UNIVERSITÀ DEGLI STUDI DI TRIESTE

XXVII CICLO DEL DOTTORATO DI RICERCA IN FISICA

Measurement of the production cross section of a Z boson in association with exactly one or at least two b jets with the CMS experiment at LHC.

SSD: FIS/04

Dottorando:
Chiara La Licata

Coordinatore:
Prof. Paolo Camerini

Supervisore di Tesi:
Prof. Giuseppe Della Ricca

Co-Supervisore di Tesi:
Dr. Vieri Candelise

Anno Accademico 2013/2014

Contents

Introduction

The Standard Model of elementary particles is one of the most successful theories of modern physics. It provides a quite satisfactory description of the properties of elementary particles and their interactions. Extensive consistency and precision tests have been performed, providing stringent constraints on the Standard Model parameters. Over the past 30 years a very extended experimental program addressed to the search of the Higgs boson was carried out in particle colliders. The discovery of a Higgs-like particle announced by the ATLAS and the CMS experiments in 2012 has solved one of the main open topics of the Standard Model: the mechanism that gives mass to the fermions and to the bosons that mediate the weak interactions. However there are many aspects that do not still find an adequate explanation inside this theory. The main aim of the LHC experiments is to answer these open questions and explore the predictions of physics beyond the Standard Model. Nevertheless the search of new physics would not be possible without a solid understanding of the Standard Model processes that could represent important backgrounds. Moreover, the accurate understanding of the Standard Model is not only crucial for searches of unknown particles and phenomena but also to test predictions of perturbative Quantum Chromodynamics calculations. The study of the production of a Z boson in association with one or more heavy quark jets (Z+b) in hadron collisions gives the opportunity to explore a wide physics phenomenology and provides a unique test to probe the perturbative Quantum Chromodynamics approximation. In this thesis two final states are analysed: the associated production of the Z boson with exactly one (Z+1b) or at least two (Z+2b) b jets. Data from proton-proton collisions at a centre of mass energy of $\sqrt{s} = 8$ TeV, delivered by LHC and recorded with the CMS experiment in 2012 (corresponding to an integrated luminosity of about 19.8 fb^{-1}), is used to perform the measurement of differential cross sections as function of different observables. This thesis is organized into six Chapters. In the first Chapter a brief description of the Standard Model is given, introducing the electroweak interaction, the Higgs mechanism and the strong interaction. This chapter ends with a review of the opened questions of the model. In the second Chapter the importance of the measurement of the production of a Z boson in association with exactly one or at least two b jets is discussed. In the first part the two different approaches (four-flavor, five-flavor schemes), adopted to describe the processes in which a b quark is involved, are described while the second part is mainly focused on the Z+b final state signatures. A general view of the CMS experiment is given in Chapter 3 where a short description of all the subdetectors is provided. Chapter 4 is dedicated to the description of the specific algorithms applied in the CMS experiment in order to reconstruct all the interested objects: electrons, muons, jets and missing energy. The jet reconstruction algorithms of interest are also reviewed and the relevant b tagging algorithms are described. More details are given on the full event reconstruction technique employed at CMS, known as the "Particle Flow", that allows a significant improvement in jet reconstruction performance over more traditional, calorimeter-based approaches. The analysis event selection is presented in Chapter 5, including a description of the data samples used both for the signal and for the Monte Carlo samples, the trigger and the preselection requirements. The different background contributions are all extracted through Monte Carlo simulations, except in the case of the top-antitop pair background whose evaluation has been performed using a data driven approach and a section is dedicated to the explanation of the procedure adopted. A particular attention has been used to extract the event b purity estimated simultaneously in the $Z + 1b$ and $Z + 2b$ samples, in order to take into account possible inefficiencies in the tagging of jets and so migrations of events from one sample to the other. Chapter 6 details the unfolding methods used in order to deconvolve the measured quantities from the detector effects. Possible sources of systematic uncertainties are investigated in this Chapter, showing for each of them the strategy adopted to estimate the final contribution. Finally the differential cross sections for the Z+1b jet and Z+2b jets are shown. These results include the comparison between experimental data and different QCD theoretical predictions with several implementations, using leading order, next-to-leading order, 4-flavor scheme, and 5-flavor scheme approaches.

Chapter 1

Standard Model

At the beginning of the 20th century the experimental evidences of protons, electrons and neutrons existence allowed to explain the composition of atoms and the periodic table of the chemical elements.

Important progresses have been made in particle physics during the last century. Moreover following the experimental evidences of new particles unveiled, in particular after the introduction of more powerful particle accelerators which allowed matter to be probed at smaller distances, the theoretical developments deeply changed the description of the fundamental matter constituents. The present knowledge about fundamental particles and their interactions is described by a quantum theory known as Standard Model (SM) [1–3]. The Standard Model is the quantum theory that describes particles and interactions in terms of the quantum field dynamic. The first step towards the Standard Model are due to Glashow [1] Salam [2] and Weinberg [3] who found a way to combine the electromagnetic and weak interactions. Later, the inclusion of the Higgs mechanism into electroweak theory gave it its modern form. Fundamental experimental confirmations such as the discovery of the W and Z particles at CERN by Rubbia and Van Der Meer at the beginning of the 80's established the SM as the reference theory for particle physics description, confirming the electroweak unification theory. Extensive consistency and precision tests were performed providing stringent constraints [4, 5] on the Standard Model parameters over a wide range of energies.

In this chapter the Standard Model will be introduced through the electroweak and strong lagrangian description.

1.1 Particles and force carriers

The Standard Model (SM) is the most comprehensive and wide theory available and, in the last forty years, many experimental evidences have confirmed the SM as the reference theory for the description of the elementary particles and their interactions. The fundamental constituents of matter interact via

different forces: the electromagnetic, the weak, the strong and gravitational forces. The SM theory has been built to describe all these interactions except the gravitational one. However, the tiny value of its coupling constant compared to those involved in the other three interactions makes its effects completely negligible in particle physics experiments. The different elementary particles described by the SM are collected in fig. 1.1 and can be divided into two different categories: fermions, particles of spin-1/2 (in units of $\hbar$), and bosons, particles with integer spin. The fermions are the two types of matter units and are structureless at the smallest distances currently probed by the highest energy experiments. This type of particles can be divided into quarks and leptons according to their interactions. Quarks can interact through the electromagnetic, weak and strong interactions even if the strong interaction is the dominant one. Six different flavours predicted by the model have been experimentally discovered: up, down, strange, charm, bottom and top.
Leptons are grouped into three different families each of them comprising a charged lepton (electron, muon, tau) and the corresponding neutrino (ν_e, ν_μ and ν_τ). The charged lepton can interact electromagnetically and weakly while in the neutrino interactions only the weak force is involved.

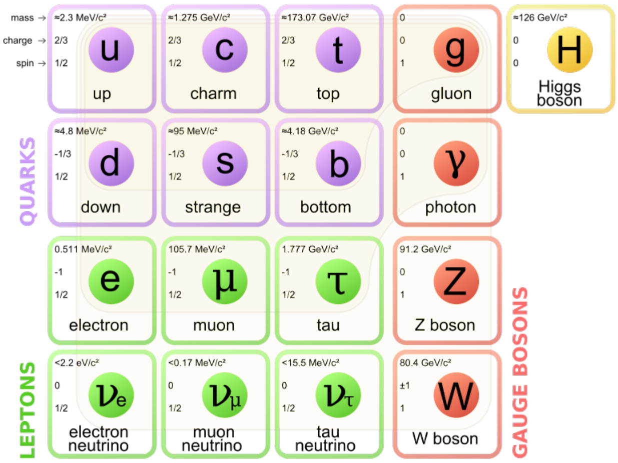


Figure 1.1: Schema of the elementary particles as described by the Standard Model: three generations of up- and down-type quarks, three generations of charged and neutral leptons, the gauge bosons (fourth column), force carriers of the electromagnetic interaction and the Higgs boson (fifth column).

The three interactions treated inside the SM are mediated by the force-carrier gauge bosons: the electromagnetic force by the photon γ, the weak force by the $W^{\pm}$ and Z, and the strong force by eight gluons. Electromagnetic and weak interactions are unified in the electroweak model that describes the two different manifestation of the same interaction $SU(2)_L \otimes U(1)_Y$. The carriers of the strong interaction are massless gauge bosons while the carriers of the electroweak interactions are both massless (γ) and massive ($M_W = 80.385 \pm 0.015$ GeV [6] and $M_Z = 91.1876 \pm 0.0021$ GeV [6]). The fact that both the charged leptons and some of the gauge bosons are not massless implies the loss of the gauge invariance of the theory. The SM, as any other Gauge Quantum Field Theory, is built under the construction principle of gauge invariance where the exchanged field quanta with spin one defines the gauge particle which must be massless in order to preserve it. For this reason it is necessary to introduce the masses by a mechanism, and it is done in the most simple way by the Higgs mechanism [7] [8], that leads to the prediction of a new massive scalar particle, the Higgs boson particle. A particle compatible with the Higgs boson has been discovered at LHC by ATLAS [9] and CMS [10] experiments in 2012.

1.2 Electroweak interaction

The weak and electromagnetic interactions can be seen as two different aspects of the same fundamental interaction and a unified description is provided by the electroweak model. The weak interaction is described in the Standard Model as a theory based on a $SU(2)_L \otimes U(1)_Y$ gauge invariant lagrangian. $SU(2)_L$ is the weak isospin group which is non abelian and has three generators $T_{1,2,3} = \tau_{1,2,3}$ where the τ_i are the Pauli matrices. $U(1)_Y$ is the weak hypercharge group which is abelian and has one generator $Y/2$. The electromagnetic group is a subgroup of the electroweak group that is $U(1)_{em} \subset SU(2)_L \otimes U(1)_Y$ and the corresponding generator is a combination of the third component of the weak isospin (T_3) and the weak hypercharge (Y):

$$Q = T_3 + \frac{Y}{2}. \tag{1.1}$$

The experiment conducted by Madame Wu [11] was the first experimental evidence of parity violation for weak charged current interactions confirming the prediction of Lee and Yang. It was observed that left-handed particles or right-handed antiparticles are involved in weak charge current interaction while right-handed particles and left-handed antiparticles are absent with a clear violation of parity invariance.

The phenomenological aspects of the weak interactions between leptons are described by the Fermi theory (interaction of the four fermions [12]) in which the Hamiltonian density of the interaction is given by the product of two currents J_μ:

$$H_I(x) = \frac{G_F}{\sqrt{2}} J_\mu^\dagger J^\mu, \tag{1.2}$$

where G_F is the Fermi's constant [1] and J_μ the charge currents whose expres-

---

[1] Experimentally the value of G_F is extracted from muon decay ($\mu \to \nu\bar{\nu}e$). The muon

sions have to take into account of the parity violation so they assume the following form:

$$J_\mu \equiv J_\mu^+ = \bar{u_\nu}\gamma_\nu\frac{1}{2}(1-\gamma^5)u_e = \nu_L\bar{\gamma_\mu}e_L, \tag{1.3}$$

$$J_\mu^\dagger \equiv J_\mu^- = \bar{u_e}\gamma_\nu\frac{1}{2}(1-\gamma^5)u_\nu = e_L\bar{\gamma_\mu}\nu_L, \tag{1.4}$$

where u_ν and u_e are the Dirac spinors respectively for the neutrino and the electron. The subscript L in the eq. 1.3 means that the left-handed spinors are involved and reminds the V-A structure of the charged currents evident in the mixture of a vector term γ_μ and the axial-vector term $\gamma^5\gamma^\mu$. Using the spinors with an established helicity (a left-handed doublet L and a right-handed singlet R):

$$\boldsymbol{\psi_L} = \begin{pmatrix} l \\ \nu_l \end{pmatrix}_L,$$

$$\boldsymbol{\psi_R} = \begin{pmatrix} l \end{pmatrix}_R,$$

with $l = e, \nu, \tau$. The charged currents can be written in a two dimensional form:

$$J_\mu^+ = \bar{\psi}_L\gamma_\nu\tau_+\psi_L, \tag{1.5}$$

$$J_\mu^- = \bar{\psi}_L\gamma_\nu\tau_-\psi_L. \tag{1.6}$$

In this expression the “step-up” and “step-down” operators $\tau_\pm = \frac{1}{2}(\tau_1 \pm \tau_2)$ are introduced where τ_i (with i=1,2,3) are the Pauli spin matrices. In order to complete the SU(2) invariance of the theory we can introduce the neutral current of the form:

$$J_\mu^3 = \bar{\psi}_L\gamma_\nu\frac{1}{2}\tau_3\psi_L. \tag{1.7}$$

The J_μ^3 current cannot be identified directly with the weak neutral current (J_μ^{nc}) because it involves only left-handed fermions and it cannot be seen as the the electromagnetic current (J_μ^{em}) that has to have both the right and left handed components and does not couple with the neutrino. In order to maintain the $SU(2)_L$ symmetry and to include in this picture both the J_μ^{nc} and the J_μ^{em} the currents are expressed as linear combinations of J_μ^3 and weak hypercharge current J_μ^Y:

$$J_\mu^{em} = J_\mu^3 + \frac{1}{2}J_\mu^Y \tag{1.8}$$

with J_μ^Y given by:

---

lifetime is only a function of the muon mass and G_F: $\Gamma = \frac{G_F^2 m_\mu^5}{192\pi^3}$. From this measurement the value obtained for this coupling constant is $G_F = 1.16639 \pm 0.0001 \times 10^{-5}GeV^{-2}$ [6].

$$J^Y_\mu = \bar{\Psi}\gamma_\mu Y \Psi. \tag{1.9}$$

The hypercharge operator Y is the generator of the symmetry group $U(1)_Y$. The electromagnetic interaction is therefore incorporated by enlarging the symmetry group to $SU(2)_L \otimes U(1)_Y$. Also the neutral weak current can be expressed as a linear combination of the J^3_μ and J^Y_μ currents.

In order to complete the unification of the weak and electromagnetic interactions, the current-current structure must be modified into an effective interaction in which massive vector bosons are the force carriers. For this pourpose we can define the weak isospin current:

$$J^i_\mu = \bar{\Psi}_L \gamma_\mu \frac{1}{2}\tau_i \Psi_L, \tag{1.10}$$

and we can introduce an isotriplet of vector fields W^i_μ coupled with strenght g to J^i_μ and a single vector field B_μ coupled to the J^Y_μ current with strenght $g^{'}/2$.
Electroweak interaction is then:

$$-ig(J^i)^\mu W^i_\mu - i\frac{g^{'}}{2}(J^Y)^\mu B_\mu, \tag{1.11}$$

The gauge fields describing massive charge bosons $W^\pm$ are:

$$W^\pm = \frac{1}{\sqrt{2}}(W^1_\pm \mp W^2_\pm), \tag{1.12}$$

while the physical state for the the neutral gauge bosons is given by the combination of W^3_μ and B_μ:

$$A_\mu = B_\mu \cos(\theta_w) + W^3_\mu \sin(\theta_w), \tag{1.13}$$

$$Z_\mu = -B_\mu \sin(\theta_w) + W^3_\mu \cos(\theta_w), \tag{1.14}$$

where θ_w, called Weinberg angle, is an important SM parameter whose value ($\sin^2(\theta_w) = 0.23120 \pm 0.00015$) has been measured using the combination of the results of SLD and LEP experiments [13].

1.2.1 Electroweak lagrangian

The Lagrangian density which describes the unified electroweak theory can be divided into two contributions, a part containing the interaction terms and a part containing kinematic terms. The introduction of the covariant derivative D_μ (eq. 1.22) replacing the ordinary derivative ∂_μ inside the kinematic term of the lagrangian ($\mathcal{L} = \bar{\Psi} i \gamma_\mu \partial_\mu \Psi$) guarantees the invariance under local tranformation and introduces the interaction term ($\mathcal{L}_\mathcal{I}$). The lagrangian density describing the interaction of the vector fields with the fermions in a non-Abelian gauge theory, Yang-Mills theory, assumes the form:

$$\mathcal{L} = \bar{\Psi} i \gamma^\mu D_\mu \Psi = \bar{\Psi} i \gamma^\mu \partial_\mu \Psi + \mathcal{L}_\mathcal{I} \tag{1.15}$$

where $\mathcal{L}_\mathcal{I}$ is the sum of three contributions: the electromagnetic (1.16), charge weak current (1.17) and neutral weak current (1.18) terms:

$$\mathcal{L}_{em} = -ieJ^{em}_\mu A^\mu, \tag{1.16}$$

$$\mathcal{L}_{cc} = -1\frac{g}{\sqrt{2}}((J^+)^\mu W^+_\mu + (J^-)^\mu W^-_\mu), \tag{1.17}$$

$$\mathcal{L}_{nc} = -i\frac{g}{\cos\theta_W} \cdot ((J^3)^\mu - \sin^2\theta_W (J^{em})^\mu) Z_\mu. \tag{1.18}$$

Introducing the following kinematic terms of the Yang-Mills fields $\vec{W}_\mu$ and B_μ:

$$E^\alpha_{\mu\nu} = \partial_\mu W^\alpha_\mu - \partial_\nu W^\alpha_\nu - g\epsilon^{\alpha\beta\gamma} W^\beta_\mu W^\gamma_\nu \tag{1.19}$$

$$F_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \tag{1.20}$$

The lagrangian density that expresses the unified $SU(2)_L \otimes U(1)_Y$ electroweak theory includes the interaction term and the kinematic term:

$$\mathcal{L}_{EWK} = \sum_f \bar{\Psi} i \gamma^\mu D_\mu \Psi - \frac{1}{4} W_{\mu\nu} W^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} \tag{1.21}$$

where the expression for the covariant derivative is:

$$D_\mu \Psi = [\partial_\mu + ig\vec{W}_\mu \cdot \frac{\vec{\tau}}{2} - i\frac{g^{'}}{2} Y B_\mu]\Psi \tag{1.22}$$

where g is the $SU(2)_L$ gauge coupling and $g^{'}$ is the $U(1)_Y$ gauge coupling. The relation between the electromagnetic coupling e and these two couplings g and $g^{'}$ is a consequence of $U(1)_{em}$ being a subgroup of $SU(2)_L \otimes U(1)_Y$:

$$g = \frac{e}{\sin\theta_W} \qquad g^{'} = \frac{e}{\cos\theta_W} \tag{1.23}$$

1.2.2 Cabibbo - Kobaiashi - Maskawa matrix

The electroweak interaction model described in the previous section, based on the gauge group $SU(2)_L \otimes U(1)_Y$, can be completed including also the quark fields. As in the lepton case the quark fields can be divided into three families (Q^α with $\alpha = 1, 2, 3$) where the left handed components of the fields are transformed as $SU(2)_L$ doublets and the right handed fields are singlets of the group:

$$\boldsymbol{Q_L} = \begin{pmatrix} u \\ d \end{pmatrix}_L$$

$$(u)_R \quad ; \quad (d)_R$$

with u = u, c, t and d = d, s, b.

The mass eigenstates for left handed doublets do not coincide with the eigenstates of the weak interaction of quarks. The quark fields enter into the charge current of the Standard Model in the form of mixed combinations. Since each electroweak eigenstate is a linear combination of the mass eigenstates, the transformation can be obtained according to the Cabibbo-Kobayashi-Maskawa (CKM) formalism [14] in which the quark mixing is parameterised by three real parameters (mixing angles) and one complex phase.

The CKM matrix transforms the mass eigenstates (d, s, and b) into weak eigenstates ($d^{'}, s^{'}$ and $b^{'}$):

$$\begin{pmatrix} d^{'} \\ s^{'} \\ b^{'} \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{1.24}$$

The quark flavours can be transformed and also the generation can change; up-type quarks convert into down-type quarks and vice versa via flavour-changing charged currents, based on the coupling to $W^{\pm}$. The transition probability between different quark flavours is described by the Cabibbo-Kobayashi–Maskawa matrix elements, and is proportional to $\mid V_{qq'} \mid^2$. Flavour-changing neutral currents with a Z exchange, not expected by the theory, have not been observed. The charged current interaction Lagrangian for quarks can be expressed in the following form in terms of the mass eigenstates:

$$\mathcal{L}^{CC} = -\frac{g}{\sqrt{2}}(\bar{u}_L, \bar{c}_L, \bar{t}_L)\gamma^\mu V_{CKM} \begin{pmatrix} d \\ s \\ b \end{pmatrix} W_\mu^\dagger + h.c. \tag{1.25}$$

In the Standard Model the CKM matrix is assumed to be unitary, so the parameters are not all indipendent because of the unitarity relation:

$$\sum_k V_{ik} V_{jk}^* = \delta_{ij}. \tag{1.26}$$

A general $n \times n$ unitary matrix ($VV\dagger = I$) requires n^2 real parameters to be specified but $2n-1$ of these parameters are not physically significant and can be eliminated redefining the phases of the quark mass eigenstates. Hence there are $(n-1)^2$ physical parameters left. A unitary matrix is also orthogonal, and as such it contains $n(n-1)/2$ parameters corresponding to the independent rotation angles between the n basis vectors; thus the remaining $(n-1)(n-2)/2$ parameters must be complex phases. For $n = 2$, i. e. two quark families there is just one mixing angle, the Cabibbo angle θ_c [15]. For $n = 3$ there are instead four physical parameters, namely three Euler angles and one phase, the latter responsible for all CP violation in meson decays in the Standard Model.
In order to determine the magnitudes V_{ij} of the elements of the CKM matrix, several processes have been studied, each one sensitive to a specific element. Combining the experimental information with the CKM unitarity condition, assuming only three generations, limits for the V_{ij} elements can be obtained. The current knowledge of the CKM matrix elements shows that transitions within the same generation are governed by the CKM matrix elements of $O(1)$, those between the first and the second generation are suppressed by CKM factors of $O(10^{-1})$, those between the second and the third generation are suppressed by $O(10^{-2})$, and the transitions between the first and the third generation are even more suppressed by CKM factors of $O(10^{-3})$.

1.2.3 Electroweak symmetry breaking

As mentioned before, the model proposed by Weinberg and Salam predicts massless particles in contradiction of what is observed so it is necessary to introduce masses by a mechanism. In fact the introduction of explicit mass terms for fermions and gauge bosons inside the electroweak lagrangian would violate the gauge symmetry leading to a non renormalizable theory loosing the lagrangian predictive power.
In order to preserve the gauge invariance, the observed fermion and boson masses, need to be explained in a different way: introducing the spontaneous symmetry breaking mechanism of $SU(2)_L \otimes U(1)_Y$ as theorized firstly by Higgs [7]. The simplest choice to perform the breakage of the symmetry is the introduction of one complex scalar $SU(2)$ doublet:

$$\mathbf{\Phi} = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}_L .$$

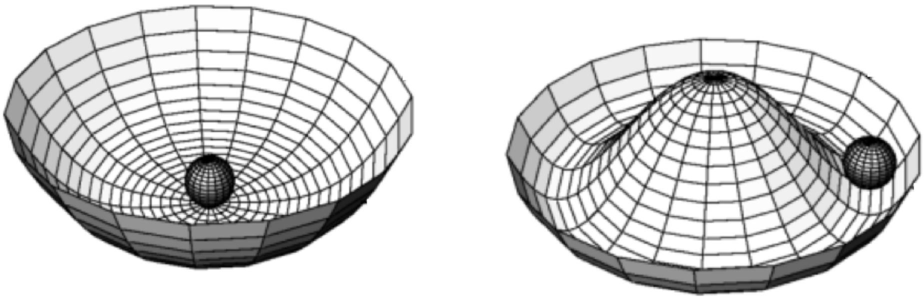
In order to introduce the Higgs mechanism within the Standard Model a Lagrangian is built including to the $\mathcal{L}_{EWK}$ the potential $V(\Phi)$, the proper covariant derivatives of Φ and the Yukawa interaction between the Higgs doublet and the fermion fields. So the lagrangian assumes the following form:

$$\mathcal{L} = \mathcal{L}_{EWK} + D_\mu \Phi D^\mu \Phi - V(\Phi), \tag{1.27}$$

where the potential $V(\Phi)$ is a quartic function of the scalar field:

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda(\Phi^\dagger \Phi)^2. \tag{1.28}$$

The fundamental state of the system is obtained through the minimization of the potential $V(\Phi)$. If $\mu^2 > 0$ (fig. 1.2(a)) the ground state is $\phi = 0$ and it preserves all the symmetries of the lagrangian. The condition on the phase $\mu^2 < 0$ gives a degenerate fundamental state as shown in fig. 1.2(b).



(a) The potential with $\mu^2 > 0$.

(b) The potential with $\mu^2 < 0$.

Figure 1.2: Higgs Potential corresponding to eq. (1.28) in the complex plane.

If one of the component of Φ takes a non void vacuum expectation value, a fundamental state can be chosen with form:

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

with $v \equiv \frac{\mu^2}{\lambda}$. In this case ($\mu^2 < 0$) the infinite degenerate vacua (minima) have the same $|< \Phi >|$ but different complex phases. The choice of a particular vacuum (complex phase) is what generates the spontaneous breaking of U(1).

The introduction of a complex scalar field in the electroweak lagrangian determines the breakage of the symmetry leaving the electromagnetic current conserved in the scheme: $SU(2)_L \otimes U(1)_Y \longrightarrow U(1)_{em}$.
The electroweak Lagrangian after the introduction of the electroweak symmetry breaking mechanism can be written as:

$$\mathcal{L}_{EWK} = \mathcal{L}_K + \mathcal{L}_N + \mathcal{L}_C + \mathcal{L}_{WWV} + \mathcal{L}_{WWVV} + \mathcal{L}_H + \mathcal{L}_{HV} + \mathcal{L}_Y. \tag{1.29}$$

Figure 1.3 shows the corresponding tree level Feynman graphs to all the terms of the lagrangian (eq. 1.29). $\mathcal{L}_K$ represents the kinetic term describing the free movement of all the fermions and bosons. The term $\mathcal{L}_N$ describes the interaction of the photon and the Z bosons with the fermions, while $\mathcal{L}_C$ is due to the interaction of the W boson with left-handed particles and right-handed antiparticles. The terms $\mathcal{L}_{WWV}$ and $\mathcal{L}_{WWVV}$ represent respectively three point and four point gauge bosons self-interactions. The terms with $\mathcal{L}_H$ take into account the Higgs self-interaction, while $\mathcal{L}_{HV}$ describes the interaction of the Higgs boson with the gauge boson. Finally the last term $\mathcal{L}_Y$ characterises the Yukawa couplings between the massive fermions and the Higgs field.

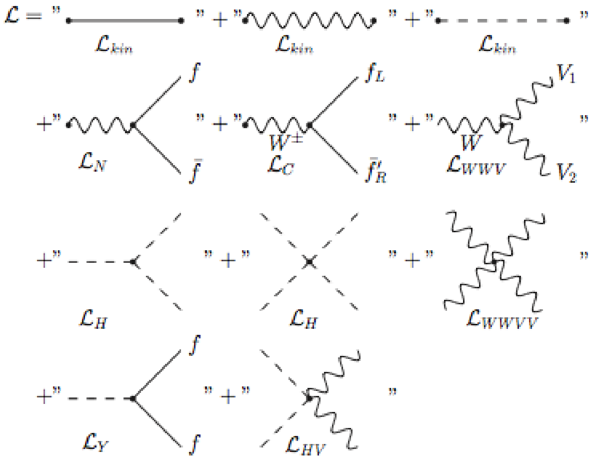


Figure 1.3: Scheme of the electroweak lagrangian as tree level Feynman diagrams.

In this framework the charged $W^\pm$ and the neutral Z boson masses that are originated by the introduction of the Higgs field in the lagrangian are related to the coupling g and to the parameter v by the following relations:

$$M_W = \frac{1}{2}gv; \qquad \frac{M_W}{M_Z} = \cos\theta_W \tag{1.30}$$

while the photon remains massless. Eq. 1.30 shows, also, the dependence of the vector boson mass with the Weinberg angle θ_w.
The Higgs mass only depends on the expectation value of the vacuum v:

$$M_H = \sqrt{\lambda}v. \tag{1.31}$$

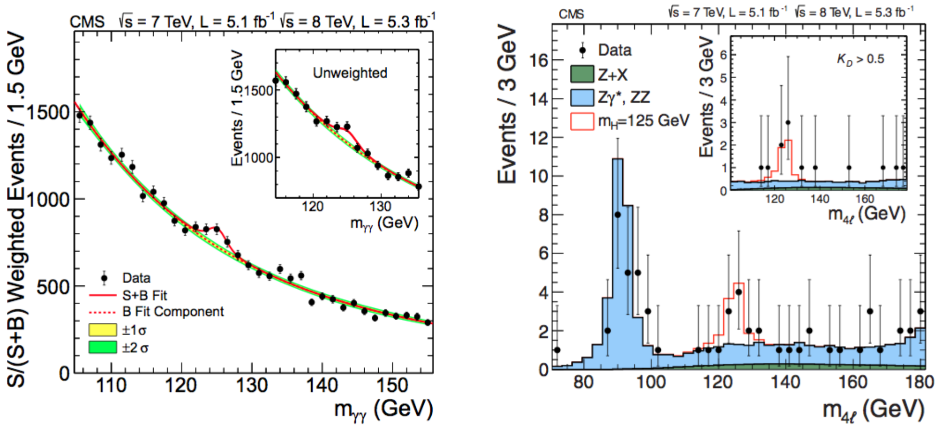
The Higgs mass cannot be predicted because the theory does not provide a value for λv. Introducing an interaction term of the fermions with the Higgs field it is possible to assign masses to the fermions. This can be done via a Yukawa coupling with coupling constant g_f, obtaining the lagrangian term:

$$\mathcal{L} = -g_f(\bar{\Psi}_L\phi\Psi_R + \bar{\Psi}_R\phi^\dagger\Psi_L). \tag{1.32}$$

The value of the coupling constant g_f is determined by the fermion masses: $g_f = \frac{m_f}{v}$.

The Higgs Mechanism leads to the prediction of a new scalar particle, called the Higgs boson. For over 50 years there have not been experimental evidences of this particle and the search of it has involved the efforts of different experiments. At the LEP collider at CERN, operating at an energy mass center around the value of the Z mass in the first phase increased to 210 GeV in a second phase, a limit to 114 GeV with a confidence level (CL) of 95% has been firstly set [16]. The latest results in the experiments CDF and D0 at Tevatron shows a significant excess of events in the mass range between 115 and 140 GeV with a local significance of 3σ at m_H =125 GeV [17]. The experimental discovery of the Higgs boson by the collaborations of ATLAS [9] and CMS [10] at the Large Hadron Collider (LHC), was announced in an international open web-conference on the 4th of July 2012. Recently, on the 8th of October 2013, Peter Higgs and Francois Englert received the Physics Nobel Prize 2013 for the proposal of this mass generation mechanism and for the prediction of the Higgs boson particle.

The CMS and ATLAS experiments have been specifically designed in order to perform the SM Higgs boson search. Different decay channels have been exploited for the Higgs search even if two of them have been particularly interesting for the discovery: $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ \rightarrow l^+l^-l^+l^-$.
Despite the small expected signal rate, the Higgs boson decay into a pair of photons, $H \rightarrow \gamma\gamma$, is a very sensitive channel in the search for a low-mass SM Higgs boson because of the distinctive signature that is a narrow peak of width determined by the instrumental resolution.
The Higgs boson decay $H \rightarrow ZZ \rightarrow l^+l^-l^+l^-$ is often referred to as the "golden channel" because of the good invariant mass resolution, the well controlled background and the clear signature. The traditional search strategy using the golden channel focuses on measuring the invariant mass spectrum of the four leptons. The great interest in this channel is also due to the possibility it gives to measure other features of the Higgs boson, for example the spin and the parity, given that four-momenta of all decay products can be reconstructed with sufficient resolution.
Fig. 1.4(a) and 1.4(b) present the results obtained in the two main channels during the first years of data collection at the LHC, shown to announce the discovery of the new particle. The CMS experiment, using the entire dataset collected during the 2011 and 2012 LHC running periods, has observed an excess of events at a mass close to 125 GeV in all the five channels analysed: $\gamma\gamma$(with a local significance of 5.7 σ [18]), ZZ (with a local significance of 6.8 σ [19]), bb(with a local significance of 2.1 σ [20]), WW(with a local significance of 4.3 σ [21]), $\tau\tau$(with a local significance of 3 σ [22]).



(a) Di-photon invariant mass $m_{\gamma\gamma}$ distribution with each event weighted by the signal to background ratio S/(S + B) value. The lines represent the fitted background and signal while the colored bands represent the ±1 and ±2 standard deviation uncertainties in the background estimate.

(b) Distribution of the four-lepton invariant mass m_{4l} for the $ZZ \rightarrow 4l$ analysis. The points represent the data, the filled histograms represent the background, and the open histogram shows the signal expectation for a Higgs boson of mass $m_H = 125$ GeV, added to the background expectation.

Figure 1.4: CMS results of the Higgs-like boson discovery. The measurements are presented for the sample of the data collected at a center-of-mass energy of 7 TeV and 8 TeV [10].

1.3 Strong interactions

In order to complete the Standard Model description it is necessary to consider also the strong interaction. The field theory of strong interactions is known as Quantum Chromodinamics (QCD) and is a gauge theory based on an unbroken $SU(3)$ symmetry.

The historical foundation of QCD dates back to the first scattering experiments that revealed that a new kind of interaction between protons and neutrons could exist and that it acted only over very short distances. The picture of the world of elementary particles started to change when many different particles were discovered and different schemes tried to explain the observed regularities in the particle zoo. Gell-Mann [23] and Ne'eman proposed a model that showed that the known hadrons could be classified as multiplets of the unitary Lie group $SU(3)$ with the isospin (I) and the hypercharge (Y=B+S) as relevant quantum numbers. When the vacant positions in the multiplets were filled by new particles, it was clear that the model was not a simple mathematical abstraction. Moreover Gell-Mann and Zweig postulated the existence of a set of three particles corresponding to the fundamental representation of $SU(3)$ that were called quarks. The quarks would constitute the basic building

blocks of all hadronic matter carrying a fractional electric charge. So all the observed particles could be described through the properties of the symmetry group $SU(3)$.
With deep inelastic lepton-nucleon scattering experiments that involved a momentum per particle greater than 1 GeV, the proton structure could be resolved for the first time. The first evidence of a sub-structure in the proton gave the support to the idea of proton as a bound state of pointlike charged quarks.
The discover of new particles, such as $|\Delta^{++}>=|uuu>$ or $|\Omega^{-}>=|sss>$, seemed to be in contradiction with the Pauli exclusion principle because these particles are bounded states of three identical quarks. To explain the existence of these particles avoiding the violation of the Pauli principle a new quantum number (colour) was proposed by Han and Nambu [24] in 1965. So it was assumed that the quarks carry an additional degree of freedom which can take three distinct values. Experimental test of the existence of this new quantum number came from the comparison of the hadron production in $e^{+}e^{-}$ annihilation and the creation of muon pairs with the same initial state. Taking the ratio R (eq. 1.33) between the two cross sections, everything except the coupling strengths cancels. Thus the ratio reduces to a direct measure of the sum of the squares of the quark charges allowing to determine the number of quark colours (N_c).

$$R = \frac{\sigma(e^{+}e^{-} \to hadrons)}{\sigma(e^{+}e^{-} \to \mu^{+}\mu^{-})} = N_c \sum_q e_q^2 = N_c \frac{11}{9}. \tag{1.33}$$

It was then assumed that colour is a charge-like quantum number conceptually similar to the electric charge or weak isospin, a source of a colour field that glues the quarks together to form the observed hadrons. The quanta of the colour field were called gluons that couple only to color charge. After the results from deep inelastic scattering many evidences were now clear:

- hadrons were bound states of fractionally charged quarks
- quarks are spin-1/2 fermions carrying the colour charge
- besides quarks, additional partons in the nucleons exist.
- those partons (gluons) do not interact with electromagnetic or weak force.

The Quantum Chromodynamics was then postulated as a quantized gauge theory with a $SU(3)_c$ symmetry in 1973 by Fritzsch [25], Gross and Wilczec [26] and Weinberg [27]. The color symmetry implies new massless gauge bosons called gluons (the eight generators of SU(3)), which mediate the strong interaction.

1.3.1 The QCD lagrangian

The QCD is a gauge theory based on the group an $SU(3)_c$. For the quark Dirac spinors fields q_i^α (where α is the flavour index, while i is the colour index) it is possible to introduce the covariant derivative:

$$(D_\mu)_{ij} = \delta_{ij}\partial_\mu + ig_s A_\mu^a(x)T_{ij}^a \tag{1.34}$$

in which $A_\mu^a(x)$ (with a = 1,2...,8) are the eight Yang-Mills gluon fields and T_{ij}^a are 3×3 matrices, generators of the $SU(3)_c$ in the fundamental rappresentation and g_s is the coupling constant. The QCD Lagrangian can then be written in the form:

$$\mathcal{L}(x) = \sum_\alpha \bar{q}_i^\alpha(i\gamma_\mu D_\mu - m_\alpha)_{ij} q_j^\alpha(x)) - \frac{1}{4}F_{\mu\nu}^a(x)F^{a\mu\nu}(x), \tag{1.35}$$

where the field strenght tensor is:

$$F_{\mu\nu}^a(x) = \partial_\mu A_\nu^a(x) - \partial_\nu A_\mu^a(x) f^{abc} A_\mu^b(x) A_\nu^c(x), \tag{1.36}$$

with f^{abc} structure constants of the group defined through the commutation relations of the $SU(3)$ generators: $[T^a, T^b] = if^{abc}T^c$.

1.3.2 QCD coupling constant

As in the case of the quantum electrodynamics, the size of the strong coupling $\alpha_s(k^2)$ (eq. 1.37) depends on the energy scale of the interaction. The trend of the coupling constant as a function of the energy scale is shown in fig. 1.5 and, after a renormalization procedure, can be expressed in the following formula at leading order (LO):

$$\alpha_s(Q^2) = \frac{\alpha_s}{1 + \frac{\alpha_s}{12\pi}(11N_c - 2N_f)\cdot \ln\frac{Q^2}{\mu^2}}, \tag{1.37}$$

where $\alpha_s = g_s/4\pi$, μ^2 is a renormalization energy scale at which $\alpha_s(Q^2)$ is supposed to be known. Differently to the quantum electrodynamic case, the sign of the logarithmic term in the fraction denominator is positive $(11N_c - 2N_f) > 0$ being $N_c = 3$ and $N_f = 6$. The equation 1.37 can be used to determine the behaviour of the interaction potential between two coloured sources as a function of their relative distance. The value of $\alpha_s(Q^2)$ is fixed to α_s if $\mu^2 = Q^2$. The running coupling is small for large momentum transfers (asymptotic freedom [29]) so for $Q^2 \gg \mu^2$, and large for small momentum transfers (color confinement). The decrease of $\alpha_s(k^2)$ with the energy scale means that at high energy quarks can be described as almost free particles. The asymptotic freedom is the basis of the perturbative approach to calculate QCD observables (pQCD). Defining Λ_{QCD}^2 as:

$$\Lambda_{QCD}^2 = \mu^2 exp[\frac{12\pi}{(11N_c - 2N_f)\alpha_s(\mu^2)}, \tag{1.38}$$

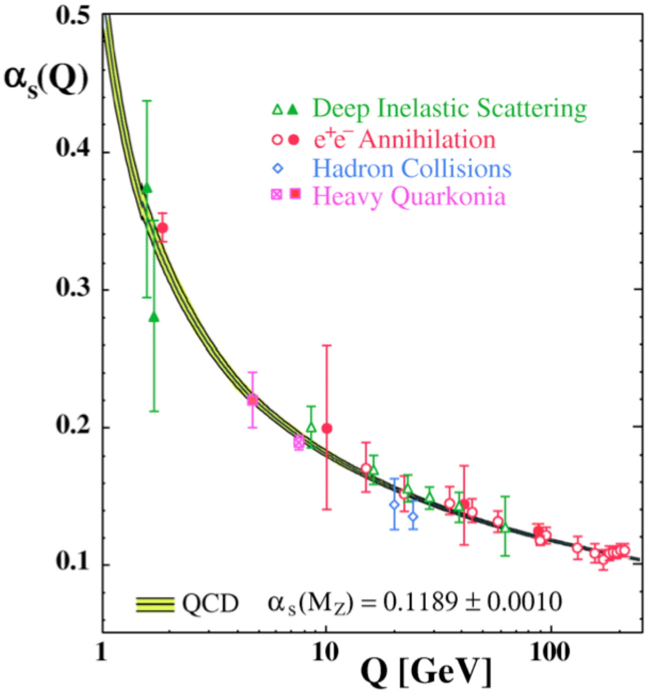


Figure 1.5: The strong running coupling constant as a function of the transferred momentum Q in different processes [28].

the running strong coupling assumes the form:

$$\alpha_s(Q^2) = \frac{\alpha_s}{(33 - 2N_f)\log(\frac{Q^2}{\Lambda^2_{QCD}})}. \tag{1.39}$$

The value assumed by Λ_{QCD} is often used to describe the hadron interactions phenomenologically. If in a proton-proton collision two partons interact with $Q^2 \gg \Lambda_{QCD}$ (hard scattering) the amplitude of the process is calculated with the pertubative QCD tools. So the value of Λ_{QCD} gives the energy at which non-perturbative effects become important. At low energy (corresponding to a large distance of the order of 1 fm), the perturbative approach is no more applicable. If a pair of quarks ($q\bar{q}$) starts to separate from each other, the exchanged gluons interact and strong coupling constant increases.

