\,\text{GeV}$ and employed PYTHIA for hadronization and extra radiation.

The Backgrounds in this analysis arise from Drell-Yan, single top, boson-boson, $t\bar{t}$ semileptonic and W+jets processes. Drell-Yan and W+jets plus $t\bar{t}$ semileptonic contributions have been estimated from data, while all other backgrounds have been estimated from MC simulations. In the case of the Drell-Yan background, the yields have been obtained with the $R_{out/in}$ method and cross checked with a template fit. The W+jet plus $t\bar{t}$ semileptonic contribution, also called Non-W/Z, have been derived from data using events with relaxed lepton requirements.

The $t\bar{t}$ cross section in the dilepton channel has been estimated using a robust cut and count method for all the three channels. The values obtained for a top quark mass of $m_t = 172.5\,\text{GeV}$ are:

$$\sigma_{t\bar{t}}\left(e^+e^-\right) = 244.3 \pm 5.2 \text{ stat.} \pm 18.6 \text{ syst.} \pm 6.4 \text{ Lumi. pb} \tag{8.1}$$

$$\sigma_{t\bar{t}}\left(\mu^+\mu^-\right) = 235.3 \pm 4.5 \text{ stat.} \pm 18.6 \text{ syst.} \pm 6.1 \text{ Lumi. pb} \tag{8.2}$$

$$\sigma_{t\bar{t}}\left(\mu^\pm e^\mp\right) = 239.0 \pm 2.6 \text{ stat.} \pm 11.4 \text{ syst.} \pm 6.2 \text{ Lumi. pb} \tag{8.3}$$

Combining the values among the channels and extrapolating to the most recent top quark mass measurement ($m_t = 173.2\,\text{GeV}$), the cross section obtained is:

$$\sigma_{t\bar{t}} = 237.5 \pm 13.1 \text{ pb,} \tag{8.4}$$

which is dominated by the cross section measured in the muon-electron channel. This result agrees with the most precise theoretical prediction $\sigma_{t\bar{t}}^{\text{NNLO+NNLL}}(8\,\text{TeV}) = 245.9 \pm 9.6\,\text{pb}$[10]. The total systematic uncertainty, detailed in Table 8.1, is dominated by the contributions coming from the $t\bar{t}$ modeling and the jet energy scale. This result provides a precise normalization for the $t\bar{t}$ contribution in all the analyses in which this process can contribute. It also allows to perform estimations of the top quark mass in the pole and $\overline{MS}$ schemes as presented in [96].

In summary, the result presented in this thesis provides an accurate estimation of the $t\bar{t}$ production cross section in proton-proton collisions at $8\,\text{TeV}$, which is consistent with the theoretical prediction at NNLO. Figure 8.1 shows the most recent results for the measurement of the $t\bar{t}$ cross section as a function of center of mass energy, including the value obtained in this thesis (open red circle). It also shows the theoretical prediction estimated at NNLO+NNLL.
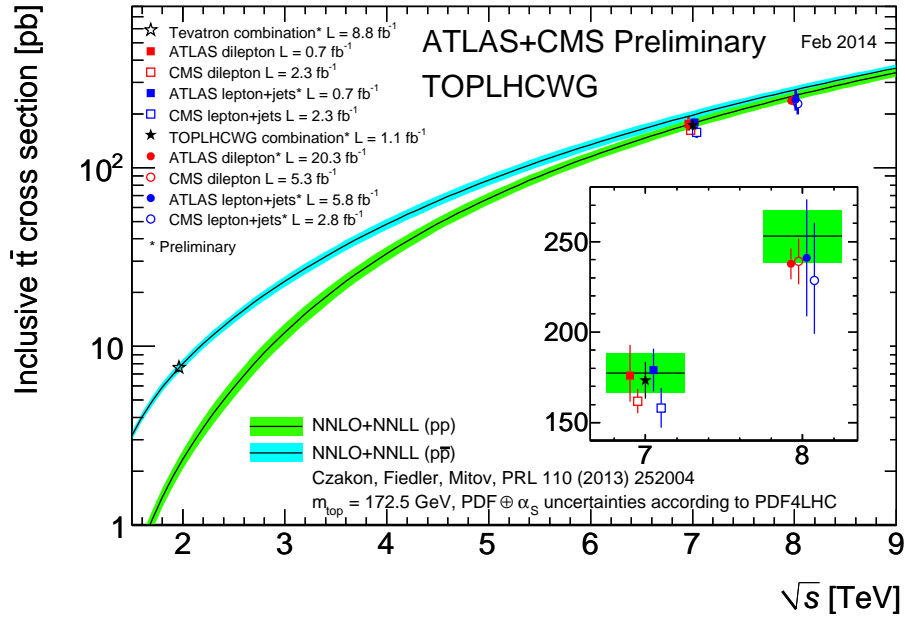
Figure 8.1: Theoretical predictions and experimental mesurements of the $t\bar{t}$ production cross section as a function of the collider center-of-mass energy ($\sqrt{s}$). The available measurements from the Tevatron collider, and from the ATLAS and CMS detectors at 1.96 TeV, 7 TeV and 8 TeV agree with the theoretical predictions.

Source	e^+e^-	$\mu^+\mu^-$	$\mu^{\pm}e^{\mp}$
	[pb]		
Trigger efficiencies	4.1	3.0	3.6
Lepton efficiencies	5.8	5.6	4.0
Lepton energy scale	0.6	0.3	0.2
Jet energy scale	10.3	10.8	5.2
Jet energy resolution	3.2	4.0	3.0
b-jet tagging	1.9	1.9	1.7
Pileup	1.7	1.5	2.0
Scale (μ_F and μ_R)	5.7	5.5	5.6
Matching partons to showers	3.9	3.8	3.8
Single top quark	2.6	2.4	2.3
VV	0.7	0.7	0.5
Drell-Yan	10.8	10.3	1.5
Non-W/Z leptons	0.9	3.2	1.9
Total systematic	18.6	18.6	11.4
Integrated luminosity	6.4	6.1	6.2
Statistical	5.2	4.5	2.6

Table 8.1: Summary of the individual contributions to the systematic uncertainty on the $\sigma_{t\bar{t}}$ measurement. The uncertainties are given in pb. The statistical uncertainty on the result is given for comparison.

CHAPTER 9

RESUMEN

La presente tesis doctoral fue realizada con los datos registrados en el detector CMS (por sus siglas en inglés, Compact Muon Solenoid) del Gran Colisionador de Protones LHC (Large Hadron Collider) del laboratorio CERN (European Organization for Nuclear Research). Dicho trabajo está dividido en dos partes: la primera de ellas está enfocada al estudio del sistema hardware de alineamiento de las cámaras de muones, y la segunda parte, corresponde a la medición de la sección eficaz de producción de pares de quarks top anti-top ($\sigma_{t\bar{t}}$) usando datos de colisiones proton-proton a una energía de centro de masa de 8 TeV.

Este resumen inicia con una breve descripción del detector CMS, seguido por los resultados más relevantes con respecto al sistema link de alineamiento y termina con la descripción del análisis de la medición de la sección eficaz de eventos top anti-top.

9.1 Detector CMS

El detector CMS [30, 31] tiene localizado en su región central un solenoide superconductor capaz de producir un campo magnético en la dirección axial de 3.8 T. CMS se puede dividir en 3 subdetectores que cumplen funciones específicas en el proceso de

reconstrucción de las partículas, ellos son, sistemas de trazas, calorímetro y detector de muones. Iniciando desde el punto más interno del detector, el primer sistema de detección con que se encuentran las partículas tras ser producidas en una colisión es el sistema de trazas. Este subdetector se encuentra dividido a su vez, en el píxel de silicio y el "strip tracker", cubriendo una región $0 < \phi < 2\pi$ en el ángulo azimutal y $|\eta| < 2.5$ en pseudorapidez, donde η es definida como $\eta = -\ln[\tan{(\theta/2)}]$, con θ siendo el ángulo polar medido en sentido contrario de las manecillas del reloj en la dirección del haz de protones. Después del detector de trazas, el siguiente sistema de detección es el calorímetro electromagnético (ECAL) y el calorímetro hadrónico (HCAL). Los cristales de plomo-tungsteno del calorímetro electromagnético y el material centelleante del calorímetro hadrónico están localizados dentro del solenoide. Por último, los muones son medidos por subdetectores de gas-ionizante localizados dentro de la estructura de acero que contiene el solenoide para retornar las líneas de campo magnético. El detector se cierra herméticamente con el fin de tener la mayor aceptancia posible y así, proveer una medida confiable del momento faltante en el plano transverso al haz de protones ($\not{E}_T$).

9.2   Alineamiento de las Cámaras de Muones

Con la finalidad de reconstruir de forma precisa los muones producidos en las colisiones, es necesario conocer con una baja incertidumbre las posiciones y orientaciones de las cámaras que conforman el espectrómetro de muones. Para este fin se diseñó y construyó un sistema de alineamiento encargado de reconstruir las posiciones y ángulos de las principales estructuras con alta precisión. Este sistema se divide en tres grupos: sistema de alineamiento del barrel, sistema de alineamiento de los endcaps y sistema link de alineamiento.