In this case, non perturbative lattice QCD approach can be used [30]. Lattice QCD is formulated on a grid of space-time points, introducing a cut-off at the order of $1/a$, where a is the lattice spacing, which regularizes the theory. Fields representing quarks are defined at lattice sites while gluons fields are defined on the link connecting close sites. At low energy scale the increasing force binds the quarks together until the energy density of the colour field between the quarks is large enough to create an additional $q\bar{q}$ pair. In the final state therefore only colourless bound state can be observed. This effect is known as “colour confinement” and justifies the non observation of free quarks and gluons.

1.4 Modelling the hard process

When two protons collide with a transfer momentum Q higher than their masses, the interaction involves their constituents, since the proton can be resolved into its partons (quarks and gluons), carrying only a fraction x of the total momentum of the proton. The calculation of production cross-section in a proton-proton collision can be seen as a combination of two different energy regimes: the short-distance or high-energy regime and the long-distance or low-energy regime. The QCD factorization theorem [31] can be used to calculate a wide variety of hard-scattering cross sections in hadron–hadron collisions. The production cross-section can therefore be expressed as a product of two terms: one relative to the parton-parton cross-section ($\sigma_{ab\to n}$) at a short distance and the other describing the internal structure of protons at a long distance. As seen in the previous section, for large momentum transfers (short distance) a perturbative approach can be used to evaluate the parton-parton interaction. For the second term, instead, perturbative QCD calculation can no longer be applied and a more phenomenological way is adopted to describe the proton structure. For this purpose parton densisty functions (PDFs) have to be introduced. Given a proton A with momentum p_A the PDF for the parton a is a function $f_{a/A}(x, Q^2)$ of the relative momentum $x = \frac{p_a}{p_A}$ of the parton in the direction of the proton momentum and depends also on the energy scale Q^2 of the scattering process. The scale at which the long-distance physics of the PDF description and the short-distance physics of the parton-parton interaction separate is called the factorisation scale and is defined as $\mu_F = Q$. The proton-proton collision can be therefore expressed as:

$$\sigma_{p_A p_B \to n} = \sum_q \int dx_a dx_b f_{a/A}(x_a, Q^2) f_{b/B}(x_b, Q^2) \sigma_{ab\to n}. \tag{1.40}$$

The functions $f_{a/A}$ and $f_{b/B}$ denote the PDFs for the partons a and b respectively in the proton A and B. Fig. 1.6 shows graphically the Drell-Yan scattering, one of the possible processes that could happen in a proton-proton collision. It is possible to expand the parton-parton cross-section in terms of the strong coupling so, the previous expression of the cross-section becomes:

$$\sigma_{p_A p_B \to n} = \sum_q \int dx_a dx_b \int f_{a/A}(x_a, Q^2) f_{b/B}(x_b, Q^2) \times [\sigma_0 + \alpha_s(\mu_R^2)\sigma_1 + ...], \tag{1.41}$$

where μ_R is the renormalization scale, the reference scale for the running of the $\alpha_s(\mu_R^2)$. σ_0 is the tree-level parton-parton cross-section and σ_1 is the first

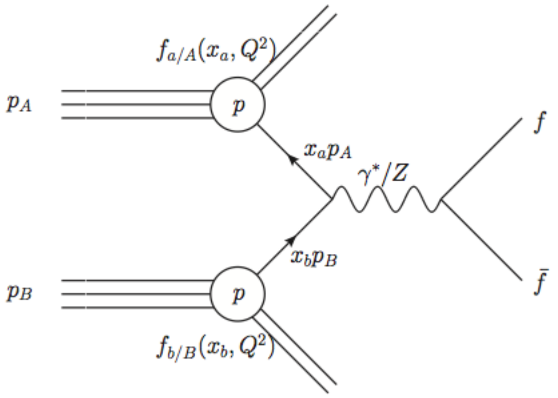


Figure 1.6: The general representation of the Drell-Yan scattering.

order QCD correction to the parton-parton cross section etc.
The eq. 1.41 can be expressed in term of the matrix elements:

$$\sigma_{p_A p_B \to n} = \sum_q \int dx_a dx_b \int d\phi_n \times f_{a/A}(x_a, Q^2) f_{b/B}(x_b, Q^2) \frac{1}{2s} \mid m_{ab\to n} \mid^2 (\phi_n), \tag{1.42}$$

where $\frac{1}{2s}$ is the parton flux, ϕ_n is the phase space of the final state and $\mid m_{ab\to n} \mid$ the matrix element for a final state n produced by the annihilation of the partons a and b.

The leading order (LO) calculation of the parton-parton annihilation cross-section can be seen as an extension of the calculation of the QED process $e^+e^- \to \mu^+\mu^-$ by including the colour factor and accounting for the fractional charge of the involved quarks.
To have a realistic model that is able to explain what is experimentally observed, also higher order QCD have to be taken into account. Additional quarks and gluons in the final state in the transverse plane can be emitted. These perturbative corrections can be calculated in the regime of small α_s. There are two classes of next-to-leading order (NLO) corrections that contribute: the virtual loop corrections and the real emission of gluon/quarks as shown in fig. 1.7.
While the virtual loop correction does not affect the transverse momentum spectrum of the final state, the real emission implies an additional parton in the final state so it enters in the balance of the momentum in the transverse plane. Currently, several leading-order calculations, such as [32,33], are available describing more than six partons in the final state. The partons cannot be detected as free particle because they carry a colour charge that has to be conserved in the process. Due to confinement, objects with a net colour charge can not exist freely. Therefore the initial partons tend to a stable configura-

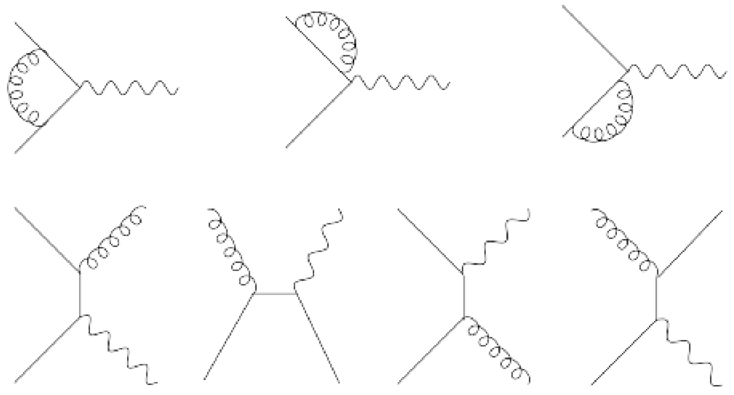


Figure 1.7: Next-to-leading order corrections: virtual loop cerrection and the real emission of gluons/quarks.

tion of uncoloured composite objects by the emission of gluons, which will in turn radiate $q\bar{q}$ pairs and so on, with each new parton nearly collinear with its parent. Final objects obtained by this hadronization process are mesons and baryons, detectable by their interaction in matter. The hadronization happens in the same direction of the initial parton that originates the chain, and the subsequent emission of particles around this direction, leads to a collimated cone of quarks and gluons referred to as a jet.

1.4.1 Parton distribution function

The parton distribution functions (PDFs) are essential in the determination of the cross section of processes in a hadron-hadron collision, giving information of the initial state. The PDFs are defined as the probability densities for finding a parton with a momentum fraction x_a of the proton momentum at a given factorization scale μ_F. These functions cannot be calculated by perturbative QCD but rather their functional form can only be determined using experimental data at a fixed scale Q^2. The PDFs are determined by global fits to data from different experiments: deep-inelastic scattering (DIS), Drell–Yan (DY) and jet production at current energy ranges. Three major groups, NNPDF [34], CTEQ [35] and MRST [36], provide updates to the parton distributions when new data or theoretical developments become available. In addition, there are also PFD available from the two HERA experiments [37]. The results from deep inelastic scattering (DIS) at HERA play an important role in determining these functions, due to the large phase space in Q^2 and the Bjorken variable x. The first LHC data has given important information about the parton distribution functions (see for example the fig. 1.8). Different processes such as Drell-Yann, W/Z + jets, W/Z boson decays and jets with a high transverse momentum (p_T) may be very useful to constraint PDFs and provide complementary information respect to those extracted from DIS experiments. For example, the measurement of associated production of a W boson and charm quark jet [38, 39] is directly sensitive to the strange quark and antiquark content of the proton. Precise measurements of this process at the Large Hadron Collider (LHC) may significantly reduce the uncertainties in the strange quark and antiquark parton distribution functions.

The fits to various sets of experimental data are performed within the DGLAP (Dokshintzer-Gribov-Lipatov-Altarelli-Parisi) [40–42] evolution scheme expressed by the following equation:

$$\frac{\partial f(x,Q^2)}{\partial logQ^2} = \frac{\alpha_s}{4\pi}\int_x^1 \frac{dz}{z}[P_{aa'}f(\frac{x}{z},Q^2)], \tag{1.43}$$

where $P_{aa'}$ is known as the Altarelli-Parisi splitting function that gives the probability to have trasformation of parton with momentum x into another with momentum z, as a consequence of the emission of one or more quarks or gluons. The PDFs evolution with the factorization scale is predicted by the the DGLAP equation (eq. 1.43) so the PDFs measured at one scale can be used to predict the results of experiments at other scales.

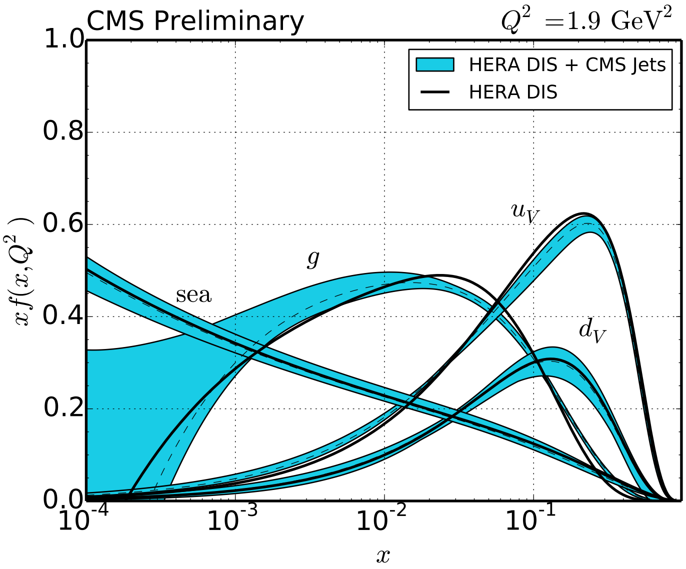
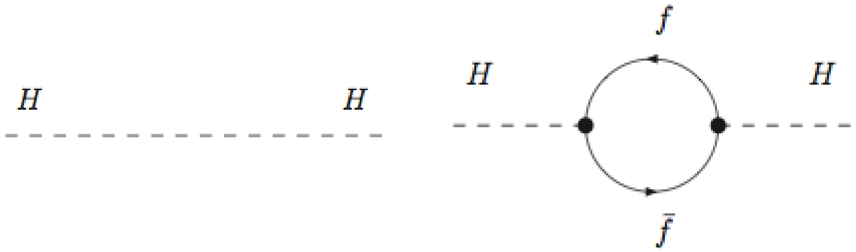


Figure 1.8: Overview of the gluon, sea, u valence, and d valence PDFs before (full line) and after (dashed line) the inclusion of the CMS inclusive jet data to DIS data from HERA into the global fit. The PDFs are shown at the starting scale $Q^2 = 1.9$ GeV^2. In addition the total uncertainty including the CMS inclusive jet data is shown as a band around the central fit [43].

1.5 Physics beyond Standard Model

Although the success in describing several particle physics processes, the Standard Model cannot be established as the ultimate theory of fundamental physics as there are different aspects of experimental observations that do not find a satisfactory explanation. The main aspects that have no answer inside the SM are here summarized:

- The hierarchy problem affects the Higgs sector. This problem can be expressed as the instability of the value of the Higgs boson mass and appears when radiative corrections are included in presence of a physical cut-off (Λ) that is placed at energies far above the electroweak scale.
  To the Higgs propagator at the tree level (fig. 1.9(a)) one can add the one-loop diagrams to correct it as shown in fig. 1.9(b) with the self-energy correction Σ_H given by the fermionic contribution Σ_H^f, and this later being easily computable from the corresponding Feynman integral over the internal fermion momentum k. Integrating over all possible loop momenta k, considering N_f degrees for fermion f with mass m_f and Yukawa coupling g_f , and keeping just the dominant terms for large cut-off Λ, the mass correction as a function of Λ is:



(a) Tree level propagator for the Higgs boson.

(b) Loop correction to the tree level propagator of the Higgs boson.

Figure 1.9: Feynman diagram of the Higgs propagator.

$$\delta m_H^2 = N_f \frac{g_f^2}{16\pi^2}(-2\Lambda^2 + 6m_f^2 \log\frac{\Lambda}{m_f} + ...). \tag{1.44}$$

  There are two dominant contributions at high values of the cut-off: the largest one that grows quadratically as Λ^2 and the other one that grows as $log\Lambda$. Choosing for example the cut-off at the Planck energy scale, one can find a huge correction that is 30 orders of magnitude larger than the tree level squared Higgs mass ($\delta M_H^2 \approx 10^{30} M_H^2$). To solve the hierarchy problem of the SM two different solutions are proposed. One assumes new interactions and the Higgs boson as a composite particle while the other assumes new symmetries and the Higgs boson is an elementary particle. At present there is not yet any experimental evidence therefore if the observed Higgs particle is elementary is still an open question.

- Cosmological observations tell us that the Standard Model explains about 4% of the energy present in the universe. Of the missing 96%, about 24% should be Dark (non luminous and non absorbing) Matter whose existence is by now well established. The observations show that Dark Matter appears to make up about 85% of the matter content of the universe. One of the indirect evidences of Dark Matter comes from measurement of galactic rotation curve. The rotation velocity v of an object on a stable Keplerian orbit with radius r around a galaxy scales following the relation: $v(r) \propto \sqrt{M(r)/r}$ where M(r) is the mass contained inside the orbit. If r lies outside the visible part of the galaxy, the velocity is expected to vary with r as $v(r) \propto 1/\sqrt{r}$. Instead in most galaxies one finds that v becomes almost constant as the value of r increases. This suggests the existence of a dark halo with mass density $\rho(r) \propto 1/r^2$. Candidates for Dark Matter must satisfy several requests: they must be stable on cosmological time scales and must interact weakly with the electromagnetic radiation. Possible candidates can be primordial black holes, axions, sterile neutrinos and weakly interacting massive particles (WIMPs). The strategies for the search of WIMPs can be divided into two groups: direct and indirect search methods. In the first case it is exploited the interaction through elastic scattering of Dark Matter particle with the ordinary matter that composes the detector. The other method consists in the indirect search of evidence in cosmic rays or at LHC where Dark Matter particle could be produced in pairs. The SM does not individualize a proper candidate for Dark Matter. Explanation of the origin of Dark Matter requires also a theory beyond the Standard Model.

- The neutrino is treated as a massless particle, and flavour mixing is not included. An explanation of the origin of non-vanishing neutrino masses and neutrino oscillations that have experimental evidence require going beyond SM. Infact the Super-Kamiokande [44], SNO [45], KamLAND [46] and OPERA [47] [48] experiments have observed that neutrinos from one family can be transformed, by oscillation, to neutrinos of another family with frequencies proportional to their mass square differences.

- Despite the Standard Model predicts that matter and anti-matter should have been created in (almost) equal amounts if the initial conditions of the universe did not involve an unbalanced ratio of matter to antimatter, the universe appears made almost entirely of matter.
  The CP violation has been observed in meson decays in the Standard

Model and it implies that the system is not symmetric under combined operations of charge conjugation (C) and parity inversion (P). However it seems not to be enough to explain the cosmological observations about this asymmetry.

- The gravitational interaction is not considered in the SM that includes only electromagnetic, weak and strong interactions. One approach to explain gravity inside the SM is to assume the existence of the graviton, a hypothetical elementary particle that mediates the force of gravitation in the framework of quantum field theory.

There is therefore the need of some new theoretical models to give solutions to these open questions. One of the theories that attempt to overcome the Standard Model is Supersymmetry (SUSY). Supersymmetry is an extension of the Standard Model and it expects an additional class of symmetries to the Lagrangian. These symmetries exchange fermionic particles with bosonic ones. Such a symmetry predicts the existence of supersymmetric particles, the sparticles, which include the sleptons, squarks, neutralinos and charginos.
As a consequence each particle of the SM would have a superpartner whose spin differs by 1/2 from the ordinary particle. SUSY solves the hierarchy problem by the additional contributions from sfermions. When computing the one-loop radiative corrections to the Higgs boson propagator in SUSY theories one has to add new contributions from the scalar fermion partners. In SUSY models an extra symmetry, called "R-parity", is added. A new quantum number (P_R) defined for each particle is conserved:

$$P_R = (-1)^{3(B-L)+2s} \tag{1.45}$$

with B, L and s the baryon number, the lepton number and the spin respectively. An important consequence of the R-parity for Dark Matter Physics is that it provides a natural particle candidate for explaining the Dark Matter: the Lightest Supersymmetric Particle (LSP) that, due R-parity, is stable.

Chapter 2

Z+b process at LHC

The Z boson, since its discovery at the Super Proton Synchrotron (SPS) [49] in the early 1980s, has been the subject of detailed measurements both at e^+e^- and hadron colliders. Many precision measurements of its properties have been performed by the ALEPH, DELPHI, L3 and OPAL experiments at the large electron-positron collider (LEP) at CERN [13]. At Tevatron the single boson production has been explored in a hadron collider with a center of mass energy $\sqrt{s} = 1.96$ TeV (Run2) by CDF [50] and D0 [51] [52] experiments. In addition precision studies of the production of Z boson in association with jets [53], with heavy flavour quarks [54] [55], become possible. The LHC collider gives the possibility to measure the associated production of Z and jets at a higher center mass energy with respect to previous experiments and the study of these types of processes allows to perform tests on perturbative Quantum Chromodynamics, and to set better constraints on the parton distribution functions. Moreover the production of Z boson with jets process is one of the dominant background for Standard Model measurement as in the Higgs boson context and also for physics beyond Standard Model. The discovery of new physics goes through a detailed understanding of the Standard Model processes.

The theoretical predictions used for the comparison with these measurements have been extended and improved. For measurements of single vector boson production in association with jets, calculation at next-to-leading order (NLO) for up to five partons in the final state are available. The magnitude of the theoretical uncertainties in these calculations are comparable to those of the experimental uncertainties [56]. The theoretical predictions for the heavy flavour production are less well known compared to ones for the inclusive Z+jets and this is a further reason why the process with a Z associated with b quarks is really interesting as it can provide important experimental constrains to improve the theoretical description of this process.

2.1 Heavy flavour description in pQCD

The treatment of events containing b quarks in the initial state is not trivial because of their heavy mass $m_b \sim 4.7$ GeV. Processes involving the production of bottom quarks are generally described in perturbative QCD employing two different schemes: 4-flavor (4F) [57] or "massive" scheme and 5-flavor (5F) [58] or "massless" scheme.

2.1.1 4-flavor scheme

The cross section of the processes involving a b quark is characterized by two different scales, the scale Q of the hard process and the mass of the b quark m_b and in perturbative calculations logarithms and power of the ratio m_b^2/Q^2 are present. In the 4-flavor scheme only 4 light quark density functions are included in the initial state while the heavy quarks do not contribute to the proton wave-function. The b-parton distribution functions are set to zero and do not enter in the computation of the running coupling constant and in the evolution of the PDFs. Bottom quarks are significantly heavier than the proton so, in high-Q^2 interactions, they can only be created as a massive particle in the final state. Calculations at next-to-leading-order (NLO) in QCD in this scheme are available for different processes relevant for hadron collider phenomenology and also the associated production of vector bosons and b quarks: $pp \to Vbb + X$ with V=W,Z [59,60]. No prediction at next-to-next-to-leading-order (NNLO) in QCD in the 4-flavor scheme are, instead, available for any relevant processes at LHC.

2.1.2 5-flavor scheme

When the typical scale Q of the process, such as the transverse momentum of the bottom quark or the mass of the Z boson, is higher than the mass of the heavy quark a natural approach to describe the production of bottom quark is the 5-flavor scheme where an initial state parton density function for a massless b quark is included in the calculation. In this case the b quark density includes the gluon splitting $g \to b\bar{b}$ where one of the b quark remains at low transverse momentum (p_T) and is integrated out, while the other is involved in the hard scattering and emerges at high p_T. Logarithms of the type:

$$\alpha_s \log \frac{Q^2}{m_b^2}, \tag{2.1}$$

arising from the splitting of the initial heavy quarks and gluons appears in the cross section calculation and are resummed in the b quark parton density function. The origin of these logarithms is due to the process of gluon splitting into a heavy quark-antiquark pair. The gluon splitting can arise both from an

initial state radiation or a final state radiation. Final state logarithms can be resummed, using techniques such as fragmentation functions.
The large initial state logarithms can be resummed into a b parton density function to all orders in perturbative QCD using the DGLAP evolution equation [40–42]. All possible multiple collinear gluon emissions from the initial state are summed up to all orders in α_s. This approach is realiable if the b quark mass can be considered small respect to the typical scale of the process. The 5-flavor scheme shows two main advantages compared to the 4-flavor scheme. The logarithms related to the collinear splitting of the initial gluons, proportional to $\log(Q^2/m_b)$ present in the four-flavor scheme, are resummed to all orders. The consequence of that is a more convergent perturbation series in this approach. The second advantage is the simplification of the calculation being the LO process in the five-calculation simpler and that implies higher-order corrections calculable more easily. In the 4-flavor scheme instead, the computation is more complicated due to the presence of massive final states and higher multiplicity. There are many differential NLO calculations available in this scheme for different processes relevant at LHC [61].

The two schemes are used in different and complementary regimes depending on the energy scale Q of the process considered with respect to the b quark mass. The two approaches are equal at all orders in perturbative theory but may give different results at finite order. The difference between the two schemes has been explored to quantify the size of such logarithms for different processes at the LHC. Recent studies [62] have shown that at LHC logarithms resummed in the PDF of the b quarks are not so large to spoil the convergence of the perturbative series of the 4-flavor scheme whose results are in general well behaved. The size of the effect of initial state collinear $\log \frac{Q^2}{m_b^2}$ has been evaluated to be always modest. The reason of this result can be ascribed to the fact that the effective scale Q which compares in the initial state collinear logarithms, proportional to the hardest scale in the process, is modified by universal phase space factors whose effect is to reduce the size of logarithms for processes taking place at hadron colliders.

2.2 Z+b process

The production of b-quarks in association with a vector boson Z is a challenging topic both under a theoretical and experimental point of view. It is interesting to study both the process with a Z boson and exactly 1 b jet (“Z+1b jet” production) and also that with a Z boson associated with at least 2 b jets (“Z+2b jets” production). Both are important especially for precision test of QCD since the two different final states involve different production mechanisms. The study of the Z+1b and Z+2b jets processes provides information to distinguish between different approaches in theoretical calculation

(4-flavor scheme, 5-flavor scheme). Moreover it is interesting to compare the measurements with theoretical predictions based on next-to leading order calculations also including full bottom quark mass effects. From a study of the leading order cross section, finite bottom-quark mass effects are expected to affect both the total and differential $Zb\bar{b}$ cross sections, mostly in the region of small $b\bar{b}$ pair invariant mass [63]. Also, the Z+1b jet process can be explored to study the b content of the proton PDF while the production of a Z boson in association with two or more b jets is of interest in the context of searches. The study of these two processes can benefit if they are estimated simultaneously because of the inefficiencies in the tagging of jets from a quark b causing the migrations of events from one sample to the other.
At present several experiments have been involved in these studies. The cross section for Z production with b jets has been measured at the Tevatron collider at Fermilab ($\sqrt{s} = 1.96$ TeV) by both the CDF [54] and D0 [55] collaborations. For the first time CDF experiment has shown the differential measurements as a function of the kinematics of the jets and Z boson and the number of jets in the event, allowing NLO QCD prediction to be tested over a wide range of final-state observables. No full NLO QCD calculations were available for this process so large variations between the theoretical predictions have been observed. The predictions generally describe the data, but the agreement is better for those predictions that use a low value for the renormalization and factorization scales. These measurements have provided a unique opportunity to test the theoretical description of heavy-quark jets at hadron colliders by performing a comparison between the Tevatron experimental data and existing theoretical predictions. Studying the same cross sections in the very different kinematic regimes available at the LHC proton proton collider is then of great interest and represents a crucial test of our understanding of QCD at high-energy colliders. Moreover at Tevatron the dominant production diagram is through $q\bar{q} \rightarrow Zb\bar{b}$ whereas at the LHC it is through $gg \rightarrow Zb\bar{b}$. The CMS experiment has already performed a measurement of the Z+b jet cross-section with exactly one b jet and with at least two b jets in the final state using an integrating luminosity of 5 fb^{-1} at an center mass energy of 7 TeV [64]. The results have been compared to MADGRAPH, aMC@NLO and MCFM predictions. The measured cross sections are found to be in fair agreement with MADGRAPH and aMC@NLO while the MCFM results differ by approximately two standard deviations from the data. Also the ATLAS experiment has performed this measurement at $\sqrt{s} = 7$ TeV using a data sample with an integrated luminosity of 4.6 fb^{-1} [65]. The cross section as a function of different observables has been measured for events with a Z boson and at least one or two b tagged jets. These results have been compared to next-to-leading order predictions provided by MCFM and aMC@NLO showing a good agreement between theoretical expectations and data. Predictions with NLO matrix element calculations have been obtained using aMC@NLO with both

the 4-flavor and 5-flavor approach. The $Z+\geqslant 1b$ data are better described by the 5FS prediction while the 4FS prediction shows a better agreement with the $Z+\geqslant 2b$ data. For the $Z+\geqslant 2b$ differential cross sections, all the predictions provide reasonable descriptions of data within the large experimental uncertainties. However there is some evidence for disagreements between predictions and data at low values of the invariant mass of the two b tagged jets and low values of the angular separation between the directions of the two b tagged jets ($\Delta R(bb)$).

2.2.1 Associated production of Z boson and a single b jet

The main leading order (LO) contribution ($O(\alpha_s)$) for the process with one Z boson and a single b jet is $gb \rightarrow Z + b$ and the corresponding tree level Feynman diagrams are shown in fig. 2.1. The subprocesses that enter in the next-to-leading order (NLO) calculation of $gb \rightarrow Z + b$ are:

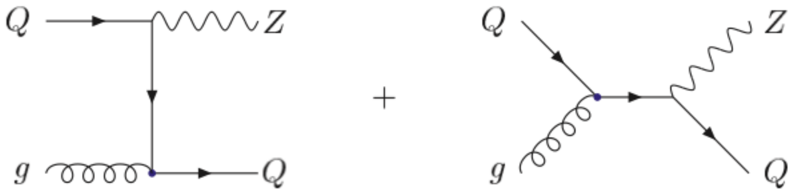


Figure 2.1: LO diagrams for the associated production of a Z boson and a single b-quark ($gb \rightarrow Z + b$).

- $qb \rightarrow Zbq$
- $gb \rightarrow Zbg$
- $gg \rightarrow Zb\bar{b}$ when one b is missed

Next-to-leading order calculations ($O(\alpha_s^2)$) for this process are available in the 5-flavor scheme approach (with $m_b = 0$) [58].
Another process that contributes to the Z production with only one b jet is $q\bar{q} \rightarrow Zb\bar{b}$, where one b is missed and with the Z boson radiated off the initial-state quarks or the final-state quarks. We can have only one b jet in the final state if one of the two b quarks produced is missed because it is outside the coverage of the detector or because the two b-quarks are collinear so they are clusterized in the same jet. This calculation is carried out at leading order with a non-zero b mass. An alternative calculational scheme is to consider $gg \rightarrow Zbb$ as the leading-order process (with $m_b \neq 0$), and to allow one b

quark to be emitted collinear to the beam, yielding a final state with Z boson with a single b quark. This approach however implies two difficulties to face: the expansion parameter in the perturbation theory has an additional factor $log(M_Z/m_b)$ that multiplies the coupling constant α_s making the theory less convergent and NLO corrections are more difficult to obtain respect to $gb \rightarrow Zb$ being one more particle in the final state.

2.2.2 Associated production of Z boson and at least two b jets

The processes that contribute to the associated production of a Z boson and at least two b quark jets (Z+$2b$ jets) are:

- $q\bar{q} \rightarrow Zb\bar{b}$
- $gg \rightarrow Zb\bar{b}$

The leading order Feynman diagrams for the subprocess $q\bar{q} \rightarrow Zb\bar{b}$ are shown in fig. 2.2 where we have the production through the gluon splitting mechanism or a Z boson emitted from initial or final fermion lines. In fig. 2.3, instead, the leading order Feynman diagrams contributing to the $gg \rightarrow Zb\bar{b}$ subprocess are presented. At hadron colliders, QCD effects are not negligible so have to be taken into account to obtain precise theoretical predictions. Since at high

Figure 2.2: Leading order Feynman diagrams for the associated production of a Z boson and at least two b jets.

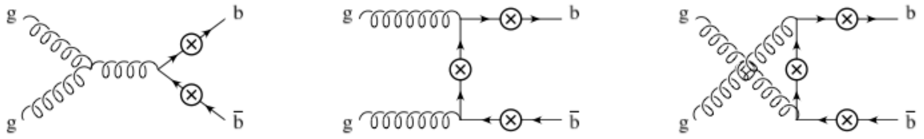


Figure 2.3: Leading order Feynman diagrams for $gg \rightarrow Zb\bar{b}$. The circled crosses correspond to all possible insertions of the Z boson, each one representing a different diagram.

energies the QCD is a perturbative quantum field theory (pQFT), QCD effects at collider energies can be calculated order by order in the strong coupling constant for the Zbb process. New production channels at higher orders of the perturbative series open in this way. The $O(\alpha_s)$ corrections to $q\bar{q} \to Zb\bar{b}$ leading order subprocesses consist of the self energy vertex and box diagrams illustrated in fig. 2.4, 2.5, 2.6 where the Z boson is emitted from the initial fermion line (q or $\bar{q}$). We have similar contributions ($O(\alpha_s)$) when the Z boson is emitted from the final fermion line. Also the corrections $O(\alpha_s)$ to $gg \to Zb\bar{b}$ have to be considered. These virtual contributions consist of the self energy and box diagrams. The production of Z boson with two b jets has been

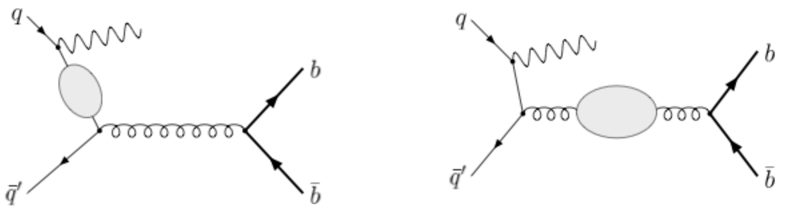


Figure 2.4: Gluon and quark $O(\alpha)$ self-energy corrections contributing to the $q\bar{q} \to Zbb$ subprocess at NLO. The shaded blobs denote standard one-loop QCD corrections to the gluon and quark propagators, respectively.

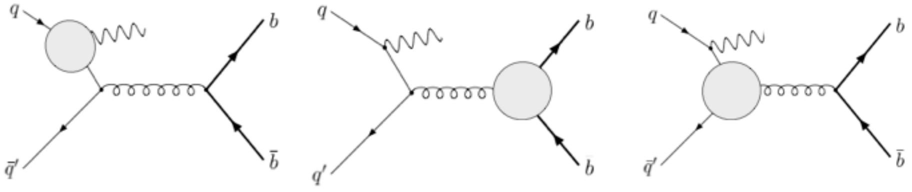


Figure 2.5: $O(\alpha_s)$ vertex corrections contributing to the $q\bar{q} \to Zbb$ subprocesses at NLO. The shaded blobs indicate the standard one-loop QCD corrections to $q\bar{q}g$, $b\bar{b}g$ and $q\bar{q}Z$ vertices.

derived at NLO in QCD using the 4-flavor scheme first in the massless b-quark approximation [66,67] and also including fully the b quark mass effects [68–70]. These studies show that in kinematical distributions the b quark mass effects can have a significant impact, particularly in phase space regions where the relevant observables are of the order of the b quark mass.

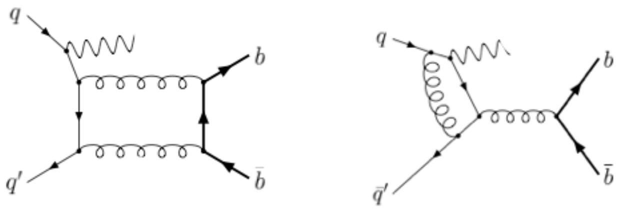


Figure 2.6: Examples of $O(\alpha)$ box diagram corrections contributing to the $q\bar{q} \to Zbb$ subprocess at NLO.

2.3 Monte Carlo Generators

The structure of events produced in high energy colliders is complex and numerical simulations are needed to describe them. Theoretical calculations illustrated in the previous sections are employed inside Monte Carlo generators that provide a detailed description of the final state so that experimental observables can be predicted and compared with experimental data, giving the possibility to test different models.
This section presents the main approaches adopted by the Monte Carlo generators to reproduce the hard scattering in a hadron collision. Inside Monte Carlo generators the complex production of realistic events is implemented subdividing into a sequence of steps that can be treated separately with the help of both analytic and numeric computations. Several event generators using different approaches and methods in the calculations are available, each with computations at different level of precision in perturbation theory.
The structure of a proton-proton collisions at the LHC as built up by event generators can be summarized by the following sequentially main steps:

- hard process
- parton shower
- hadronization
- underlying event
- unstable particle decays

A schematic representation of the different steps that are implemented in event generators is shown in fig. 2.7.

2.3.1 Matrix Element algorithm

The first step in the generation of events is the description of the hard processes cross section. The most natural way to study the hard scattering process is

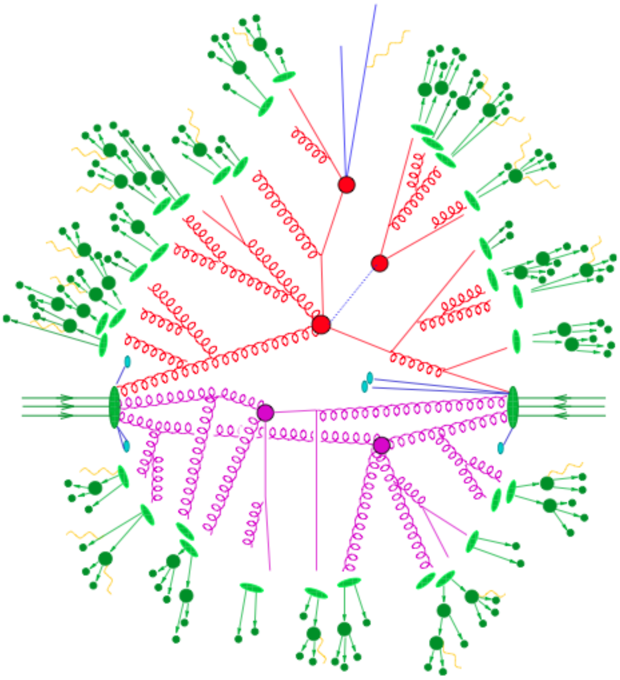


Figure 2.7: Schematization showing the structure of a proton-proton collision, in which the different colours indicate the different stages involved in event generation. The event represented contains the hard event followed by hard decays of two heavy unstable particles (red), and two more hard parton interactions (purple). The partons are dressed with secondary radiation as well, before the parton ensemble is transformed into primary hadrons which then decay further (green).

to compute the squared amplitudes of the corresponding Feynman diagrams, order by order, integrating them over the appropriate phase space. The Matrix Element method consists in the exact calculation of the matrix elements applying the Feynman rules in perturbative QCD. In principle this approach could provide an analytical and complete solution of the problem. However it is evident the complexity of this approach, so predictions obtained at the leading order of the perturbative expansion are most common. Loop calculations from a computational point of view are complex and they are available only for a limited number of processes. The Matrix Element calculation reliably describes the fundamental parton interaction, characterised by hard momentum scales, but does not correctly handle the emission of soft and collinear radiation. There are two problems to face that may cause the divergence of the cross section calculated with the Matrix Element method: the presence of a high number of soft partons with low transverse momentum (soft divergences) and the situation in which a radiated parton is collinear to the emitted parton (collinear divergences). At leading order it is not possible to treat these types of divergences without loop corrections so the only solution is to perform the matrix element cross section calculations away from soft and collinear divergences. The only Matrix Element method is therefore inadequate to produce realistic events, phase-space regions omitted from the matrix element calculations need to be recovered. In order to cover also regions where partons become soft and colliner, the Matrix Element method has to be integrated with the Parton Shower.

2.3.2 Parton Shower algorithm

Parton Shower is an approximate method to remove the soft and collinear divergences at each order in perturbation theory. The deletion of the divergences is mimed by using the Sudakov factors. The starting point is to "factorize" a complex 2 → n process (where n represent the number of the partons in the final state) into a simpler process 2 → 2 convoluted with showers to add the remaining partons as shown in fig. 2.8. The parton showers are described by this algorithm as a sequence of splittings of the type $a \to bc$ (where a, b and c are generic partons) for each of the coloured partons in the initial and final state where each event can happen with a certain probability. At each branching, part of the energy of the primary parton is carried away, and the momentum of the partons in the shower becomes softer. Each parton has a certain probability of emitting another one parton carrying out a fraction z

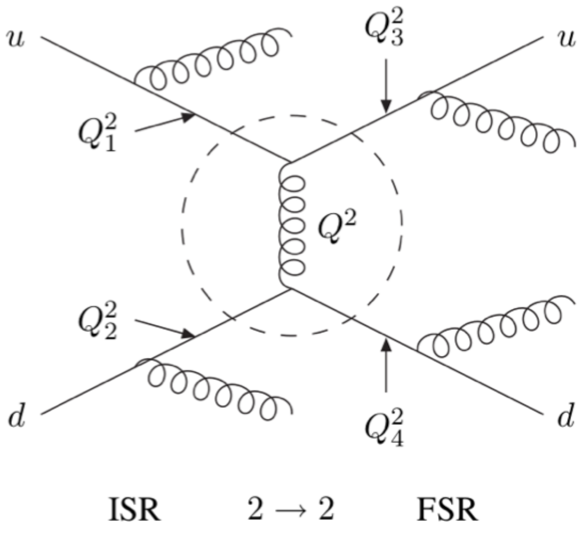


Figure 2.8: The "factorization" of a complex 2 → n process.

of its energy according to the Altarelli-Parisi splitting functions $P_{ab}(z)$. Both Final State Radiation (FSR) and Initial State Radiation (ISR) must be simulated by the event generator with the Parton Shower algorithm. The first consists in the production of further radiation by the final state partons while the initial-state shower is initiated by the partons inside the incoming hadrons and is terminated at the scale at which the hard scattering occurs. The structure of initial-state radiation is more complicated than that of FSR, since the complex structure of the incoming hadrons which enter in the hard process. A proton is made of three valence quarks, uud, plus the gluons that bind them together. However gluons are continuously emitted and absorbed by quarks, and each gluon may temporarily split into two gluons or into a qq pair.

Matching Matrix Element and Parton Shower algorithms

Matrix Element (ME) and Parton Shower (PS) calculations show different applicability limits. The ME well describes separated parton configurations, in regions of the phase space free of the soft and collinear approximations and calculations are exact to a given order in perturbation theory. Parton Shower method is, instead, an approximation derived from QCD that is valid in the collinear and soft limits so favoured to describe the structure of jets, but may fail in describing well separated parton configurations.

The two techniques are therefore complementary and a combination of the two approaches would be the best solution for the extraction of theoretical predictions to compare to hadron collider data. The main problem to face in the merging of ME and PS is the definition of the separation between the hard component of the process, that is solved with matrix element calculation, and the soft and collinear emission that is treated in parton shower. There is an intrinsic ambiguity: an event with N jets can be produced either from the showering of the N final state partons, or from $N-1$ final state partons plus an additional hard emission during the shower evolution.

To avoid double counting in the phase space regions where there is an overlap between the Matrix Element calculation and the Parton Shower a specific prescription, *matching scheme*, is introduced. The simple solution is to apply a cutoff, known as matching scale. All the branchings occurring at a scale harder than the matching scale are treated by the Matrix Element, while softer radiation emissions are left to the parton shower. Several matching schemes have been developed, such as the MLM and the CKKW [71, 72].

2.3.3 Hadronization

The perturbative theory is valid until the partons reach the energies of about 1 GeV that is where the strong coupling constant becomes so strong that non-perturbative effects become important. One of the most important non perturbative effects to take into account is the hadronization which converts colour partons into uncoloured composite objects (mesons and baryons). The physics mechanisms explained so far, implemented in the event generators, are mainly related to the partonic level while, from an experimental point of view, the final objects are hadrons. So also a hadronization phase, where all the outgoing partons end up confined inside hadrons has to be implemented inside generator events. Several phenomenological models, inspired to QCD, have been developed to describe this non-perturbative process. The main approaches in use are cluster fragmentation and string fragmentation.

- Cluster fragmentation.
  It starts with the gluon splitting $g \rightarrow q\bar{q}$ after the Parton Shower and all quarks are clustered in color singlets with a mass of a few GeV. The clusters formed are characterized by their mass and flavor quantum numbers and are treated as resonances that decay into a pair of hadrons proportionally to their phase space. Only the very lightest clusters are converted directly into hadrons, whereas heavier clusters may either convert into heavy hadrons or decay into lighter clusters.

- String fragmentation.
  In this model the QCD potential between two quarks that form the simplest color singlet system $(q\bar{q})$ is parameterized at low scale by field lines compressed to a tube-like region by self interactions among soft gluons. The potential is given by:

  $$V(r) = \frac{4\pi}{3r} + kr, \tag{2.2}$$

  where r is the radial distance between the two quarks and k is the string constant, i.e. the amount of energy per unit length whose value is $k \approx 1$ GeV/fm from hadron spectroscopy. Neglecting the first term that represents the Coulomb potential, the energy stored by the quark pair increases linearly with the separation between them. The string represents the field line connecting two quarks and the force field is linearly increasing with the distance according to the QCD potential shown in eq. 2.2. Therefore with the separation of the two quarks the potential energy grows linearly until it reaches a level beyond which the string breaks forming a new quark pair $q^{'}\bar{q}^{'}$ as shown in fig. 2.9. Two new pairs are then formed, $q\bar{q}^{'}$ and $q^{'}\bar{q}$, and if the invariant mass of either of these pairs is large enough the strings can break again, otherwise the process stops.

2.3.4 Multiparton interaction model

In Monte Carlo modelling contexts, multiple parton interactions (MPI) denote the possibility of having multiple partonic interactions occurring within the same bunch crossing. A consequence of MPI is thus the possibility of observing

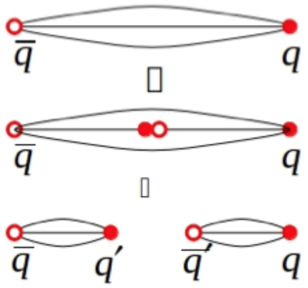


Figure 2.9: Illustration of string breaking by quark pair creation in the string field.

several distinct (i.e., hard) parton-parton interactions in some fractions of hadron-hadron events. The first Monte Carlo model to implement the MPI was proposed by Sjostrand and Van Zijl [73], it remains the basis for most modern implementations as described in [74]. An intuitive way of arriving at the idea of multiple interactions is to view hadrons simply as 'bunches' of incoming partons. No physical law then prevents several distinct pairs of partons from undergoing scattering processes within the same hadron-hadron collision.

2.3.5 PYTHIA

PYTHIA [75] is a Monte Carlo event generator capable of modelling a large number of hard scattering processes, including a set of physics models for the evolution from a few-body hard process to a complex multiparticle final state. This generator has been optimized for 2 → 1 and 2 → 2 hard scattering processes but it does not allow the simulation of $2 \rightarrow n$, with $n > 3$, hard scattering processes. No higher order loop corrections are included in the event simulation; however, the Parton Shower allows PYTHIA to approximate higher order behavior e.g. the processes 2 → 3, 2 → 4 etc...
It uses the Matrix Element algorithm for the cross section calculation at the first order in perturbation theory while for higher order diagrams, like ISR and FSR gluon emission, it uses the Parton Shower algorithm. In each event generated by PYTHIA the following steps are carried out:

- incident beam particles provided
- initial state radiation (ISR) performed (if needed)
- hard scattering performed
- final state radiation (FSR) performed (if needed)

- colored particles are hadronized into color singlet states using the Lund string model [76]
- unstable particles decay

At the end of the steps above, a list of stable particles is provided for each event.

Generator Tunes

The achievable accuracy in Monte Carlo generators models depends on the sophistication of the simulation. An important ingredient for that is the development of improved theoretical models but it also depends crucially on the available constraints on the remaining free parameters. For some Monte Carlo generators such as PYTHIA, different parameters have to be adjusted to better fit some aspects of the data. A specific set of these parameters is referred to as "tune". A phenomenon that needs to be considered in hadron colliders is the underlying event (UE), that is everything in single particle collision except the hard process of interest. Components of UE are: initial and final state radiations, beam-beam remnants and multiple parton interactions. The increased activity in the underlying event of a hard scattering collision over that observed in soft collisions cannot be explained solely by ISR. This is originated by secondary interactions between beam remnants and between partons that do not participate in the hard interaction. The generation of the underlying event is a complicated process and several phenomenological models have been developed to describe it. A hard scattering is more likely to occur when the beam hadrons overlap and this is also when the probability of a MPI is the greatest. The transverse region is sensitive to the underlying event and the MPI parameters have to be tuned to fit the data. Besides the underlying event, also the hadronization step has to be tuned to data.

In CMS, a series of tunes for PYTHIA have been developed on the basis of the results obtained in the pre-LHC era [77,78] but after the start of LHC, a series of new versions of PYTHIA (version 6.4) tunes have been provided. For this work the $Z2^*$ tune has been used both for signal and background Monte Carlo samples. The $Z2^*$ tune is derived from previous tunes: $Z2$ and $Z1$ [79] which uses the CTEQ5L parton distribution set, whereas $Z2^*$ adopts CTEQ6L [80]. An important set of parameters that has to be tuned is related to the regularization of the $1/\widehat{p}_T^4$ divergence of the leading order partonic scattering amplitude that arises when the final state parton transverse momentum $\widehat{p}_T$ approaches 0. PYTHIA performs this regularization through the replacement $1/\widehat{p}_T^4 \to 1/(\widehat{p}_T^4+\widehat{p}_{T0}^4)$ where $\widehat{p}_{T0}$ is a transverse momentum cutoff parametrized as:

$$\widehat{p}_{T0}(\sqrt{s}) = \widehat{p}_{T0}(\sqrt{s_0}) \cdot (\sqrt{s}/\sqrt{s_0})^{\varepsilon}, \tag{2.3}$$

where $\sqrt{s_0}$ is the reference energy at which $\widehat{p}_{T0}$ is determined and ε is a parameter that describes the energy evolution. A single value of $\widehat{p}_{T0}$ is used to regularize both the hard scattering and MPI. For smaller values of $\widehat{p}_{T0}$ and ε more MPI activity is predicted. The $Z2^*$ adopts optimized values for $\widehat{p}_{T0} = 1.921$ GeV and its energy rescaling $\varepsilon = 0.227$.

2.3.6 MadGraph

MadGraph [81] is a leading order matrix element generator which does not generate the whole event but it only simulates the hard scattering of a collision event. So it provides a precise calculation of the Matrix Element but it needs to be completed with the Parton Shower done by other programs, for example PYTHIA, to produce the whole event. The advantage in using MadGraph is that it allows to simulate $2 \to n$ (with $n > 2$) hard scattering processes, while the case with $n > 3$ is not possible in PYTHIA as we have already discussed. It can generate matrix elements at the tree-level for any Lagrangian based model. The generator will simulate matrix elements for decays of heavy particles like Z or $W^{\pm}$ and $2 \to n$ hard scattering processes simultaneously creating the Feynman diagrams that correspond to these processes. The Matrix Elements simulated by MadGraph are at tree level so virtual emission and loop diagrams are not included in the calculation of the hard scattering process. First, the initial and final state particles of a process in terms of initial and final state particles have to be specified and as a result, MadGraph generates all Feynman diagrams for that process. The parton level process (for example $q\bar{q} \to Zb$) is defined and MadGraph calculates the Born amplitude while Parton Shower, clustering partons into jets, hadronization and decays are carried out by the Parton Shower. Using the MadGraph event generator it is possible to simulate the associated production of Z boson and one or more b quarks at tree level. Currently both calculations with the two approaches, 4-flavor scheme and 5-flavor scheme, are available within MadGraph.

2.3.7 POWHEG

The POWHEG-BOX [82] (Positive Weight Hardest Emission Generator) is another approach to match NLO matrix element computations with Parton Shower simulations.

The basic idea in POWHEG-BOX is to generate the hardest radiation first and then to let the simulation of the event to any Parton Shower generator for subsequent softer radiation. In shower generators ordered in transverse momenta, the hardest emission is always the first, and in this case POWHEG-BOX simply replaces the hardest emission with its own NLO emission while in angular ordered showers, instead, the hardest radiation could not be the first in the shower chain. In this second case, in fact, the largest angle emission takes place earlier in the shower development, in order to account for coherent soft gluon emission from a bunch of collinear partons and the interface with POWEHG-BOX has to be treated with the so called “truncated showers” algorithm [83]. Differently from the case of MadGraph in which both the 4-flavor scheme and the 5-flavor scheme approches are applied in the calculation, within POWHEG-BOX the $Z + b$ process is calculated only in the 5-flavor scheme.

2.4 Z+b final state

Accurate modelling of the Z boson produced in association with b-quark jets is of great importance in a collider physics program. The study of that allows to improve our knowledge on the heavy-quark content in the proton and to better understand the difference between the two approaches (massive and massless b-quark) in calculations. Measuring different kinematic observables in $Z + b$ final state topology is a powerful tool to test QCD predictions with different implementations (LO multileg+PS, NLO, NLO+PS) and also it allows to tune constants and parameters chosen to describe the process in the Monte Carlo. Moreover, the production of a Z boson with one or two b jets is an irreducible background for several Higgs boson production channels at the LHC and also for various beyond Standard Model searches. In this section the relevant processes that involve the associated production of the Z boson with b quark both in the Standard Model and in the searches context are described.

2.4.1 Zb polarization asymmetry

The Standard Model of particle physics has been very successful in describing three out of the four fundamental interactions. It contains several parameters which have been precisely measured and compared to their predictions in order to show the consistency of the theory. Several electroweak precision tests have been carried out by the LEP experiments at CERN [84]. Most of the theoretical SM predictions have been confirmed by the related experimental measurements but at least one experimental result still appears in some sizable contradiction with the SM. The measurement of the forward-backward asymmetry of bb production at the Z peak A^b_{FB} [85] shows, in fact, a discrepancy of the order of 2.8σ. The production of a $b\bar{b}$ pair through the reaction $e^+e^- \to b\bar{b}$ (fig.2.10) has been studied at LEP at the Z^0 resonance. Defining as θ the angle between the incoming electron and the outgoing b quark, it is possible to express the differential cross section in the following way:

$$\frac{d\sigma}{d\cos\theta} \propto 1 + \cos^2\theta + \frac{8}{3}A^b_{FB}\cos\theta, \tag{2.4}$$

with

$$A^b_{FB} = \frac{3}{4}\left(\frac{2g^e_V g^e_A}{(g^e_V)^2 + (g^e_A)^2}\right)\left(\frac{2g^b_V g^b_A}{(g^b_V)^2 + (g^b_A)^2}\right), \tag{2.5}$$

where g^f_V and g^f_A are the weak vector and axial couplings between the Z and the fermion f defined by:

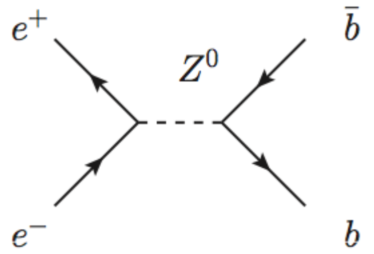


Figure 2.10: Feynman diagram of a positron-electron annihilation with the production of a $b\bar{b}$ pair at $\sqrt{s} = m_Z$, Z resonance.

$$g_V^f = T_3^f - 2\frac{q_f}{\mid e \mid}\sin^2\theta_W, \tag{2.6}$$

$$g_A^f = T_3^f, \tag{2.7}$$

with T_3^f the longitudinal component of the weak isospin of the fermion f. Defining the forward and backward cross section (respectively σ_F^f and σ_B^f) with respect to the polar angle θ:

$$\sigma_F = \int_0^1 \frac{dN}{d\cos\theta} \qquad\qquad \sigma_B = \int_{-1}^0 \frac{dN}{d\cos\theta}. \tag{2.8}$$

The forward-backward asymmetry of the fermion A_{FB}^f can be directly evaluated knowing the dependence by σ_F^f and σ_B^f:

$$A_{FB}^f = \frac{\sigma_F^f - \sigma_B^f}{\sigma_F^f + \sigma_B^f}. \tag{2.9}$$

The forward backward asymmetry defined can also provide information about the Z boson coupling to fermions at its pole being:

$$A_{FB}^f = \frac{3}{4}A_e \cdot \frac{g_A g_V}{g_A^2 + g_V^2} = \frac{3}{4}A_e A_f. \tag{2.10}$$

The asymmetry A_{FB}^b has been measured at LEP and independent measurements were also performed at the Stanford Linear Collider (SLC) using polarized electron beams. The results for A_{FB}^b are the following [6]:

$$A_{FB,exp}^b = 0.0992 \pm 0.0016 \qquad\qquad A_{FB,th}^b = 0.1037 \pm 0,008, \tag{2.11}$$

which gives a deviation between theory and experiment of 2.8σ. As the A_{FB}^b parameter shows the biggest deviation with respect to the SM expectation hence it is necessary to provide an independent measurement identifying a specific observable defined at LHC. We shall consider the process of associated production of a Z boson and a single b-quark, shown in fig.2.11, defined at parton level by subprocesses $bg \to bZ$ involving two Born diagrams with bottom quark exchange in the s-channel and in the u-channel. In this process

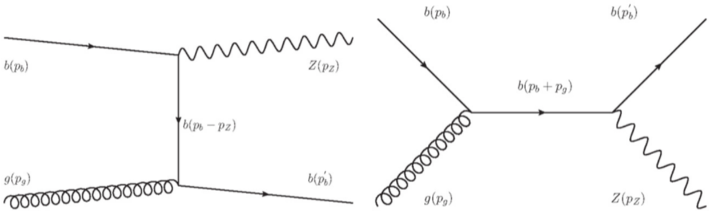


Figure 2.11: Feynman diagrams for the production of a Z boson and one b quark.

the vertex Zbb is involved and it is possible to build an experimental observable (A_Z) that enters with the same combination of the couplings as in the asymmetry measured at LEP/SLD.:

$$A_Z = \frac{\sigma_{Z_R} - \sigma_{Z_L}}{\sigma_{Z_R} + \sigma_{Z_L}}, \tag{2.12}$$

where σ_{Z_R} is the cross-section for the production of a right-handed Z boson, so the projection of the Z boson spin on its momentum direction is positive. The Standard Model expression of A_Z contains A_b, a quantity that appears in the expression of A^b_{FB} given by the eq.2.10 (where $f = b$).
Measuring the A_Z parameter in the Z+b process, a proportionality coefficient between A_Z and A_b can be extracted and so we can obtain an indirect measurement of A^b_{FB}. Recent studies [86] indicate that A_Z is almost free from theoretical uncertainties related to QCD scale variations as well as to pdf set variations. This property strongly suggests that a measurement of A_Z at LHC could be a promising candidate to clarify the discrepancy observed between SM expectation and the measurement of A^b_{FB} at LEP1.

2.4.2 Standard Model Higgs decay into b$\bar{\text{b}}$

The coupling constant between the Higgs boson and the fermion f (with mass m_f) can be written in terms of the Fermi constant G_F as:

$$G_{f\bar{f}H} = [\sqrt{2}G_F]^{1/2} m_f. \tag{2.13}$$

From this expression it is evident that the Standard Model Higgs boson couples with a fermion f with a strength that is proportional to its mass. The Higgs boson partial decay width to fermions can then be expressed as:

$$\Gamma(H \to f\bar{f}) = K\frac{G_F M_H}{4\sqrt{2}\pi} m_f^2(M_H^2)(1 - 4\frac{m_f^2}{M_H^2})^{3/2}, \tag{2.14}$$

where $K = N_c$ number of colors if f is a quark, $K = 1$ if f is a charged lepton. The decay width expressed in eq. 2.14 shows that the dominant decay process for a Higgs boson at 125 GeV is $H \to b\bar{b}$ (with a branching fraction of about 0.56 [87]) as the decay into $t\bar{t}$ pair is only possible for an heavier Higgs boson. For this decay channel, in a hadron collider, the inclusive signal is overwhelmed by QCD production of bottom quarks. In order to suppress the large QCD background also the search for the associated production of the Higgs boson with a W or Z boson is performed (fig. 2.12). In this case one of the main backgrounds arise from the production of vector bosons in association with heavy quarks jets.

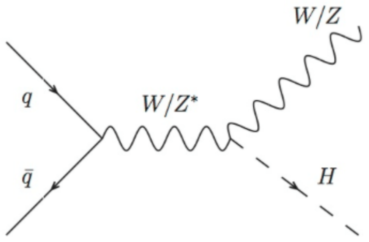


Figure 2.12: Feynman diagrams for the production of the Higgs boson in association with a vector boson (W,Z)

2.4.3 Fourth generation of heavy quarks

The Standard Model currently describes the presence of three fermion families, however the number of generation is not fixed by the theory. The asymptotic freedom constraint from QCD only sets an upper limit to the number of families equal to 9. Measurements of the hadronic production cross section at the LEP experiment [13] predicted the number of light neutrino families. Neutrino counting at the Z pole shows that the number of generations with light neutrinos is equal to 3. However additional heavier neutrinos cannot be excluded. Therefore it is meaningful to search for a possible fourth SM familiy. Thus the possible existence of a fourth generation remains an important subject for experimental investigation because of the significant impact on the Standard Model parameters and possible solution to solve open problems of the Standard Model. Additional massive quarks may, indeed, provide an answer to the problem of the matter-antimatter asymmetry in the Universe [88].
Different models propose the extension of the number of generations of quarks

and leptons adding a new doublet with a $t^{'}$ and $b^{'}$. The Standard Model production of Z boson and b quark is important because it represents the main background for the process (shown in fig.2.13):

$$b^{'} \rightarrow Zb, \tag{2.15}$$

Searches for new fourth generation heavy quarks have been performed both by ATLAS [89] and CMS experiments. ATLAS experiment has presented a search for pair production of heavy down-type quarks in data corresponding to an integrated luminosity of 1.04 fb^{-1} from proton-proton collisions at a mass center energy of 7 TeV. A heavy down-type quark with mass less than 480 GeV has been excluded at the 95% confidence level. A search for pair-produced heavy vectorlike charge-2/3 quarks, T, in proton-proton collisions at a center of mass energy of 7 TeV, is performed with the CMS detector. A T quark with a mass less than 475 GeV is excluded at the 95% confidence level.

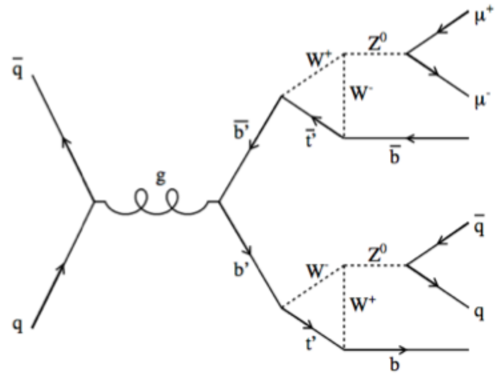


Figure 2.13: Decay chain in $q\bar{q}$ anihilations of the $b^{'}$, 4th generation of b quarks.

Chapter 3

The CMS Experiment at LHC

The Standard Model of particle physics has successfully undergone many precision tests at high energy colliders. Extensive consistency and precision tests have been performed, providing stringent constraints on the Standard Model parameters over a wide range of energy. However, as already discussed, there are still missing elements and different aspects do not find a satisfactory explanation within this theory. In order to seek answers to fundamental particle physics questions new energy regimes have to be explored. The Large Hadron Collider (LHC) has been built up to make these new territories accessible in high energy physics. LHC is a proton-proton collider and offers the possibility to span over a wider energy spectrum and also to reach higher production rates. The accelerator principles and characteristics will be briefly reviewed in the first part of this Chapter. In the second part, the experiment this work is involved in, the Compact Muon Solenoid (CMS) is described in more detail.

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [90] is the largest and most powerful hadron collider ever built and it has been designed to run for the next two decades. It has been installed in the underground tunnel which housed the Large Electron Positron Collider (LEP) [13] in operation until 2000 at CERN laboratories in Geneva, Switzerland. The transition from a leptonic collider to a hadronic collider has brought the advantage to build a machine with the same size of the previous one, but that can reach a higher energy in the center of mass frame because of a lower amount of energy loss through synchrotron radiation emitted by the accelerated particles. Two counter circulating proton beams flow in the 27 km LHC ring, at a depth varying from 50 m to 175 m. LHC has been designed to operate at a center mass energy of $\sqrt{s}$ =14 TeV, therefore providing parton-parton collisions up to energy of about 1 TeV, and with a very high collision rate in order to favour the study of the production of rare particles. These features are particularly important for a machine dedicated to the discovery of "new" physics as well as to test the limits of the current Standard Model of Particle Physics. Also heavy ions (lead nuclei) may be accelerated by LHC at an energy of 2.76 TeV/nucleon in the center of mass frame. To reach such a high energy, the proton beams acceleration is performed in separated stages. The proton source is a duo-plasmatron: the protons are produced by removing electrons from a source of hydrogen gas and then sent to the LINAC2 [91], a linear accelerator 36 m long that, using Radio Frequency Quadrupoles (RFQ) and focusing quadrupole magnets, generates a 30 μs pulsed beam with an energy and intensity respectively equal to 50 MeV and 180 mA. The beam is subsequently sent to the Proton Synchrotron Booster (PSB) [91] the first circular accelerator of the series, consisting of four superimposed synchrotron rings of circumference equal to 157 m, in which the energy of the protons is increased to the value of 1.4 GeV. Then the protons are injected into the Proton Synchrotron (PS) [91], a single synchrotron ring four times long the PSB (628 m) where the energy is increased to the value of 25 GeV. The sequential combination of these two synchrotrons allows also to create a series of proton bunches interspersed by 25 ns (i.e. at the frequency of 40 MHz) as required for the final correct operation of LHC. The final injection stage of protons at the LHC is the Super Proton Synchrotron (SPS) [91], a synchrotron with a circumference of approximately 7 km in which the energy of the protons is brought to the value of 450 GeV. Subsequently protons are extracted and injected into the LHC ring via two transmission lines, in order to form two beams that run in opposite directions and that are accelerated up to the energy of interest. The LHC is composed of two synchrotron storage rings of superconducting magnets, with a common cryogenic system but independent vacuum pipes, and opposite magnetic dipole fields. Along them two beams of protons (or heavy ions such as Pb) circulate in opposite directions, intersecting at four interaction points where the experiments are located: ALICE [92], ATLAS [93], CMS [94] and LHCb [95]. In fig. 3.1 a schematic view of the accelerator complex installed at CERN is given while in fig. 3.2 a picture of the area occupied by the LHC ring is shown.

The collision rate is quantified in terms of the instantaneous luminosity (number of collisions per unit time and transverse section of the beams), which depends only on the beam parameters and can be written for a Gaussian beam distribution as:

$$L = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi \epsilon_n \beta^*} F, \tag{3.1}$$

where N_b is the number of particles per bunch, n_b the number of bunches per beam, f_{rev} the revolution frequency, γ_r the relativistic gamma factor, ϵ_n the normalized transverse beam emittance, β^* the beta function at the collision point and F the geometric luminosity reduction factor due to the crossing angle at the interaction point. The number of events per second generated in the LHC collisions is given by:

$$N_{event} = L\sigma_{event}, \tag{3.2}$$

where σ_{event} is the cross section for the event under study and L is the machine luminosity.

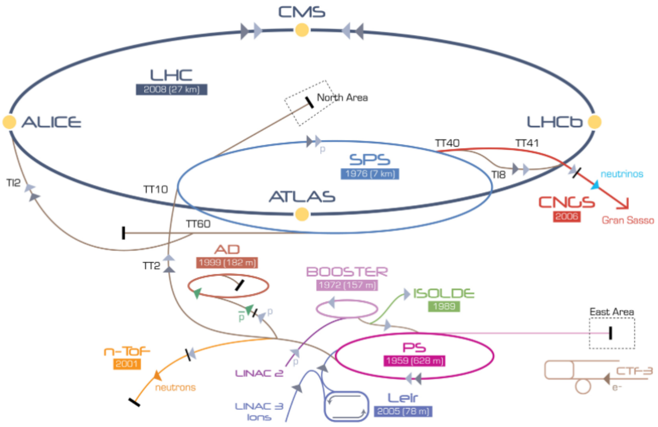


Figure 3.1: Schematic description of the accelerator complex installed at CERN.

Figure 3.2: Panoramic view of the area occupied by the Large Hadron Collider ring.

3.2 The CMS Detector

The Compact Muon Solenoid (CMS) [94] experiment is one of the two general-purpose detectors installed at the LHC. It aims to explore the phenomena of electroweak symmetry breaking, to perform QCD precision measurements and also to search for possible new physics beyond the Standard Model, such as supersymmetry, extra dimensions etc. The wide physics program has imposed stringent requirements and the design of CMS had to meet them:

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

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

- Good electromagnetic energy resolution, good diphoton and dielectron mass resolution (about 1% at 100 GeV), wide geometric coverage ($\mid \eta \mid < 2.5$), measurement of the direction of photons and correct localization of the primary interaction vertex, π^0 rejection and efficient photon and lepton isolation at high luminosities.

- Good E_T^{miss} and dijet mass resolution, requiring hadron calorimeters with a large hermetic geometric coverage ($\mid \eta \mid <5$) and with fine lateral segmentation ($\Delta\eta \times \Delta\phi < 0.1 \times 0.1$).

These characteristics allow to detect cleanly the diverse signatures from the several channels studied, identifying and precisely measuring muons, electrons, photons and hadrons over a large energy range. In a high energy and high luminosity machine like LHC, the design of the detector has been a really challenging task. At the designed characteristics of LHC, the number of collision events observed by the detector will be approximately 10^9 events/s. The on-line event selection system must be designed to reduce it to a reasonable size, 100 events/s. Moreover an important problem to face is the number of minimum bias events (pileup events) superimposed to the event of interest. To reduce the pileup effect the detector must have high granularity and good time resolution. In addition, the detector needs to be radiation-hard as the large flux of particles emerging from interaction point leads to high radiation levels. The structure of CMS is typical of general purpose collider detectors. It consists of several cylindrical detecting layers, coaxial with the beam direction (barrel region), closed at both ends with disks (endcap region). In fig. 3.3 a schematic view of the CMS detector is provided. From the inner region to the outer one, the various components of CMS (briefly described in the following sections) are:

- Silicon Tracker: it is placed in the region $r < 1.2$ m and $\mid \eta \mid < 2.5$ It consists of a silicon pixel vertex detector and a surrounding silicon microstrip detector, with a total active area of about 215 m^2. It is used to reconstruct charged particle tracks and vertices;

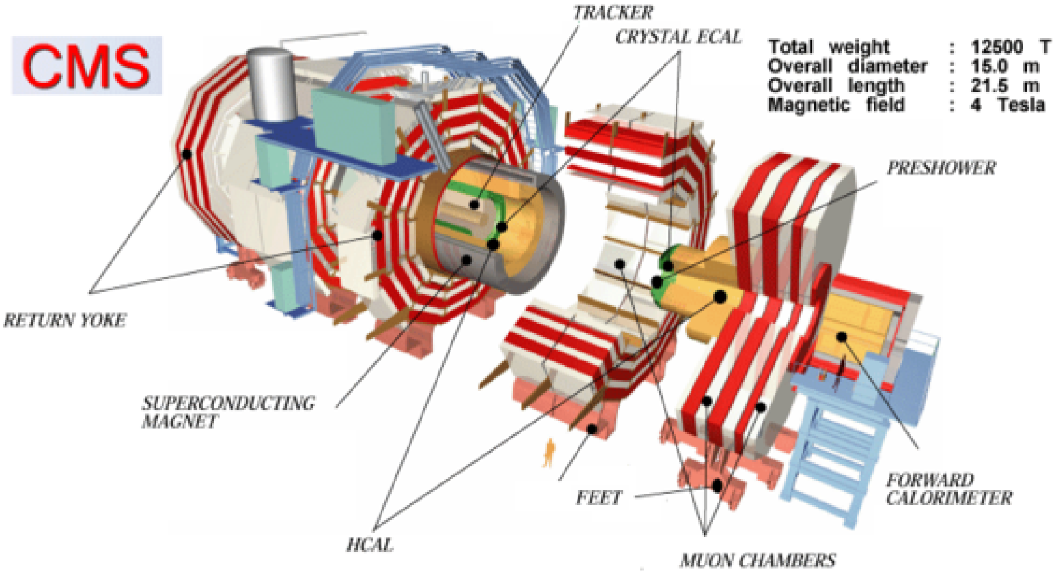


Figure 3.3: A view of the CMS detector with its subdetectors labeled.

- Electromagnetic Calorimeter (ECAL): it is placed in the region 1.2 m $< r <$ 1.8 m and $|\eta| < 3$. It consists of scintillating crystals of lead tungstate ($PbWO_4$) and it is used to measure the trajectory and the energy released by photons and electrons;
- Hadron Calorimeter (HCAL): it is placed in the region 1.8 m $< r <$ 2.9 m and $|\eta| < 5$. It consists of brass layers alternated with plastic scintillators and it is used to measure the direction and the energy released by the hadrons produced in the interactions;
- Superconducting Solenoidal Magnet: it is placed in the region 2.9 m $< r <$ 3.8 m and $|\eta| < 1.5$. It generates an internal uniform magnetic field of 3.8 T along the direction of the beams, necessary to deflect the charged particles in order to allow a measurement of their momentum through the curvature observed in the tracking system. The magnetic field is closed with an iron yoke 21.6 m long with a diameter of 14 m, where a residual magnetic field of 1.8 T is present, in the opposite direction with respect to the 3.8 T field;
- Muon System: it is placed in the region 4 m $< r <$ 7.4 m and $|\eta| < 2.4$. It consists of Drift Tubes (DT) in the barrel region and Cathode Strip Chambers (CSC) in the endcaps. A complementary system of Resistive Plate Chambers (RPC) is used both in the barrel and in the endcaps. This composite tracking system for muons is used to reconstruct tracks

released by muons that pass through it. The muon chambers are housed inside the iron (high magnetic permeability) structure of the return yoke that encloses the magnetic field.

3.2.1 Coordinate System

The coordinate system adopted by CMS has its origin situated at the nominal collision point, the y axis is directed vertically upward, the x axis points radially toward the center of the LHC while the x axis points along the beam direction toward the Jura mountains from LHC Point 5. The cylindrical symmetry of CMS design and the invariant description of proton-proton physics suggest the use of a pseudo angular reference frame, given by the triplet (r, ϕ, η), where r is the distance from the z axis, the azimuthal angle ϕ is measured from the x axis in the $x - y$ plane, η is the pseudorapidity defined as:

$$\eta = -\ln(\theta/2). \tag{3.3}$$

The polar angle θ is measured from the z axis. The pseudorapidity, for high energies, is a good approximation of the rapidity (y) of a particle, defined as:

$$y = \frac{1}{2}\ln(\frac{E + p_L}{E - p_L}), \tag{3.4}$$

where E is the particle's energy and p_L is the component of its momentum projected along the beam axis. The rapidity is a useful variable at a hadron collider, as differences in rapidity (Δy) are invariant under Lorentzian boosts along the axis direction. It follows from its definition that the pseudorapidity is null for θ =0 (see fig. 3.4) and increases in absolute value when approaching the beam pipe, asymptotically reaching infinity at $\theta = \pi/2$ (z axis) The collision in a proton-proton collider involves the partons inside the protons. The total four-momentum conservation would require the knowledge of the initial four-momentum of the quarks inside the protons. However the longitudinal component of the quark momentum inside the proton is not known

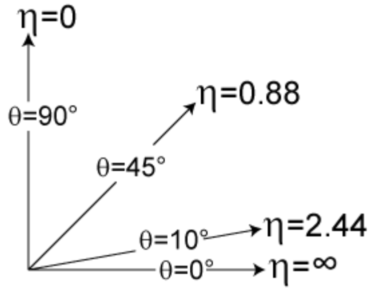


Figure 3.4: Variation of the pseudorapidity as a function of the polar angle θ. As θ approaches zero, pseudorapidity tends towards infinity.

a priori. Instead, in the transverse plane it is possible to to apply the four momentum conservation as the sum of the transverse momenta of the proton constituents is zero. A complete description of the particles kinematics is so provided defining the transverse kinematic variables:

$$p_T = p \cdot \sin(\theta), \tag{3.5}$$

$$E_T = E \cdot \sin(\theta), \tag{3.6}$$

where p_T is the component of a particle's momentum in the plane transverse to the beam axis and E_T its transverse energy.

3.3 Inner Tracking System

The Silicon Tracker [96] is the subdetector closest to the interaction point and covers the region $\mid \eta \mid < 2.5$ with a r <120 cm. It is formed by a Silicon Pixel detector surrounded by a Silicon Microstrip detector. Its purpose is to reconstruct the charged tracks with the highest possible resolution and to allow a precise determination of the position of secondary vertices, basic requirement to study processes that involve particles with short mean life (in particular, hadrons containing the b quark that decays after a few hundred μm). The large number of pileup events that overlaps with the event of interest, especially at the designed luminosity of operation ($10^{34}\text{cm}^{-2}\text{s}^{-1}$), can make pattern recognition a complex problem. In order to facilitate the pattern recognition two requirements are fundamental:

- low detector occupancy
- large hit redundancy

The first is achieved with a high granularity that assures clean track identification, reducing the occupancy and the overlap probability of the hits. In principle, the other requirement can be satisfied using a large number of detecting layers but it is needed to find a compromise between the redundancy of measured points per track and the amount of material composing the tracker. The inner silicon tracker must comply with severe material budget constraints in order not to degrade its momentum resolution and also not to reduce the performance of all the detectors. In fig. 3.5 the material budget of the tracker is reported in units of radiation length t/X_0 and in units of nuclear interaction length t/λ_I as a function of η. Moreover one of the most important problems arised during the planning of the tracking system was to design a detector able to operate in a harsh radiation environment. Radiation is the main reason of bulk damage that compromises the operation, in particular, of the pixel detector that is the innermost part of it, closest to the interaction point.

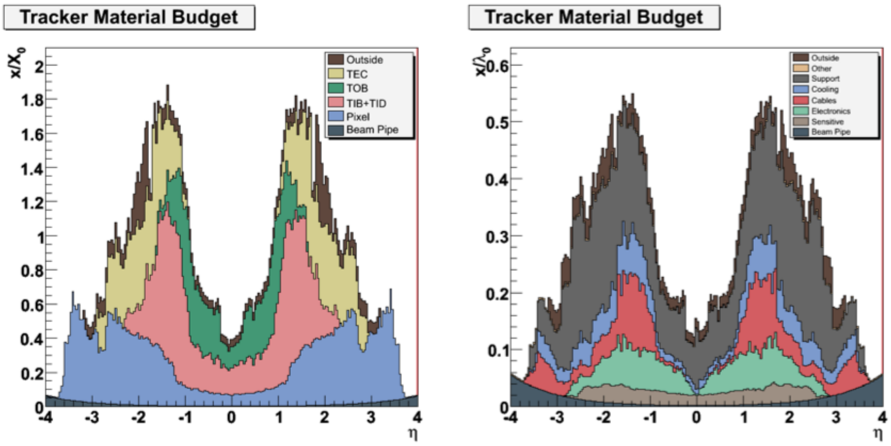


Figure 3.5: Material budget plots of the tracker for each subdetector in units of radiation length t/X_0 and in units of nuclear interaction length t/λ_I as a function of η, for the nominal geometry of the detector.

3.3.1 The pixel Detector

The silicon pixel detector(fig. 3.6) is the innermost active section of CMS. It is mainly used in CMS as a starting point for the reconstruction of the tracks providing on average three spatial measurement points that are used as initial seeds. It plays also a key role in the reconstruction of the primary vertex and of any possible secondary vertices. It consists of three barrel layers with two endcap disks on each side on them. The 3 barrel layers are located at mean radii of 4.4 cm, 7.3 cm and 10.2 cm, and have a length of 53 cm. Each half-cylinder is composed of ladders and half ladders that serve as support and cooling structure for the modules of pixels, with each ladder containing 8 modules. In total, the barrel is composed of 768 modules. The 2 end disks are placed on each side at $\mid z \mid$ =34.5 cm and 46.5 cm and extend from 6 to 15 cm in radius in such a way that each track included in the detector acceptance passes through at least two layers. Each disk is divided into 24 segments, on each of which 7 modules of different sizes are mounted, for a total of 672 modules on all the endcaps. Each module is composed of several units that contain a highly integrated and segmented silicon sensor with a thickness of 250 μm. The sensors implement the n-in-n concept, with pixels consisting of n^+ implant over a n-type substrate sensor, with a size of 100 × 150 μm and are combined with analog signal readout to profit of charge sharing effects among pixels and improve position resolution by interpolation. The choice of using pixel sensors guarantees the required high spatial precision, allowing resolution in the hit reconstruction of about 10 - 15 μm in the barrel and about 15 μm in the endcaps.

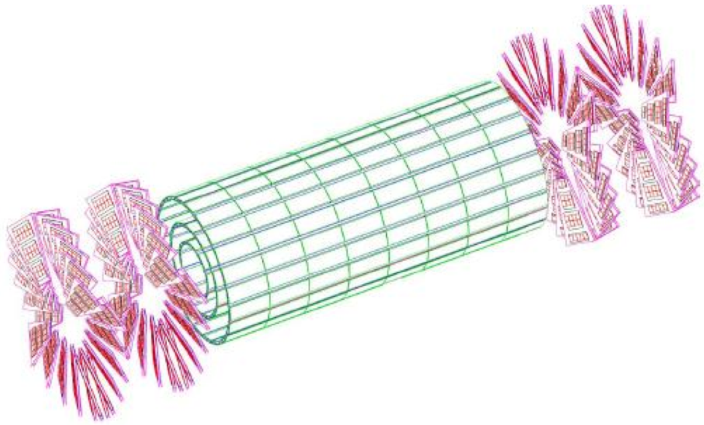


Figure 3.6: A schematic view of the pixel vertex detector.

3.3.2 The Silicon Strip Tracker

The silicon strip detector covers the radial region between 20 and 116 cm, and consists of three systems: the Tracker Inner Barrel (TIB) and Disks (TID), the Tracker Outer Barrel (TOB), and the two Tracker EndCaps (TEC+, TEC-). In fig. 3.7 a schema of the silicon strip tracker is provided. The TIB extends to a radius of 55 cm, and is composed by four cylindrical layers completed by the three TID disks at each end. The TIB covers up to $\mid z \mid < 65$ cm while, with six layers, TOB extends in -110 cm $< z <$ 110 cm region. The amount of the radiation received by each subsystem changes as a function of the distance to the interaction point. Many different silicon micro-strip sensor designs have been used for each of the them, adapted to the particle flux, to the required spatial resolution, and to the surface that needs to be covered by instrumentation at the corresponding radial distances from the interaction region. The inner part implements micro-strips with a pitch of 80 μm in the two innermost layers, 120 μm in the two outermost layers, and between 100 and 141 μm in the disks. Micro-strips with larger pitches, up to about 185 μm, have been installed in the TOB and in the TEC detectors. The $r-\phi$ resolution for a single point ranges from 23-34 μm in TIB (which has smaller sensors) and 35-52 μm in TOB. In z direction, the single-point resolution is 23 and 52 μm in TIB and TOB, respectively. Each TEC is composed of 9 disks that extend into the region 120 cm $<\mid z \mid<$ 280 cm, and each TID consists of three small disks that fill the gap between the TIB and the TEC. The whole tracking system is positioned in a container with 5.4 m length and 2.4 m diameter and operates at a temperature of about -20 ^{o}C. At low temperature, the radiation damages on the silicon sensors are frozen. Hence the detector quality will not gradually decrease.

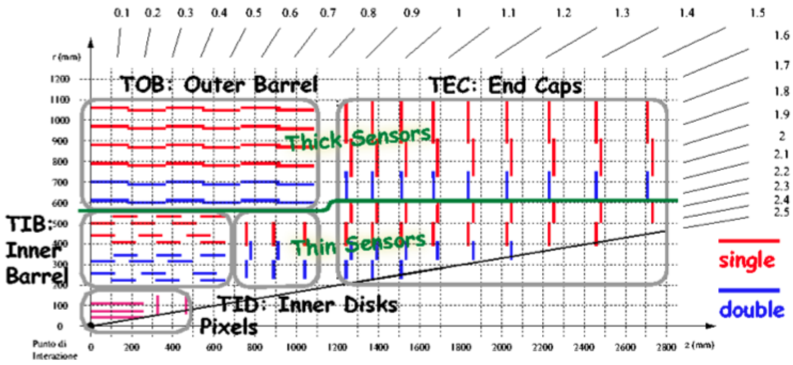


Figure 3.7: An $r - z$ schematic view of a sector of the Silicon Strip Tracker. The location of single sided and double sided detectors is put into evidence.