La monitorización de las posiciones de las cámaras de muones con los sistemas de alineamiento se realiza de forma continua debido a que con relativa frecuencia, y sobre todo, tras cada cambio de campo magnético se producen pequeños desplazamientos en las estructuras del detector. El sistema link de alineamiento es el encargado de unificar la información de los demás sistemas en el sistema coordenado global de CMS. El proceso se lleva a cabo con un complejo sistema óptico-mecanico que se encarga de medir las posiciones de las estructuras más relevantes por medio de laseres, sensores 2D, distanciómetros, inclinómetros y otros instrumentos de alta precisión. Estas mediciones son luego usadas para realizar la correspondiente reconstrucción

geométrica del sistema por medio de el programa COCOA [44]. Tras cada reconstrucción de la geometría del detector, las posiciones y orientaciones de las siguientes estructuras son estimadas: $YE \pm 1$, $YB \pm 2$, $ME1/1$ y las cámaras $ME1/2$, MABs y los LD.

El mayor cambio en la geometría del detector debido a las variaciones del campo magnético, es un desplazamiento de aproximadamente 14 mm de los discos $YE \pm 1$ en dirección al centro del detector. Junto a este desplazamiento, los discos también sufren una leve desformación debido a su interacción con los "Z-stop", estructuras diseñadas específicamente para limitar la compresión del detector. Como resultado, se produce una desformación de los discos $YE \pm 1$ de aproximadamente 4 mrad, como lo muestra la figura 4.9.

Regularmente son llevadas a cabo reconstrucciones de las geometría del espectrómetro de muones para monitorizar cualquier cambio en el sistema. Los resultados consignados en las tablas 4.3 y 4.4 corresponden a las reconstrucciones llevadas a cabo en el periodo de febrero a agosto de 2010.

9.3 Medición de la Sección Eficaz de Producción de pares top anti-top

La precisa medición de la sección eficaz de producción de pares top anti-top ($\sigma_{t\bar{t}}$) puede ser usada para poner a prueba la teoría cuántica de la cromodinámica (QCD por sus siglas en inglés Quantum Chromodynamics) a un nivel de precisión de NNLO (del inglés *next-to-next-to leading order*). Esta medición de la $\sigma_{t\bar{t}}$ también puede ser usada para estudiar los fits de las funciones de distribución partonicas (PDF) a NNLO, y la estimación de $\alpha_s(M_Z)$ como se describe con más detalle en [96, 97]. Adicionalmente la producción de quarks top es una importante fuente de fondos en muchas de las búsquedas de física más allá del modelo standard y en los estudios del recién descubierto bosón de Higgs [7]. En el Modelo Standard (SM), la producción de quarks top está dominada por la producción de parejas top ati-top vía interacción fuerte. Con respecto a su decaimiento, el quark top casi el $\sim 100\%$ de las veces decae en dos bosones W y dos quarks bottom. Una gran muestra de eventos con quarks top ha sido recogida por el gran acelerador de hadrones (LHC), permitiendo realizar múltiples estudios en diferentes canales de decaimiento así como búsquedas de desviaciones con respecto a las predicciones del Modelo Standard.

En esta tesis se presenta la medición de la sección eficaz de producción de parejas de quarks top anti-top en canales de decaimiento dileptónico en colisiones protón-protón con una energía de centro de masa de $\sqrt{s} = 8\,\text{TeV}$ usando los datos correspondientes a una luminosidad de $5.3\,\text{fb}^{-1}$. Los datos fueron tomados por el detector CMS durante el periodo de abril a junio del 2012.

La medición de la sección eficaz fue realizada con eventos que contienen un par de leptones con carga eléctrica opuesta, momento faltante asociado a los neutrinos del decaimiento del bosón W y dos jets de partículas resultantes de la hadronización de los dos quarks bottom.

9.3.1   Simulaciones de Monte Carlo

Muchos generadores de eventos Monte Carlo han sido usados para simular la señal y el fondo: MADGRAPH (v. 5.1.4.8) [79], POWHEG (r1380) [81] y PYTHIA (v. 6424) [80] dependiendo del proceso considerado. El generador MADGRAPH con "spin correlation" es usado para modelar los eventos $t\bar{t}$ considerando una masa del quark top de $m_t = 172.5\,\text{GeV}$. Para simular las cascadas partónicas y la hadronización, se ha empleado la interfase de PYTHIA. MADGRAPH es también usado para simular W+jets, WW y procesos Drell-Yan mientras la producción de "single-top" es simulada con POWHEG. La producción inclusiva de dibosones WZ y ZZ es simulada con PYTHIA. El decaimiento de los leptones taus ha sido procesado por medio del TAUOLA (v. 2.75). En el análisis, las contribuciones de los fondos de dibosones VV (WW, WZ y ZZ) y la producción de "single-top" son tomadas de simulaciones de MC con un nivel de precisión NLO en sus secciones eficaces. Todos los demás fondos son estimados de muestras de control extraídas de los mismos datos de colisiones.

La más reciente estimación teórica de la sección eficaz de la producción de parejas top anti-top es $\sigma_{t\bar{t}} = 252.9^{+6.4}_{-8.6}(\text{scale}) \pm 11.7(\text{PDF} + \alpha_s)\,\text{pb}$. Esta es estimada con el programa TOP++ [20] a NNLO en QCD perturbativa, incluyendo soft "gluon resummation" a NNLL [10] y asumiendo una masa del quark top de $m_t = 172.5\,\text{GeV}$. La incertidumbre asociada a la escala de interacción viene de las variaciones independientes de la escala de los factores de renormalización, μ_F y μ_R, mientras la contribución del PDF+α_s fue estimada siguiendo el procedimiento PDF4LHC [98].

9.3.2   Selección de Eventos

La selección de eventos para separar señal y fondo de la muestra de datos empieza con la pre-selección que usan los triggers elegidos para el análisis. Así, los eventos seleccionados deben tener dos electrones, dos muones o un muón y un electrón con momento transverso $p_T > 17\,\text{GeV}$ para el primer leptón y para el segundo un $p_T > 8\,\text{GeV}$. La eficiencia de los triggers dileptónicos escogidos es aproximadamente 90%-93% para los 3 estados finales. Usando la medida de la eficiencia de los trigger obtenida de los datos y la simulación de la eficiencia en MC, los eventos simulados se corrigen por unos factores de escala (SF) en bines de p_T, los cuales tienen un valor medio de 0.96 y una incertidumbre de 1 a 2%.

Tras esta primera pre-selección basada en el trigger, los eventos son filtrados requiriendo que ambos leptones tengan carga eléctrica opuesta, $p_T > 20\,\text{GeV}$ y estén dentro del rango de $|\eta| < 2.5$ para electrones y $|\eta| < 2.1$ para muones. Si algún evento tiene más de un par de leptones que pasen los cortes antes mencionados, se seleccionara como pareja a aquella que tenga el mayor p_T combinado. Los eventos que tengan taus contribuyen a la señal sí y solo sí su decaimiento es a electrones o muones, y estos últimos, pasan los cortes de selección.

Los leptones candidatos provenientes del decaimiento del bosón W están usualmente aislados de cualquier otra partícula en el evento. Para cada muón o electrón candidato, un cono de $\Delta R < 0.3$ es construido al rededor de la traza del leptón, donde ΔR esta definido como $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ y $\Delta\eta$ y $\Delta\phi$ son las diferencias en pseudorapidez y en el ángulo azimutal entre cualquier deposito de energía y el eje del cono. La suma escalar del p_T de todas las partículas reconstruidas con el algoritmo PF [11] provenientes del vértice primario y contenidas por este cono es calculada, excluyendo la energía del leptón al que se le esta estimando su aislamiento. El discriminador a usar es el "aislamiento relativo" (I_{rel}) el cual esta definido como el ratio entre esta suma del p_T de las partículas en el cono y el p_T del leptón candidato. Un candidato a leptón es rechazado si su aislamiento relativo es $I_{rel} > 0.15$.

La eficiencia de la selección de leptones ha sido medida usando el método "tag and probe" en eventos dileptónicos de una muestra enriquecida con bosones Z. La medida obtenida para los valores de eficiencia de identificación (ID) y de aislamiento (ISO) son del orden de 96% para muones y 90% para electrones. Comparando las eficiencias de selección de leptones medidas en datos y simulaciones MC, los eventos simulados

son escalados con $SF_{ID,ISO}$ en bines de p_T y η. Los $SF_{ID,ISO}$ tienen un valor medio de 0.99 y una incertidumbre del 1 al 2%.

Eventos que contienen dos leptones con una masa invariante menor a $m_{\ell\ell} < 20\,\text{GeV}$ son eliminados para suprimir el fondo de resonancias de "heavy flavors" así como las contribuciones de eventos de Drell-Yan con baja masa. Adicionalmente, para los canales ee y $\mu\mu$ se aplica un veto en la región de masa invariante $76\,\text{GeV} < m_{\ell\ell} < 106\,\text{GeV}$ para reducir también la contribución de Drell-Yan.

Los jets son reconstruidos con los candidatos de partículas obtenidos con el algoritmo PF usando el algoritmo de agrupamiento de partículas llamado "anti-k_t" [66]. Los eventos seleccionados deben tener por lo menos 2 jets reconstruidos con $p_T > 30 GeV$ dentro del rango en $\eta < 2.5$.

La energía faltante en el plano transverso al haz de protones, $\not{E}_T$, esta definida como el negativo de la suma del momento de todas las partículas reconstruidas en el plano transverso al haz de protones. Un valor de $\not{E}_T > 40\,\text{GeV}$ es requerido en los canales ee y $\mu\mu$ para suprimir eventos Drell-Yan, en los cuales no se espera $\not{E}_T$. Para el canal μe no es necesario este corte ya que tiene poco fondo.

Dado que los eventos de $t\bar{t}$ contienen jets provenientes de la hadronización de los quarks b, exigir la presencia de estos b-jets disminuye la cantidad de fondo presente en la muestra. Los b-jet son identificados usando el algoritmo llamado CSV (por sus siglas en inglés, Combined Secondary Vertex). El corte que se establece para realizar esta selección corresponde a una eficiencia de identificación de 85% y al rededor de 10% de probabilidad de identificar otros jets provenientes de u, d, s, c quarks y gluones como b-jet. La selección de eventos usada en esta tesis exige la presencia de por lo menos un b-jet.

La figura 9.1 muestra la distribución de p_T del leptón con mayor momento y del jet de mayor momento para la combinación de los tres canales finales. En esta y en las demás figuras, se considera una masa del quark top de aproximadamente $m_t = 172.5\,\text{GeV}$. Las regiónes sombreadas corresponden a la incertidumbre estadística que se espera de la simulaciones. El ratio entre los datos y la suma de las simulaciones y las predicciones basadas en los mismos datos para la señal y para los fondos es presentada en la parte baja de las figuras. El número de jets y b-jets seleccionados está en la figura 9.2 para el canal muón-electrón (μe), el cual, tiene la menor contribución

de fondos. Un acuerdo similar se obtiene en los canales muón-muón ($\mu\mu$) y electrón-electrón (ee).
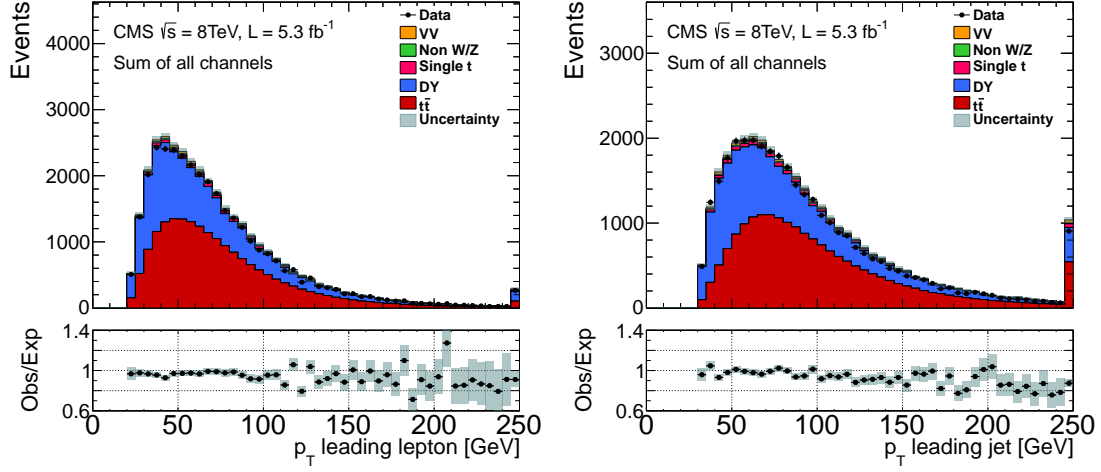
Figure 9.1: Distribución del momento transverso (p_T) para el leptón (izquierda) y el jet (derecha) con mayor p_T después de aplicar el corte en el número de jets (por lo menos dos jets). La distribución para los eventos t$\bar{t}$ ha sido normalizada al valor esperado de la sección eficaz, $\sigma_{t\bar{t}} = 252.8\,\mathrm{fb}^{-1}$. En la parte baja del plot se muestra el ratio entre los datos y las predicciones. La región sombreada corresponde al error estadístico.

9.3.3   Determinación del fondo

Los fondos en este análisis provienen de eventos single-top, Drell-Yan y bosón-bosón, en los cuales, al menos dos leptones son producidos de los decaimientos de los bosones Z o W. Otras fuentes de fondos como los eventos t$\bar{t}$ y W+jets con decaimientos en leptón más jets y donde al menos uno de los jets es reconstruido erróneamente como leptón (lo cual pasa usualmente para electrones) o se considera un leptón del decaimiento de un hadrón que contenga quarks bottom o charm (lo cual pasa usualmente para muones) están agrupados en la categoría de "Non-W/Z". El número de eventos de single-top y bosón-bosón son estimados con simulaciones MC mientras los fondos Drell-Yan y Non-W/Z son estimados de los datos.

El fondo de Drell-Yan es estimado usando el método $R_{out/in}$ [99] en el cual, el número de eventos fuera de la ventana del Z es estimado por medio de la normalización de los eventos simulados. El factor usado en la normalización se basa en el ratio del número de eventos observados dentro y fuera de la ventana del bosón Z en la simulación.
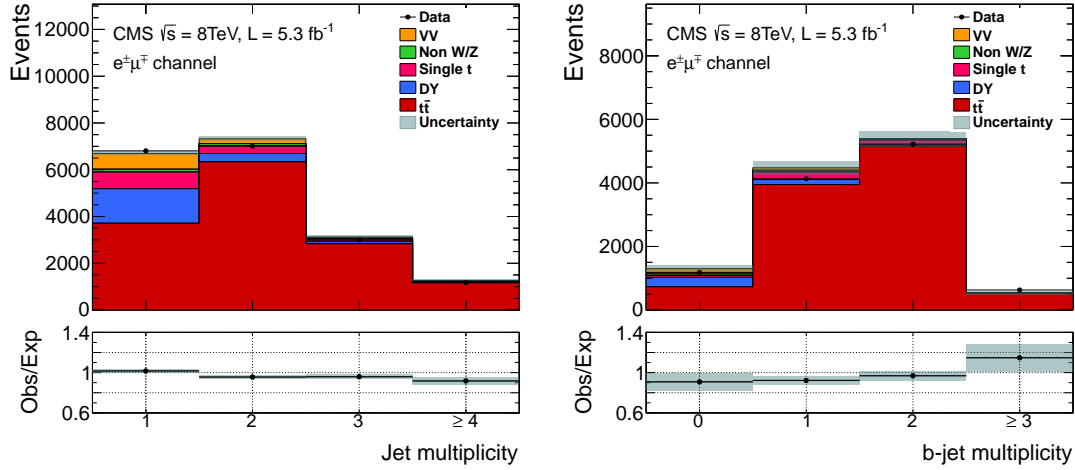
Figure 9.2: Distribución del número de jets (izquierda) y el número de b-jets (derecha) tras aplicar todos los cortes de selección, excepto el corte en b-jets. La distribución para los eventos t̄t ha sido normalizada al valor esperado de la sección eficaz, $\sigma_{t\bar{t}} = 252.8\,\text{fb}^{-1}$. En la parte baja del plot se muestra el ratio entre los datos y las predicciones. La región sombreada corresponde al error estadístico.

Para los canales ee y $\mu\mu$ los SF encontrados son respectivamente 1.7 ± 0.5 y 1.6 ± 0.5, mientras para el canal es μe 1.3 ± 0.4. Los valores obtenidos son compatibles con los estimados usando un método alternativo basado en un fit de la masa dileptónica que incluye todos los procesos (t̄t, Drell-Yan, etc). La tabla 6.13 compara los resultados entre ambos métodos.

Para la estimación del fondo Non-W/Z es necesario primero clasificar los leptones en dos clases: leptones "prompt" y "non-prompt". Los leptones que provienen del decaimiento de bosones W o Z, y que por ello, suelen estar aislados y bien identificados se les conoce con el nombre de leptones "prompt". Los cortes empleados para la selección de dichos leptones son los mismos presentados en la sección 9.3.2. El numero de leptones "prompt" es extraído de una muestra de datos enriquecida con eventos que contienen bosones Z. Por otra parte, los leptones que provienen de decaimientos de mesones o quarks pesados, mala identificación de jets o debido a efectos de la resolución finita del detector se les conoce con el nombre de leptones "non-prompt". El numero de leptones non-prompt es extraído de una región de control de datos enriquecida con eventos QCD. Dichos leptones son seleccionados con requerimientos menos estrictos en la identificación y el aislamiento.

El ratio entre el número de leptones "prompt" y "non-prompt" es estimado en bines de p_T y η de los mismos leptones. Para obtener la estimación final del fondo Non-W/Z en la región de señal, se estima un peso a cada evento de una muestra de dileptones en la cuales ambos leptones fueron seleccionados con un criterio más relajado en la identificación y en el aislamiento. Las incertidumbre sistemática asociada es de un 40% sobre los eventos estimados.

9.3.4 Fuentes de Incertidumbres Sistemáticas

Los eventos simulados son escalados de acuerdo a las correcciones debidas a eficiencias de leptones (usualmente muy cercanas a uno) medidas usando muestras de control de datos, dando como resultado un 1 a 2% de incertidumbre en la eficiencia de eventos $t\bar{t}$.

El impacto de la incertidumbre en la escala de energía del jet (JES) y la resolución en su energía (JER) es estimada desde la variación en el número de eventos seleccionados de $t\bar{t}$ en MC después de variar el momento de los jets dentro de la incertidumbre del JES y en el caso del JER, aplicando correcciones que dependen de η con un valor medio de 10%. Para los canales $\mu\mu$ y ee, estas incertidumbres son también propagadas a la estimación de la $\not{E}_T$ dando como resultado una incertidumbre mayor que en el canal μe, donde no hay corte sobre la $\not{E}_T$.