3.4 Electromagnetic Calorimeter

The investigation of electroweak symmetry breaking is one of main searches in the broad physics programme of the CMS experiment. The direct search of the Higgs Boson has been conducted in different channels in a wide mass range. The two-photon decay mode has been one of the most sensitive and important channel in the discovery of this particle. The Electromagnetic Calorimeter (ECAL) [97] has been designed to provide high resolution measurements of electron and photon energy to enhance the capability to detect the Higgs boson decay into two photons. ECAL (fig. 3.8) is a hermetic, fine grained, homogeneous calorimeter, comprising 75,848 lead tungstate scintillating crystals, located inside the CMS superconducting solenoidal magnet. It is divided into an Ecal Barrel (EB) with 61200 crystals and two Ecal Endcaps (EE), each containing 7324 crystals. The barrel inner radius is 129 cm, while the

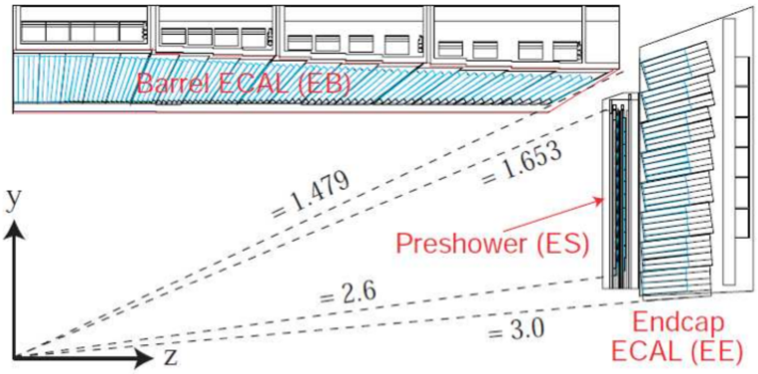


Figure 3.8: Schematic view of ECAL layout.

length is 630 cm and the pseudorapidity extent is $\mid \eta \mid < 1.479$. It consists of 36 supermodules, each one with a length equal to the half of the barrel length. Each supermodule is a 20 × 85 crystal matrix on the plane $\phi - \eta$. Supermodules are divided into 4 modules along the η direction and each module is in turn divided into submodules. Each submodule, consisting in a 5×2 crystal arrays mounted on a glass fiber structure, forms the elementary unit of EB. Crystals in the barrel region have a truncated pyramidal shape with a front face cross-section of about 22×22 mm^2 with a length of 23 cm, corresponding to 24.7 radiation length (X_0). The granularity of a single crystal is equal to $\Delta\eta \times \Delta\phi = 0.0175\times\ 0.0175$ (about 1^0). Crystals are grouped into 5×5 matrices called trigger towers whose information is used by the trigger system. To avoid that cracks might align with particle trajectories, the crystal axes are tilted by 3^0 with respect to the direction from the interaction point both in ϕ and η. Each endcap covers the region $1.479 <\mid \eta \mid < 3$ and consists of two halves called Dees. All the crystals have the same shape and they are grouped in structures of 5×5 crystals called supercrystals. The $\Delta\eta \times \Delta\phi$ granularity in the endcaps varies from 0.0175× 0.0175 to 0.05× 0.05. Unlike in the barrel, where the crystals are arranged in a $\eta - \phi$ geometry, the endcap crystals are arranged in a $x - y$ geometry. To ensure good hermeticity, the outer perimeter of the ECAL endcaps has been studied in order to give an overlap of half crystal between the barrel and the endcaps. In order to improve the π^0/γ separation and vertex identification, two preshower detectors are placed on the inner side of the endcaps. The preshower, which covers the region $1.653 <\mid \eta \mid < 2.6$, is a sampling calorimeter consisting of two lead converters (2 X_0 and 1 X_0 thick respectively) followed by silicon strips with a pitch of less than 2 mm. The strips following the two absorbers are disposed in orthogonal way. The presence of a preshower (a total of 3 X_0 of lead) in the endcap region allows the use of slightly shorter crystals (22 cm), keeping the total radiation length more than 26 X_0. The choice of the $\mathrm{PbWO_4}$ [98] as active medium in ECAL is due to several reasons. The high-density ($\rho = 8.3\ \mathrm{g/cm^3}$), the short radiation length ($X_0 = 0.89$ cm) and the reduced Moliere radius ($R_M = 2.2$ cm) allow to build a compact and high granularity calorimeter. Moreover, the 15 ns decay scintillation time allows to collect about 80% of the emitted light during the 25 ns that exist between two consecutive beam interactions in the LHC design conditions. Finally, the $\mathrm{PbWO_4}$ crystals have a good intrinsic radiation hardness and therefore they can operate for years in the hard LHC environment, with a modest deterioration in performance. The disadvantage of this type of crystals is the low light yield (about 10 photoelectrons/MeV) which makes an internal amplification for the photodetectors necessary. Silicon avalanche photodiodes (Avalanche Photodiodes, APD) are used in the barrel while single stage photomultipliers (Vacuum Photo-Triode, VPT) are used in the endcaps. The ECAL barrel energy (E) resolution for electrons in beam tests has been measured to be [99]:

$$\frac{\sigma_E}{E} = \frac{2.8\%}{\sqrt{E(\mathrm{GeV})}} \oplus \frac{12\%}{E(\mathrm{GeV})} \oplus 0.3\%, \tag{3.7}$$

where the three contributions correspond to the stochastic, noise, and constant terms. This result was obtained reconstructing the showers in a matrix of 3×3 crystals where the electron impact point on the calorimeter was tightly localized in a region of 4×4mm to give maximum containment of the shower energy within the 3×3 crystal matrix. The noise term of 12% at 1 GeV corresponds to a single-channel noise of about 40 MeV, giving 120 MeV in a matrix of 3×3 crystals. The constant term, which is the dominant contribution to the energy resolution for high-energy electron and photon showers, depends on non-uniformity of the longitudinal light collection, energy leakage from the back of the calorimeter, single-channel response uniformity and stability. The beam test setup was without magnetic field, no inert material in front of the calorimeter, and accurate equalization and stability of the single-channel response (better than 0.3%) [100]. During CMS operation, the contributions to the resolution due to detector instabilities and to the channel-to-channel response spread must be kept to within 0.5%, in order to retain the excellent intrinsic resolution of the ECAL.

3.5 Hadronic Calorimeter

The calorimetric system of CMS is completed with a sampling hadronic calorimeter [101]. It surrounds the ECAL completely, covering a pseudorapidity range up to $\mid \eta \mid$=5.2. Its purpose is to contain the showers of hadronic particles, and therefore measure jet four-momenta and the missing transverse energy of events. The two key features to meet these demands are a high hermeticity and a good transverse granularity. Furthermore, a good energy resolution and a sufficient longitudinal containment are also important. A longitudinal section of a quarter of the hadronic calorimeter is shown in fig. 3.9. The design of the HCAL has been constrained strongly by the size of the solenoid magnet surrounding it. It is formed by two separate detectors: a central (HCAL) covering the pseudorapidity range $\mid \eta \mid$<3 and an outer hadron calorimeter (HF) placed outside the solenoid complementing the barrel calorimeter. HCAL is divided into two subdetectors: a barrel ($\mid \eta \mid$<1.3) and two endcaps (1.3<$\mid \eta \mid$<3). It is a sampling calorimeter, with brass used as absorber and plastic scintillators as active material. The plastic scintillator is divided into 16 η sectors, resulting in a segmentation of $(\Delta\eta, \Delta\phi)$ = (0.087, 0.087). The energy resolution of the HCAL, for pions of energy E, is parametrized as:

$$\frac{\sigma}{E} = \frac{100\%}{\sqrt{E(\mathrm{GeV})}} \oplus 8\%. \tag{3.8}$$

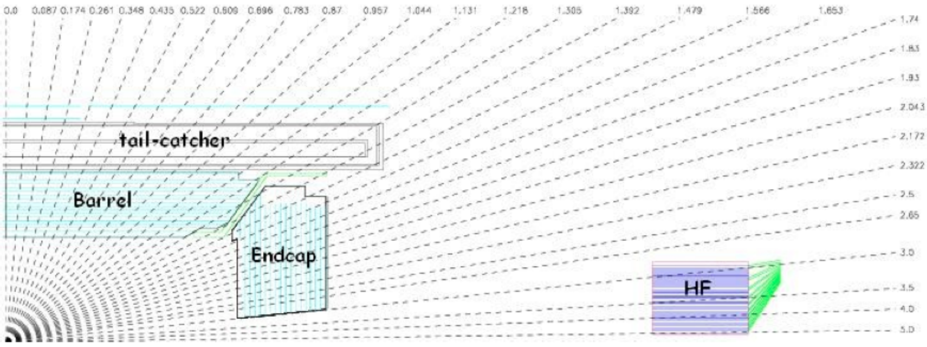


Figure 3.9: Longitudinal section of a quarter of the CMS hadronic calorimeters:HCAL is visible on the left, HF far away from the interaction point on the right. Some values of pseudorapidity are marked.

It has a total thickness of 7-10 interaction lengths (λ_i). An additional layer of active material is added behind the solenoid, as a complete containment of a high energetic shower is not possible in $7\lambda_i$. This layer increases the total effective thickness by about $3\lambda_i$ improving the energy resolution for 300 GeV pions of about 10%. The requirement to have a high hermetic calorimeter has driven the placement of an additional calorimeter (HF) outside the magnet yoke, 11 m away from the interaction point, on both sides. It covers the forward pseudorapidity region $3<|\ \eta\ |<5$. In order to sustain the very high doses of radiation and high particle multiplicities expected for such a region, sampling calorimeters made of iron and quartz fibers as the active medium, to detect the Cherenkov light emitted by showers, have been chosen. It is essentially a cylindrical steel structure with an outer radius of 130 cm. Its front face is situated 11.2 m from the interaction point. In addition, this subdetector plays a crucial role in the determination of luminosity, since coincidences between both directions along the z axis will indicate a collision has taken place.

3.6 The Muon System

The CMS Muon System [102] is dedicated to the identification and measurement of high p_T muons, in combination with the Tracker. The system is placed outside the magnetic coil, embedded in the return yoke, to exploit the 1.8 T return flux fully. The system consists of three independent subsystems, as shown in fig. 3.10:

- Drift Tubes (DT) are placed in the barrel region, where the occupancy is relatively low ($<$ 10 Hz/m^2).
- Cathode Strip Chambers (CSC) are in the endcaps, where the occupancy

is higher (> 100 Hz/m^2);

- Resistive Plate Chambers (RPC) are both in the barrel and in the endcaps.

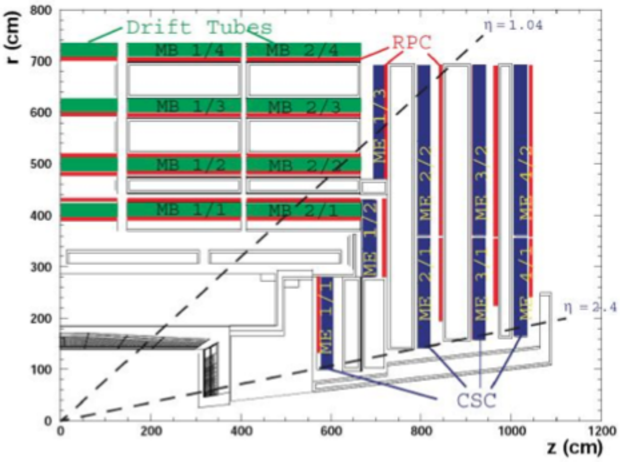


Figure 3.10: A schematic view of a quadrant of the CMS Muon System.

The Drift Tube system is placed in the barrel region $\mid \eta \mid < 1.2$, where the particle flux is low enough (about 10 Hz/cm^2) and the magnetic field is enough weak and homogeneous (1.8 T). The Drift Tube system layout follows the yoke segmentation and consists of 5 iron wheels composed of 12 azimuthal sectors, covering an angular region of approximatively 30^0 each. Every wheel consists of four concentric rings of chambers, called stations. Each station is formed by 12 DT chambers. The basic detector element of the DT system is a rectangular drift tube cell with a transverse size of 13×42 mm^2 and whose length varies from 2 to 4 m. Chambers are filled with a gas mixture of Ar (85%) and CO_2 (15%) and are grouped in the parallel direction to form detection layers. Groups of four layers are assembled to form a superlayer. In each superlayer two chambers have anode wires parallel to the beam axis, two have perpendicular wires. Thus, each superlayer can provide two measurements of the $r - \phi$ coordinate and two measurements of the z coordinate of the track hit positions. A schematic layout of a single DT chamber is shown in fig. 3.11. The position resolution is about 100 μm in both $r - \phi$ and $r - z$. The high magnetic field (up to 3.5 T) and the particle rate expected in the muon system endcaps do not allow to use drift tube detectors to perform measurements at large η values. Therefore, a solution based on Cathode Strip Chamber detector has been adopted. The CSC system is made of multiwire proportional chambers filled with a gas mixture of Ar (40%), CO_2 (50%), CF_4 (10%) in which the plane cathode is segmented into strips oriented radially

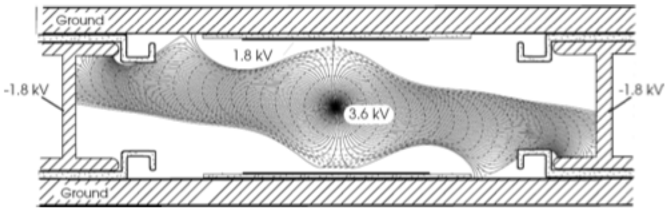


Figure 3.11: A schematic representation of a Drift Tube chamber. Drift lines in presence of magnetic field are also shown.

and transversely with respect to the anode wires (fig. 3.12). This allows the simultaneous measurement of two coordinates (r through the wires, ϕ using the strips). The drift path of the charge carriers is shorter with respect to

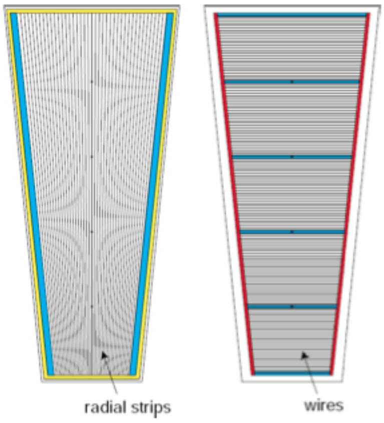


Figure 3.12: A schematic representation of CSC cathode panel (left) and anode panel (right).

the drift tubes, therefore these detectors can be placed in regions with higher flows of charged particles and less homogeneous magnetic fields. The CSC is composed of 4 superimposed disks (called stations), mounted on the iron disks of the return yoke and orthogonal to the direction of the beams. Each station is formed by two rings (three for the innermost station), divided into 18 or 36 CSC with trapezoidal shape. Each CSC used in the system consists of six layers of sensitive wires. The strip spatial resolution is about 80-85 μm while r can be determined with a precision of 0.5 cm. The CSC coverage is $0.8 < | \eta | < 2.4$. Resistive Plate Chambers (fig.3.13) are used both in barrel and endcaps, complementing DT and CSC systems, in order to ensure robustness and redundancy to the muon spectrometer. The RPC system is made of gaseous detectors characterized by a coarse spatial resolution, but are able to perform precise time measurements, comparable with the ones provided by scintillators.

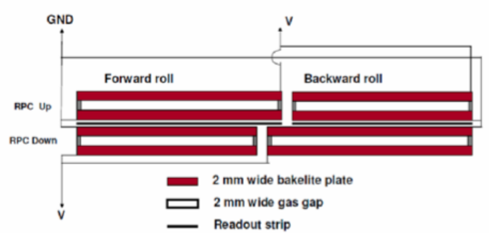


Figure 3.13: Schematic layout of a double-gap barrel RPC chamber composed of two sub-units, called rolls. The readout strip plane is also shown.

This ensures precise bunch crossing identification to the muon trigger system. CMS uses double-gap RPC chambers composed of 4 bakelite planes alternated to form two gas gaps 2 mm thick. The planes are coated by graphite to make electrodes that are set at 9.5 kV of potential difference. The central part of the chamber is equipped with insulated aluminum strips, used to collect the signal generated by crossing particles. In the barrel the strips are rectangularly segmented (12.1 to 41 cm wide and 80 to 120 cm long) and run along the beam axis, whereas the endcaps are equipped with trapezoidal shaped strips covering approximatively the range $\Delta\phi = 5 - 6^0$, $\Delta\eta$=0.1. The gap between the plates is filled with a mixture of $C_2H_2F_4$ (94.5%) and i-C_4H_{10}. In order to sustain higher rates, the detector operates in avalanche mode. The RPC coverage is $\mid \eta \mid < 2.1$.

3.7 Trigger system

LHC produces interactions at a frequency of up to 40 MHz, but only a small fraction of this enormous amount of data but can be written on disk. Moreover the vast majority of interactions are soft collisions, which are not interesting for the search-oriented CMS physics program. Therefore a complex trigger system analyses each bunch crossing and applies a first but rather sophisticated event selection, which reduces the rate to few hundreds of Hz. CMS has chosen a two-level trigger system, consisting of a Level-1 Trigger (L1) [103] and a High Level Trigger (HLT) [104]. Level-1 Trigger runs on dedicated processors, and accesses coarse level granularity information from calorimetry and muon system.Its task is to reduce the data flow from 40 MHz to 100 kHz. The L1 output rate is then further reduced down to the target rate of 300 Hz by the High Level Trigger. The HLT code can access the full granularity information of all the subdetectors. The design goal of the trigger system as a whole is to have a reduction rate capability of 10^7.

Level-1 Trigger

The Level-1 trigger reduces the rate of selected events down to 50-100 kHz. The full data are stored in pipelines of processing elements, while waiting for the trigger decision. The L1 decision whether taking or discarding data from a particular bunch crossing has to be taken in 3.2 μs; if the L1 accepts the event, the data are moved to be processed by the HLT. The time available to L1 trigger is not enough to read data from the whole detector therefore it employs the calorimetric and muon information only, since the tracker algorithms are too slow for this purpose. The L1 trigger is organized into a Calorimeter Trigger and a Muon Trigger, whose information is transferred to the Global Trigger which takes the final accept-reject decision. It accesses the information from the muon and from the ECAL systems to perform a first elementary muon and electron identification. For the electrons, it reads out the energy deposits in 5×5 matrices of ECAL crystals, the trigger towers, and compares their sum to a given threshold. A rudimentary isolation requirement can also be applied, taking into account the energy deposits in neighbouring towers and in the corresponding HCAL cells. Muons provide a clean signature, and can be reconstructed with good precision even at the hardware trigger stage. Hits in the muon stations are detected by the L1 electronics, and trajectory candidates are built in the DT and CSC systems using a fast track finding algorithm. The four candidates with highest p_T and best quality are combined with the information from the RPC to build a L1 global muon object, which has to pass a given p_T threshold. As well as for electrons, additional requirements on the isolation can be applied, using the information from the calorimeters.

HLT Trigger

The High-Level Trigger reduces the output rate to about 100 Hz. The HLT software is organized in a set of algorithms (known as HLT “paths”) which are designed to select specific event topologies. It starts from the L1 objects, and exploits high resolution data from all subdetectors to perform a partial reconstruction of the event in three sequential stages:

- Level 2: accesses only the muon and calorimetric information;
- Level 2.5: the data of the pixel seeds are added;
- Level 3: makes use of the full information from all the tracking detectors.

To limit the CPU usage, the most intensive steps, such as tracker hit reconstruction and tracking, are applied only in the limited detector regions around the L1 identified candidates. Interaction vertices and high level objects, such as τ leptons and b quark jets, are identified using simplified and faster versions of the algorithms used in the offline data processing.

Chapter 4

Event Reconstruction

In a proton-proton collider, in addition to a hard parton scattering with large transverse momentum transfer, also underlying events contribute to the final state. The complexity of the collision events has imposed several constraints in the features of each subdetectors in order to guarantee high efficiency in the reconstruction of the final physics objects. A large magnetic field in order to maximise the separation between charged and neutral hadrons, a calorimetry with an excellent resolution and granularity, an efficient tracking system and also a particular attention to the material budget in front of the calorimeters are therefore necessary characteristics. An efficient object reconstruction implies also the development of sophisticated reconstructed algorithms providing a complete description of the final state. In this Chapter the technique adopted by the CMS experiment in order to provide a detailed knowledge of the final objects is described. A section has been dedicated to the description of the specific algorithms applied in order to reconstruct all the interested objects: electrons, muons. The jet reconstruction algorithms of interest will be reviewed and the relevant b tagging algorithms will be described.

4.1 Particle Flow

In order to give a complete picture of the final state, an algorithm called *Particle Flow* [105] [106] has been developed by the CMS experiment and is nowadays employed in most analyses. It allows to identify and reconstruct each particle arising from the LHC proton-proton collision combining the information provided by each subdetector with an optimal determination of their direction, energy and type. Charged hadrons, photons, neutral hadrons, muons, and electrons are identified and can then be used to construct a variety of higher-level particle based objects and observables such as jets, missing transverse energy (MET), lepton and photon isolation, τ identification and b jet tagging. The CMS Particle Flow algorithm relies on efficient track reconstruction, on clustering algorithms that are able to properly treat the overlap between different showers, using a specific procedure to connect together the deposits of each particle in the sub detectors. The main features of these algorithms can be summarized as follows. The muons are identified beforehand by combining the tracker and muon chamber information so that their track does not give rise to a charged hadron. The matching between a cluster in the hadronic calorimeter and a track constitutes a charged hadron and the building bricks are not considered anymore in the rest of the algorithm. Specific track reconstruction methods [107] are used for the electrons because the Bremsstrahlung photon emission obliged to a dedicated algorithm able to merge the photon clusters to the electron avoiding energy double counting. Once all the tracks have been treated, the remaining clusters are supposed to come from photons if the cluster is formed in the Electromagnetic calorimeter, or from neutral hadrons if the cluster is situated in the Hadronic calorimeter. The information from all the subdetectors is then combined to extract a measurement of the four-momentum of each particle once all the deposits have been associated. The elements used in the Particle Flow algorithms are therefore charged particle tracks, calorimeter clusters, and muon tracks that must be delivered with a high efficiency and a low fake rate. Advanced tracking and clustering algorithms have been developed to fulfil this task.

4.1.1 Track reconstruction

The presence of multiple interactions (pileup) and the finite resolution of the detector imply the contribution of prior or later bunch crossings. Reconstructing tracks in such a high-occupancy environment is really challenging and it is difficult to have a high track-finding efficiency, while keeping the fraction of misidentified reconstructed tracks (fake tracks) small. A good reconstruction efficiency is needed for tracks with very low p_T (about 100 MeV) in order to obtain optimum jet energy resolution with particle flow techniques. Furthermore, excellent impact parameter resolution is needed for a precise measurement of the positions of primary pp interaction vertices as well as for identifying b-quark jets [108]. The track reconstruction gives estimates for the momentum and position parameters of the charge particles starting from the hits in the tracking system. The tracking software adopted in CMS is called *Combinatorial Track Finder* (CTF) that is an extension of the combinatorial Kalman filter [109] [110]. An iterative method is used to reconstruct tracks [111]. The idea is to reconstruct first the tracks easier to find (those with large p_T and near the interaction region) removing the hits associated with these tracks. In this way the combinatorial complexity is reduced and the subsequent iterations are simplified, improving the search of more difficult classes of tracks (those with low p_T or greatly displaced tracks). First, tracks are seeded (2 or 3 hits) and reconstructed with very tight criteria, obtaining a negligible fake rate and a moderate tracking efficiency. The hits assigned to the tracks found are then removed and the looser track seeding criteria are taken. This approach increases the tracking efficiency while the fake rate is kept low thanks to the hit removal. In the first three iterations, tracks originating from within a thin cylinder 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 fourth and fifth iterations allow to reconstruct secondary charged particles originating from photon conversions and nuclear interactions in the tracker material and decay products of long-lived particles (K_S^0 or Λ) requiring relaxed constrains on the origin vertex. This iterative procedure allows to reconstruct charged particles with a small number of hits (3), a low transverse momentum less than 150 MeV and an origin vertex more than 50 cm distant from the beam axis, with a fake rate of the order of a percent.

4.1.2 Cluster reconstruction

The cluster algorithms have different requirements to satisfy: they have to detect and measure the energy and direction of stable neutral particles (photons and neutral hadrons), separate these neutral particles from energy deposits from charged hadrons, reconstruct the electrons recovering the energy loss because of the Bremsstrahlung photons emission, improve the energy measurement of charge hadrons, especially if an optimal estimation of the track parameters was not possible (high p_T particles). The clustering is performed separately in each subdetector: ECAL barrel, ECAL endcaps, HCAL barrel, HCAL endcap, PS first layer and PS second layer. The clustering algorithm starts with the identification of “cluster seeds” that are calorimetric cells with an energy greater than a given threshold. Then “topological clusters” are built from the seeds adding the cells with at least one side in common with a cell already in the cluster and if their energy is greater then a given value. This thresholds represent two standard deviations of the electronic noise in ECAL (800 MeV in the barrel and 300 MeV in the endcap) and in HCAL (800 MeV for all the region of the calorimeter).

4.1.3 Link algorithm

Once the tracks and the calorimetric clusters have been reconstructed, these elements need to be connected together, avoiding double counting as a particle can give rise to several particle flow elements. Each reconstructed track is extrapolated from its last measured hit in the tracker to calorimetric cluster. The matching between the track and the cluster is done if the extrapolated position in the corresponding calorimeter is within the cluster boundaries. The energy loss due to Bremsstrahlung photons emission is recovered extrapolating tangents to the track in the intersection points between the track and the tracker layer. If the tangent position is within the boundaries of a cluster, it is considered as Bremsstrahlung photons energy and so it is matched to the track. A muon track in the muon system is linked to the track in the tracker with a fitting procedure and the final object is called *global muon*. If the matching is possible with several tracker tracks, only the global muon that returns the smallest χ^2 is retained.

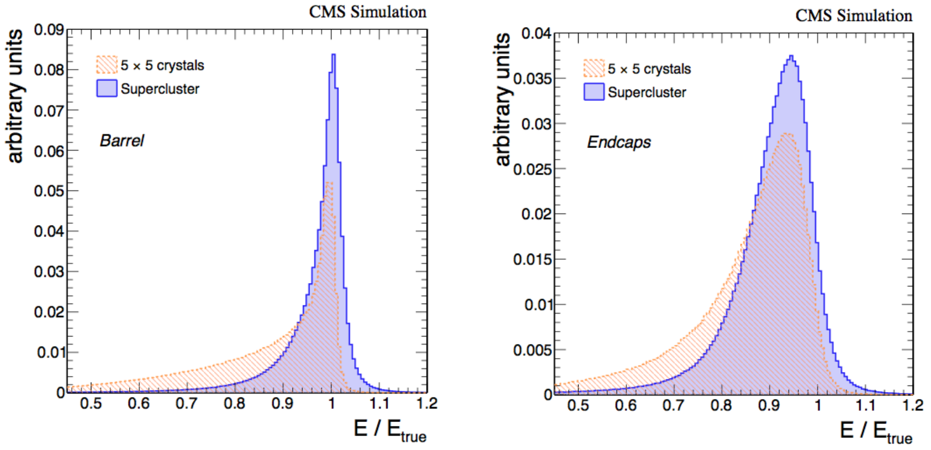
4.2 Electrons

The electron reconstruction and selection is of primary importance for many of the physics analyses performed with the CMS experiment, for this reason, a great effort has been dedicated to the implementation of dedicated algorithms [112]. Electrons are reconstructed by associating a track in the tracker with energy deposits in the electromagnetic calorimeter.

Clustering of the electron energy deposits in the ECAL

The energy released in the calorimeter is spread into several crystals, especially when the electron looses much energy because of the Bremsstrahlung photons emission. Two different algorithms have been developed in order to reconstruct the electron cluster in ECAL [113]. In the barrel the “hybrid” algorithm is used, it considers clusters formed from windows of 5 crystals in η around the most energetic crystal (seed) and a variable window in ϕ direction (up to 35 crystals wide). In the endcap, clustering is performed with the “multi5x5” algorithm that merges together matrices of 5x5 crystals around the most energetic crystal. In both cases the clusters are then grouped into superclusters if the energy is greater than a threshold. The supercluster energy is the sum of the energies of all the clusters contained in it while its position is obtained by taking a weighted mean calculated using the logarithm of the crystal energy, applying a correction with respect to the depth of the shower at the same time. In fig. 4.1 the effect of superclustering is shown in the comparison between the reconstructed energy using a matrix of 5x5 crystals around the seed and the one reconstructed within the supercluster.

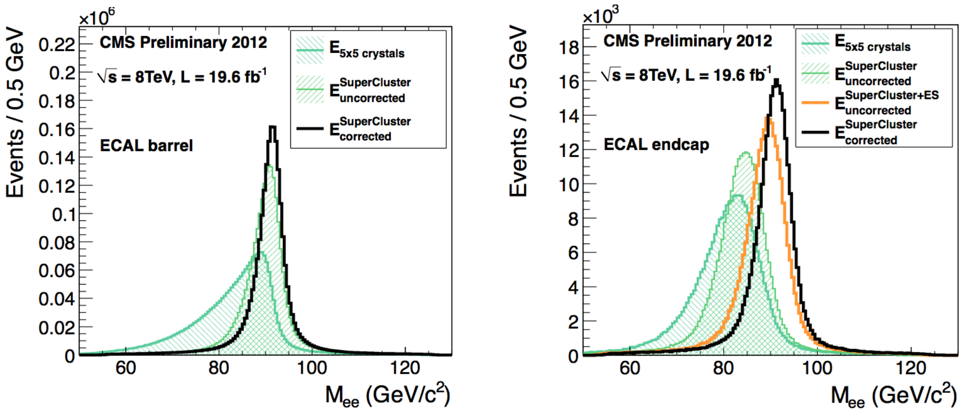
Several procedures are used to calibrate the energy response of individual crystals that enter in the clustering steps [114]. After clusterisation, the energy of the electron candidate is estimated from the sum of signal amplitudes (A_i) [115] of the individual channels that belong to the considered cluster, weighted with channel dependent coefficients to consider the variation in time of crystal transparency and photo-detector response ($L_i(t)$) and corrected to consider the intrinsic differences in response of crystals and photodetectors (c_i), the calibration of the ADC to energy conversion (G) and the presence of geometry effects ($F_{e,\gamma}$). Figure 4.2 shows the invariant mass of electron pairs in the $Z \to e^+e^-$ dacays using different levels of the energy reconstruction: fixed array 5x5, the raw supercluster energy and the supercluster energy including algorithmic corrections.



(a) Energy reconstructed within the supercluster compared to that obtained using a simple matrix 5×5 around the most energetic crystal in the barrel case.

(b) Energy reconstructed within the supercluster compared to that obtained using a simple matrix 5×5 around the most energetic crystal in the endcap case.

Figure 4.1: Ratio of the reconstructed raw energy over the generated energy, using the super- clustering compared to an energy collection in a matrix of 5×5 crystals, for simulated electrons from Z boson decays. No energy correction is applied on both the distributions.



(a) Invariant mass of electron pairs in the $Z \to e^+e^-$ decays when the electrons are reconstructed in the barrel.

(b) Invariant mass of electron pairs in the $Z \to e^+e^-$ decays when the electrons are reconstructed in the endcap.

Figure 4.2: The two plots show the impact on the $Z \to e^+e^-$ energy scale and resolution from the incorporation of more sophisticated clustering and cluster correction algorithms.

Electron Momentum Regression

The raw supercluster energy has to be corrected in order to take into account ~~possible inefficiencies of the clustering algorithm or energy lost in the tracker~~

(case of soft charged particles do not reach ECAL because of the magnetic field). Energy and position corrections that exploit the interaction mechanism of the electrons upstream of ECAL and the CMS geometry are needed. A multivariate regression technique based on a Boosted Decision Tree (BDT) implementation [114] is used in order to extract these corrections. These methods take as input some variables related to the shower shape and the global coordinate of the supercluster (η, ϕ). The shower shape variables considered are: the ratio between the 5x5 crystal energy and the raw supercluster energy (called R_9) that provides a way to identify electrons with little radiation in the tracker, the number of basic clusters, the ratio of hadronic energy behind the SC over the electromagnetic energy of the cluster. Other variables as the energy and position of the seed cluster and a number of crystal energy ratios are also included in order to take into account the degree of showering in the material. The η and ϕ variables of the seed crystal, the position of the seed cluster with respect to the crystal center and the seed cluster energy ratios allow to obtain an estimate of the amount of energy lost in the crystal and module gaps and drive the level of local containment corrections predicted by the regression. Finally, the number of primary vertices and the estimate of an average energy ρ in the event are included in order to correct residual energy scale effects from pileup. The electron-tuned regression is trained on a Drell-Yan sample and is validated by comparing the Monte Carlo to data performance for electrons using the decay of the Z boson.

Electron tracking

Electron tracks are not reconstructed using the standard Kalman Filter (KF) track reconstruction procedure. Because of the large radiative losses in the tracker material, this algorithm is not optimal for the electron case. Infact, it could lead to a reduction in the hit collection efficiency: hits can be lost because of a too large change of curvature due to Bremsstrahlung photons emission. For these reasons a dedicated tracking procedure (see fig. 4.3) is used for electrons in order to preserve optimal determination of the track parameters even if a large fraction of energy is lost by Bremmstrahlung. The Gaussian Sum Filter Algorithm (GSF) [116] models the Bremsstrahlung energy loss distribution by a Gaussian mixture rather than a single Gaussian. In the GSF track fit a sudden curvature radius change is properly taken into account but it requires also a very high CPU-time consumption. So it is slower than the KF and consequently it cannot be used for tracking all the charged particles but it can be run on a limited number of seeds.

4.2.1 Seeding strategies

In CMS there are two electron reconstruction algorithms whose difference is related to the seeding procedure: "Ecal driven" and "Tracker driven" proce-

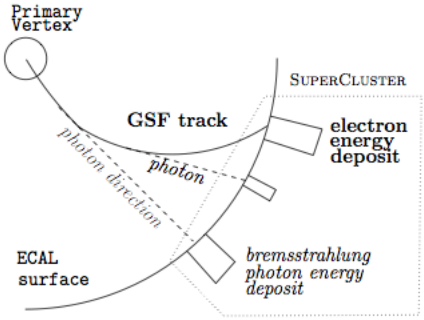


Figure 4.3: Schematical view of an electron reconstructed in CMS. The track is reconstructed using the GSF algorithm in order to take into account the change in the trajectory due to energetic bremsstrahlung photon emission. The energy deposits belonging to the emitted photons are collected together to the electron cluster by the clustering algorithm.

dures.

- Ecal driven: This procedure starts from the ECAL measurement and then a track is associated to it. The basic idea of this method is that the position of the barycentre of all the energy deposits of an electron is on the helix of the initial electron energy propagated through the magnetic field without any radiation emission. So the supercluster energy and positions are used to extrapolate the electron trajectory towards the vertex. The performance of this method depends on the ability to reconstruct the supercluster from all Bremsstrahlung energy photon and electron energy deposits. This algorithm is more suitable for isolated and high p_T electrons as, if an electron is inside a jet, the determination of the barycentre of all deposits from the electron is biased by deposits from other particles. Moreover, because of the high track multiplicity, a back-propagated track could be compatible with several track seeds from other charged particles.

- Tracker driven: This algorithm starts with the track reconstructions applying the KF algorithm. Initially the tracks are reconstructed using a standard KF algorithm. As we have already discussed, a trajectory can be accurately reconstructed with the KF approach if electron bremsstrahlung radiation is negligible. In that case KF track collects hits up to the ECAL, matching the closest PF cluster, and the momentum is measured with good precision. Each KF track which has a direction compatible with the position of the closest PF cluster and that verifies the condition $E_{p_{th}} < E/p < 3$ has its seed selected for further electron track reconstruction (where $E_{p_{th}}$ is set to 0.65 for electrons with $2 < p_T < 6$ GeV and 0.75 for electrons with $p_T \geq 6$ GeV). Some tracks that do not satisfy that condition because a non-negligible bremsstrahlung radiation may have occurred are then refitted using the Gaussian Sum Filter. This algorithm is more suitable to reconstruct electrons with low energy (below 10 GeV) or non isolated electrons.

Electron track-cluster association

The electron candidates are built from the matching of a supercluster with a GSF track. For the ECAL driven seeded electrons, the supercluster associated to the track is simply the one reconstructed by the hybrid or the multi5×5 algorithm from which the seed was found. For the electrons seeded using the tracker driven approach the supercluster is composed by several PF clusters that correspond to the electron and the emitted bremsstrahlung photons along the trajectory. The electron PF cluster is the one matched to the track while the clusters due to Bremsstrahlung photons are identified using the tangent to the electron track at each point of intersection with the tracker layer. Most of the bremsstrahlung photons are recovered in this way. For those photons, a specific procedure is performed, selecting displaced KF tracks with a multivariate criterium associating them with PF clusters.

4.2.2 Electron Identification

Several strategies have been developed in order to distinguish the identified primary isolated signal electrons from background sources originated from electrons coming from the semileptonic decays of b or c jets and from jets faking electrons. A set of discriminating variables is used in order to have a better identification of isolated and high transverse momentum electrons. These variables provide some discriminating power and can be grouped into three main categories. Observables related to the comparison between the measurements obtained from ECAL and those from the tracker. The comparison is between the ECAL supercluster energy and the track momentum. Purely calorimetric observables are particularly useful to distinguish real electrons (signal electrons or electrons from photon conversions) from fake electrons (like jets having a large electromagnetic component). The variables used are based on the transverse shape of electromagnetic showers in ECAL and exploit the fact that electromagnetic showers are narrower than hadronic ones. Electromagnitic showers are shorter and fully contained in the ECAL so the ratio of the energy in the HCAL behind the supercluster energy is an important observable. Purely tracking observables are finally employed to improve the separation between electrons and charged hadrons, exploiting the difference between the information obtained from the Kalman Filter and the Gaussian Sum Filter track fits.

4.2.3 $Z \rightarrow e^+e^-$ selection

A Z boson decaying into an electron pair is reconstructed from its decay products. Two opposite charge particle flow electrons have to be identified with a transverse momentum greater than 20 GeV and in the region individuated by $\mid \eta \mid < 2.4$. The pair is selected if the reconstructed invariant mass is compatible with the Z boson mass so if it is in the range [71, 111] GeV. The electron selection is performed using some selection requirements on the following variables:

- $\Delta\eta$: the distance in η between the supercluster position and the track extrapolation to the point of closest approach to the supercluster position
- $\Delta\phi$: difference angle in the transverse plane between the supercluster position and and the track extrapolation to the point of closest approach to the supercluster position
- H/E: the ratio between the Hadronic calorimetric tower energy behind seed crystal of the seed basic cluster (H) and the energy of the electromagnetic supercluster (E)
- $\sigma_{i\eta i\eta}$: this shower shape variable represents the covariance of the energy deposits in the η direction. Hadronic deposits in ECAL have a larger spread in the η direction with respect to pure electromagnetic deposits
- $\mid 1/E_{SC} - 1/p \mid$: E being the supercluster electron energy and p the track momentum at vertex
- d_0 and d_z: the impact parameter of the electron defined as the transverse and longitudinal distances respectively to the primary vertex at the point of closest approach in the transverse plane. These variables are used in order to reject those electrons coming from the conversions. So the dielectron vertex position is compared to the closer primary vertex
- *missing hits*: the number of hits used for the track reconstruction. It is required that the number of hits missing in the pixel layers to be below a cut threshold
- *PF iso*: isolation variables are used in order to reject the semileptonic decay of b, c quarks inside a jet. the Particle Flow isolation is defined as a combination of the activity of charged, neutral particles around the direction of the considered electron after that the electron deposit has been properly removed:

$$I = \frac{1}{p_T^e}[I_{ch} + max(I_{nh} + I_\gamma - 0.5 I_{chPU}, 0)]_{\Delta R < 0.3}. \qquad (4.1)$$

Where p_T^e is the transverse momentum of the electron, I_{ch} is the sum of the transverse energy of all the charge particles within the cone of $\Delta R < 0.3$, excluding the electron, coming from primary vertex, I_{nh} and I_γ are the sums of the transverse energy of all the neutral hadron within the cone and the photons respectively. I_{chPU} is an estimate charged particle deposits not from primary vertex.