La incertidumbre asociada a los SF usados para la identificación de b-jets en los eventos $t\bar{t}$ es aproximadamente 2% para b-jets y un 10% para jets mal identificados, los cuales dependen del p_T del jet. estas incertidumbres son propagadas a la eficiencia de los cortes en la selección de eventos $t\bar{t}$.

La incertidumbre asociada a la simulación del "pileup" suma un 0.8%, obtenida variando la sección eficaz inelástica por un 5%.

La incertidumbre asociada a la medida de la luminosidad integrada suma un 2.6% de la sección eficaz medida.

El efecto sistemático relacionado con los diagramas de mayor orden que no se tuvieron en cuenta en la simulación de MadGraph fueron estimados de dos formas. La incertidumbre en la aceptancia de la señal es determinada variando las escalas de renormalización y factorización simultáneamente arriba y abajo por un factor de 2

usando MADGRAPH, y la incertidumbre que se toma es la máxima diferencia después de los cortes de selección. El efecto calculado en la sección eficaz t$\bar{t}$ es 2.3%, el cual es usado como incertidumbre en el análisis. Este valor es corroborado comparándolo con la predicción de los generadores MADGRAPH y POWHEG, donde se a usado PYTHIA para la hadronización y la radiación extra. El valor obtenido es 2.2%, valor comparable con el número obtenido anteriormente.

El "matching" entre la matriz de elementos (ME) y la cascada partónica (PS) es llevada a cabo siguiendo el procedimiento MLM [95]. Cambiando los umbrales que controlan el matching de los partones de la ME con aquellos del PS, se obtiene un 1.6% de variación en la eficiencia de selección de eventos t$\bar{t}$.

La incertidumbre proveniente del modelo de hadronización afecta principalmente el JES y la fragmentación de los b-jets. Como la eficiencia de b-jets y mistag son tomadas de los datos, ninguna incertidumbre adicional que incluya la hadronización es necesaria para esta contribución. En el caso de la incertidumbre del JES, esta ya incluye la contribución de la incertidumbre del modelo de hadronización. Esta incertidumbre también es determinada comparando muestras de eventos generados con POWHEG donde la hadronización es modelada en un caso con PYTHIA y en otro con HERWIG. El efecto observado en la sección eficaz de t$\bar{t}$ es de 1.4%, el cual esta contenido, como se indico anteriormente, en la incertidumbre del JES.

La incertidumbre en el número de eventos de fondo seleccionados de single-top y bosón-bosón es calculada siguiendo los mismo procedimiento antes mencionados. Adicionalmente, otra fuente de incertidumbre asociada al error que se tiene en las mediciones de las secciones eficaces de estos fondos es incluida. Este error es tomado de las ultimas medidas experimentales hechas, que corresponde aproximadamente a 20%.

La tabla 9.1 resume la magnitud de la incertidumbre sistemática en la sección eficaz de producción de eventos t$\bar{t}$ para las todas las fuentes antes descritas.

9.3.5   Resultados

La sección eficaz de producción de eventos t$\bar{t}$ es medida contando los eventos que pasan los cortes de selección mencionados anteriormente, por medio de la expresión:

Fuente	e^+e^-	$\mu^+\mu^-$	$\mu^\pm e^\mp$
		[pb]	
Eficiencia de Trigger	4.1	3.0	3.6
Eficiencia de Leptón	5.8	5.6	4.0
Escala de Energía del Leptón	0.6	0.3	0.2
Escala de Energía del Jet	10.3	10.8	5.2
Resolución de Energía del Jet	3.2	4.0	3.0
Identificación de b-jet	1.9	1.9	1.7
Pileup	1.7	1.5	2.0
Escala (μ_F and μ_R)	5.7	5.5	5.6
Matching	3.9	3.8	3.8
Single top	2.6	2.4	2.3
Bosón-Bosón	0.7	0.7	0.5
Drell-Yan	10.8	10.3	1.5
Non-W/Z	0.9	3.2	1.9
Sistemático Total	18.6	18.6	11.4
Luminosidad Integrada	6.4	6.1	6.2
Estadístico	5.2	4.5	2.6

Table 9.1: Resumen de cada una de las contribuciones a la incertidumbre de la medición de la sección eficaz ($\sigma_{t\bar{t}}$). Las incertidumbres están dadas en pb.

$$\sigma(\text{pp} \to t\bar{t}) = \frac{N - N_{Bkg}}{\underbrace{(\mathcal{A} \times \mathcal{E} \times \text{Br})}_{\mathcal{A}_{\text{total}}} \int \mathcal{L}\, dt} \tag{9.1}$$

Donde:

N es el número de eventos obtenidos en la muestra de datos.

N_{Bkg} es el número de eventos de fondo extraídos de las simulaciones y de los datos.

$\mathcal{A}$ es la aceptancia de los cortes de selección.

$\mathcal{E}$ es la eficiencia del detector.

Br es el ancho de decaimiento de eventos $t\bar{t}$ en dileptones (Br).

$\mathcal{L}$ es la luminosidad, $5.3\,\mathrm{fb}^{-1}$.

La tabla 9.2 muestra el número total de eventos observados en datos, el número de eventos de señal y los fondos tomados directamente de MC y aquellos estimados de los datos. La tabla 9.3 muestra la aceptancia multiplicada por la eficiencia de la selección y el ancho de decaimiento (Br) para los estados finales con dos leptones. También muestra la sección eficaz para los tres canales, electrón-electrón (ee), muón-muón ($\mu\mu$) y muón-electrón (μe), los cuales tienen valores compatibles entre si. Debido a la contribución de incertidumbres adicionales que tienen los canales ee y $\mu\mu$ provenientes de la propagación de algunos sistemáticos a la $\not{E}_T$, la precisión en estos canales es menor que la obtenida en μe.

Usando el método BLUE [89] para combinar los resultados de los tres canales se obtiene una medición de la sección eficaz de $\sigma_{t\bar{t}} = 239 \pm 2(\mathrm{stat.}) \pm 11(\mathrm{syst.}) \pm 6(\mathrm{Lumi.})$ pb para una masa del quark top de $m_t = 172.5\,\mathrm{GeV}$.

En este análisis, la dependencia de la aceptancia como función de la masa del quark top fue parametrizada con la función cuadrática:

$$\frac{\sigma_{t\bar{t}}(m_t)}{\sigma_{t\bar{t}}(m_t^o)} \;=\; 1.00 - 0.009 \times (m_t - m_t^o) - 0.000168 \times (m_t - m_t^o)^2 \qquad (9.2)$$

donde la m_t esta en GeV. Asumiendo el ultimo valor medido de la masa del top, $m_t = 173.2\,\mathrm{GeV}$ [3], la sección eficaz resultante es $\sigma_{t\bar{t}} = 237.5 \pm 13.1$ pb.

La figura 9.3 muestra la distribución para la masa invariante de los leptones, el $\not{E}_T$, y la diferencia de los ángulos azimutales de los leptones y sus correspondientes ratios con las distribuciones simuladas para el canal μe, el cual domina la combinación.

Muestra	Número de Eventos		
	e^+e^-	$\mu^+\mu^-$	$\mu^\pm e^\mp$
Drell-Yan	386±116	492±148	194±58
Non-W/Z	25±10	114±46	185±72
Single top	127±28	157±34	413±88
Bosón-Bosón	30±8	39±10	94±21
Total de Fondos	569±120	802±159	886±130
Señal t$\bar{\text{t}}$ dileptónico	2728±182	3630±250	9624±504
Total MC	3297±218	4432±296	10510±520
Datos	3204	4180	9982

Table 9.2: Número de eventos esperados y observados que superan todos los cortes de selección. En la tabla se presenta cada contribución de fondo por separado. El valor esperado de eventos t$\bar{\text{t}}$ esta normalizado a una sección eficaz de $\sigma_{t\bar{t}} = 252.9$ pb, la cual corresponde a una masa del quark top de 172.5 GeV. La incertidumbre para cada entrada corresponde a la suma cuadrática del error sistemático y estadístico.

Canal	$\mathcal{A}_{\text{total}}[\%]$	$\sigma_{t\bar{t}} \pm$ Estad. $\pm$ Sist. $\pm$ Lumi.$[\,\text{pb}\,]$
e^+e^-	0.203 ± 0.012	$244.3 \pm 5.2 \pm 18.6 \pm 6.4$
$\mu^+\mu^-$	0.270 ± 0.017	$235.3 \pm 4.5 \pm 18.6 \pm 6.1$
$\mu^\pm e^\mp$	0.717 ± 0.033	$239.0 \pm 2.6 \pm 11.4 \pm 6.2$

Table 9.3: La primera columna tiene la aceptancia total ($\mathcal{A}_{\text{total}}$) que esta definida como el producto de la aceptancia de los cortes, la eficiencia del detector y el ancho de decaimiento dileptónico para el proceso t$\bar{\text{t}}$. Estos valores fueron estimados para una masa del quark top de 172.5 GeV. La segunda columna tiene los valores de sección eficaz de producción de t$\bar{\text{t}}$ para los tres canales estudiados con su correspondiente error estadístico, sistemático y el error asociado a la luminosidad.
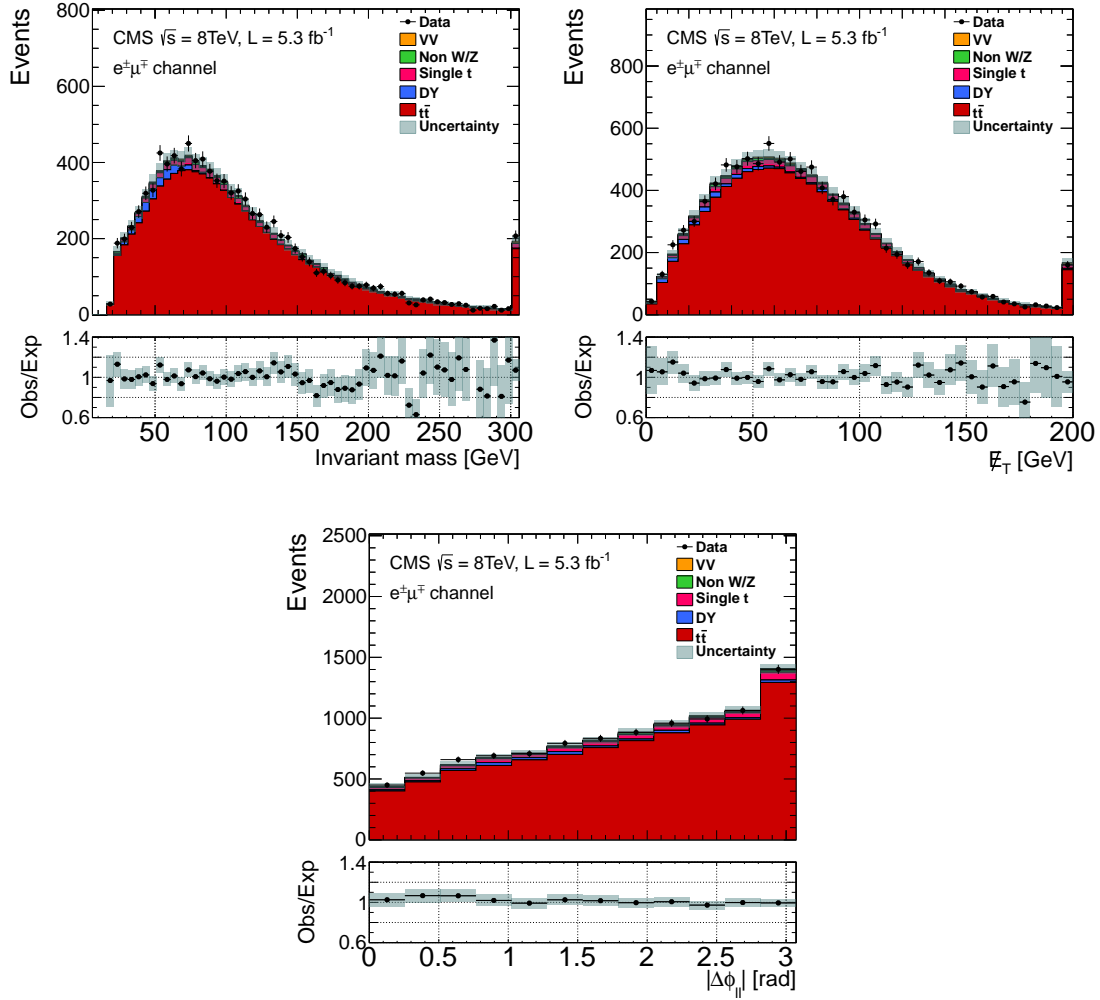
Figure 9.3: Distribución de la masa invariante del sistema dileptónico (superior-izquierda), $\not{E}_T$ (superior-derecha) y la diferencia de los ángulos azimutales de los leptones (inferior), después de aplicar todos los cortes de selección en el canal muón-electrón. La distribución de señal $t\bar{t}$ ha sido normalizada al valor de la sección eficaz medida en este trabajo. En la parte baja del plot se muestra el ratio entre los datos y las predicciones. La región sombreada corresponde al error estadístico.