4.3 Muons

Muons with enough energy can pass the whole CMS detector, thus tracks of muons can be reconstructed both in the tracker and in the muon system. In CMS tracks are first reconstructed independently in the inner tracker (*tracker track*) and in the muon system (*standalone-muon track*). The muon reconstruction is then based on these two independent measurements, using two different approaches: “Global Muon” reconstruction (outside-in) and “Tracker Muon” reconstruction (inside-out).

- Global Muon reconstruction: for each track reconstructed in the muon chambers a matching with a tracker track is looked for comparing the parameters of the two tracks propagated onto a common surface. Using a Kalman filter technique a global muon track is fitted, combining the hits from the tracker track and standalone-muon track. At high transverse momentum (p_T >200 GeV) the global muon fit can improve the momentum resolution respect to the one obtained using a tracker only fit. [117,118].
- Tracker Muon reconstruction: this method starts from all tracker tracks with $p_T > 0.5$ GeV and total momentum $p > 2.5$ GeV that are considered as possible muon candidates and are extrapolated to the muon system taking into account the effects produced by multiple Coulomb scattering in the detector material, the magnetic field and the average expected energy losses. If at least one muon segment (a short track stub made of DT or CSC hits) is geometrically matched to the extrapolated track, the corresponding tracker track is taken as Tracker Muon.

Calibration of track momentum

The determination of muon track parameters, as the transverse momentum, is highly sensitive to the precise alignment of the silicon sensors of the tracker and of the muon chambers, to the material budget in the detector, to the detailed map of the magnetic field inside and outside the solenoid volume, and in general to the performance of tracking detectors. A correction to the track momentum is then needed in order to take into account our imperfect knowledge and modeling of the sub-detectors alignment, of the detector material, and of the magnetic field. The calibration of track momentum and the investigation of systematic effects introduced by these sources rely on the use well known resonances (J/Ψ, Υ or Z). The strategy of the algorithm developed is to extract all the information by combining the reconstructed kinematics of the muon pair with the parent's invariant mass. A probabilistic approach is necessary in order to relate the difference between expected and observed mass with a hypothetical bias on the measured parameters of either or both daughter tracks. A multi-parameter likelihood fit is the best approach to the solution of this problem. A set of functions describing the dependence on track kinematics of those biases (offsets from the true values) and measurement resolutions are defined. Their features have a direct correspondence with the possible imperfect description of the detector material, of the alignment of the sensitive devices, of the magnetic field map. Different scenarii are considered:

- Tracker misaligned
- Muon System (DT and CSC) misaligned
- Both Muon System (DT and CSC) and Tracker misaligned
- B distorted by 2 permill everywhere in the detector
- B distorted by 2 permill in the solenoid, 2 percent in the barrel yoke, 5 percent in the endcaps

a multivariate likelihood approach is then applied. The best estimate of the parameters of those functions can be determined from a likelihood minimization. The momentum scale corrections applied in this analysis are provided by the MuScleFit (Muon Scale Fit) package [119].

4.3.1 $Z \rightarrow \mu^+\mu^-$ selection

Particle Flow muons with opposite charge and with and invariant mass within the range [71, 111] GeV have to be select in order to identify the Z boson. The reconstruction of the muons considered in this analysis is based on the Global Muon reconstruction. However additional identification criteria are applied on the reconstructed muons in order to reduce the muon fake rate while keeping the selection efficiency high. The following set of variables is used:

- χ^2 of the global fit, normalised to the number of degrees of freedom ($\chi^2/d.o.f$): it ensures a good quality of the global fit which is used to estimate the muon parameters, and allows to reduce mismatchings of tracker tracks and standalone muons
- N_{hits}: minimum number of hits in the tracker, considering both the pixel and the strip hits. This requirement allows to reduce the combinatorics reducing the probability of wrong matching and ensuring an accurate transverse momentum measurement

- minimum number of hits in the muon system. This is an additional protection against punch-through hadrons and ensures a good muon parameter estimate
- d_0: distance of closest approach with respect to the beam line in the $x-y$ plane. The cut on this variable allows to reject cosmic muons. Also many decay-in-flight muons are rejected. A remaining contribution from decays of b or c hadrons cannot be excluded
- at least 1 hit in the pixel detector: with this requirement decay-in-flight muons are further suppressed and the trajectory close to the interaction region is enhanced
- *PF iso*: the isolation variable is defined inside a cone in the $\eta-\phi$ space around the muon with a $\Delta R = 0.3$. It is given combining tracker and calorimetric measurements:

$$I_{\mu} = \frac{1}{p_T^e}[I_{ch} + max(I_{nh} + I_{\gamma} - 0.5I_{chPU}, 0)]_{\Delta R<0.3}. \qquad (4.2)$$

4.4 Jet Reconstruction

The QCD confinement implies that the partons, produced in a proton-proton collision, form colorless hadrons which are collimated particle showers moving along the direction of the originating parton, called jets. The jets play a key role at LHC because, due to their large production cross section, they allow studies of new kinematic regimes, confronting predictions of perturbative QCD and probing physics processes within and beyond the Standard Model. The understanding of the energy calibration and resolution of the jets becomes a crucial point and it represents one of the leading source of systematic uncertainty when the processes studied have jets in the final state. The combination of the contributions of all the subdetectors are used in CMS in order to reconstruct jets. Three different clustering algorithms have been implemented: Calorimeter jets, Jet-Plus-Track (JPT) jets, Particle-Flow (PF) jets.

- Calorimeter jets: they are reconstructed using the combination of the measurement of the electromagnetic and hadronic calorimeters. One or more HCAL cells and the geometrically corresponding ECAL crystals are combined together in a calorimeter tower. In the barrel region the unweighted sum of one single HCAL cell and 5x5 ECAL crystals form a projective calorimeter tower. The association between the HCAL cells and ECAL crystal is instead more complex in the endcap region. In order to reduce the effects of calorimeter readout electronics noise, the cells considered to build the calorimetric towers must have a minimum threshold in energy and also a cut with the missing transverse energy is applied. Moreover an additional cut is applied on the transverse energy ($E_T < 0.3$ GeV) to suppress contribution from event pile-up.

- Jet-Plus-Tracks: algorithm [120] uses not only the measurement from the electromagnetic and hadronic calorimeters but also the excellent performance of the CMS tracking detectors is exploited in order to improve the resolution of calorimeter jets. The first step of this algorithm consists in the reconstruction of calorimeter jets as described above. Then charged particle tracks are geometrically associated with each jets considering the separation in $\eta - \phi$ between the jet axis and the track direction. The tracks extrapolated to the calorimeter surface are then classified as "in-cone", if they point to within the jet cone around the jet axis, or "out-cone", if the the magnetic field has bent the track out of the jet cone. The momenta of both in-cone and out-cone tracks are then added to the energy of the associated calorimeter jet. Based on the momentum measurement of the track, an expected average energy deposition in the calorimeter is then subtracted for the in-cone tracks. The direction of the axis of the original calorimeter jet is also corrected by the algorithm.
- Particle Flow (PF) jets. As already discussed, the Particle Flow algorithm allows to identify all the stable particles of the final state of an event combining the information from all the subdetectors. In this way electrons, muon, photons, charged and neutral hadrons are reconstructed. In particular charged hadrons are identified by the matching of tracks and calorimeter clusters not reconstructed as electrons. Neutral hadrons are reconstructed from energy clusters in the electromagnetic and hadron calorimeters. Clusters separated from the extrapolated position of tracks in the calorimeters constitute a clear signature of these neutral particles. PF jets are then reconstructed from the resulting list of final state particles as represented in fig. 4.4. The PF jets show an improved momentum and spatial resolution respect to the calorimeter jets. This is due to the excellent ECAL granularity and the good performance of the tracking detector that combined together provide precise measurement of charged hadrons and photons inside a jet whose energy is composed for the 90% of these types of particles.

4.4.1 Jet clustering algorithm

Several approaches are available for the jet clustering and all of them can be grouped into two different classes: "cone algorithms" [121] and "sequential recombination" [122]. The first one is based on the idea of defining a jet as a cone around the direction of dominant energy flow. To find directions of dominant energy flow, cone algorithms usually take some of the event particles as seeds. Then for each seed a list of particles in a cone is established. The

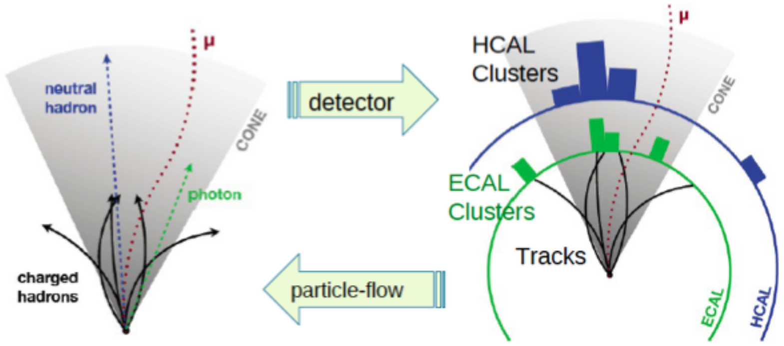


Figure 4.4: Schematic view of the Particle Flow algorithm. Particles are obtained from the combination of tracks and energy deposits. The PF algorithm attempts to fully reconstruct an event using the information from all CMS subdetectors.

sum of the four momenta of all the particles considered in the cone is evaluated and the resulting four-momentum is used as a new trial direction for the cone. This procedure is iterated until the cone direction no longer changes, and so until one has a “stable cone”. The second type of clustering algorithm is instead based on the definition of a distance between pairs of particles, performing successive recombinations of pair of closest particles and stopping when all resulting objects are too far apart. The algorithms that belong to this class differ for the definition of the distance. An important characteristic that jet clustering algorithms must have is the InfraRed and Collinear safety (IRC-safe), in order to be independent of infrared and collinear corrections in perturbative QCD. The infrared corrections concern the emission of a gluon with an infinitely low energy while the collinear corrections concern the splitting of a hard particle into two collinear particles. The algorithm is IRC-safe if the number of reconstructed jets is not affected. The standard IRC-safe algorithms used in CMS are the SISCone [123] (cone algorithm) and the k_t and anti-k_t algorithms [124].

anti-k_T algorithm

The k_t and anti-k_t algorithms introduce a distance d_{ij} between two particles (i and j) and d_{iB} that is the distance between the particle i and the beam:

$$d_{ij} = min(k_{ti}^{2p}, k_{tj}^{2p})\frac{\Delta_{ij}^2}{R}, \tag{4.3}$$

$$d_{iB} = k_{ti}^{2p}, \tag{4.4}$$

where R is the radius parameter that characterizes the method (in this analysis the anti-k_t is used with a R=0.5). $\Delta^2_{ij} = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, k_{ti}, y_i and ϕ_i are respectively the transverse momentum, the rapidity and azimuth of the particle i while p is a parameter that distinguishes among different distance definitions. The distances d_{iB} and d_{ij} for each particle i are calculated and the smallest distance between d_{ij} and d_{iB} is identified. If the condition $d_{ij} < d_{iB}$ is verified then the particles i and j are recombined, while if $d_{ij} > d_{iB}$, i is considered as a jet and removed from the list of the final particles. The distances are then evaluated again and the procedure is repeated until no particles are left. The parameter p that is present in the definition of the distances can have different values, identifying different jet clustering algorithms: $p = 1$ defines the k_t algorithm, the case of $p = 0$ corresponds to the inclusive Cambridge/Aachen algorithm [125] while $p = -1$ refers to the anti-k_t algorithm. The anti-k_t algorithm is the one adopted in the clusterization of jets in this work. The functionality of the anti-k_t algorithm is clear if one assumes to have an event with few well separated hard particles with transverse momenta k_{t1}, k_{t2}, ... and many soft particles. The distance d_{1i} between the hard particle 1 and the soft particle i is dominated by the transverse momentum of the hard particle k_{t1} and by the Δ_{1i} separation. The d_{ij} separation between softer particles results to be much larger. Therefore soft particles will tend to cluster with hard ones long before they cluster among themselves. If no hard particles are within a distance $2R$ respect to another hard particle, then it will be clustered with all the soft particles inside a circle of radius R obtaining a perfect conical jet. If two hard particles satisfy the condition $R < \Delta_{12} < 2R$ then there will be two hard jets. In this case it is not possible for both the jets to be perfectly conical. If $k_{t1} \gg k_{t2}$ then jet 2 will be partly conical because it will miss the part overlapping jet 1 that, instead, will be conical. If $k_{t1} = k_{t2}$ neither jets will be conical and the part of overlap will be equally divided between them. Finally if two hard particles (1 and 2) are within a distance $\Delta_{12} < R$ they are joined to form a single jet that could be conical centred on k_1 (if $k_{t1} \gg k_{t2}$) or have a more complex shape (if $k_{t1} \sim k_{t2}$). From what has been discussed a key feature appears evident: the soft particles do not modify the shape of the jet, while hard particles do, the jet boundary in this algorithm is resilient with respect to soft radiation, but flexible with respect to hard radiation. The behaviours of different jet algorithms are illustrated in fig. 4.5.

4.4.2 Jet energy calibration

The real parameters of the original parton, originating the jet, reflect on the measured jet parameters. However the jet energy measured in the detector generally differs from the corresponding particle jet energy. The main cause for this energy mismatch is the non-uniform and non-linear response of the CMS calorimeters. Furthermore, electronics noise and additional proton-proton in-

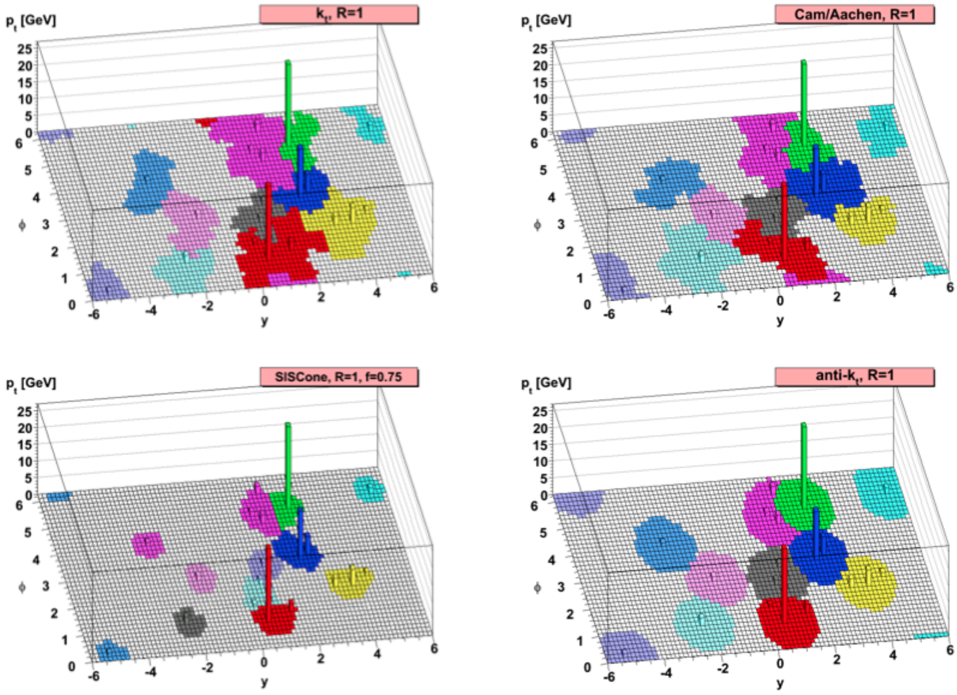


Figure 4.5: A sample parton-level event (generated with Herwig), together with many random soft "ghosts", clustered with four different jets algorithms, illustrating the "active" catchment areas of the resulting hard jets. For kt and Cam/Aachen the detailed shapes are in part determined by the specific set of ghosts used, and change when the ghosts are modified [124].

teractions in the same bunch crossing (pileup event) can lead to extra unwanted energy. The jet energy corrections (JEC) allow to relate, on average, the energy measured in the detector for the jet to the energy of the corresponding true particle jet. A true particle jet results from the clustering of all stable particles originating from the parton fragmentation and of the particles from the underlying event activity. The correction is applied as multiplicative factor C to the each component of the raw measurement of the transverse momentum of the jet (p_μ^{raw}):

$$p_\mu^{corr} = C \cdot p_\mu^{raw}. \tag{4.5}$$

The factor C corresponds to three subsequent corrections: offset, relative and absolute corrections and this sequence for the jet energy correction can be expressed by:

$$C = C_{offset}(p_T^{raw}) \times C_{Rel}(\eta, p_T^{''}) \times C_{Abs}(p_T^{'}), \tag{4.6}$$

where p''_T is the transverse momentum of the jet corrected for offset and $p'_T = p''_T \times C_{Rel}(\eta, p''_T)$ is the transverse momentum of the jet corrected for offset and pseudorapidity dependence. $C_{offset}(p_T^{raw})$, C_{Rel} and C_{Abs} represent the three corrections applied.

- $C_{offset}(p_T^{raw})$ (LV1 - offset) is the offset correction; the first correction applied in the chain of factorized corrections. Its purpose is to correct the excess of measured energy due to electronic noise and to the presence of pileup events. Three methods have been developed in CMS to extract the offset correction: the jet area, the average offset and the hybrid jet area methods. A description of the jet area [126] [127] method is given as it has been adopted in this work. The idea of this approach is to estimate an average p_T density ρ per unit area for each event. This density ρ allows to describe the soft jet activity and includes the contributions from the underlying event, the electronic noise and the pileup. A large number of soft four momentum vectors (soft enough not to change the properties of the true jets) are artificially added in the event and clustered by the jet algorithm together with the true jet components. In order to extract this correction two quantities are needed: the active area A_j and the ρ density. The region in the $y - \phi$ space occupied by the soft particles clustered in each jet defines the active area A_j. An average p_T-density is estimated applying the k_T algorithm with a distance parameter $R = 0.6$ that naturally clusters a large number of soft jets in each event. So an event-by-event and jet-by-jet pileup correction factor can be defined:

  $$C(p_T^{raw}, A_j, \rho) = 1 - \frac{(\rho - <\rho_{UE}>) \cdot A_j}{p_T^{raw}}, \tag{4.7}$$

  where $< \rho_{UE} >$ is the p_T-density component due to the UE and electronic noise and is measured in events with exactly one reconstructed primary vertex (no pileup).

- C_{Rel} (LV2 - relative) is the relative correction that ensures a flat response in jet psudorapidity removing variations in jet response versus jet η, fixing the jet p_T. To derive the relative energy corrections from collider data, the dijet balance technique is used [128]. The idea is to use p_T balance in back to back dijet events where one jet is in the central ($|\eta| < 1.3$) control region of the calorimeter (barrel jet) while the other (probe jet) is at an arbitrary η. The central region is chosen as reference because of the uniformity of the detector, and because it has highest jet transverse momentum reach. The balance quantity B is defined as:

  $$B = \frac{p_T^{probe} - p_T^{barrel}}{p_T^{ave}}, \tag{4.8}$$

where p_T^{ave} is the average p_T of the two leadings jets:

$$p_T^{ave} = \frac{p_T^{barrel} + p_T^{probe}}{2}. \tag{4.9}$$

The balance B is recorded in bins of η^{probe} and p_T^{ave}. The average value $< B >$, in a given p_T^{ave} and η^{probe} bin is used to determine the relative response $R_{rel}(\eta^{probe, p_T^{ave}})$:

$$R_{rel}(\eta^{probe, p_T^{ave}}) = \frac{2 + < B >}{2 - < B >}. \tag{4.10}$$

The R variable is an estimator of the relative response.

- C_{Abs} (LV3 - absolute) is the absolute correction that removes variations in jet response versus jet transverse momentum. It is measured with the “missing transverse energy projection fraction” (MPF) method using γ/Z+jets events. This method is based on the fact that these events have no intrinsic missing transverse energy (MET) and, at parton level, the γ or the Z is perfectly balanced by the hadronic recoil in the transverse plane:

  $$p_T^{\gamma,Z} + p_T^{recoil} = 0. \tag{4.11}$$

  For reconstructed objects, this equation can be re-written as:

  $$R_{\gamma,Z} p_T^{\gamma,Z} + R_{recoil} p_T^{recoil} = -MET, \tag{4.12}$$

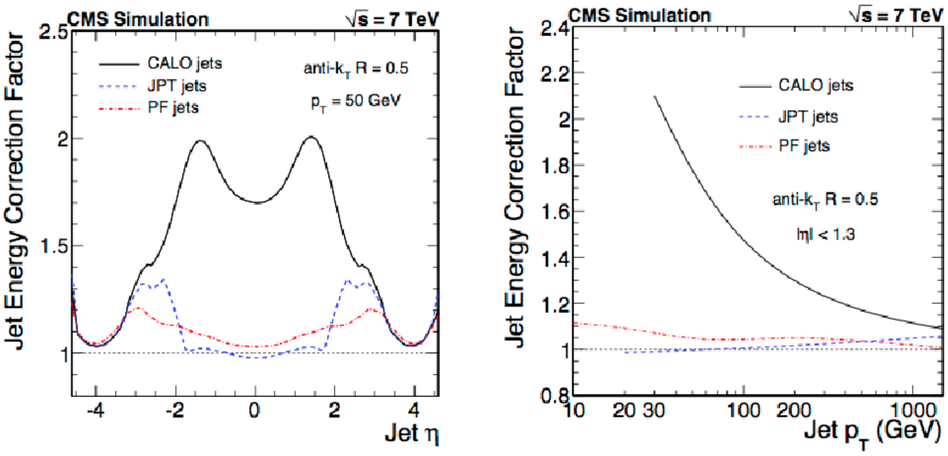
  where $R_{\gamma,Z}$ and R_{recoil} are the detector responses to the γ or Z and the hadronic recoil respectively. The R_{recoil} is then given by:

  $$R_{recoil} = R_{\gamma,Z} + \frac{MET \cdot p_T^{\gamma,Z}}{(p_T^{\gamma,Z})^2} \equiv R_{MPF}. \tag{4.13}$$

  This equation forms the definition of MPF response R_{MPF}. The last step is to extract the jet energy response from the measured MPF. In general, the recoil consists of additional jets, beyond the leading one, soft particles and unclustered energy. $R_{leadjet} = R_{recoil}$ can be assumed in good approximation if the particles, that are not clustered into the leading jet, have a response similar to the ones inside the jet, or if these particles are in a direction perpendicular to the photon axis.

These three corrections have been derived using Monte Carlo simulation that can provide an estimate of the particle jet energy. For data events however, additional corrections (LV2LV3Res - residual calibration) are applied in a fourth step, to account for differences on the detector responses in simulation and reality. For this reason they are applied on data only. They are p_T and η dependent and follow the approach of the L2 and L3 corrections. The residual correction is defined by the difference of these corrections on simulated and data events.

The non linear response of the calorimeters mainly affect the calorimetric jets as only calorimetric measurements are used to reconstruct this type of jets. The track-based jet types (JPT and PF) require much smaller correction factors because the charged component of the jet shower is measured accurately in the CMS tracker as shown in fig. 4.6. Corrections dependence versus η is more stable for PF jets than for CALO and JPT jets.



(a) Jet energy correction factors obtained from a Monte Carlo simulation at $\sqrt{s} = 7$ TeV reconstructing CALO, JPT, and PF jets as a function of η (p_T=50 GeV).

(b) Jet energy correction factors obtained from a Monte Carlo simulation at $\sqrt{s} = 7$ TeV reconstructing CALO, JPT, and PF jets as a function of jet transverse momentum.

Figure 4.6: Jet energy correction factors obtained from a Monte Carlo simulation at $\sqrt{s} = 7$ TeV reconstructing CALO, JPT, and PF jets as a function of jet η and jet p_T. Jets are reconstructed using the anti-k_T algorithm [129].

4.4.3 Jet identification

In this thesis the jets used are PF jets clustered with the anti-k_T algorithm with a distance parameter $R = 0.5$. The measured jet energy is corrected with the scheme described above. To reduce the jet fake rate, additional identification criteria are necessary to select only real hadronic jets. However they are kept at low thresholds to maintain high selection efficiency. All jets have to fulfil the following requirements:

- Neutral hadronic fraction below 99%
- Neutral electromagnetic fraction below 99%

- At least two constituents
- Charged hadronic fraction above 0%
- Charged electromagnetic fraction below 99%
- At least one charged constituent

4.5 Flavor identification

A wide number of processes that involve the jets arising from bottom quark hadronization (b jets) is present in the physics program of the CMS experiment. The accurate identification of b jets becomes a crucial point in several analysis. For this reason a considerable effort is devoted to the identification of b flavored jets (b jet tagging) trying to reduce background from processes involving jets from gluons (g) and light-flavor quarks (u, d, s), and from c quark fragmentation. The b tagging algorithms exploit the distinct properties of b hadrons which are the product of b quark fragmentation:

- The relatively long life-time of b hadrons, $\tau \sim 1.5ps$ corresponding to an average decay length of $\lambda = (\beta\gamma)c\tau$ =450μm. It implies a displaced vertex (secondary vertex) that can be reconstructed relying on the excellence of the CMS inner tracking system.
- The final state of a B hadron decay contains 5 tracks on average, all having a sizable impact parameter (IP), which is defined as the distance from the PV to the track at their point of closest approach in space.
- The B hadron semileptonic decay can become a useful discriminator to perform a b tagging, using the presence of a soft lepton among the B decay products.

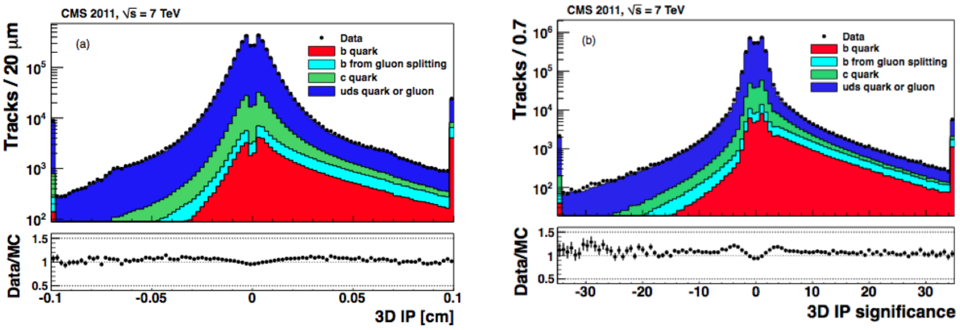
In order to discriminate between b and light-parton jets, several reconstructed objects, such as tracks, vertices and identified leptons, can be used to build discriminating variables. Some algorithms use just a single observable, while others combine several of these objects to improve the discrimination power. Each of these algorithms produces a single discriminator value for each jet. The minimum thresholds imposed on these discriminators, define different working points: loose (“L”), medium (“M”) and tight (“T”) corresponding to misidentification probability for light-parton jets of 10%, 1%, and 0.1%, respectively, at an average jet p_T of about 80 GeV. The b jet tagging algorithms are applied to clustered jets and exploit the measured kinematic properties of charged particles, including identified leptons, inside it. Tracks are the most powerful ingredients for b tagging algorithms and only those fulfilling the following criteria are used:

- angular distance between track and jet axis $\Delta R < 0.3$
- at least 2 pixel hits and 8 tracker hits (including pixel)
- distance smaller than 0.2 cm (17 cm) in the transverse plane (along the beam axis) between the track and the primary vertex at the point of closest approach of the trajectory to the PV in the transverse plane
- transverse momentum greater than 1 GeV
- χ^2 normalised to the number of degrees of freedom in the fit less than 5 in order to ensure a good fit
- distance to jet axis < 0.07 cm, defined as the spatial distance between the trajectory and the jet axis at their point of closest approach, where the jet axis is reconstructed with respect to the primary vertex
- decay length < 5 cm, defined as the spatial distance between the PV and the point of closest approach between the track trajectory and the jet axis.

Several algorithms have been developed in order to perform the b tagging and two categories can be distinguished depending on the observables exploited: impact parameter (IP) or secondary vertices (SV). The impact parameter is defined as the distance from the primary vertex (PV) to the track at their point of closest approach in space. Examples of algorithms based on the measurement of the impact parameters are the Track Counting (TC) that sorts tracks in a jet by decreasing values of the IP significance and the Jet Probability (JP) that uses an estimate of the likelihood that all tracks associated to the jet come from the primary vertex. Secondary vertices (SV) within jets are reconstructed using the adaptive vertex fitter [130]. The resulting list of vertices is then subject to a cleaning procedure, rejecting SV candidates that share 65% or more of their tracks with the primary vertex. The distance in space from the primary vertex to the secondary vertex, the flight distance, can be then used to identify long-lived particles and its significance must exceed 3σ. SV candidates with a radial distance from the primary vertex greater than 2.5 cm whose associated tracks form an invariant mass compatible with the K^0 mass are rejected.

4.5.1 Identification using track impact parameters

The impact parameter of a track with respect to the primary vertex is a useful observable to distinguish the decay products of a B hadron from prompt tracks. The sign of the impact parameter allows to discriminate among the prompt tracks and those coming from the B hadron. The sign of the IP is the same of the scalar product of the vector pointing from the primary vertex to the point of closest approach with the jet direction. Tracks originating from the decay of particles travelling along the jet axis will tend to have positive IP values. In contrast, the impact parameters of prompt tracks can have positive or negative IP values. The track impact parameter can be calculated either in the plane normal to the beam axis (transverse impact parameter) or in three dimensions (3D impact parameter). The first has the advantage of being less sensible to the uncertainty on the primary vertex position due to the small size of the LHC beam spot in the transverse plane (less than 20 $\mu m \times 20\ \mu m$). The 3D impact parameter is affected by the larger error on the primary vertex position in the z direction. The experimental resolution is taken into account using the track impact parameter significance S_{IP} defined as the ratio of the IP to its estimated uncertainty. In fig. 4.7 the distributions of IP values and their significance are shown.



(a) Distributions of the 3D impact parameter or all the selected tracks measured with 7 TeV collision data in 2011 by CMS.

(b) Significance of the 3D impact parameter for all the selected tracks measured with 7 TeV collision data in 2011 by CMS.

Figure 4.7: Distributions of the 3D impact parameter and the significance of the 3D impact parameter for all the selected tracks measured with 7 TeV collision data in 2011 by CMS. Different flavors compositions are presented in different colours showing the tagging power of the tagger.

The impact parameter significance has discriminating power between the decay products of b and non b jets. The Track Counting (TC) b tagging algorithm is based on the high multiplicity of charged tracks produced in the decay of a b hadron (about 5 tracks). This algorithm firstly sorts tracks in a jet by decreasing values of the IP significance. The jet is tagged if there are enough tracks with an impact parameter significance exceeding a given cut. A natural extension of the TC algorithms is the combination of the IP information of several tracks in a jet. Two b tagging discriminators can be built from the impact parameter calculation and used in dedicated algorithms called: *Jet Probability* (JP) and *Jet B Probability* (JBP). The JP algorithm uses an estimate of the likelihood that all the tracks associated to the jet come from the primary vertex. The JBP algorithm gives more weight to the tracks with the highest IP significance, up to a maximum of four such tracks, which matches the average number of reconstructed charged particles from B hadron decays. The likelihood P_{jet} is defined as:

$$P_{jet} = \prod \cdot \sum_{i=0}^{N-1} \frac{(-ln \prod)^i}{i!}, \tag{4.14}$$

with

$$\prod = \prod_{i=1}^{N} max(P_i, 0.005), \tag{4.15}$$

where N is the number of tracks under consideration and P_i is the estimated probability for track i to come from the primary vertex.

4.5.2 Identification using secondary vertices

Another way to discriminate between b jets and non b jets is the reconstruction of a secondary vertex and the kinematic variables associated with this vertex. Since this b tagging algorithm uses not only the presence of a secondary vertex, but also topological and kinematical variables related to the SV, it is desirable to reconstruct the decay vertex as completely as possible, to increase the discriminating power. The following quality cuts are applied to the resulting vertices to select secondary vertex candidates coming from b hadron decays:

- The distance L_t from the primary vertex to the secondary vertex in the transverse plane has to be within 100 $\mu m < L_t <$ 2.5 cm, and to have a significance $L_t/\sigma_{L_t} > 3$;
- The invariant mass of charged particles associated to the vertex must be compatible with a b quark, and thus not exceed 6.5 GeV;
- The vertex must not be compatible with a light neutral meson decay

After the reconstruction and selection of the secondary vertex three categories can be individualized:

- Reco Vertex:
  At least one secondary vertex has been reconstructed and the selection requirements have been satisfied. If there is more than one accepted secondary vertex all the tracks associated to them are used to build vertex related variables.
- Pseudo Vertex:
  If no reconstructed secondary vertex candidate has been found, a so-called Pseudo Vertex is created from charged particle tracks not compatible with the event primary vertex, having a signed transverse impact

parameter significance greater than two, if at least two such tracks are present in the jet.

- No Vertex:
  This is the case of those events in which neither a Reco Vertex nor a Pseudo Vertex candidates have been found. In this case this category reverts to track-based variables that are combined in a way similar to that of the JP algorithm.

An optimal performance is obtained by the combination of several topological and kinematical variables related to the secondary vertex reconstruction and to the the impact parameter significances of charged particle tracks. The choice of the variables used in the combination depends on the vertex category. If a Reco Vertex has been reconstructed the following variables are defined:

- The invariant mass of charged particles associated to the secondary vertex. If a jet comes from the hadronization of a b quark, the vertex mass will be significantly above the mass of charm hadrons, suppressing efficiently this background.
- The charged track multiplicity. A larger track multiplicity is expected for a B hadrons decay than for a charm hadron decay.
- The distance between the primary vertex and the secondary vertex in the transverse plane divided by its error. This is sensitive to the large flight path of B hadrons.
- The energy of charged particles associated to the secondary vertex divided by the energy of all charged particles associated to the jet. This quantity is sensitive to the hard fragmentation function of b and c quarks.
- The rapidities of charged particle tracks associated to the secondary vertex with respect to the jet direction: $y = \frac{1}{2} \cdot \ln(\frac{E+p_{||}}{E-p_{||}})$. If n is the charged particle multiplicity of the secondary vertex, this variable enters for those n tracks. For b jets, the rapidities are on average smaller than for c jets because of the large mass of b hadrons resulting in larger angles of the charged particle tracks relative to the jet axis.

In the case of the category 2, the significance of the flight distance cannot be used as there is no attempt to fit the geometrical position of the "PseudoVertex", but all the other variables are used as input variables still offering some discriminating power between b and non b jets. All these variables are then combined into a single discriminant using a Likelihood function defined as:

$$\mathcal{L}^{b,c,q} = f^{b,c,q}(\alpha) \times \prod_i f_{\alpha}^{b,c,q}(x_i), \tag{4.16}$$

where α indicates the categories ($\alpha = 1, 2, 3$), x_i is the individual variable, q refers to u, d, s quark or gluon jets, $f^{b,c,q}(\alpha)$ is the probability for flavor b, c, q to fall into a category α and $f_\alpha^{b,c,q}(x_i)$ is the probability density function for variable x_i and the flavor b, c, *light*. The discriminating variable is then defined as:

$$d = f_{BG}(c) \times \frac{\mathcal{L}^b}{\mathcal{L}^b + \mathcal{L}^c} + f_{BG}(q) \times \frac{\mathcal{L}^b}{\mathcal{L}^b + \mathcal{L}^q}, \tag{4.17}$$

where $f_{BG}(c, q)$ is the expected bayesian *a priori* probability for the charm and light content in non b jets ($f_{BG}(c)$+$f_{BG}(q)$=1) The probability density functions are extracted from a statistically independent sample of simulated QCD events, and depend on the transverse jet energy and pseudorapidity.

4.5.3 Tagging efficiency

The measurement efficiency for each algorithm to select genuine b jets is really crucial. Several techniques that can be applied to CMS data have been developed in order to measure the efficiency *in situ*. Some efficiency measurement are performed using samples that include a jet with a muon within $\Delta R = 0.4$ from the jet axis ("muon jet"). As the semileptonic branching fraction of B hadrons is significantly larger than that for other hadrons, jets containing a muon are more likely to arise from the hadronization of b quarks than from another flavor. Moreover muons are identified very efficiently in the CMS detector. These muons can be used to perform the b tagging efficiency because the efficiencies of the muon- and lifetime-based b jet identification techniques are largely uncorrelated. To determine the b tagging efficiency, two different methods are used, based on kinematic properties of the muon-jets:

- "PtRel" method:
  As a consequence of the large b quark mass, the momentum component of the muon transverse to the jet axis, p_T^{Rel}, is large for muons from the B hadron decay than for muons in light-parton jets or from charm hadrons. This variable is then used as discriminant. The discriminating power of this variable depends on the muon jet p_T. The muon p_T^{Rel} distributions provide better separation for jets with p_T smaller than about 120 GeV.

- "IP3D" method:
  The impact parameter of the muon track, calculated in three dimensions, is larger for B hadrons than for other hadrons. In this case the IP distribution has a greater discriminating power for jets with $p_T > 120$ GeV.

The muon-jets are separated into tagged and untagged subsamples by a discriminator working point whose efficiency is to be measured. For the two subsamples separately, the spectra of muon-jets p_T^{Rel} or IP3D are fitted using templates of b, c, and *udsg* jets derived from simulation. From each fit the fractions of b jets (f_b^{tag}, f_b^{untag}) are extracted from the data. With these fractions and the total yields of tagged and untagged muon jets (N_{data}^{tag}, N_{data}^{untag}) the number of b jets in data these samples are calculated, and the efficiency ε_b^{tag} for tagging b jets in data is inferred:

$$\varepsilon_b^{tag} = \frac{f_b^{tag} \cdot N_{data}^{tag}}{f_b^{tag} \cdot N_{data}^{tag} + f_b^{untag} \cdot N_{data}^{untag}}. \tag{4.18}$$

The b tagging efficiency for the CSVM criterion as a function of jet p_T and jet $\mid \eta \mid$ is shown in fig. 4.8.

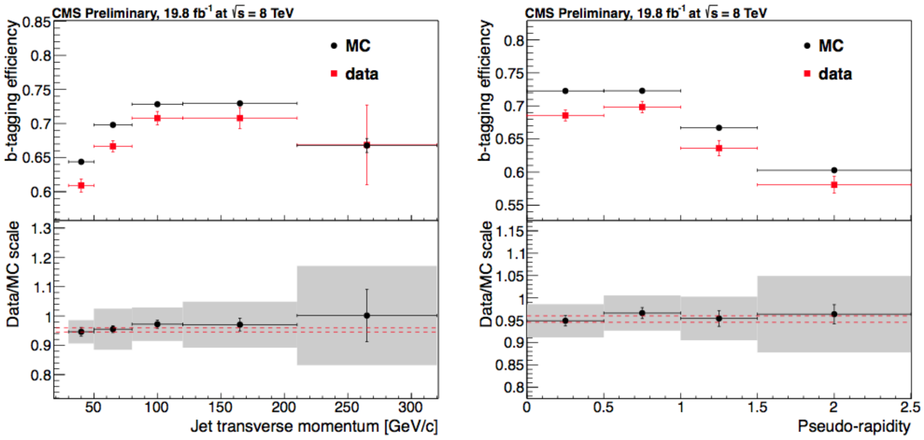


Figure 4.8: b tagging efficiencies and data/MC ratio as a function of p_T jet (jeft) and jet $\mid \eta \mid$ (right). The grey filled area represents the total statistical and systematic uncertainties [131].

An overview of the individual and combined measurements of the b tagging efficiency scale factor for the CSVM criterion obtained using various methods is shown in fig. 4.9.