APPENDIX A

COMPARISON OF THE t̄t SAMPLES

In this appendix are collected the comparison plots for the different t̄t samples. As was explained in Section 6.1.1, the t̄t selected, due to the level of agreement with data, was MadGraph with spin correlation. This improvement is notorious in the tails of the $\Delta\phi_{\ell\ell}$ distribution. In the invariant mass range of $\sim 20 - 60\,\text{GeV}$ is also better the description obtained using this MonteCarlo sample.
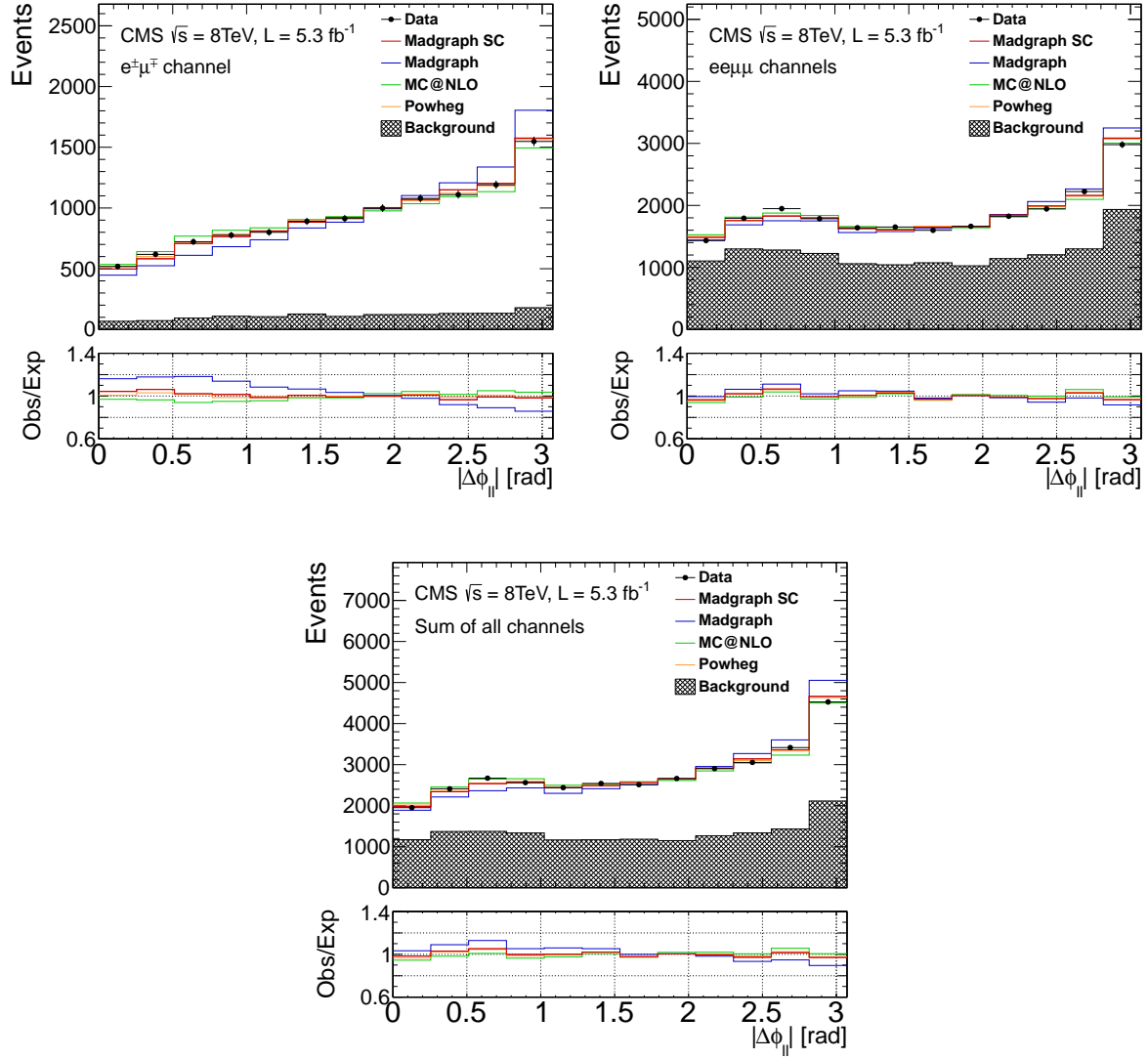
Figure A.1: Difference of the azimuthal angle between the two selected leptons $\Delta\phi_{\ell\ell}$ for the different signal generators after the 2 jets cut for μe, $\mu\mu$ + ee and all channels.
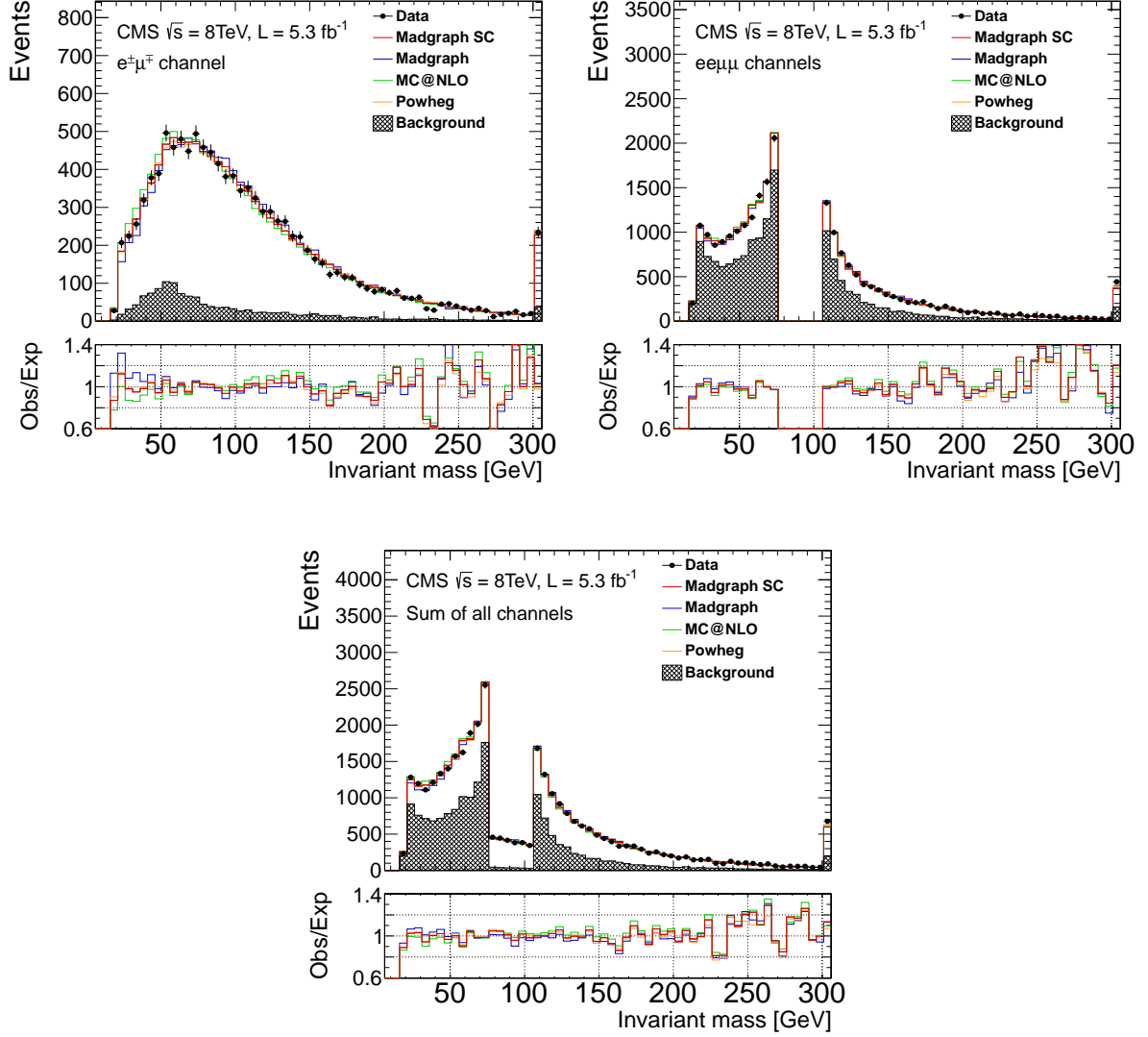
Figure A.2: Invariant mass distribution $(m_{\ell\ell})$ for the different signal generators after the 2 jets cut for μe, $\mu\mu$ + ee and all channels.
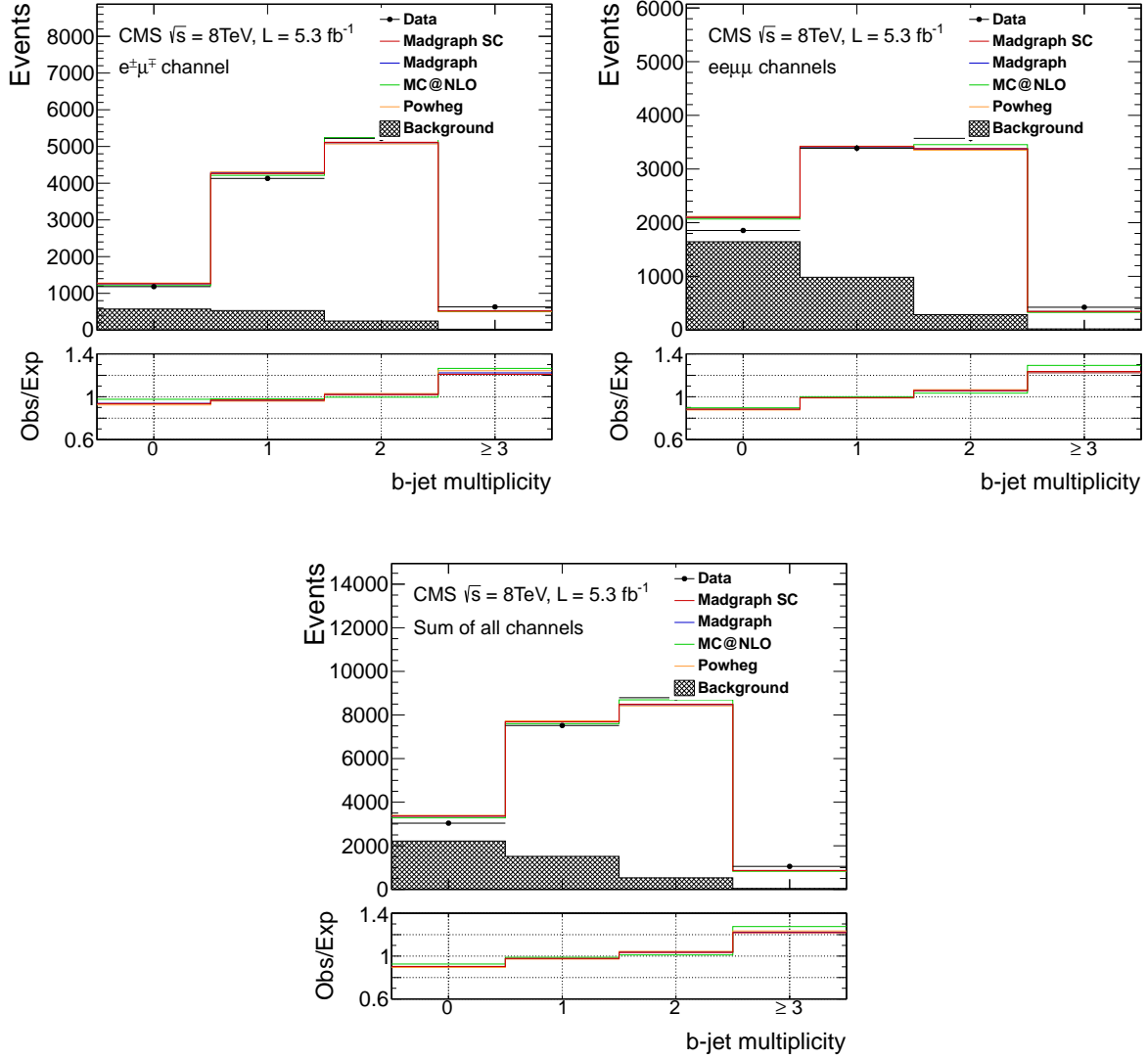
Figure A.3: b-jet multiplicity for the different signal generators after the $\not{E}_T$ cut for μe, $\mu\mu$ + ee and all channels.

APPENDIX B

BLUE METHOD

The BLUE method allows to combine n measurements for N observables. Let's consider the three measurements of the $\sigma_{t\bar{t}}$ for each channel ($\mu\mu$, ee and μe), in this case $n = 3$ and $N = 1$. The measurements are:

$$\left.\begin{array}{c} \sigma_{t\bar{t}}^{\mu\mu} + \delta\sigma_{t\bar{t}}^{\mu\mu} = \sigma_1 + \delta\sigma_1^{\text{stat.}} + \delta\sigma_1^{\text{syst.}} \\ \sigma_{t\bar{t}}^{ee} + \delta\sigma_{t\bar{t}}^{ee} = \sigma_2 + \delta\sigma_2^{\text{stat.}} + \delta\sigma_2^{\text{syst.}} \\ \sigma_{t\bar{t}}^{\mu e} + \delta\sigma_{t\bar{t}}^{\mu e} = \sigma_3 + \delta\sigma_3^{\text{stat.}} + \delta\sigma_3^{\text{syst.}} \end{array}\right\} \text{Measurements} \qquad (B.1)$$

An estimator, $\hat{\sigma}$, for the cross section can be written as:

$$\hat{\sigma} = \sum_i w_i \sigma_i = w_1 \sigma_1 + w_2 \sigma_2 + w_3 \sigma_3 \qquad (B.2)$$

which is constructed as a linear combination of the individual measurements scaled by a weight w satisfying the normalization condition of:

$$\sum_i w_i = 1 \tag{B.3}$$

As the BLUE method consist of deriving the weights such the variance is minimal, the next step is to define the *variance* of the estimator as a function of the weights:

$$\text{Var}[\hat{\sigma}] = \sum \sum w_i M_{ij} w_j \tag{B.4}$$

with **M** as the *covariance matrix*:

$$\mathbf{M} = \begin{pmatrix} \delta\sigma_1^2 & 0 & 0 \\ 0 & \delta\sigma_2^2 & 0 \\ 0 & 0 & \delta\sigma_3^2 \end{pmatrix}_{\text{stat.}} + \begin{pmatrix} \delta\sigma_1^2 & \rho\delta\sigma_1\delta\sigma_2 & \gamma\delta\sigma_1\delta\sigma_3 \\ \rho\delta\sigma_1\delta\sigma_2 & \delta\sigma_2^2 & \beta\delta\sigma_2\delta\sigma_3 \\ \rho\delta\sigma_1\delta\sigma_3 & \beta\delta\sigma_2\delta\sigma_3 & \delta\sigma_3^2 \end{pmatrix}_{\text{syst.}} \tag{B.5}$$

Where it is assumed that all statistical errors are uncorrelated, while systematic uncertainties are correlated by a factor ρ, γ and β. These correlation factors across the channels are equal to 1, except in the case of $SF_{tr,ID,ISO}$ (as was explained in Section 7.1.1) and in the data-driven methods, in which case, were taken as uncorrelated errors.

Finally, Lagrange Multipliers are used to derive the weight matrix **w**:

$$\mathbf{w} = \left[\mathbf{U}^T\mathbf{M}^{-1}\mathbf{U}\right]^{-1}\left[\mathbf{U}^T\mathbf{M}^{-1}\right] \quad \text{with} \quad \mathbf{U} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \tag{B.6}$$

This matrix provide all the weight necessaries to perform the combination of the $\sigma_{t\bar{t}}$ among the channels.

BIBLIOGRAPHY

[1] DØ Collaboration, "Precision measurement of the top-quark mass in lepton+jets final states", `arXiv:1405.1756`.

[2] CDF, DØ Collaboration, "Combination of CDF and DØ results on the mass of the top quark using up to 8.7 fb^{-1} at the Tevatron", `arXiv:1305.3929`.

[3] Particle Data Group, "The Review of Particle Physics", *Phys. Rev. D* **86** (2012) 010001, `doi:10.1103/PhysRevD.86.010001`.

[4] CDF Collaboration: F. Abe et al., "Observation of Top Quark Production in $p\bar{p}$ Collisions with the Collider Detector at Fermilab", *Phys.Rev.Lett.* **74(14)** (1995) 2626–2631, `doi:10.1103/PhysRevLett.74.2626`.

[5] DØ Collaboration: S. Abachi et al., "Observation of the Top Quark", *Phys.Rev.Lett.* **74(14)** (1995) 2632–2637, `doi:10.1103/PhysRevLett.74.2632`.

[6] ATLAS , CDF , CMS , DØ Collaboration, "First combination of Tevatron and LHC measurements of the top-quark mass", `arXiv:1403.4427`.

[7] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC", *Phys.Lett.* **B716** (2012) 30–61, `doi:10.1016/j.physletb.2012.08.021`, `arXiv:1207.7235`.

[8] 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** (2012) 1–29, `doi:10.1016/j.physletb.2012.08.020`, `arXiv:1207.7214`.

[9] CMS Collaboration, "Measurement of the $t\bar{t}$ production cross section in the dilepton channel in pp collisions at $\sqrt{s}$ = 8 TeV", *JHEP* **1402** (2014) 024, `doi:10.1007/JHEP02(2014)024`, `arXiv:1312.7582`.

[10] M. Czakon, P. Fiedler, and A. Mitov, "Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\frac{4}{S})$", (2013). `arXiv:1303.6254`.

[11] CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET", *CMS-PAS* **PFT-09-001** (2009).

[12] The Super-kamiokande Collaboration, "Three flavor neutrino oscillation analysis of atmospheric neutrinos in Super-Kamiokande", *Phys. Rev. D* **74** (2006) 032002, `doi:10.1103/PhysRevD.74.032002`.

[13] I. Aitchison and A. Hey, "Gauge theories in particle physics: A practical introduction. Vol. 1: From relativistic quantum mechanics to QED". Institute of Physics Publishing Bristol and Philadelphia, 2003.

[14] I. Aitchison and A. Hey, "Gauge theories in particle physics: A practical introduction. Vol. 2: Non-Abelian gauge theories: QCD and the electroweak theory". Institute of Physics Publishing Bristol and Philadelphia, 2004.

[15] F. Englert and R. Brout, "Broken Symmetry and the Mass of Gauge Vector Mesons", *Phys. Rev. Lett.* **13** (1964) 321–322, `doi:10.1103/PhysRevLett.13.321`.

[16] P. W. Higgs, "Broken Symmetries and the Masses of Gauge Bosons", *Phys. Rev. Lett.* **13** (1964) 508–509, `doi:10.1103/PhysRevLett.13.508`.

[17] Quang Ho-Kim, Xuan-Yem Pham, "Elementary Particles and Their Interactions". Springer, Berlin Heidelberg, 1998.

[18] J. H. Kuhn and A. Penin, "Sudakov logarithms in electroweak processes", `arXiv:hep-ph/9906545`.

[19] A. Martin et al., "Parton distributions for the LHC", *Eur.Phys.J.* **C63** (2009)

189–285, `doi:10.1140/epjc/s10052-009-1072-5`, `arXiv:0901.0002`.

[20] M. Czakon and A. Mitov, "Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders", `arXiv:1112.5675`.

[21] M. Mangano, "Top mass and its interpretation: open TH issues - Anatomy of top production and decay, and impact on top-quark mass measurement", in *Talk given at TOP-LHCWG meeting*. 2012.

[22] CMS Collaboration, "Measurement of the top-quark mass in all-jets $t\bar{t}$ events in pp collisions at $\sqrt{s}$=7 TeV", *Eur.Phys.J.* **C74** (2014) 2758, `doi:10.1140/epjc/s10052-014-2758-x`, `arXiv:1307.4617`.

[23] CMS Collaboration, "Measurement of the top-quark mass in $t\bar{t}$ events with dilepton final states in pp collisions at $\sqrt{s} = 7$ TeV", *Eur.Phys.J.* **C72** (2012) 2202, `doi:10.1140/epjc/s10052-012-2202-z`, `arXiv:1209.2393`.

[24] CMS Collaboration, "Measurement of the top-quark mass in $t\bar{t}$ events with lepton+jets final states in pp collisions at $\sqrt{s} = 7$ TeV", *JHEP* **1212** (2012) 105, `doi:10.1007/JHEP12(2012)105`, `arXiv:1209.2319`.

[25] CMS Collaboration, "Measurement of the $t\bar{t}$ production cross section in the τ+jets channel in pp collisions at $\sqrt{s} = 7$ TeV", *The European Physical Journal C* **73** (2013), no. 4, `doi:10.1140/epjc/s10052-013-2386-x`.

[26] CMS Collaboration, "Standard Model Cross Section for CMS at 8 TeV", SM Cross Sections twiki page. `twiki.cern.ch/twiki/bin/viewauth/CMS/StandardModelCrossSectionsat8TeV`.

[27] O. Bruning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock, "LHC design report", *CERN Reports* **volume I, the LHC main ring** (2004).

[28] ALICE Collaboration, "The ALICE experiment at the CERN LHC", *Journal of Instrumentation* **3** (2008), no. 08, S08002.

[29] ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *Journal of Instrumentation* **3** (2008), no. 08, S08003.

[30] CMS Collaboration, "CMS Physics: Technical Design Report Volume 1: Detector

Performance and Software". Technical Design Report CMS. CERN, Geneva, 2006.

[31] CMS Collaboration, "CMS Physics Technical Design Report, Volume II: Physics Performance", *Journal of Physics G: Nuclear and Particle Physics* **34** (2007), no. 6, 995.

[32] LHCb Collaboration, "The LHCb Detector at the LHC", *Journal of Instrumentation* **3** (2008), no. 08, S08005.

[33] LHCf Collaboration, "The LHCf detector at the CERN Large Hadron Collider", *Journal of Instrumentation* **3** (2008), no. 08, S08006.

[34] TOTEM Collaboration, "The TOTEM Experiment at the CERN Large Hadron Collider", *Journal of Instrumentation* **3** (2008), no. 08, S08007.

[35] CMS Collaboration, "The CMS magnet project: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 1997.

[36] V. Karimaki et al., "The CMS tracker system project: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 1997.

[37] CMS Collaboration, "The CMS electromagnetic calorimeter project: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 1997.

[38] CMS Collaboration, "The CMS hadron calorimeter project: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 1997.

[39] CMS Collaboration, "The CMS muon project: Technical Design Report". Technical Design Report CMS. CERN, Geneva, 1997.

[40] A. Behrens and A. Maurisset, "CMS - YB+/-1/2. Position of YBs Targets at 0.0T and 3.8T", Technical Report CMS-SG-UR-0167, CERN, Geneva, (Feb, 2011).

[41] V. Gori, "The CMS High Level Trigger", in *Talk given at International Workshop on High Energy Physics (IHEP)*. 2013.

[42] CMS Collaboration, "High Level Trigger", HLT twiki page. `https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideHighLevelTrigger`.

[43] C. Kohler et al., "Construction process and read-out electronics of amorphous silicon position detectors for multipoint alignment monitoring", *Nucl.Instrum.Meth.* **A608** (2009) 55–67, `doi:10.1016/j.nima.2009.06.058`.

[44] CMS Collaboration, "COCOA: CMS object-oriented code for optical alignment", *CMS IN 2002/060* (2002) 193–194.

[45] CMS Collaboration, "CMS Object-oriented Code for Optical Alignment", COCOA page. `http://cmsdoc.cern.ch/cms/MUON/alignment/software/COCOA/doc/cocoa.html`.

[46] M. S. S. nudo, "CMS detector geometry reconstructed with the Link alignment system". PhD thesis, University of Cantabria, 2009.