Corrections can be applied to simulated events using a scale factor SF_b, defined as the ratio of the efficiency measured with collision data to the efficiency found in the equivalent simulated samples, using Monte Carlo generator level information to identify the jet flavor. The scale factors in general depend on the jet flavor, jet pseudorapidity and jet transverse momentum. An event weight can be calculated using the scale factors and applying it to Monte Carlo events. A simpler example, in which only the leading two jets are taken for b tagging, can be considered. An event in which both of the leading two jets are b tagged can contribute to events with 2, 1, and 0 b tags with the following event weights:

$$w(2|2) = SF_1 SF_2, \tag{4.19}$$

$$w(1|2) = (1 - SF_1)SF_2 + (1 - SF_2)SF_1, \tag{4.20}$$

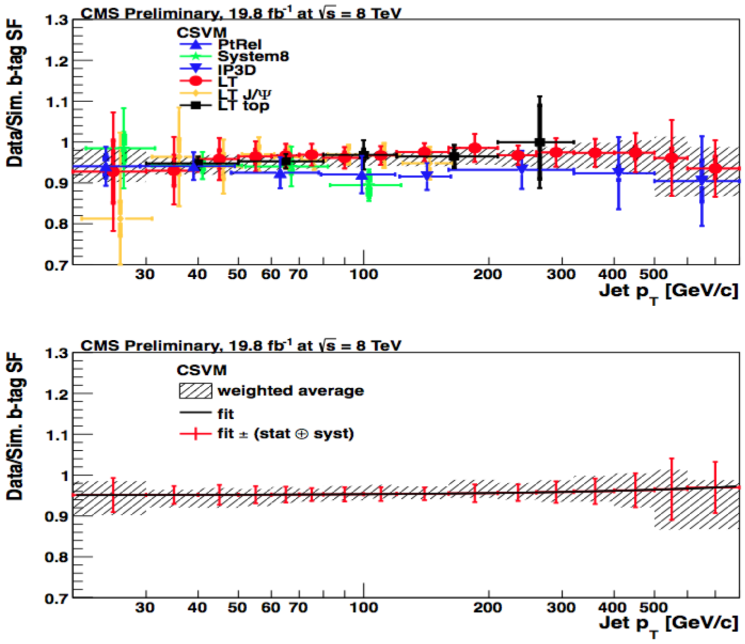


Figure 4.9: Individual (top) and combined (bottom) measurement of The Combined Secondary Vertex Scale Factors calculated with different techniques employed to evaluate the b tagging efficiency (coloured lines) as a function of the jet p_T with 8 TeV data [131].

$$w(0|2) = (1 - SF_1)(1 - SF_2), \tag{4.21}$$

where $w(n|m)$ $(n \leq m)$ is the event weight for an event with m b tags to contribute to events with n b tags and SF_1 and SF_2 are the scale factors for the leading and second leading jet, respectively. Similarly, an event in which only one of the leading two jets is b tagged can contribute to events with 1 and 0 b tags with the following event weights:

$$w(1|1) = SF, \tag{4.22}$$

$$w(0|1) = (1 - SF), \tag{4.23}$$

while event with 0 b tagged jets can only contribute to events with 0 b tags with an event weight of 1:

$$w(0|0) = 1. \tag{4.24}$$

The method can be naturally extended to more complicated jet configurations. The event weight corresponding to events with at least 1 b tag for which the event weight is defined as:

$$w(\geq 1|n) = 1 - w(0|n), \tag{4.25}$$

where

$$w(0|n) = \prod_{i=1}^{n}(1 - SF_i). \tag{4.26}$$

Another, slightly more complicated, case corresponds to events with at least 2 b tags for which the event weight is defined as:

$$w(\geq 2|n) = 1 - w(0|n) - w(1|n), \tag{4.27}$$

with

$$w(1|n) = \sum_{j=1}^{n} \cdot [\prod_{i=1, i \neq j} (1 - SF_i)] SF_j. \tag{4.28}$$

The b tagging efficiency and scale factor as a function of the operating point value is illustrated in fig. 4.10

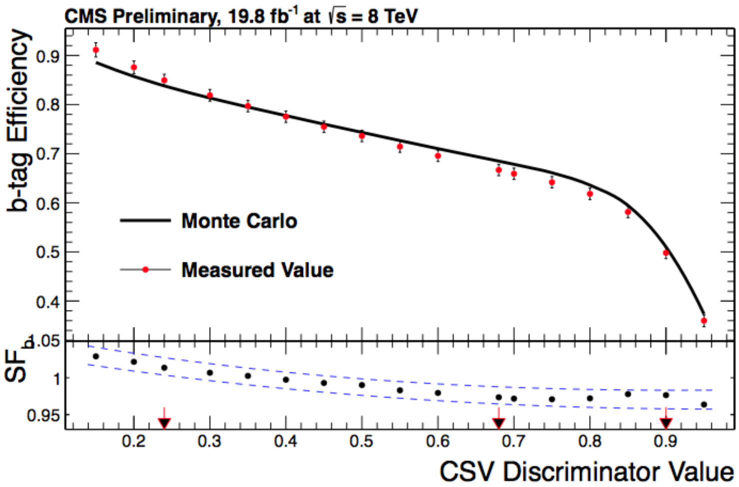


Figure 4.10: b jet tagging efficiency as a function of the discriminator threshold for the CSV algorithm. The efficiency measured in data and predicted by simulation are shown in the upper panel. The scale factor SF_b is shown in the lower panel, where the blue dashed lines represent the combined statistical and systematic uncertainties. The arrows indicate the standard operating points [131].

4.6 Missing transverse energy

Neutral weakly interacting particles, such as neutrinos, do not produce any direct response in the detector elements. The consequence is that these types of particles are not detectable in a typical collider detector. Nevertheless the presence of these particles can be inferred from the imbalance of total momentum. The vector momentum imbalance in the plane perpendicular to the beam direction is known as missing transverse momentum ($\vec{\not{E}}_T$) and its magnitude is called missing transverse energy and is denoted as $\not{E}_T$ (MET).

$$\vec{E}_T^{total} = \sum_{i \in X} \vec{E}_{T_i} = -\vec{\not{E}}_T, \tag{4.29}$$

where $\vec{E}_{T_i} = (E_{x_i}, E_{y_i})$ is the measured transverse momentum of the i^{th} reconstructed object, X is the set of reconstructed objects (such as PF particles) used to calculated $/E_T$. The equation 4.29 ($\vec{E}_T^{total}$) defines the *raw* missing transverse energy that is systematically different from the *true* missing transverse energy. The magnitude of the $\vec{E}_T^{total}$ can be underestimated or overestimated for a variety of reasons as the minimum energy thresholds in the calorimeters, inefficiencies in the tracker and non linearity of the response of the calorimeter for hadronic particles. These effects need to be taken into account in order to have a better estimate of true $\vec{E}_T^{total}$, so four types of corrections are applied.

- **Type-I**:
  It consists in the propagation of the jet energy corrections (JEC), described in the previous section, to the MET. The Type-I correction replaces the vector sum of transverse momenta of particles which can be clustered as jets with the vector sum of the transverse momenta of the jets to which JEC is applied. Classifying all the stable particles of the event into two disjoint sets: clustered as jets or unclustered, the missing transverse energy can be written as:

  $$\vec{\not{E}}_T^{raw} = -\sum_{i \in jets} \vec{p_{T_i}} - \sum_{i \in uncl} \vec{p_{T_i}}. \tag{4.30}$$

  The sum $\sum_{i \in jets} \vec{p_{T_i}}$ is the same as the vector sum of p_T of all jets without JEC application:

  $$\sum_{i \in jets} \vec{p_{T_i}} = \sum_{jets} \vec{p}_{T_{jet}}^{raw}. \tag{4.31}$$

  The Type-I correction replaces the raw jet $\vec{p_T}$ with the corrected jet $\vec{p_T}$ so it can be written as the difference between the two vector sums:

  $$\vec{C}_T^{Type-I} = \sum_{jets} \vec{p}_{T_{jet}}^{raw} - \sum_{jets} \vec{p}_{T_{jet}}^{JEC}. \tag{4.32}$$

- **Type-II**:
  The Type-II correction corrects the energy deposits not clustered in any jet, scaling them by a constant scale factor:

  $$\vec{C}_T^{Type-II} = (1 - C^{uncl}) \sum_{i \in uncl} \vec{p}_{Ti}. \tag{4.33}$$

- **Type-0**:
  The Type-0 correction is used to mitigate the degradation of the MET reconstruction due to pileup interactions. The pileup interactions produce few invisible particles (as neutrinos from the Kaon decays) so that has an effect on the MET reconstruction that shows a degradation as the number of pileup interactions increases. The correction removes charged hadrons originating from the vertices for pileup interactions and also an estimate of neutral pileup contributions. The particles of final state can be classified as whether they are produced in the hard scattering of interest (HS) or in pileup interactions (PU). Moreover a further classification can be done dividing the pileup particles into neutral particles (neuPU) and charged particles (chPU). So the raw missing transverse energy can be expressed as:

  $$\vec{\not{E}}_T^{raw} = -\sum_{i\in HS} \vec{p}_{T_i} - \sum_{i\in neuPU} \vec{p}_{T_i} - \sum_{i\in chPU} \vec{p}_{T_i}. \tag{4.34}$$

  where the last two sums are taken over the neutral and charged particles produced in the pileup interactions. The charged pileup particles can be identified from the vertices so in principle the last sum can be measured. The charged and neutral pileup particles are produced in the same interactions with a very little true MET. So at the true level it can be assumed:

  $$\sum_{i\in neuPU} \vec{p}_{T_i}^{\,true} + \sum_{i\in chPU} \vec{p}_{T_i}^{\,true} = 0. \tag{4.35}$$

  The Type-0 correction also assumes that we can measure charged pile-up particles perfectly thanks to the strong magnetic field and the fact that pileup particles have a low p_T so:

  $$\sum_{i\in chPU} \vec{p}_{T_i}^{\,true} = \sum_{i\in chPU} \vec{p}_{T_i}. \tag{4.36}$$

  Furthermore, the Type-0 correction assumes that we can measure the directions of neutral pileup particles perfectly and measure their energies systematically off by the same factor:

  $$\sum_{i\in neuPU} \vec{p}_{T_i} = R^0 \sum_{i\in neuPU} \vec{p}_{T_i}^{\,true}. \tag{4.37}$$

  With all these assumptions the contributions from the neutral pileup particles, can be estimated as:

  $$\sum_{i\in neuPU} \vec{p}_{T_i} = -R^0 \sum_{i\in chPU} \vec{p}_{T_i}. \tag{4.38}$$

  Then the Type-0 correction can expressed as:

  $$\vec{C}_T^{Type-0} = (1 - R^0) \sum_{i\in chPU} \vec{p}_{T_i}. \tag{4.39}$$

- **xy-Shift**:
  The xy-Shift correction reduces the MET ϕ modulation. This correction is also a mitigation for the pileup effects. The distribution of the true MET should not have a dependence on the ϕ variable because of the symmetry in ϕ of the collisions around the beam axis. However due to anisotropic detector responses, inactive calorimeter cells, the detector misalignment, the displacement of the beam spot, a dependence on ϕ is observed in the reconstructed MET. Moreover this modulation increases linearly with the number of pileup interactions. Shifting the origin of the coordinate in the transverse momentum plane:

  $$\vec{p}_{Ti} \rightarrow \vec{p}_{Ti} - \vec{c}, \tag{4.40}$$

  where $\vec{c}$ is the shift. So, with this shift, the missing transverse energy is:

  $$\vec{\not{E}}_T^{xy} = -\sum_{i \in all} (\vec{p}_{Ti} - \vec{c}) = \vec{\not{E}}_T^{raw} + n \cdot \vec{c}, \tag{4.41}$$

  where n is the number of particles. The total shift $n \cdot \vec{c}$ does not depend on the value of MET. The xy-Shift correction is:

  $$\vec{C}_T^{xy} = n\vec{c} \tag{4.42}$$

Applying all the corrections described above, the corrected value of the MET will be:

$$\vec{\not{E}}_T^{corr} = \vec{\not{E}}_T^{raw} + \Delta, \tag{4.43}$$

where Δ is the sum of the four corrections:

$$\Delta = \vec{C}_T^{Type-I} + \vec{C}_T^{Type-II} + \vec{C}_T^{Type-0} + \vec{C}_T^{xy}. \tag{4.44}$$

The missing transverse energy is an important observable in several analyses as it allows to discriminate signature containing neutrinos from backgrounds that do not have the contribution from this type of particle. So it plays a key role in many physics analyses at the LHC. A nonzero value of the missing transverse energy is not only related to the production of neutrinos in the collision but other causes can also be identified such as: measurement resolution, reconstruction inefficiencies and instrumental defects. In order to identify those events where the reconstructed $\vec{\not{E}}_T$ is consistent only with contributions from particle measurement resolutions and reconstruction inefficiencies the $\vec{\not{E}}_T$ significance, S can be evaluated. Three quantities, for each reconstructed object, are relevant for the derivation of the significance: true transverse momentum $\vec{e}_{Ti}$, the reconstructed transverse momentum $\vec{E}_{Ti}$ and their difference $\vec{\epsilon}_{Ti} = \vec{E}_{Ti} - \vec{e}_{Ti}$. The likelihood of observing a total transverse momentum $\vec{\epsilon}$ under the null hypothesis, for the two object case, is given by:

$$L(\vec{\epsilon}) = \int P_1(\vec{\epsilon}_1|\vec{e}_{T1}) P_2(\vec{\epsilon}_2|\vec{e}_{T2}) \delta(\vec{\epsilon} - (\vec{\epsilon}_1 + \vec{\epsilon}_2)) d\vec{\epsilon}_1 d\vec{\epsilon}_2 \tag{4.45}$$

For an arbitrary number of input objects, the full likelihood function can be generated by a recursive application of eq.4.45. The significance is defined as the log-likelihood ratio:

$$S \equiv 2 \cdot ln(\frac{L(\vec{\epsilon} = \sum \vec{\epsilon_i})}{L(\vec{\epsilon} = 0)}) \tag{4.46}$$

The significance gives the likelihood that the observed $\not{E}_T$ is consistent with zero given the reconstructed content of the event and known measurement resolutions.

Chapter 5

Z+b jet selection

As we have already discussed in Chapter 2, the associated production of the Z boson with one or at least two b jets is a really promising channel both in the search context and to perform pQCD tests. It represents an irreducible background for several Standard Model and beyond SM processes, and for this reason it is of great interest to characterize accurately the Z+b process. Moreover, it could provide precious information to improve our knowledge about the production mechanism of processes in which a b quark is involved, allowing a more accurate description of theoretical calculations. In this thesis two processes are studied: the associated production of the Z boson with exactly one b jet and with at least two b jets. It is interesting to explore separately the two processes because different production mechanisms are involved, but, at the same time, studying them in the same analysis is fundamental in order to take into account possible inefficiencies in the tagging of the jets originating from b quarks and consequently migrations of events from one sample to the other.

In this Chapter the details of the strategy adopted to identify the events where a Z boson is produced in association with exactly one (referred as Z+1b from now on) or at least two b jets (referred as Z+2b) are described. The characteristics of data and Monte Carlo simulation samples used are discussed and a description of the main steps of the analysis are provided. This chapter is organized as follows. The first part is dedicated to the description of the data samples used both for the signal and for the Monte Carlo samples. Then the procedure adopted to select the sample events both the online and the offline selection will be analysed. Details on the methods adopted to extract the background contributions are provided with a dedicated section for the $t\bar{t}$ background. A particular attention has been used to describe the extraction of the b purity of selected events, which has been estimated simultaneously in the Z+1b and Z+2b samples.

5.1 Data and Monte Carlo samples

The study of Z+1b jet and Z+2b jets processes, detailed in this Chapter, has been performed using the full 2012 proton-proton collision data at a center mass energy of 8 TeV. The data have been collected by the CMS experiment at LHC from April 4th to December 16th, 2012. Data must be certified before they can be used for analysis to ensure the good operational status of all subdetectors at the time of data taking. Therefore there is a difference in the integrated luminosity delivered by the LHC machine and the one available for data analysis. In 2012 the luminosity delivered by LHC was about 23.30 fb^{-1} [132], while 21.79 fb^{-1} were recorded by the CMS detector and 19.79 fb^{-1} certified as good for physics analysis during stable beams as shown in fig.5.1.

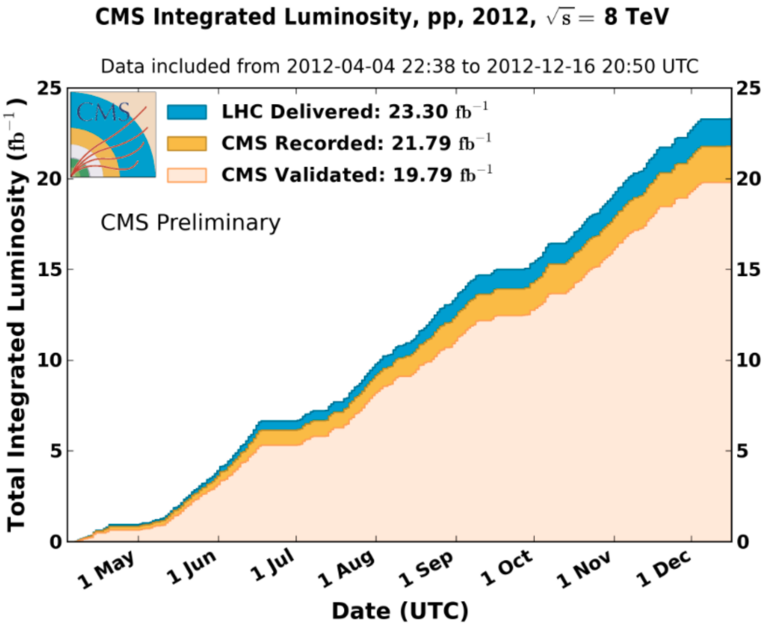


Figure 5.1: The Luminosity delivered by LHC, recorded by CMS and certified as good for physics analysis during stable beams in 2012 at a center mass energy of 8 TeV.

The data samples used in the Z+b analysis correspond to an integrated luminosity L = 19.78 fb^{-1} in the $Z \to e^+e^-$ selection and L = 19.75 fb^{-1} in the $Z \to \mu^+\mu^-$ selection. Due to different physics interests, all data collected by the CMS detector are subdivided into Primary Datasets (PD) each with specific trigger requirements. The analysis relies on the so-called `DoubleElectron`, `DoubleMu` PDs in which a lepton pair is selected by the trigger with specific requirements that will be discussed later. The 2012 data taking has been performed in four different run periods (named A, B, C and D) that refer to different detector and beam conditions. In table 5.1 details

about each period and the corresponding PDs are provided, while tables 5.2 and 5.3 summarize the integrated luminosity available for each PDs used in the analysis for $Z \to e^+e^-$ and the $Z \to \mu^+\mu^-$ selection.
In this analysis the primary dataset (`MuEG`) is used in order to evaluate the $t\bar{t}$ background using a data-driven approach.

Run range	**Primary datasets**
Run2012A: 190456-193621	/DoubleElectron/Run2012A-22Jan2013-v1 /DoubleMu/Run2012A-22Jan2013-v1
Run2012B: 193833-196531	/DoubleElectron/Run2012B-22Jan2013-v1 /DoubleMuParked/Run2012B-22Jan2013-v1
Run2012C: 198022-203742	/DoubleElectron/Run2012C-22Jan2013-v1 /DoubleMuParked/Run2012C-22Jan2013-v1
Run2012D: 203777-208686	/DoubleElectron/Run2012D-22Jan2013-v1 /DoubleMuParked/Run2012D-22Jan2013-v1

Table 5.1: List of Primary Datasets (PD) used in the analysis for each run period both for the electron and muon final state.

5.1.1 Monte Carlo samples

In order to predict the behaviour for both the signal and background, different Monte Carlo samples have been used in this study. The samples have been simulated at 8 TeV center of mass energy to obtain consistent comparison with data collected in 2012. In table 5.4 the Monte Carlo samples used are summarized for the signal and the different background processes specifying the Generator used and the name of each dataset. The leading order (LO) cross sections, used as internal cross section during the generation step of the simulation, are shown for each sample. All samples have been rescaled to the next-to leading order (NLO) or next-to-next to leading order (NNLO) calculations when available, as shown in table 5.5. The cross section of the Drell-Yan process includes the Z boson branching fraction. After the generation step, events are passed through the full CMS detector simulation obtained with GEANT 4 [133]. This toolkit allows to implement the geometry of all the

Primary Dataset	$\int Ldt$ (fb^{-1})
/DoubleElectron/Run2012A-22Jan2013-v1	0.889
/DoubleElectron/Run2012B-22Jan2013-v1	4.429
/DoubleElectron/Run2012C-22Jan2013-v1	7.152
/DoubleElectron/Run2012D-22Jan2013-v1	7.318

Table 5.2: The integrated luminosity available for each PD used in the analysis for the electron final state.

Primary Dataset	$\int Ldt$ (fb^{-1})
/DoubleMuParked/Run2012A-22Jan2013-v1	0.889
/DoubleMuParked/Run2012B-22Jan2013-v1	4.429
/DoubleMuParked/Run2012C-22Jan2013-v1	7.152
/DoubleMuParked/Run2012D-22Jan2013-v1	7.318

Table 5.3: The integrated luminosity available for each PD used in the analysis for the muon final state.

subdetectors that form CMS and also to simulate the interaction of particles with the detector.

Signal

Three different samples have been chosen to describe our signal allowing the comparison between data and theoretical predictions. They have been simulated using MadGraph [81] and Powheg [82] generators:

- Drell-Yan Z + jets events: the hard scattering process, with the emission of up to four additional partons, is calculated at Leading Order by MadGraph 5 interfaced with PYTHIA 6 (tune $Z2^{*}$) [75], for the hadronization. For the calculation of the matrix element the 5-flavor scheme is adopted so in the massless b quark approximationAs seen in section 2.1.2, in this approach an initial state parton density function for a massless b quark is introduced including the gluon splitting $g \to b\bar{b}$ in it.

- Drell-Yan $Z + b\bar{b}$ events: in this case the matrix element is calculated by MadGraph 5 but a different description of the b quark production is considered, the 4-flavor scheme. In this approach only four massless quark densities are considered in the initial state and non zero mass b quarks appear only in final states through gluon splitting. Also in this case the hadronization is left to PYTHIA 6 with tune $Z2^{*}$.

- Drell-Yan Z + jets events: for the hard scattering process calculation POWHEG [82] generator is employed. Only one jet is generated from the matrix element for the Drell-Yan at the next-to-leading-order while the other jets of any flavour are obtained with the Parton Shower provided by PYTHIA 6, tune $Z2^{*}$.

Background

Many Standard Model processes show a similar topology like the processes with a Z boson and b jets in the final state. A complete knowledge of all the background contributions is needed also to perform a comparison between

Process	**Dataset**	**Generator**
DY Z + jets	$DYJetsToLL\_M - 50\_TuneZ2Star\_8TeV - madgraph - tarball/Summer12\_DR53X - PU\_S10\_START53\_V7A - v1$	MadGraph 5FS
$Z + b\bar{b}$	$/Zbb\_4F\_8TeV\_madgraph/Summer12\_DR53X - PU\_S10\_START53\_V19 - v1$	MadGraph 4FS
W + jets	$/WJetsToLNu\_TuneZ2Star\_8TeV - madgraph - tarball/Summer12\_DR53X - PU\_S10\_START53\_V7A - v2$	MadGraph
$t\bar{t}$	$/TTJets\_MassiveBinDECAY\_TuneZ2star\_8TeV - madgraph - tauola/Summer12\_DR53X - PU\_S10\_START53\_V7A - v1$	MadGraph
ZZ	$/ZZ\_TuneZ2star\_8TeV\_pythia6\_tauola/Summer12\_DR53X - PU\_S10\_START53\_V7A - v1$	PYTHIA
WZ	$/WZ\_TuneZ2star\_8TeV\_pythia6\_tauola/Summer12\_DR53X - PU\_S10\_START53\_V7A - v1$	PYTHIA
WW	$/WW\_TuneZ2star\_8TeV\_pythia6\_tauola/ * Summer12\_DR53X - PU\_S10\_START53\_V7A - v1$	PYTHIA
single top t^+ (s-channel)	$/Tbar\_s - channel\_TuneZ2star\_8TeV - powheg - tauola/Summer12\_DR53X - PU\_S10\_START53\_V7A - v1$	POWHEG
single top t^+ (t-channel)	$/Tbar\_t - channel\_TuneZ2star\_8TeV - powheg - tauola/Summer12\_DR53X - PU\_S10\_START53\_V7A - v1$	POWHEG
single top tW	$/Tbar\_tW - channel - DR\_TuneZ2star\_8TeV - powheg - tauola/Summer12\_DR53X - PU\_S10\_START53\_V7A - v1$	POWHEG
single top t^- (s-channel)	$/T\_t - channel\_TuneZ2star\_8TeV - powheg - tauola/Summer12\_DR53X - PU\_S10\_START53\_V7A - v1$	POWHEG
single top t^- (t-channel)	$/T\_s - channel\_TuneZ2star\_8TeV - powheg - tauola/Summer12\_DR53X - PU\_S10\_START53\_V7A - v1$	POWHEG
single top tW	$/T\_tW - channel - DR\_TuneZ2star\_8TeV - powheg - tauola/Summer12\_DR53X - PU\_S10\_START53\_V7A - v1$	POWHEG

Table 5.4: Signal and background simulated samples. For each sample the generator used and the name of the dataset are specified.

Process	σ_{LO} (pb)	σ_{norm} (pb)
Drell-Yan Z + jets	2950	3503,71 (NNLO)
$Z + b\bar{b}$	76.75	76.75 (LO)
W + jets	30400	36703.2 (NNLO)
$t\bar{t}$	136.3	225.197 (NLO)
ZZ	5.196	8.059 (NLO)
WZ	12.63	33.21 (NLO)
WW	34.01	54.838 (NLO)
single top t^+ (s-channel)	1.57	1.76 (NNLO)
single top t^+ (t-channel)	25	30.7 (NNLO)
single top tW	10.7	11.1 (NNLO)
single top t^- (s-channel)	2.82	3.79 (NNLO)
single top t^- (t-channel)	47	56.4 (NNLO)
single top tW	10.7	11.1 (NNLO)

Table 5.5: The Leading Order cross section σ_{LO} and the cross section used for the normalization σ_{norm} are shown for each Monte Carlo sample.

data and Monte Carlo. All the backgrounds are evaluated from Monte Carlo simulation except for $t\bar{t}$ case in which a data driven approach is adopted. Some background samples ($W + jets$ and $t\bar{t}$) are generated with the combination of MadGraph 5 and PYTHIA 6, the diboson contributions (WW, ZZ, WZ) are described with PYTHIA 6 while the single top samples are provided by the POWHEG generator interfaced with PYTHIA6.

5.1.2 Pile-up reweighting

The LHC detectors have to operate in a very harsh environment due to a high frequency of bunch crossing and high event rate. The total event rate is so high that several soft interactions overlap to the high transverse momentum interaction (pileup). In a single bunch crossing a large number of interactions may occur. The mean number of interactions per bunch crossing during 2012 is around 21, as shown in fig. 5.2. Two different types of pileup have to be considered:

- "in-time pileup": overlap of events due to the interaction between partons of the same bunch crossing
- "out-of-time pileup": caused by the events of different bunch crossing

In order to have a realistic description of a process, pileup events have to be generated and superimposed to the signal in the Monte Carlo. The number of in- and out-of-time pileup interactions is generated in the Monte Carlo samples with a Poisson distribution. This number is meant to roughly

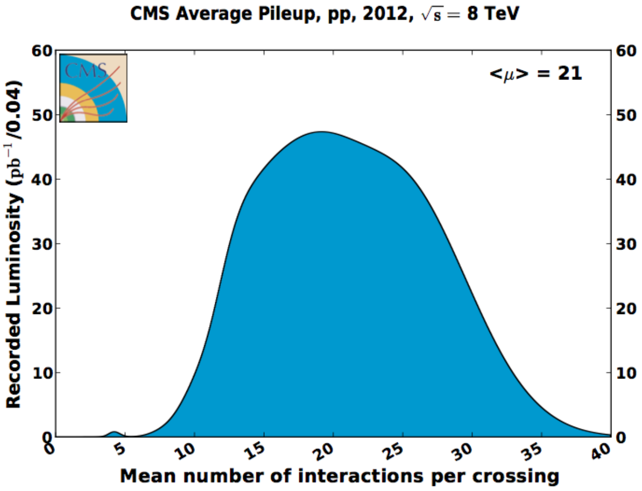
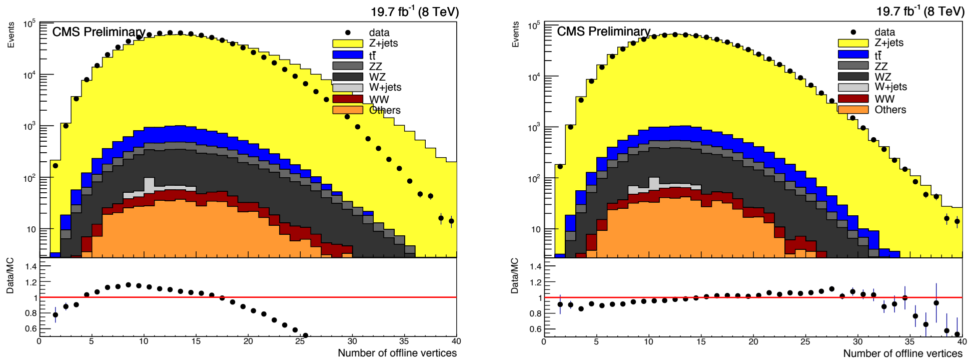


Figure 5.2: Mean number of interactions per bunch crossing (average pile-up) in 2012 data taking.

cover, though not exactly match, the conditions expected for each data-taking period. In order to reproduce the expected number of the pileup interactions correctly, the Monte Carlo events have thus to be reweighted. The Monte Carlo reweighting procedure is done by assigning weights to a Monte Carlo event to reweight the number of pileup interactions. Two inputs are thus necessary for reweighting Monte Carlo events to match the data sample used in the analysis. The first is the histogram of the pileup distribution corresponding exactly to the luminosity distribution in the data sample and the second is the distribution of the number of pileup events in the Monte Carlo samples. The bin-per-bin ratio w_{PU_i} between the data distribution (N_i^{DATA}) and the input distribution from Monte Carlo (N_i^{MC}):

$$w_{PU_i} = \frac{N_i^{DATA}}{N_i^{MC}} \tag{5.1}$$

is then used to extract a weight factor for each of the input events. In order to validate the reweighting procedure applied, the number of reconstructed vertices distribution can be used as a check. Figure 5.3 shows the distributions of the reconstructed vertices in the Z+jets sample in which a Z boson is selected through its decay modes into electrons, $Z \to e^+e^-$, (similar results are obtained also in the $Z \to \mu^+\mu^-$ selection) with the requirement of at least one jet. The distribution shows the comparison between data and different contributions of background before (fig. 5.3(a)) and after (fig. 5.3(b)) the Monte Carlo pileup reweighting. The discrepancy between the data ad simulation evident in the left plot is significantly reduced after the reweighting procedure is applied.



(a) Number of reconstructed vertices before the pileup reweighting of the Monte Carlo.

(b) Number of reconstructed vertices after the pileup reweighting of the Monte Carlo.

Figure 5.3: Number of reconstructed vertices in the Z +jets sample.

5.2 Triggers

As already seen the CMS trigger and data acquisition are based on two levels of online event selection. Dedicated triggers menu with specific requirements have been used to select the datasets considered in the analysis. The HLT trigger path used to select the "DoubleElectron" dataset is:

- $HLT\_Ele17\_CaloIdT\_CaloIsoVL\_TrkIdVL\_TrkIsoVL$ $\_Ele8\_CaloIdT\_CaloIsoVL\_TrkIdVL\_TrkIsoVL$

Two electrons are required, one with a transverse momentum greater than 17 GeV and at least another electron with a minimum transverse momentum of 8 TeV. The selections consist also in tight requirements on $CaloId$ variables for both electrons, i.e. $CaloIdT$ which implies cuts on shower shape variables such as:

- H/E: defined as the ratio between the Hadronic calorimetric tower energy behind seed crystal of the seed basic cluster (H) and the energy of the electromagnetic supercluster (E)
- $\sigma_{\eta\eta}$: useful to study the lateral development of the shower; it is defined as:

$$\sigma^2_{\eta\eta} = \frac{\sum_i^{5x5} w_i(\eta_i - e\bar{t}a_{5x5})^2}{\sum_i^{5x5} w_i}, \tag{5.2}$$

  where the index i runs over the crystals that form the 5x5 matrix around the crystal seed, η_i is the pseudorapidity of the i^{th} crystal and $\eta\bar{_{5x5}}$ is defined as

$$\eta\bar{_{5x5}} = \frac{\sum_i^{5x5} w_i\eta_i}{\sum_i^{5x5} w_i} \tag{5.3}$$

and with w_i, weight for the i^{th} crystal:

$$w_i = max(0, 4.7 + \ln \frac{E_i}{E_{5x5}}), \tag{5.4}$$

where E_i and E_{5x5} represent the energy of the i^{th} and the $5x5$ block of crystals respectively.

Very Loose (VL) isolation cuts are applied on the variables, $IsoECAL$, $IsoHCAL$ ($CaloIsoVL$) and $IsoTRK$ ($TrkIsoVL$), and finally a Very Loose (VL) threshold is applied ($TrkIdVL$) on the following variables:

- $\Delta\eta$: distance in the η direction between the track and the SuperCluster
- $\Delta\phi$: distance in the ϕ direction between the track and the SuperCluster

In table 5.6 a summary of the trigger identification and isolation cuts is shown both for the barrel and the endcap case. The electrons are then matched by ΔR to the trigger objects ($\Delta R < 0.3$), only one match per trigger object is allowed and the best match found per reco object is taken.

variable	**criterion**
$CaloIdL$	$H/E < 0.15$ (0.10) $\sigma_{\eta\eta} < 0.014$ (0.035)
$CaloIsoVL$	$IsoECAL/E_T < 0.2$ (0.2) $IsoHCAL/E_T < 0.2(0.2)$
$TrkIsoVL$	$IsoTRK/E_T < 0.2$ (0.2)
$TrkIdVL$	$\mid \Delta\eta \mid < 0.01$ (0.01) $\Delta\phi < 0.15$ (0.10)

Table 5.6: Triggers Identification and isolation cuts for the barrel (endcap).

A dedicated trigger is used to select muons:

- HLT_Mu17_Mu8

in which a muon with a transverse momentum greater than 17 GeV and one with a transverse momentum greater than 8 TeV are required. A similar matching to the trigger objects as in the electron case is performed also for muons.

The electron and muon of the "MuEG" dataset, used to estimate the top-antitop background, are required to pass the HLT trigger menu defined by at least one of the following conditions:

- $HLT\_Mu17\_Ele8\_CaloIdT\_CaloIsoVL\_TrkIdVL\_TrkIsoVL$

- $HLT\_Mu8\_Ele17\_CaloIdT\_CaloIsoVL\_TrkIdVL\_TrkIsoVL$

whose requirements are: a muon with $p_T > 17$ GeV and an electron with $p_T > 8$ GeV and with a tight cut on $CaloId$, Very Loose isolation cuts and a Very Loose threshold on $TrkIdVL$ or a muon with $p_T > 8$ GeV and an electron with $p_T > 17$ GeV.

5.3 Event Selection

The first event selection is performed at the *online* level; it consists of the trigger requirements described in the previous section. After the reconstruction of the physics objects using the Particle Flow algorithm, the data are analysed *offline* to select the interesting objects and to distinguish the processes in which the Z boson is produced in association with b jets. The Z boson candidates are reconstructed looking for two well identified, isolated and high energy opposite charge leptons (electrons or muons). The dielectron or dimuon pairs passing the identification requirements are used to reconstruct the $Z \to e^+e^-$ or $Z \to \mu^+\mu^-$ candidates. The events are selected if the reconstructed Z boson has an invariant mass within the range $[71, 111]$ GeV.

This section is dedicated to the description of the analysis selection criteria for electrons, muons and jets.

5.3.1 Electrons

Once the analysis objects are reconstructed using the Particle Flow algorithm as already discussed, a series of preselection cuts are applied before the final selection requirements. The preselection cuts have been optimized to obtain a first rejection of those events that reasonably will not pass the following selection cuts. These requirements are looser than the selection cuts, in order to avoid a loss in efficiency and also to avoid the introduction of bias in the the analysis procedure. Therefore the electrons that matched the trigger objects have to satisfy requirements on three different sets of variables:

- electron identification variables: η_{SC}, $\Delta\eta$, $\Delta\phi$, $\sigma_{i\eta i\eta}$, H/E
- Particle Flow isolation variable I (5.5), computed from the flux of Particle Flow candidates found within a cone of $R = 0.3$ built around the electron direction applying also the “delta beta” ($\Delta\beta$) correction:

$$I = \frac{1}{p_T^e}[I_{ch} + max(I_{nh} + I_\gamma - 0.5 I_{chPU}, 0)]. \tag{5.5}$$

  Where I_{ch} is the sum of the transverse energy of all the charge particles within the cone, excluding the electron, coming from primary vertex, I_{nh} and I_γ are the sums of the transverse energy of all the neutral hadron within the cone and the photons respectively. I_{chPU} is an estimate charged particle deposits not from primary vertex. The $\Delta\beta$ correction is applied in order to take into account the pileup contamination. Charged particles from vertices other than the primary vertex can enter in the isolation cone and so contribute to the isolation value. With Particle Flow, these pileup particles can be sorted out and just not counted. But there are also neutral particles from pileup. Since there are no tracks associated to these neutral particles, it is not possible to assign them to vertices. So as an estimate, we subtract half of the sum of the transverse momentum of the charged particles in the cone of interest but with particles not originating from the primary vertex.

- conversion rejection variables to reject electrons coming from conversions: d_0, d_z and $miss_{hits}$

The selection efficiency obtained with the cuts chosen is called *working point* (WP). Table 5.7 shows the full set of electron identification and isolation criteria applied in the preselection corresponding to a WP80 that means we have an efficiency of 80% in the electron identification.

variable	**cut**	
	Barrel	Endcap
identification variables		
η_{SC}	< 1.442	> 1.566 & < 2.5
$\Delta\eta$	< 0.004	< 0.007
$\Delta\phi$	< 0.06	< 0.03
$\sigma_{i\eta i\eta}$	< 0.01	< 0.03
H/E	< 0.12	< 0.10
PF isolation variable		
I	< 0.15	< 0.15
conversion rejection variables		
d_0	< 0.02 cm	< 0.02 cm
d_z	< 0.1 cm	< 0.1 cm
$miss_{hits}$	≤ 1	≤ 1

Table 5.7: Electron selection criteria and cut thresholds for the barrel and the endcap cases.

The electron momentum regression corrections described in section 4.2 are applied. These corrections, optimized using a multivariate regression technique based on a Boosted Decision Tree (BDT), are needed to improve the determination of the electron energy.

The last steps of the selection are the requirements:

- to have a transverse momentum p_T greater than 20 GeV

- to be inside the tracker coverage: $\eta < 2.4$

In order to correct the simulation for the efficiencies the so called scale factors, resulting from the ratio between efficiencies in data and in Monte Carlo, are applied. The two highest energy opposite charge electrons selected are considered to calculate the invariant mass identifying Z boson candidates.

5.3.2 Muons

Two Particle Flow muons are selected with opposite charge in the Z mass range. Also in this case, some identification criteria are imposed on the reconstructed muons in order to reduce the muon fake rate. A tight selection is applied on muons that have to be isolated to reduce the contamination from b or c semileptonic decays. A muon is considered isolated if, in a cone of size $R = 0.4$ centred on the muon direction, the sum of the transverse momentum of the tracks and of the transverse energy of the calorimeter hits is less than 20% of the muon p_T. The "delta beta" correction is applied to correct for pileup contamination (eq. 5.5). The two leptons are also required to be reconstructed as "global muons", meaning that a valid fit exists between the track from the central tracking system and the track from the muon system and, in order to ensure a good quality of the global fit used for the muon parameters estimate, a cut on the χ^2 normalised to the number of degrees of freedom is applied. Table 5.8 summarizes all the identification criteria applied and cuts used to reject cosmic muons. The tight ID working point cuts are imposed on a set of variables already described in the Sec. 4.3.1. The muon momentum have to be corrected in order to take into account the effects due to the reconstruction capability and limited knowledge of the physical configuration of the detector. The momentum scale corrections applied are provided by the MuScleFit (Muon Scale Fit) package [119]. As in the case of the electrons, considering the highest energy opposite charge muons with:

- transverse momentum greater than 20 GeV
- pseudorapidity variable comprised in the range $[-2.4, 2.4]$

the invariant mass is calculated to identify the Z boson candidate. Scale factors, defined as the ratio between the efficiencies measured in data and in Monte Carlo, are used to rescale the simulation in order to match data.

5.3.3 Jets

In the selection of jets produced in association with the Z boson that has passed the selection described above, many aspects have to be taken into account. One problem in the jet reconstruction is that there can be ambiguities: a reconstructed muon or an high energy electron could be a jet constituent.

variable	**cut**
identification variables	
χ^2/f	< 10
N layer with measurement	> 5
Number of pixel hits	> 0
Number of muon hits	> 0
d_0	< 0.02
segments matched	> 1
PF isolation variable	
I	< 0.2

Table 5.8: Tight muon selection criteria and cut thresholds.

In order to resolve these ambiguities, the particles are reconstructed following a fixed order. First the charged hadronic pileup particles, that are not associated to the considered primary vertex, are subtracted. Only about 20% pileup charged hadrons remain untouched. This compromise is necessary to avoid over-subtracting high p_T tracks from jets. The muons are reconstructed before the electrons reconstruction and the jet clusterization as they are the only particles which can be detected in the muon system. All the information assigned to the isolated muons, as the hits in the tracking system, is then projected out of the event and no longer available for the reconstruction of the other objects. Only isolated electrons are considered in the analysis and thus their entries from the tracking system and from the electromagnetic calorimeter are removed. Only after the removal of isolated muons and electrons, jets can be reconstructed using the *anti-kt* algorithm with a clusterization parameter of $R = 0.5$ already described in Sec. 4.4.1 . A set of corrections are applied on the jet energy in order to take into account different effects achieving:

- the removal of the energy coming from pileup events ($L1$ correction)
- a uniformity in the pseudorapidity of the jet response ($L2$ correction)
- a flat response as a function of the transverse momentum ($L3$ correction)
- a fix for the residual differences between data and and Monte Carlo simulation. Differently from the other corrections, this one is applied only in data. ($LV2LV3Res$)

The jet energy resolution (JER) in data is worse than in the simulation. The reconstructed momentum of jet in simulated events is therefore smeared applying a correction factor depending on the pseudorapidity value of the Monte Carlo jet. Several cuts (summarized in tab. 5.9) are applied on the energy composition of PF-jets with at least two clusterized particles. Jets are required to have charge hadron fraction $CHF > 0$ if within the tracking fiducial region of $\eta < 2.5$, neutral hadron fraction $NHF < 0.99$, charged electromagnetic fraction (electrons) $CEF < 0.99$ and neutral electromagnetic (photon) fraction $NEF < 0.99$. These requirements remove fake jets arising from spurious energy depositions in a single sub-detector.

variable	**cut**
Charged Multiplicity	> 0
Number of constituents	> 1
CHF	> 0
CEF	< 0.99
NEF	< 0.99
NHF	< 0.99

Table 5.9: Cuts applied on the energy composition of PF-jets.

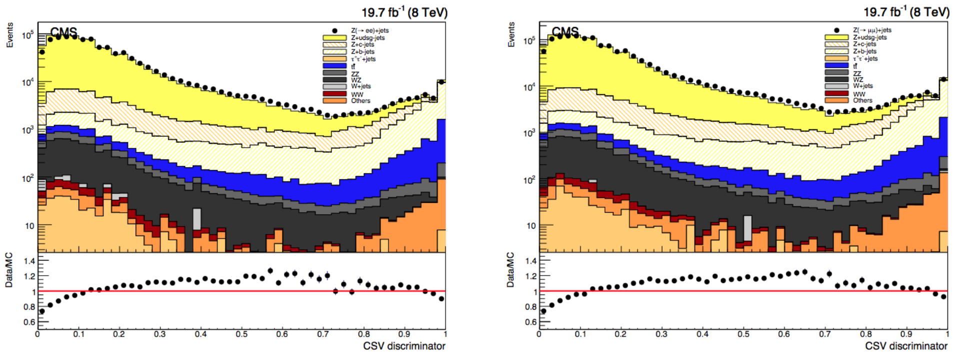
The final selection on the jets requires:

- a minimum threshold on the jet transverse momentum p_T of 30 GeV to reduce the contamination from the underlying event and increase jet energy resolution
- η variable within the range [-2.4,2.4], inside the tracker acceptance to ensure a good quality of the tracking information
- a minimum distance of 0.5 in ΔR between the jet and the two leptons identified as decay products of the Z boson. This cut is applied in order to avoid the clusterization of residual final state radiation from the leptons.

5.3.4 b-tagging

After the selection of the jets collection, for each of them a b tagging procedure is applied in order to identify those jets that are originated from the hadronization of the b quarks. The b tagging algorithm chosen is based on inclusive secondary vertex reconstruction. Several topological and kinematical secondary vertex related variables as well as information from track impact parameters are combined into a single tagging variable to discriminate between jets originating from b quarks and the ones from other sources. This algorithm is the *Combined Secondary Vertex* (CSV) [134](already described in sec. 4.5), used in this analysis with a a tight working point: $CSV > 0.895$. In fig. 5.4 the Combined Secondary Vertex discriminant distribution, after the Z and jet selection, is shown for the inclusive sample.
Two different exclusive samples are selected, one with a Z boson reconstructed and exactly one b tagged jet (Z+1b) and the other one with a Z boson associated with at least two b tagged jets (Z+2b). The Monte Carlo events with

at least one b tagged jet are reweighted with scale factors in order to take into account the difference between the b tagging efficiencies measured in data and those predicted by the simulation. The scale factors applied are defined as the ratio between the b tagging efficiency estimated in data and in simulation: $SF = \frac{\varepsilon_{data}}{\varepsilon_{MC}}$ (see section 4.5.3). In the CSV discriminant shown in fig. 5.4 the scale factors are not applied indeed the plot shows a deviation between data and simulation evident in the ratio plot.



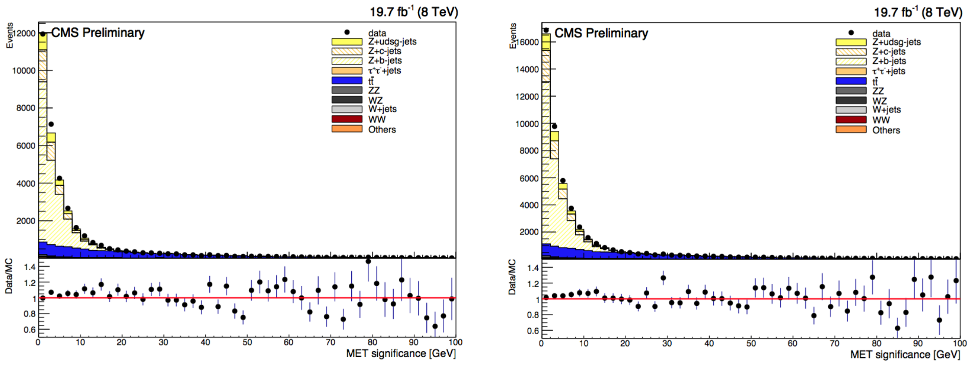
(a) Combined Secondary Vertex Discriminant in the Z+jets inclusive sample in the dielectron final state.

(b) Combined Secondary Vertex Discriminant in the Z+jets inclusive sample in the dimuon final state.

Figure 5.4: Combined Secondary Vertex Discriminant in the inclusive sample with a Z boson, selected in the electron or in the muon decay mode, and at least one jet of any flavour.

5.3.5 Missing transverse energy

For each event an additional cut on the missing transverse energy significance ($MET^{sign} < 30$ GeV) is applied to reduce the $t\bar{t}$ background. This variable already defined in section 4.6 is an estimate of the compatibility of the reconstructed missing transverse energy with zero. The *Type-0*, *Type-1* and *Type-2* corrections are applied in order to reduce the degradation of the MET reconstruction due to the pileup interactions, to propagate the jet energy corrections (JEC) to MET and to correct the remaining energy deposits not clustered in any jet. Finally the *xy-Shift* correction reduces the MET φ modulation. This correction is also a mitigation for the pile-up effects. Fig. 5.5 shows the MET significance distribution obtained after all the corrections have been applied in the sample with a Z boson and at least one b tagged jet.



(a) Missing transverse energy significance distribution in the electron final state for the sample with a Z+b sample.

(b) Missing transverse energy significance distribution in the muon final state for the sample with a Z+b sample.

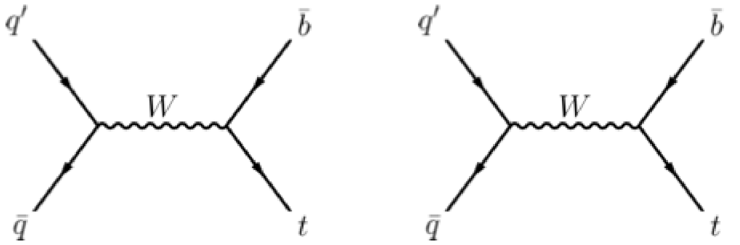
Figure 5.5: Missing transverse energy significance distribution for the Z with at least one b tagged jet sample.

5.4 Backgrounds

Several physics final state have signatures that mimic the Z+1b jet and Z+2b jets processes having the same final state. The background contributions come from the production of a pair of vector bosons Z or W in association with jets and from the top-antitop ($t\bar{t}$) production. The latter is the dominant background especially for the Z+2b process. The diboson events that could provide a similar Z+b final state are:

- ZZ+jets$\rightarrow 4l^{\pm}$+jets
- ZZ+jets$\rightarrow 2l^{\pm} + q\bar{q}$
- $W^{+}W^{-}$+jets$\rightarrow 2l^{\pm} + 2\nu$+jets
- $W^{\pm}Z$+jets $\rightarrow 3l^{\pm} + \nu$+jets
- $W^{\pm}Z$+jets$\rightarrow l^{\pm} + \nu + q\bar{q}$

Another source of contamination for the $Z + b$ final state comes up from the single top production in the s-channel, in the t-channel and when the top quark is produced in association with a W boson. In the first two cases the single top is produced with a b quark (see fig. 5.6) so in the final state we can have an isolated lepton and at least two b jets as the top quark decays into a W boson and a b quark. However the dominant contribution comes from the associated production of a W boson with a top quark. Also the Drell-Yan $Z \rightarrow \tau^{+}\tau^{+}$ production in association with two or more jets and the W+jets with a misidentified lepton can contribute to the backgrounds. All

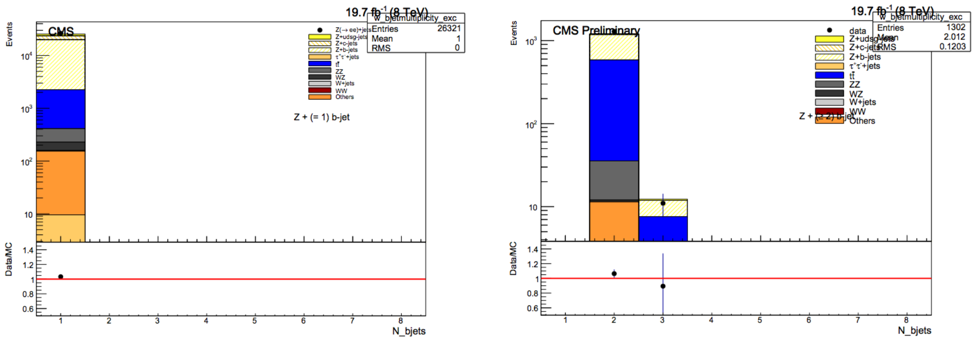


(a) Leading order Feynman diagrams for single t production in the s-channel.

(b) Leading order Feynman diagrams for single t production in the t-channel.

Figure 5.6: Feynman diagrams for single t production in the s-channel and in the t-channel.

these processes are modelled using dedicated Monte Carlo samples. In order to provide a better estimate of the $t\bar{t}$ background a data driven approach is used and the strategy will be described in a dedicated subsection. After the selection applied, the different background contributions are shown in the exclusive b jet multiplicity distribution in the Z+1b and Z+2b samples in fig. 5.7.



(a) b jet multiplicity in the sample with exactly one b jet produced in association of a Z boson in the dielectron decay mode.

(b) b jet multiplicity in the sample with at least two b jets produced in association of a Z boson in the dielectron decay mode.

Figure 5.7: Exclusive b jet multiplicity for the Z+1b and Z+2b samples. These distributions provide an easy way to illustrate the different background contributions in the two samples.

5.4.1 Top-antitop pair background

The $t\bar{t}$ background is the main contribution for the Z+b analysis in particular when studying the final state with a Z boson produced in association with at least two b jets. The *top* quark decays, in fact, into a W boson and a b quark with a branching fraction of about 100% due to the fact that $V_{tb} \sim 1$ in the CKM matrix. When a $t\bar{t}$ pair is produced, two opposite charge leptons and two b tagged jets in the final state may mimic our signature (see fig. 5.8). In order to perform a precise estimate of the $t\bar{t}$ background a data-driven approach has been chosen. A dataset, `MuEG` (from now on referred as '$e\mu$'), with a preselection in which a muon and an electron are selected with specific cuts, as previously seen, is used. The same selection adopted for the Z+1b or Z+2b processes is applied on the $e\mu$ sample requiring that the invariant mass reconstructed with the selected muon and electron lies in the mass range $[71, 111]$ GeV. In order to evaluate the shape of the $t\bar{t}$ contribution, all the measured observables studied in the Z+b are built using the selected events in the $e\mu$ sample. The normalization of the $e\mu$ sample is obtained by fitting the invariant mass sidebands ($M_{ee,\mu\mu}$ <85 GeV and $M_{ee,\mu\mu}$ >100 GeV) of the Z+b distribution (built without the cut on the MET significance in order to enrich the Z+b sample with $t\bar{t}$ events) with the sidebands of the invariant mass reconstructed in the $e\mu$ sample. Fig. 5.9 and 5.10 show the fit of the invariant mass sidebands in the electron and muon final state. The coefficient c_t extracted by the fit is then used to normalized the shape form the $e\mu$ sample before the background subtraction.

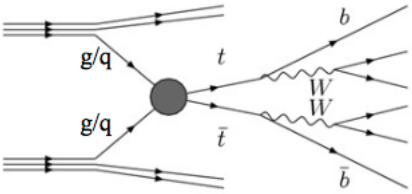
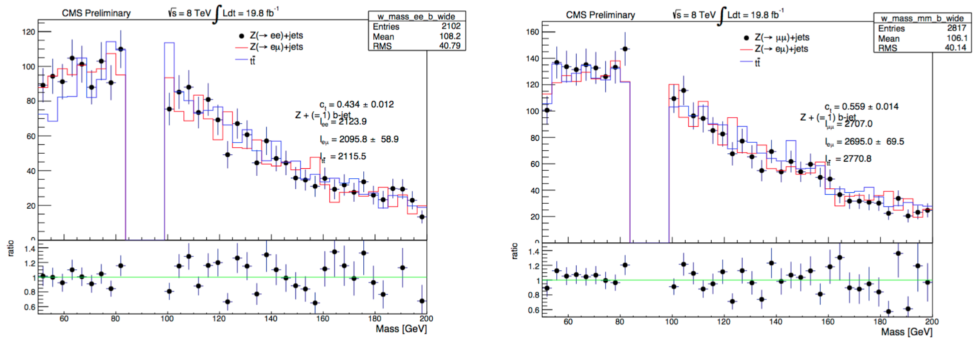


Figure 5.8: Feynman diagram of the $t\bar{t}$ pair production. In the final state two opposite sign leptons and two b jets can be observed.

5.5 Flavor composition of the b tagged sample

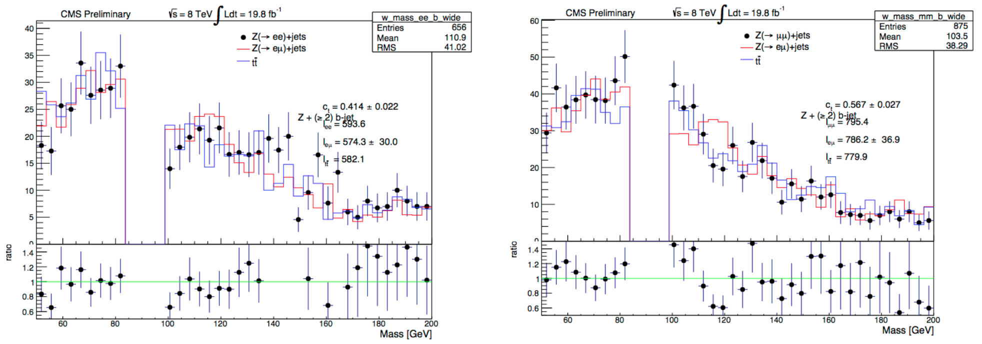
In the previous section, the methods used to extract the different background contributions (from diboson, $t\bar{t}$ and single top events) have been discussed. However a not negligible contribution comes from Z+jets events where charm and light/gluon jets are misidentified as b jets. To quantify this contribution and evaluate the real fraction of b jets (b purity) in the selected events, the best procedure is to choose a distribution in which the shapes of the quark components are different and so it is easier to discriminate among them. The identification of jets originated from lighter quarks is based on the information provided by the Monte Carlo. Templates can be built from Monte Carlo,



(a) The sidebands of the invariant mass distribution (in the case of $Z(\rightarrow e^+e^-)$+1b jet) fitted with the sidebands of the $e\mu$ invariant mass distribution.

(b) The sidebands of the invariant mass distribution in the case of $Z(\rightarrow \mu^+\mu^-)$+1$b$ jet) fitted with the sidebands of the $e\mu$ invariant mass distribution.

Figure 5.9: The sidebands (85 GeV < $M_{ee,\mu\mu}$ and $M_{ee,\mu\mu}$ > 100 GeV) of the invariant mass distribution (in the case Z+1b) fitted with the sidebands of the $e\mu$ invariant mass distribution. Black dots represent data, red line represents the $e\mu$ sideband shapes and blue line is the $t\bar{t}$ Monte Carlo that is shown in this plot only as reference as it is not used to estimate the top-antitop background.



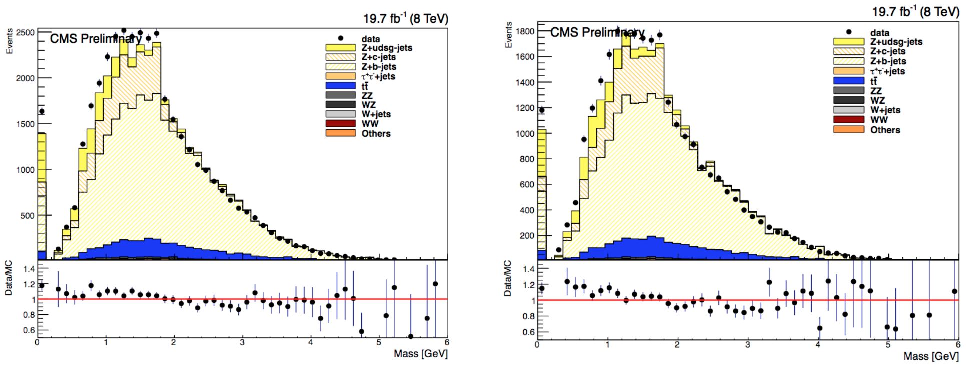
(a) The sidebands of the invariant mass distribution (in the case of $Z(\rightarrow e^+e^-)$+2b jet) fitted with the sidebands of the $e\mu$ invariant mass distribution.

(b) The sidebands of the invariant mass distribution in the case of $Z(\rightarrow \mu^+\mu^-)$+2$b$ jet) fitted with the sidebands of the $e\mu$ invariant mass distribution.

Figure 5.10: The sidebands (85 GeV < $M_{ee,\mu\mu}$ and $M_{ee,\mu\mu}$ > 100 GeV) of the invariant mass distribution (in the case Z+2b) fitted with the sidebands of the $e\mu$ invariant mass distribution. Black dots represent data, red line represents the $e\mu$ sideband shapes and blue line is the $t\bar{t}$ Monte Carlo that is shown in this plot only as reference as it is not used to estimate the top-antitop background.

matching the flavor of the reconstructed jet to the parton flavor. In order to know what the expected contents of truly b flavored jets and mistagged jets are, a fit of the shape of the different contributions to data is performed. Through this fit the scale factors c_{guds}, c_c and c_b are extracted and then used to rescale the Monte Carlo components. Different distributions with a high discrimination power can be chosen for this purpose:

- The Secondary Vertex Mass (*SVTX Mass*) is defined as the invariant mass of all the tracks at the secondary vertex (assuming the neutral pion mass for all secondary tracks). The discrimination power of this variable is due to the difference of the secondary vertex mass values when a jet is originated from the hadronization of a b, c or light quark. Fig. 5.11 shows the distribution of the SVTX mass requiring events with a Z boson with at least one b tagged jet. The first bin in the distribution contains the events of the Pseudo Vertex and No Vertex categories explained in Sec. 4.5.2. The Drell-Yan Monte Carlo is decomposed into the three different contributions from b, c and light (u,d,s) quarks. One can observe the edge related to the D meson around 2 GeV and the slope of the region pure in B meson decays, between 2 and 5 GeV, correctly reproduced.



(a) SVTX mass distribution after the Z+b selection with a Z decaying into electrons.

(b) SVTX mass distribution after the Z+b selection with a Z decaying into muons.

Figure 5.11: Secondary vertex mass distributions for the leading b tagged jet in the events with a Z boson with at least one b jet.

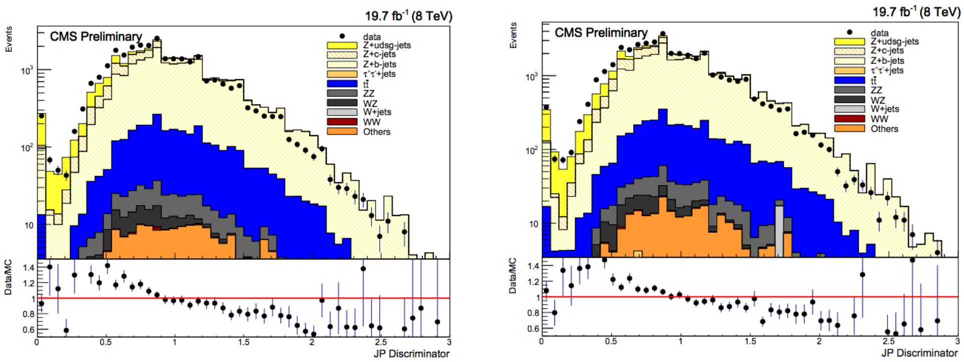
- The Jet Probability (*JP*) is a discriminator that exploits the long B hadron lifetime combining the information coming from all the secondary vertex tracks. For each track, the probability that a given track comes from the primary vertex is computed and these probabilities are combined together to provide the jet probability defined as:

$$P_{jet} = \prod \cdot \sum_{j=0}^{N-1} \frac{(-\ln \prod)^j}{j!}, \tag{5.6}$$

where

$$\prod = \prod_{i=1}^{N} P_{tr}(i) \tag{5.7}$$

with N the number of tracks, and P_{tr} the estimated probability for track i to come from the primary vertex. This tagger uses as calibration the negative impact parameter significance distributions, which are extracted from data, to calculate the probability of a track to come from the primary vertex. Fig. 5.12 shows the distribution of the JP requiring events with a Z boson in association with at least one b tagged jet.



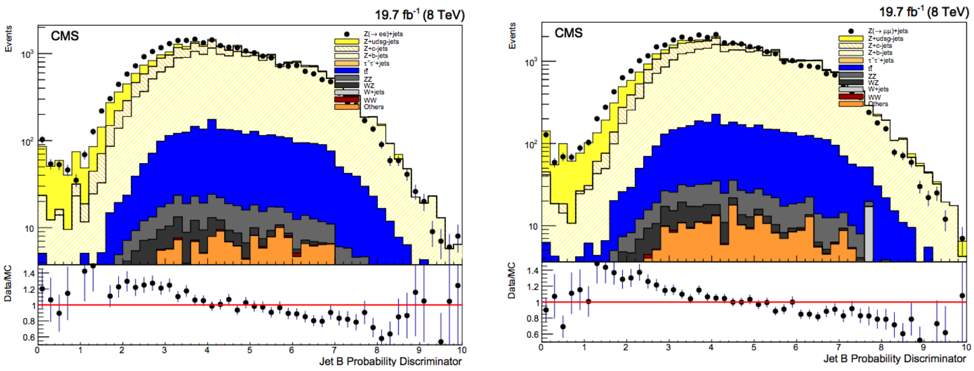
(a) Jet Probability distribution after the Z+b selection with a Z decaying into electrons.

(b) Jet Probability distribution after the Z+b selection with a Z decaying into muons.

Figure 5.12: Jet Probability distributions in the events with a Z boson with at least one b jet.

- The B Jet Probability (*BJP*) is another discriminator defined as JP with the difference that it uses only the four most displaced tracks. It estimates how likely it is that the four most displaced tracks are compatible with the primary vertex. This choice is supported by the average charged track multiplicity in weak b hadron decay that is approximately equal to 5. In fig. 5.13 the distributions of this discriminator are presented for the sample selected with a Z boson and at least one b jet.

For both the Z+1b jet and Z+2b jets the fraction of real b jets is estimated performing a fit to the Secondary Vertex Mass. The estimate of the b purity has to be done simultaneously in the Z+1b jet and Z+2b jets samples in order to take into account possible inefficiencies in the tagging of jets and so migrations of events from one sample to the other.



(a) B Jet Probability distribution after the Z+b selection with a Z decaying into electrons.

(b) B Jet Probability distribution after the Z+b selection with a Z decaying into muons.

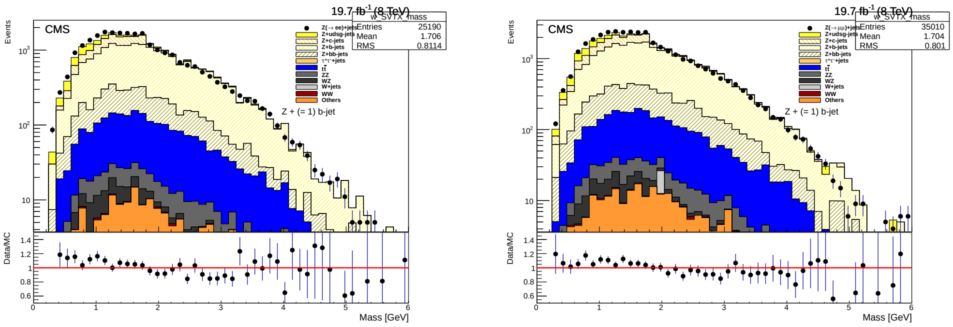
Figure 5.13: B Jet Probability distributions in the events with a Z boson with at least one b jet.

5.5.1 Z+1b jets purity

Fig. 5.14 shows the Secondary Vertex Mass distribution in the Z+1b jet sample that has been obtained requiring the presence of a Z boson and only one b tagged jet ($CSV > 0.895$) in the final state. Because of inefficiencies of the b tagging algorithm, the Z+1b sample selected may contain events with c or light jets wrongly identified as b jets and also events with two or more b jets in which only one is tagged as b jet. The Monte Carlo information has been exploited to estimate the proper fraction of these events. Templates derived from the Drell-Yan Z + jets (based on the MadGraph 5 interfaced with Pythia 6 in the 5-flavour scheme approximation) are used to describe the different flavor contributions:

- Z+c jet
- Z+ light jet
- Z+$\geq 2b$

The templates are built matching the flavor of the reconstructed jet to the parton flavor. In fig. 5.14, the different background contributions are shown, including the Z+c (region with orange stripes) and Z+light quark jets (yellow region) components. Moreover, also the contribution of Z+$\geq 2b$ events (region with green stripes) with two b jets in which only one has been tagged ($CSV > 0.895$) while the other has been mistagged are shown. The template reproduce the shape of all the components but we need the scale factors to normalize them. In order to extract the scale factors c_b, c_c, c_{udsg} for the Z+1b, Z+c and Z+light respectively, an unconstrained fit of the templates

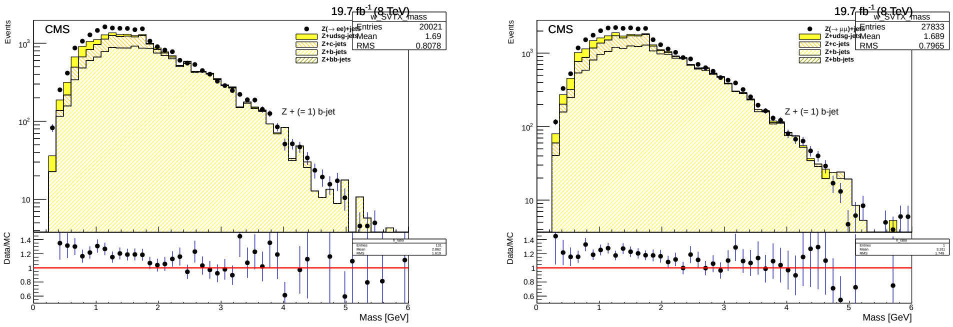


(a) Secondary Vertex Mass distribution after the Z+1b jet selection with a Z decaying into an electron pair.

(b) Secondary Vertex Mass distribution after the Z+1b jet selection with a Z decaying into a muon pair.

Figure 5.14: Secondary Vertex Mass distributions in the events with a Z boson with exactly 1 b tagged jet.

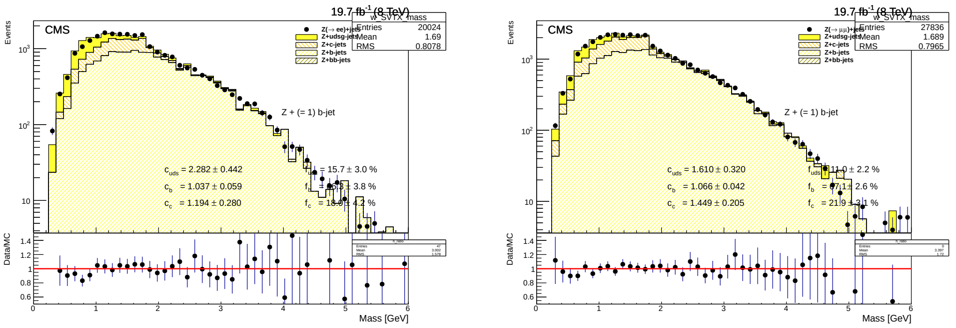
to the measured distribution is performed, after the subtraction of all the backgrounds: diboson, $t\bar{t}$, single top and the $Z+\geq 2b$ jets. The fit performed applies the minimum chi-square method. The $Z+\geq 2b$ is subtracted after being properly rescaled (more details are provided in sec. 5.5.3). After the backgrounds have been subtracted (fig. 5.15), the templates with the different flavor components are fitted to data (fig. 5.16) in order to obtain the real fraction of the b jet in the selected samples and also the c_b, c_c, and c_{udsg} (see tab. 5.10) factors that are used to rescale the Monte Carlo templates.



(a) Secondary Vertex Mass distribution after the Z+1b jet selection with a Z decaying into an electron pair after the background subtraction.

(b) Secondary Vertex Mass distribution after the Z+1b jet selection with a Z decaying into a muon pair after the background subtraction.

Figure 5.15: Secondary Vertex Mass distributions in the events with a Z boson with exactly 1 b tagged jet after the background subtraction.



(a) Secondary Vertex Mass distribution after the Z+1b jet selection with a Z decaying into an electron pair.

(b) Secondary Vertex Mass distribution after the Z+1b jet selection with a Z decaying into a muon pair.

Figure 5.16: Secondary Vertex Mass distributions in the events with a Z boson with exactly 1 b tagged jet after the background subtraction. The fraction of events containing b quarks is shown with yellow stripes while c fraction is represented by orange stripes. Results of the fit to the different templates are shown in the plot as f_b, f_c and f_{guds} for the different flavour percentage inside the sample and with the scaling factors c_b, c_c and c_{guds}.

Distribution	c_{guds}	c_b	c_c
Secondary Vertex Mass (ee)	2.28 ± 0.44	1.03 ± 0.05	1.19 ± 0.28
Secondary Vertex Mass ($\mu\mu$)	1.61 ± 0.32	1.07 ± 0.04	1.45 ± 0.21

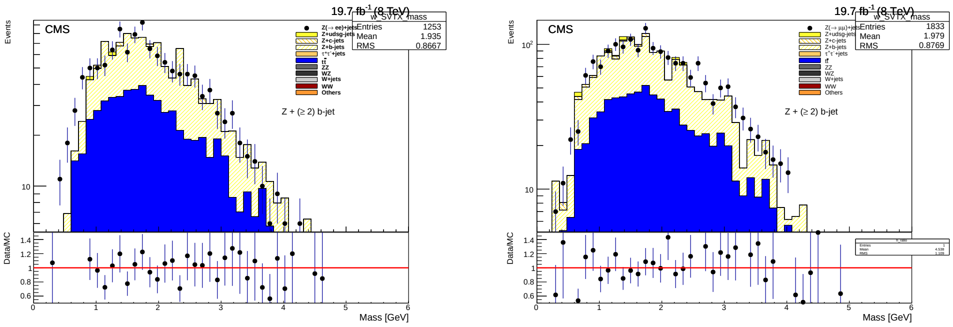
Table 5.10: Coefficients c_{guds}, c_c and c_b extracted form the fit of the Secondary Vertex Mass distributions in the electron and muon final state.

Distribution	f_{guds}	f_b	f_c
Secondary Vertex Mass (ee)	15.7 ± 3 %	66.3 ± 3.8 %	18.0 ± 4.2 %
Secondary Vertex Mass ($\mu\mu$)	11.0 ± 2.2 %	67.1 ± 2.6 %	21.9 ± 3.1 %

Table 5.11: fractions of the Z+light, Z+b and Z+c samples extracted form the fit of the Secondary Vertex Mass distributions in the electron and muon final state.

5.5.2 Z+2b jets purity

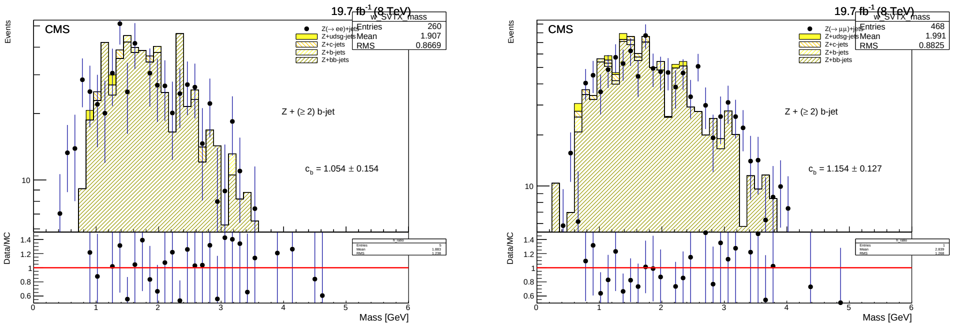
In order to distinguish the different jet flavour components and to estimate the coefficients to rescale the Monte Carlo, the SVTX Mass distribution is used (fig. 5.17) also in the case of the Z+2b sample. The main background contribution comes from $t\bar{t}$ (blue region) events with a small contamination due to c (region with orange stripes), light quarks or gluons (yellow region) jets misidentified as b jets. Subtracting the predicted background as described, an almost pure Z+2b jets sample is obtained as can be seen in fig. 5.18.



(a) Secondary Vertex Mass distribution after the Z+2b jets selection with a Z decaying into an electron pair.

(b) Secondary Vertex Mass distribution after the Z+2b selection with a Z decaying into a muon pair.

Figure 5.17: Secondary Vertex Mass distributions in the events with a Z boson with at least two b tagged jets.



(a) Secondary Vertex Mass discriminator distribution for the Z(ee)+2b final state. After the background subtraction the sample is almost pure (region with green stripes).

(b) Secondary Vertex Mass distribution for the $Z(\mu\mu)$+2b final state. After the background subtraction the sample is almost pure (region with green stripes).

Figure 5.18: Secondary Vertex Mass distribution for the Z+2b final state after the background subtraction.

5.5.3 Migrations between the two samples

The estimate of the event purity in the Z+1b and Z+2b samples need a particular attention and care due to possible migrations of events to one sample to the other because of jet flavor misidentification. In the Z+2b sample, after the background subtraction, no relevant migrations from the Z+1b sample have been observed and an almost pure sample is obtained with the selection applied. The sample with exactly one b tagged jet ($CSV > 0.895$) in the final state produced in association with a Z boson is, instead, affected by a contamination of those events with two b jets in which only one has been tagged

($CSV > 0.895$) while the other has been mistagged. In order to identify these events and subtract them, the Monte Carlo information is used by matching the flavor of the reconstructed jet to the parton flavor. We have checked the presence of further b jets in the whole jet collection in the Z+1b sample. These events are then treated as a background and subtracted after being rescaled for a factor evaluated by fitting the Secondary Vertex Mass in the Z+2b jets case. The Secondary Vertex Mass distribution (fig. 5.18), in the Z+2b case, after background subtraction, is fitted to data in order to extract the scale factor c_b.

5.6 Detector level distributions

In this section distributions of several observables are presented both for the Z+1b sample and for the Z+2b sample. For each of them the experimental data with the statistical error and also the contributions of different backgrounds are shown. The contributions coming from Z+jets events where charm and light/gluon jets are misidentified as b jets are corrected for the scale factors extracted from the Secondary Vertex Mass fitting procedure (explained in sec. 5.5). The Monte Carlo signal used for this comparison is the Drell-Yan Z+jets where the hard scattering process, with the emission of up to four additional partons, is calculated at Leading Order by MadGraph 5 interfaced with PYTHIA 6 for the hadronization. For the calculation of the matrix element the 5-flavor scheme is used. To estimate the agreement between data and Monte Carlo simulation, a ratio plot between them is present in the bottom part of each distribution. The distributions here presented are called "detector level" distributions as no unfolding technique, to deconvolve the detector effects, is applied. The unfolding procedure will be discussed in the next section and the "particle level" distributions will be presented. To characterize both the Z+1b and Z+2b final states, the following distributions have been exploited:

- p_T^{jet}: transverse momentum of the leading b tagged jet (fig. 5.21)
- η^{jet}: pseudorapidity variable of the leading b tagged jet (fig. 5.22)
- p_T^Z: the transverse momentum of the Z boson (fig. 5.28)
- H_T: defined as the sum of the transverse momenta of the jets of any flavor when only one b tagged jet is present in the event (Z+1b case) or when at least two b jets have been tagged in the final state of the event (Z+2b case) (fig. 6.34)
- $\Delta\phi_{Zb}$: the angular separation between the directions of the Z boson and the leading b jet in the transverse plane (fig. 5.25)

5.6.1 Z+2b angular distributions

For the Z+2b sample other observables are also considered: angular variables useful to reach a better understanding of the production mechanism of Z+2b and invariant mass distributions that are particularly interesting in the context of searches. The angular correlations between the b tagged jets and between the b jets and the Z boson are exploited to study the Z+2b final state and are described by the following variables:

- ΔR_{bb}: the angular separation between the directions of the two b tagged jets in (η, ϕ) plane (fig. 5.31). It is defined as $\Delta R_{bb} = \sqrt{(\Delta\phi_{bb})^2 + (\Delta\eta_{bb})^2}$, where $\Delta\phi_{bb}$ and $\Delta\eta_{bb}$ are the azimuthal and pseudorapidity separations. This variable is interesting because it gives the possibility to test the modelling of the different production mechanisms of $Zb\bar{b}$ final state. This quantity allows the identification of the contributions $qq \to Zb\bar{b}$ and $qg \to Zb\bar{b}X$ where the the $b\bar{b}$ pair is produced via gluon splitting (ΔR_{bb} <1) and the production modes $gg \to Zb\bar{b}$ and $qq \to Zb\bar{b}$ with the emission of a Z boson from one of the final state b quark (ΔR_{bb} >3).

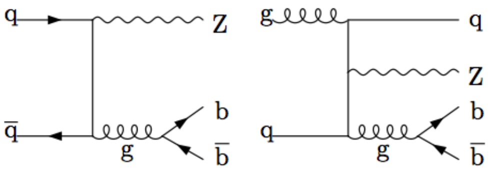


Figure 5.19: Leading order Feynman diagrams for the $Zb\bar{b}$ final state involving $g \to b\bar{b}$.

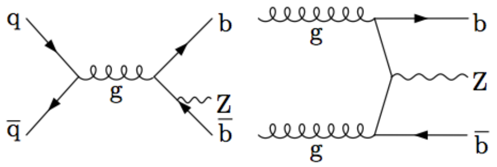


Figure 5.20: Leading order Feynman diagrams for the $Zb\bar{b}$ final state produced by the processes $q\bar{q} \to Zb\bar{b}$ with the emission of a Z boson from a b quark and $gg \to Zb\bar{b}$.

- $\Delta\phi_{bb}$: the angular separation between the directions of the two b jets in the transverse plane (fig. 5.32). It is an important observable to study the back-to-back configuration of the b quarks.
- ΔR_{Zb}^{min}: the angular separation between the Z boson and the closest b tagged jet in the (η, ϕ) plane (fig. 5.33). It allows to identify events with

the Z in the vicinity of one of the two b jets, so it is useful for testing NLO corrections in which a Z is radiated from a quark.

- ΔR_{Zb}^{max}: the angular separation between the Z boson and the further b tagged jet in the (η, ϕ) plane (fig. 5.34)
- A_{Zbb}: the asymmetry between the b jet direction and the Z direction using a combination of the two variables above defined ΔR_{Zb}^{min} and ΔR_{Zb}^{max} (fig. 5.35):

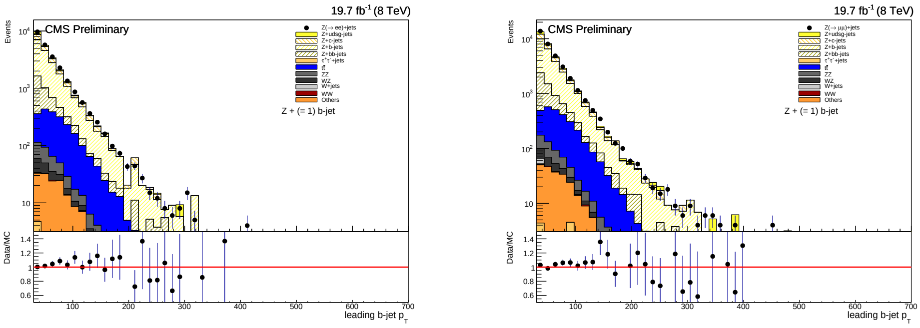
  $$A_{Zbb} = \frac{(\Delta R_{Zb}^{max} - \Delta R_{Zb}^{min})}{(\Delta R_{Zb}^{max} + \Delta R_{Zb}^{min})}. \tag{5.8}$$

  The A_{Zbb} asymmetry can provide an indirect test of quantum chromodynamics (QCD) validity at higher orders of the perturbative series. A nonzero value of A_{Zbb} is, in fact, related to the emission of additional gluon radiation in the final state while values of A_{Zbb} close to zero, in fact, identify configurations in which the two b jets are emitted symmetrically with respect the Z direction.