[47] CMS Collaboration, "CMS, the magnet project: Technical design report", *CERN-LHCC-97-10* (1997).

[48] CMS Collaboration, "The CMS computing model", *Nucl.Phys.Proc.Suppl.* **172** (2007) 53–56, `doi:10.1016/j.nuclphysbps.2007.07.030`.

[49] M. Aderholz et al., "Models of networked analysis at regional centres for LHC experiments (MONARC). Phase 2 report",.

[50] I. Antcheva et al., "ROOT: A C++ framework for petabyte data storage, statistical analysis and visualization", *Comput.Phys.Commun.* **180** (2009) 2499–2512, `doi:10.1016/j.cpc.2009.08.005`.

[51] CMS-HCAL Collaboration, "Design, performance, and calibration of CMS forward calorimeter wedges", *Eur. Phys. J.* **C 53** (2008) 139–166, `doi:10.1140/epjc/s10052-007-0459-4`.

[52] 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).

[53] CMS Collaboration, "Luminosity Based on Pixel Cluster Counting - Summer 2013 Update", *CMS-PAS* **LUM-13-001** (2013).

[54] CMS Collaboration, "Track Reconstruction in the CMS tracker", *CMS-PAS* **TRK-09-001** (2009).

[55] W. Adam et al., "Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC", *Journal of Physics G: Nuclear and Particle Physics* **31** (2005), no. 9, N9.

[56] CMS Collaboration, "PF2PAT + PAT Tutorial", Work Book twiki page. `https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookPF2PAT`.

[57] CMS Collaboration, "Track and vertex reconstruction in CMS", *Nucl.Instrum.Meth.* **A582** (2007) 781–784, `doi:10.1016/j.nima.2007.07.091`.

[58] R. Fruhwirth, W. Waltenberger, and P. Vanlaer, "Adaptive Vertex Fitting", Technical Report CMS-NOTE-2007-008, CERN, Geneva, (Mar, 2007).

[59] CMS Collaboration, "Electromagnetic physics objects commissioning with first LHC data", Technical Report CMS-PAS-EGM-10-001, CERN, Geneva, (2010).

[60] CMS Collaboration, "Electron reconstruction and identification at sqrt(s) = 7 TeV", Technical Report CMS-PAS-EGM-10-004, CERN, Geneva, (2010).

[61] CMS Collaboration, "Commissioning of the Particle-Flow Event Reconstruction with the first LHC collisions recorded in the CMS detector", *CMS PAS* **PFT-2010/001** (2010).

[62] A. Calderón, S. Folgueras, and S. Xie, "Isolation Studies", in *Talk given at Muon DPG-PH Meeting*. 2012.

[63] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems", *Transactions of the ASME–Journal of Basic Engineering* **82** (1960), no. Series D, 35–45.

[64] R. Frühwirth, "Application of Kalman filtering to track and vertex fitting", *Nucl. Instrum. Methods Phys. Res., A* **262** (Jun, 1987) 444. 19 p.

[65] CMS , "Baseline muon selections", Muon selection twiki page. `https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideMuonId`.

[66] M. Cacciari, G. P. Salam, and G. Soyez, "The Anti-kT jet clustering algorithm", *JHEP* **0804** (2008) 063, `doi:10.1088/1126-6708/2008/04/063`, `arXiv:0802.1189`.

[67] S. Catani et al., "Longitudinally invariant k_t clustering algorithms for hadron hadron collisions", *Nucl.Phys.* **B406** (1993) 187–224, `doi:10.1016/0550-3213(93)90166-M`.

[68] G. P. Salam and G. Soyez, "A Practical Seedless Infrared-Safe Cone jet algorithm", *JHEP* **0705** (2007) 086, `doi:10.1088/1126-6708/2007/05/086`, `arXiv:0704.0292`.

[69] A. Jafari, "Measurement of the b-tagging efficiency in the CMS experiment with the first LHC collisions". PhD thesis, Vrije Universiteit Brussel, 2011.

[70] A. Rizzi, F. Palla, and G. Segneri, "Track impact parameter based b-tagging with CMS", Technical Report CMS-NOTE-2006-019, CERN, Geneva, (Jan, 2006).

[71] C. Weiser, "A Combined Secondary Vertex Based B-Tagging Algorithm in CMS", Technical Report CMS-NOTE-2006-014, CERN, Geneva, (Jan, 2006).

[72] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", *JINST* **8** (2013) P04013, `doi:10.1088/1748-0221/8/04/P04013`, `arXiv:1211.4462`.

[73] T. Speer et al., "Vertex Fitting in the CMS Tracker", Technical Report CMS-NOTE-2006-032, CERN, Geneva, (Feb, 2006).

[74] CMS Collaboration, "Results on b-tagging identification in 8 TeV pp collisions", *CMS-PAS* **DP-2013-005** (2013).

[75] CMS Collaboration, "Physics Analysis Toolkit (PAT)", SW Guide twiki page. `https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuidePAT`.

[76] CMS Collaboration, "Physics Analysis Toolkit (PAT)", Work Book twiki page. `https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookPAT`.

[77] CMS Collaboration, "PAT: the CMS Physics Analysis Toolkit", Technical Report CMS-CR-2010-203, CERN, Geneva, (Oct, 2010).

[78] CMS Collaboration, "Particle Flow in PAT (PF2PAT)", SW Guide twiki page. `https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuidePF2PAT`.

[79] J. Alwall et al., "MadGraph 5 : Going Beyond", *JHEP* **1106** (2011) 128,

`doi:10.1007/JHEP06(2011)128, arXiv:1106.0522.`

[80] T. Sjostrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 Physics and Manual", *JHEP* **0605** (2006) 026, `doi:10.1088/1126-6708/2006/05/026,` `arXiv:hep-ph/0603175.`

[81] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method", *JHEP* **0711** (2007) 070, `doi:10.1088/1126-6708/2007/11/070, arXiv:0709.2092.`

[82] CMS Collaboration, "TOP Reference Selections and Recommendations", Top Quark twiki page. `https://twiki.cern.ch/twiki/bin/view/CMS/TWikiTopRefEventSel.`

[83] CMS Collaboration, "Conversion Rejection", EgammaPOG twiki page. `https://twiki.cern.ch/twiki/bin/view/CMS/ConversionBackgroundRejection.`

[84] CMS Collaboration, "Electron Identification Using Multivariate Methods", Technical Report CMS-AN-11-413, CERN, Geneva, (2011).

[85] CMS Collaboration, "b Tag & Vertexing Physics Object Group", b-tag POG twiki page, (2012). `https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagPOG#Recommendation_for_b_c_tagging_a.`

[86] J. A. et al., "Measurement of the top dilepton cross section using b-tagging at $\sqrt{s} = 7$ TeV with 2.2 fb^{-1} in pp collisions", *Analysis Note* **AN-11-477** (2011).

[87] H. Bakhshian et al., "Computing the contamination from fakes in leptonic final states.", *CMS AN-2010/261* (2010).

[88] H. Bakhshian et al., "Lepton fake rates in dilepton final states.", *CMS AN-2010/397* (2010).

[89] L. Lyons, D. Gibaut, and P. Clifford, "How to combine correlated estimates of a single physical quantity", *Nucl. Instrum. Meth.* **A270-110** (1988) `arXiv:10.1016/0168-9002(88)90018-6.`

[90] CMS Collaboration, "Lepton energy scale", Systematic Uncertainties twiki page. `https://twiki.cern.ch/twiki/bin/viewauth/CMS/TWikiTopRefSyst.`

[91] M. D. et al., "Updated measurements of the new Higgs-like boson at 125 GeV in the two photon decay channel", *Analysis Note* **AN-12-374** (2012).

[92] CMS Collaboration, "Jet Energy Corrections: Official Software Tools for applying JEC Corrections and Uncertainties", Work Book twiki page. `https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookJetEnergyCorrections`.

[93] CMS Collaboration, "Jet energy scale", Systematic uncertainties twiki page. `https://twiki.cern.ch/twiki/bin/viewauth/CMS/TWikiTopRefSyst#Jet_energy_scale`.

[94] CMS Collaboration, "Jet energy resolution", Systematic uncertainties twiki page. `https://twiki.cern.ch/twiki/bin/viewauth/CMS/TWikiTopRefSyst#Jet_energy_resolution`.

[95] M. L. Mangano et al., "Matching matrix elements and shower evolution for top-quark production in hadronic collisions", *JHEP* **0701** (2007) 013, `doi:10.1088/1126-6708/2007/01/013`, `arXiv:hep-ph/0611129`.

[96] CMS Collaboration, "Determination of the top-quark pole mass and strong coupling constant from the $t\bar{t}$ production cross section in pp collisions at $\sqrt{s} = 7$ TeV", *Phys.Lett.* **B728** (2014) 496–517, `doi:10.1016/j.physletb.2013.12.009`, `arXiv:1307.1907`.

[97] M. Czakon et al., "Constraints on the gluon PDF from top quark pair production at hadron colliders", *JHEP* **1307** (2013) 167, `doi:10.1007/JHEP07(2013)167`, `arXiv:1303.7215`.

[98] M. Botje et al., "The PDF4LHC Working Group Interim Recommendations", `arXiv:1101.0538`.

[99] W. Andrews et al., "A method to measure the contribution of DY$\rightarrow \ell\ell$ to a dilepton + MET selection.", *CMS AN-2009/023* (2009).

INDEX