5.6.2 New physics related distributions

Two invariant mass distributions, particular interesting in the context of the search of new physics, are also considered for the Z+2b sample:

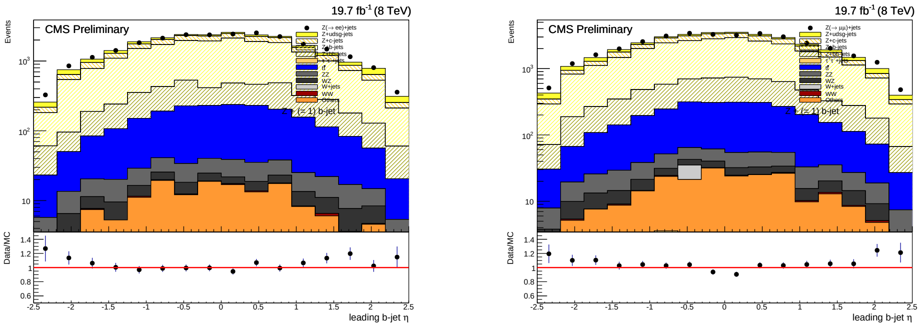
- M_{bb}: the invariant distribution of the two most energetic b tagged jets (fig. 5.36),
- M_{Zbb}: the invariant mass distribution of the system composed by the Z boson and the two most energetic b tagged jets(fig. 5.37).



(a) The transverse momentum of the leading jet in the Z+1b jet sample in the dielectron final state.

(b) The transverse momentum of the leading jet in the Z+1b jet sample in the dimuon final state.

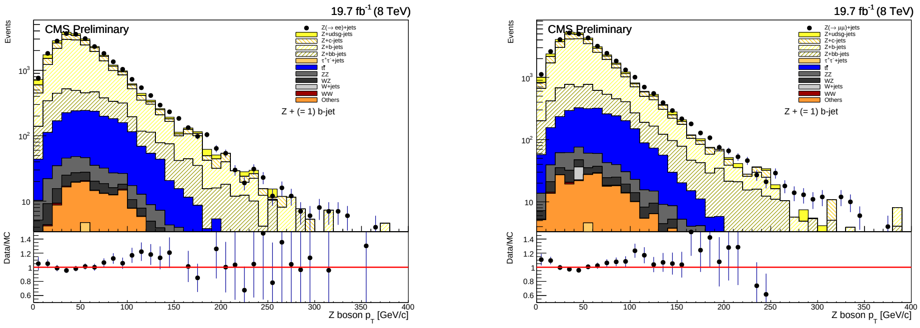
Figure 5.21: The transverse momentum of the leading jet in the Z+1b jet sample in the dielectron and dimuon final state.



(a) The psudorapidity variable of the leading jet in the Z+1b jet sample in the dielectron final state.

(b) The pseudorapidity variable of the leading jet in the Z+1b jet sample in the dimuon final state.

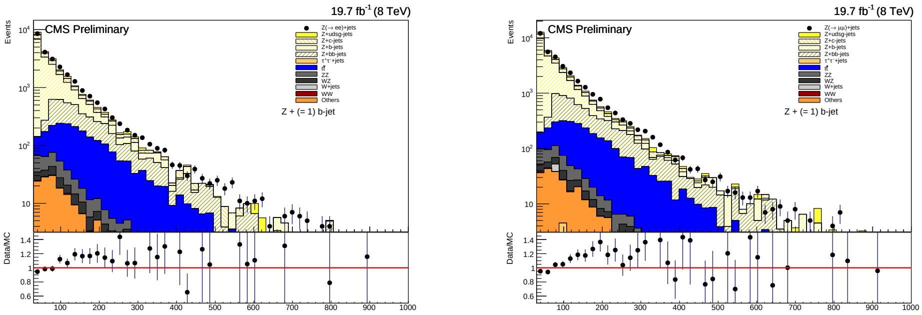
Figure 5.22: The pseudorapidity variable of the leading jet in the Z+1b jet sample in the dielectron and dimuon final state.



(a) The transverse momentum of the Z boson in the Z+1b jet sample in the dielectron final state.

(b) The transverse momentum of the Z boson in the Z+1b jet sample in the dimuon final state.

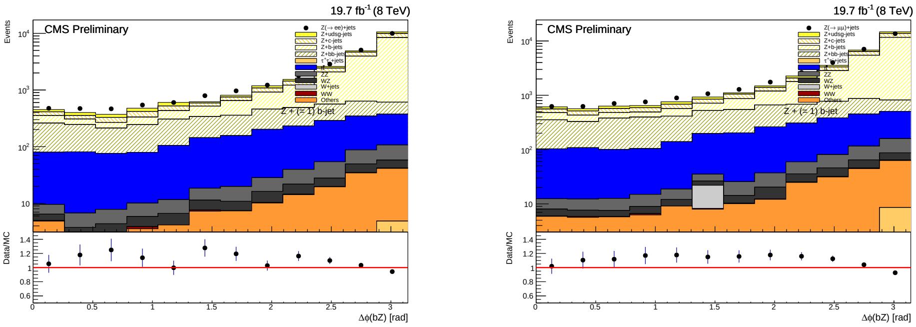
Figure 5.23: The transverse momentum of the Z boson in the Z+1b jet sample in the dielectron and dimuon final state.



(a) The H_T variable in the Z+1b jet sample in the dielectron final state.

(b) The H_T variable in the Z+1b jet sample in the dimuon final state.

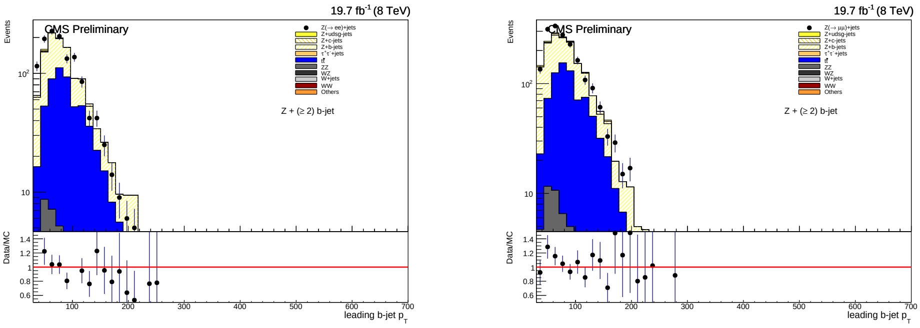
Figure 5.24: The H_T variable in the Z+1b jet sample in the dielectron and dimuon final state.



(a) The $\Delta\phi_{Zb}$ variable in the Z+1b jet sample in the dielectron final state.

(b) The $\Delta\phi_{Zb}$ variable in the Z+1b jet sample in the dimuon final state.

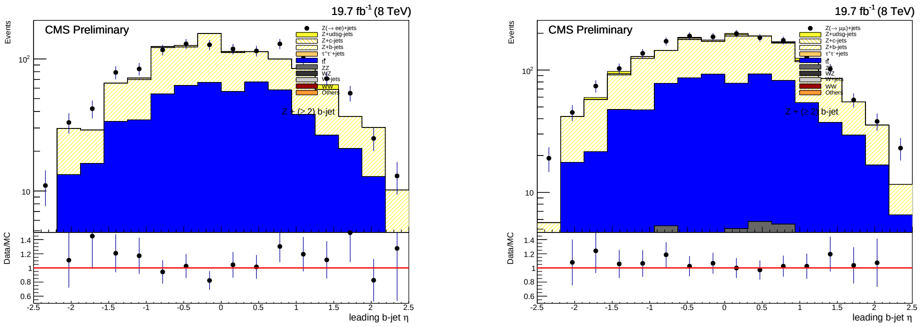
Figure 5.25: The $\Delta\phi_{Zb}$ variable in the Z+1b jet sample in the dielectron and dimuon final state.



(a) The transverse momentum of the leading jet in the Z+2b jet sample in the dielectron final state.

(b) The transverse momentum of the leading jet in the Z+2b jet sample in the dimuon final state.

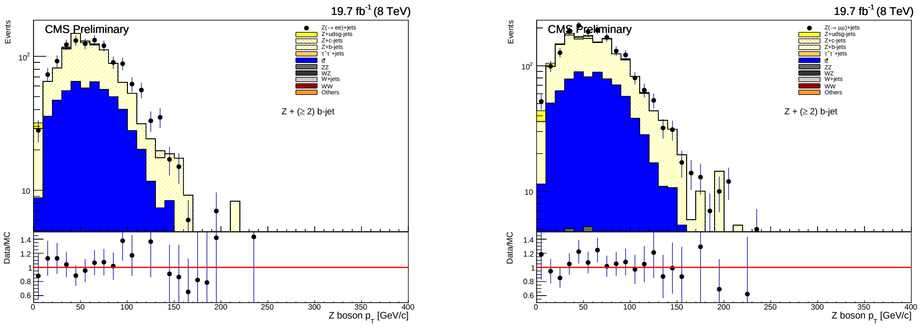
Figure 5.26: The transverse momentum of the leading jet in the Z+2b jet sample in the dielectron and dimuon final state.



(a) The pseudorapidity variable of the leading jet in the Z+2b jet sample in the dielectron final state.

(b) The pseudorapidity variable of the leading jet in the Z+2b jet sample in the dimuon final state.

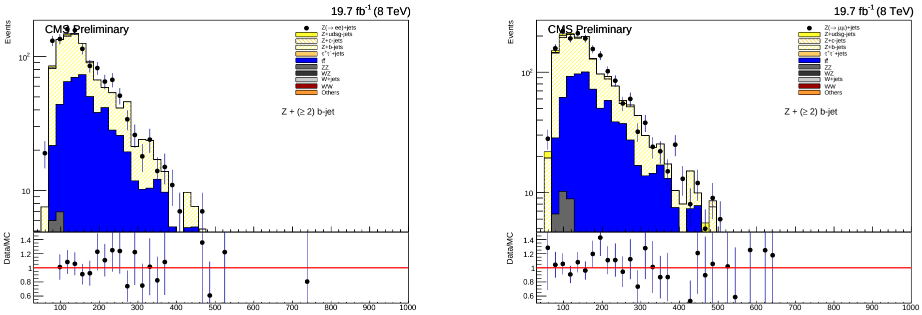
Figure 5.27: The psudorapidity variable of the leading jet in the Z+2b jet sample in the dielectron and dimuon final state.



(a) The transverse momentum of the Z boson in the Z+2b jet sample in the dielectron final state.

(b) The transverse momentum of the Z boson in the Z+2b jet sample in the dimuon final state.

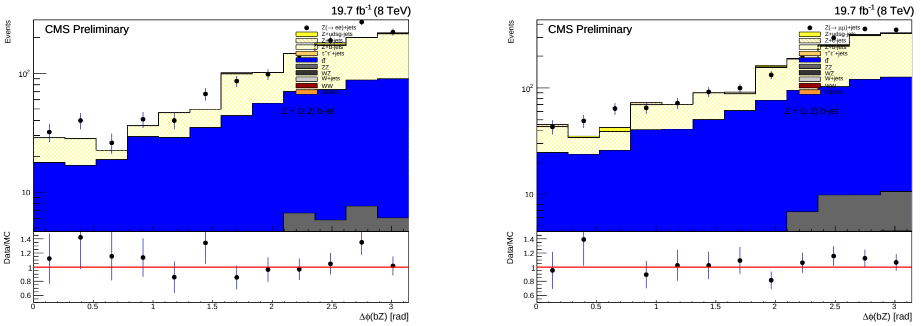
Figure 5.28: The transverse momentum of the Z boson in the Z+2b jet sample in the dielectron and dimuon final state.



(a) The H_T variable in the Z+2b jet sample in the dielectron final state.

(b) The H_T variable in the Z+2b jet sample in the dimuon final state.

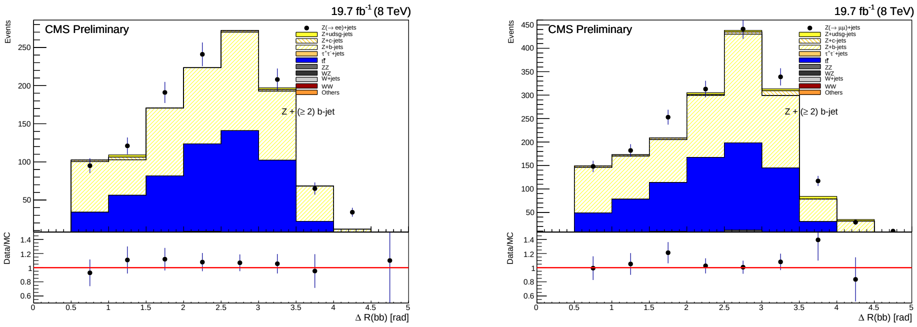
Figure 5.29: The H_T variable in the Z+2b jet sample in the dielectron and dimuon final state.



(a) The $\Delta\phi_{Zb}$ variable in the Z+2b jet sample in the dielectron final state.

(b) The $\Delta\phi_{Zb}$ variable in the Z+2b jet sample in the dimuon final state.

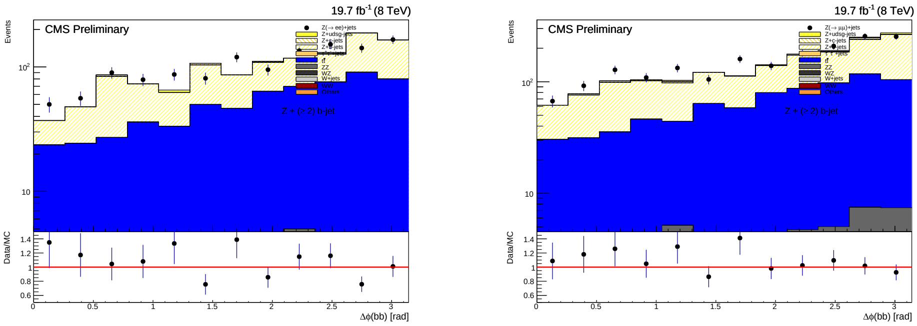
Figure 5.30: The $\Delta\phi_{Zb}$ variable in the Z+2b jet sample in the dielectron and dimuon final state.



(a) Angular separation between the directions of the two b tagged jets in (η, ϕ) plane in the Z+2b jets sample with the Z boson selected in the electron decay mode.

(b) Angular separation between the directions of the two b tagged jets in (η, ϕ) plane in the Z+2b jets sample with the Z boson selected in the muon decay mode.

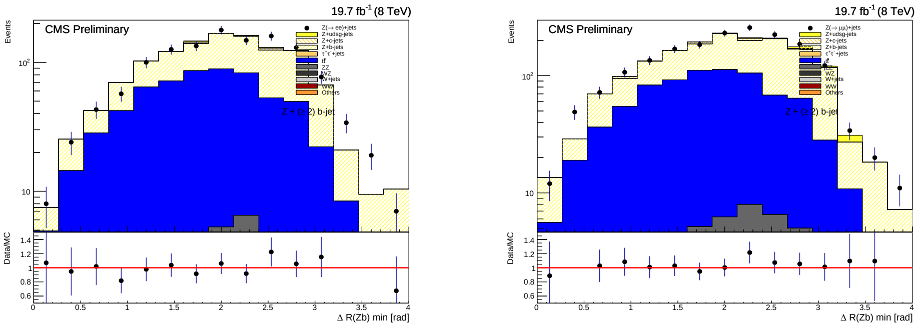
Figure 5.31: Angular separation between the directions of the two b tagged jets in (η, ϕ) plane in the Z+2b jets sample.



(a) The angular separation between the directions of the two b jets in the transverse plane in the Z+2b jets sample with the Z boson selected in the electron decay mode.

(b) The angular separation between the directions of the two b jets in the transverse plane in the Z+2b jets sample with the Z boson selected in the muon decay mode.

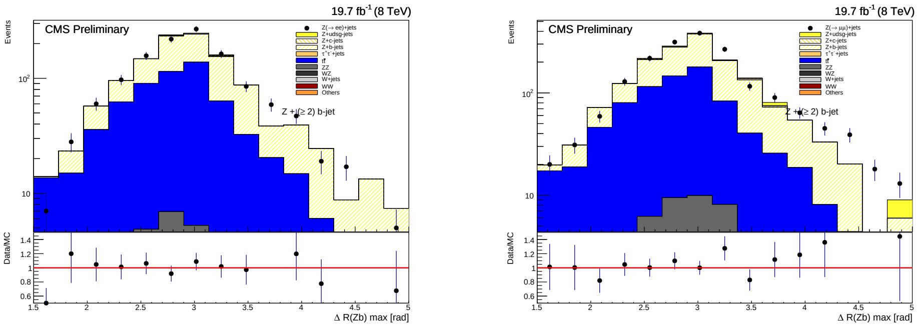
Figure 5.32: The angular separation between the directions of the two b jets in the transverse plane in the Z+2b jets sample.



(a) Angular separation between the Z boson and the closest b tagged jet in the (η, ϕ) plane in the Z+2b jets sample with the Z boson selected in the electron decay mode.

(b) Angular separation between the Z boson and the closest b tagged jet in the (η, ϕ) plane in the Z+2b jets sample with the Z boson selected in the muon decay mode.

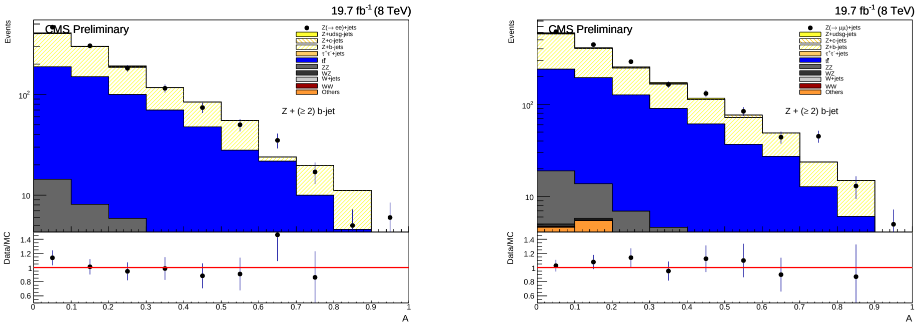
Figure 5.33: Angular separation between the Z boson and the closest b tagged jet in the (η, ϕ) plane in the Z+2b jets sample.



(a) Angular separation between the Z boson and the further b tagged jet in the (η, ϕ) plane in the Z+2b jets sample with the Z boson selected in the electron decay mode.

(b) Angular separation between the Z boson and the further b tagged jet in the (η, ϕ) plane in the Z+2b jets sample with the Z boson selected in the muon decay mode.

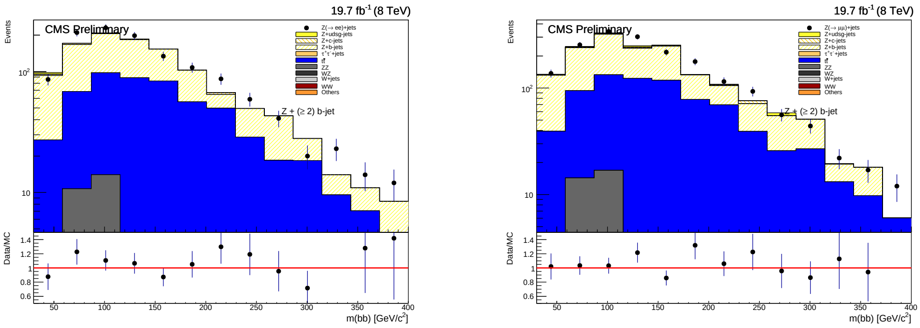
Figure 5.34: Angular separation between the Z boson and the further b tagged jet in the (η, ϕ) plane in the Z+2b jets sample.



(a) Distribution of the asymmetry between the b jet direction and the Z direction in the Z+2b jets sample with the Z boson selected in the electron decay mode.

(b) Distribution of the asymmetry between the b jet direction and the Z direction in the Z+2b jets sample with the Z boson selected in the muon decay mode.

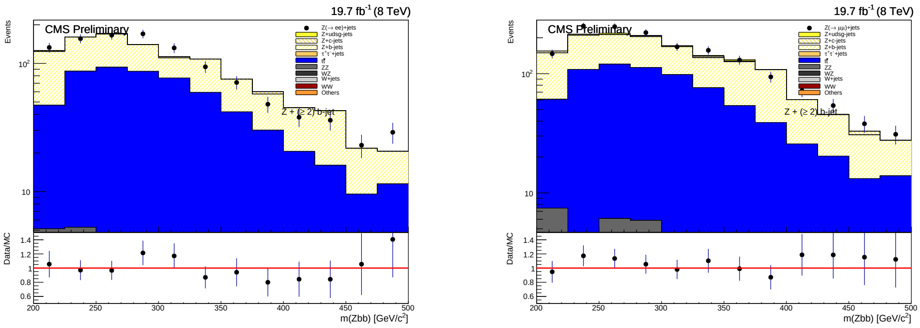
Figure 5.35: Distributions of the asymmetry between the b jet direction and the Z direction in the Z+2b jets.



(a) Invariant mass distribution of the of the b tagged jets in the Z+b jets sample with dielectron final state.

(b) Invariant mass distribution of the of the b tagged jets in the Z+b jets sample with dimuon final state.

Figure 5.36: Invariant mass distribution of the of the b tagged jets in the Z+b jets sample.



(a) Invariant mass distribution of the of the system composed of the b tagged jets and the Z boson in the Z+b jets sample with dielectron final state.

(b) Invariant mass distribution of the of the system composed of the b tagged jets and the Z boson in the Z+b jets sample with dimuon final state.

Figure 5.37: Invariant mass distribution of the of the system composed of the b tagged jets and the Z boson in the Z+b jets sample.

Chapter 6

Cross section measurement

In the previous Chapter the details about the data samples used in the analysis and the requirements applied in order to select the events with a Z boson produced in association with exactly one b jet or at least two b jets have been presented. A description of the procedure adopted to estimate the different background contributions and in particular the one coming from $t\bar{t}$ has been given. A section has been dedicated to the explanation of the strategy used in order to quantify the b purity that is the real fraction of b jets in the selected events. Finally distributions of different observables have been presented. For each of them the experimental data with the statistical error and also the contributions of different backgrounds are shown. Their ratio is also given to estimate the agreement between the data and Monte Carlo simulation. The distributions shown are called "detector level" distributions as these results are not deconvolved from the all the possible effects induced by the finite resolution.
In this Chapter the unfolding technique adopted will be explained. This procedure is needed to correct the distributions for experimental resolution effects and to compare the experimental results to theoretical expectations. The different systematic sources that enter in this analysis will be discussed and the method used to quantify their effects will be described. Finally the differential cross sections as a function of the different observables deconvolved from the detector effects and corrected for the efficiency will be shown. The results are compared to theoretical predictions obtained using different approaches (massive vs massless b-quarks, 4/5-flavor approaches).

6.1 Unfolding procedure

Under ideal conditions, with a perfect detector one could have a "true" measurement of physics quantities. However the measured distribution of a physical observable is in general distorted respect to the true one. This is due to several effects such as the finite resolution, a non linear response and limited acceptance of the detector. For this reason it is often very difficult to have a direct comparison between the data obtained using different detectors to each other and to various theoretical expectations. An important step of the experimental method is therefore to infer the true distribution from the observed one, correcting the observed spectrum from distortions. One way to solve this problem is the so called *bin-to-bin* correction: an efficiency calculated as the ratio between the number of events in a certain bin of the reconstructed variable and the number of events in the same bin of the "true" variable. This efficiency is then used to estimate the number of true events from the number of events observed in that bin. A limit of this method is that it cannot take into account large migration of events from one bin to the others and correlations between adjacent bins.

Ideally one would like to have a function to describe the response of the detector, so that the measured distribution can be considered as a convolution of this function with the true one. Let the distribution of a measured observable be stored in a vector $\mathbf{y}$ of dimension n_y, where the i^{th} coordinate of the vector contains the number of entries of the i^{th} bin of the histogram. Be $\mathbf{x}$ the vector (with n_x dimension) of the corresponding true distribution. In an ideal condition with a detector that does not introduce distortion in the distributions, the two vectors $\mathbf{y}$ and $\mathbf{x}$ would be exactly the same. The real detector instead induces a distortion meaning that each event from the true distribution may find itself in a range of adjacent bins. This is expressed by the equation:

$$A\mathbf{x} = \mathbf{y}, \tag{6.1}$$

where A is a matrix $n_y \times n_x$ that describes the detector effects. The matrix element A_{ij} represents the probability to have a migration of events belonging to the j-th bin of $\mathbf{x}$ to the i-th bin of $\mathbf{y}$. Given a distribution of a measured variable to obtain the true distribution we need to invert the eq. 6.1 but the inversion of the matrix A is not a trivial procedure and in principle there is no reason why this matrix should exist. In this section two different methods (*unfolding* techniques) using different approaches will be discussed: the *Singular Value Decomposition* (SVD) [135] and the *Bayesian* unfolding [136]. The first is the one used as default in the analysis while the other is used as a cross check. The two methods have been implemented using the RooUnfold package [137].

6.1.1 Singular Value Decomposition method

Under certain conditions the system defined in eq. 6.1 may be reduced to a diagonal system that can be easily solved, otherwise the problem becomes ill-determined. In this second case conventional methods of solving linear systems do not work and the exact solution is not obtainable. The singular value decomposition (SVD) of a real $m \times n$ matrix A consists of its factorization as expressed in the formula:

$$A = USV^T, \tag{6.2}$$

with U an orthogonal matrix $m \times m$, V an orthogonal matrix $n \times n$, while S is a diagonal matrix $m \times n$ with non-negative diagonal elements. After this decomposition of the matrix A, it becomes easier to manipulate it. The system 6.1 represents the solution of the following least square problem:

$$\sum_{i=1}^{n_y}(\sum_{j=1}^{n_x} A_{ij}x_j - y_i)^2 = min. \tag{6.3}$$

If the measurement errors in the vector y do not vary from bin to bin the above equation is not adequate and the following expression has to be minimized:

$$\sum_{i=1}^{n_y}(\frac{\sum_{j=1}^{n_x} - y}{\Delta y_i})^2 = min, \tag{6.4}$$

where Δy_i is the error in y_i. So the general case of this expression is:

$$(Ax - y)^T B^{-1}(Ax - y) = min, \tag{6.5}$$

with B the covariance matrix of the measured vector y. Being B a covariance matrix, it should be symmetric and positive-definite, so its SVD can be written as:

$$B = QRQ^T, \qquad R_{ii} \equiv r_i^2 > 0, \qquad B^{-1} = Q^T R^{-1} Q. \tag{6.6}$$

Rescaling the matrix A and the vector $\mathbf{y}$ as:

$$\widetilde{A}_{ij} \equiv \frac{1}{r_i}\sum_m Q_{im}A_{mj} \qquad \widetilde{b} \equiv \frac{1}{r_i}\sum_m Q_{im}b_m. \tag{6.7}$$

Substituting in eq. 6.5 it is obtained the solution:

$$\sum_{j=1}^{n_x} \widetilde{A}_{ij}x_j = \widetilde{y}. \tag{6.8}$$

These transformations change the appearance of the system but the main problem to face still remains and it appears with small singular values. The exact solution of eq. 6.8 will again most certainly lead to a rapidly oscillating distribution, which may have a smaller amplitude but is still useless. The oscillatory component should be suppressed by adding the regularization or stabilization term to the expression to be minimized that now becomes:

$$(\widetilde{A}x - \widetilde{y})^T(\widetilde{A}x - \widetilde{y}) + \tau(Cx)^T Cx = min. \tag{6.9}$$

C is a matrix which defines the condition imposed on the solution, while the value of the regularization parameter τ determines the relative weight of this condition. While the optimal value of τ, an important part of the procedure, depends on the problem we deal with, the explicit form of the matrix C should be chosen from general considerations. One choice for the C matrix could be the one that suppresses the solutions x having large curvatures where the curvature is defined as $\sum_j [(x_{j+1} - x_j) - (x_j - x_{j-1})]^2$:

$$C = \begin{pmatrix} -1 & 1 & 0 & 0 & ... & & \\ 1 & -2 & 1 & 0 & ... & & \\ 0 & 1 & -2 & 1 & ... & & \\ & ... & & & ... & & \\ & ... & & 1 & -2 & 1 \\ & ... & & & 1 & -1 \end{pmatrix} \tag{6.10}$$

Minimization of eq. 6.9 leads to a new linear system, which has n_x additional equations:

$$\begin{pmatrix} \widetilde{A} \\ \sqrt{\tau} \cdot C \end{pmatrix} x = \begin{pmatrix} \widetilde{b} \\ 0 \end{pmatrix} \tag{6.11}$$

A method (called *damped least squares* [138]) which allows to express the solution of (6.11) for any τ through the solution of the initial non-regularized problem corresponding to $\tau = 0$ can be used. For $\tau = 0$ the system (6.11) is equivalent to (6.8), if the inverse C^{-1} exists and can be calculated. It is not possible to invert the C matrix defined above. For this reason a small ψ component (10^{-3}, 10^{-4}) is added to all the diagonal elements in order to obtain a symmetric non-singular matrix. So the system can be solved in the case of $\tau = 0$ applying the singular value decomposition expressed in eq. 6.2. Once this solution has been obtained using the *damped least squares* method, it is possible to calculate the solution for each value of the regularization parameter τ.

6.1.2 Bayes method

The Bayes' theorem could represent a way to unfold the experimental distributions obtaining the true ones. The procedure of Bayesian unfolding can be explained in terms of causes C_i ($i = 1, ..., n_C$) that correspond to the true value of a variable and in terms of effects E_j ($j = 1, ..., n_E$) referring to the variable values. Each cause can produce different effects, but for a given effect the exact cause is not known. We can assume to know the initial probability $P(C_i)$ and the conditional probability of the i-th cause to produce the effect $P(E|C_i)$. The goal is to estimate the probability $P(C_i|E)$ that different causes C_i were responsible for the observed effect E. The Bayes formula provides a solution for this problem:

$$P(C_i|E) = \frac{P(E|C_i) \cdot P(C_i)}{\sum_{k=1}^{n_C} P(E|C_k) \cdot P(C_k)}. \tag{6.12}$$

This formula has the following meaning: if we observe a single event (effect), the probability that it has been due to the i-th cause is proportional to the probability of the cause multiplied by the probability of the cause to produce the effect. The probability $P(C_i|E)$ depends on the knowledge of the probability $P(C)$. Its knowledge can increase with the increasing number of observations. The outcome of a measurement has possible effects E_j $(j = 1, ..., n_E)$ for a given cause C_i. For each E_j we can define the probability $P(C_i|E_j)$ using the Bayes' theorem 6.13. We can refer to the conditional probabilities $P(C_i|E_j)$ as a matrix that describes cell to cell migration. The probability $P(C_i|E_j)$ depends on $P(E_j|C_i)$ and this probability can be calculated from the matrix A previously introduced 6.1. If $n(E_j)$ events with effect E_j are observed, the expected number of events associated to each cause is:

$$n(C_i) = \sum_{j=1}^{n_E} n(E_j) \cdot P(C_i|E). \tag{6.13}$$

This formula can be rewritten in terms of the unfolding matrix M_{ij}:

$$n(C_i) = \sum_{j=1}^{n_E} M_{ij} \cdot n(E_j), \tag{6.14}$$

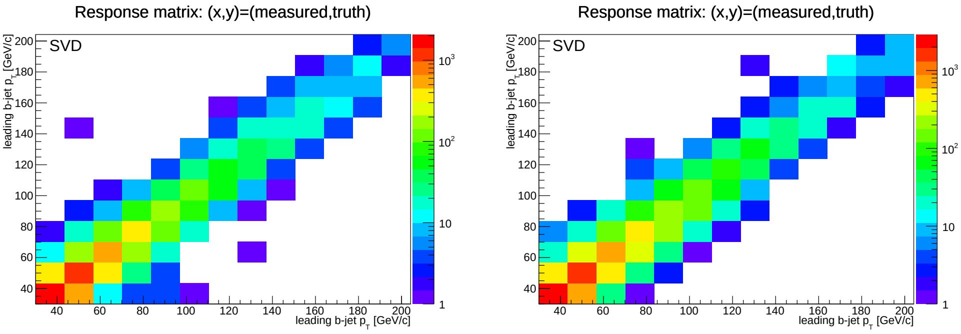
with M_{ij} expressed by:

$$M_{ij} = \frac{P(E_j|C_i) \cdot P(C_i)}{[\sum_{k=1}^{n_E} P(E_k|C_i)] \cdot [\sum_{l=1}^{n_C} P(E_j|C_l) \cdot P(C_l)]}. \tag{6.15}$$

This method is also dependent on a regularization parameter that represents the number of the iterations.

6.1.3 Response matrix calculation

As already explained, because of the limited acceptance and the finite resolution of the detector, migration, distortions and transformations can affect the distributions of the measured variables. To infer the true distributions starting from the measured ones a response matrix, describing the migrations among the bins, has to be built. This is a crucial step both in the SVD and in the Bayesian approach. The response matrix maps both the information at reconstructed level and the information at generator (*truth*) level requiring that reconstructed level objects are associated to the corresponding particles at event generator level. The generator level (*truth*) and reconstructed (*reco*) Monte Carlo spectra are therefore needed. For this purpose the Monte Carlo distributions, simulated by MadGraph 5-flavor scheme with PYTHIA 6 as a default, are used. The truth Monte Carlo spectra are obtained with a generator level analysis that should mimic as closely as possible the analysis done on the real data and reconstructed MC (described in the previous chapter). In this case, according to the reconstructed selection, at generator level we require events with a Z boson that decays into electrons or muons and exactly one b jet or at least two b jets applying the same selection cut thresholds. Further details of the generator level analysis will be given later in the dedicated section 6.4 . In the case of the response matrix of the transverse momentum and eta variable of the b jet, the matching between the reconstructed jet and the generator jet is done requiring a minimum angular separation between the two b jets in the (η, ϕ) plane ($\Delta R < 0.5$). Some examples of response matrices for the leading b jet momentum in the Z+1b and Z+2b are shown in fig. 6.1 -6.2. In Appendix 6.6 all the other matrices are available.



(a) Response matrix for the leading b jet momentum in the electron final state with the Z+1b selection.

(b) Response matrix for the leading b jet momentum in the muon final state with the Z+1b selection.

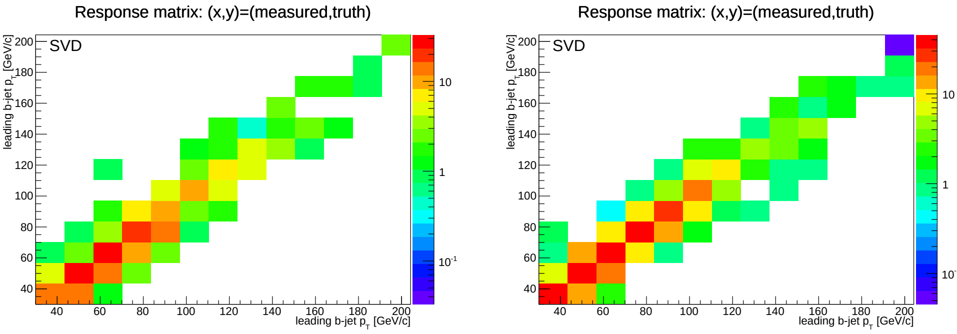
Figure 6.1: Response matrix for the leading b jet momentum with the Z+1b selection.

6.1.4 Validation of unfolding method

Once the response matrix is obtained for each measured observable, it is applied to the data distribution using the SVD algorithm as default. Two validation methods have been used as a check of the whole unfolding procedure: the *identity* test and the *closure* test.

Identity test

The identity test consists in the application of the response matrix to the same reconstructed Monte Carlo distribution used to build the matrix itself. This check allows to understand if some biases have been introduced in the matrix calculation. The resulting unfolded distribution is expected to coincide exactly with the Monte Carlo truth distribution. In fig. 6.3 an example of the identity test performed on the pseudorapidity variable of the b jet distribution

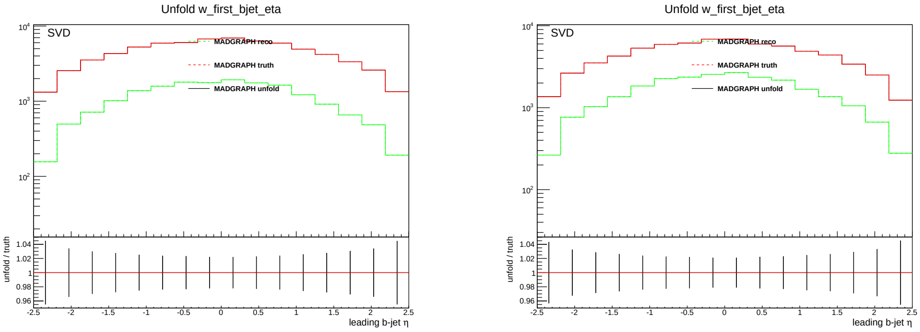


(a) Response matrix for the leading b jet momentum in the electron final state with the Z+2b selection.

(b) Response matrix for the leading b jet momentum in the muon final state with the Z+2b selection.

Figure 6.2: Response matrix for the leading b jet momentum with the Z+2b selection.

is provided. The plot shows the distribution of the pseudorapidity of the reconstructed b jet (red line) and of the b jet at generator level (green line) that coincides with the unfolded distribution as can be checked looking at the ratio between the unfolding output and the generator level Monte Carlo that is exactly one.



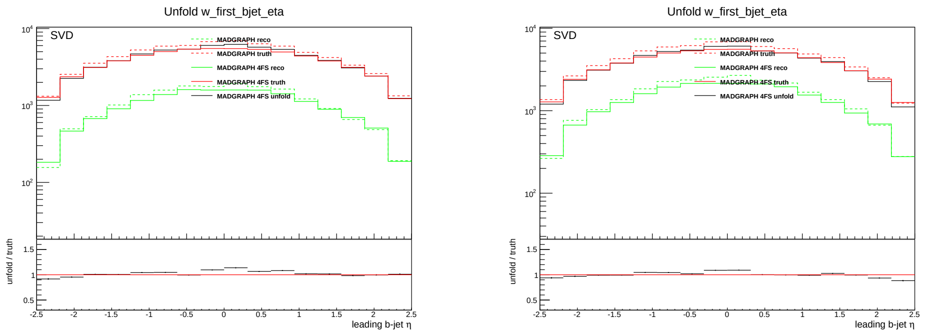
(a) The identity test performed on the pseudorapidity variable of the b jet distribution in the Z+1b case and electron final state.

(b) The identity test performed on the pseudorapidity variable of the b jet distribution in the Z+1b case and muon final state.

Figure 6.3: The identity test performed on the pseudorapidity variable of the b jet distribution in the Z+1b case.

Closure test

The identity test has been performed for all the distributions and the results are consistent with our expectations in all the cases. Another test to demonstrate that the unfolding algorithm does not introduce additional biases in the distributions is the closure test. It is performed applying the response matrix to the reconstructed distribution obtained with another Monte Carlo with respect to the one used to derive the response matrix. As example, in fig. 6.4 the closure test performed on the pseudorapidity of the b jet distribution of the b tagged jet is shown. In that case the MadGraph 5-flavor scheme is used to calculate the matrix that is then applied to the distribution obtained with Madgraph 4-flavor scheme. A good agreement between unfolded and truth distribution is obtained showing that there is not an explicit dependence on the type of Monte Carlo used in the unfolding procedure.



(a) The closure test performed on the pseudorapidity variable of the b jet distribution in the Z+1b case and electron final state.

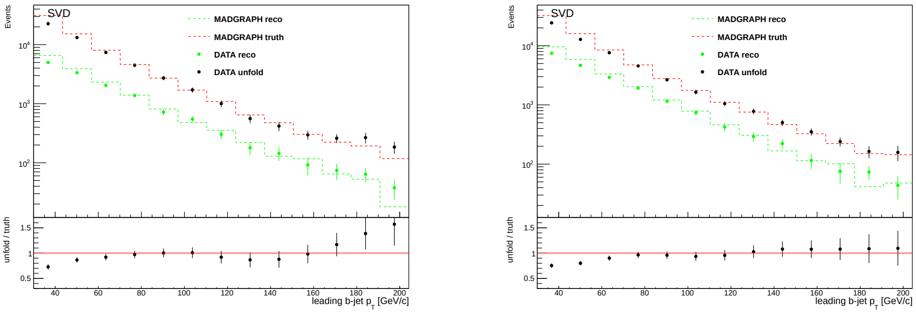
(b) The closure test performed on the pseudorapidity variable of the b jet distribution in the Z+1b case and muon final state.

Figure 6.4: The closure test performed on the pseudorapidity variable of the b jet distribution in the Z+1b case.

6.1.5 Unfolded distributions

In this analysis the unfolding has been implemented using the RooUnfold package [137] that provides a framework for different unfolding algorithms. The SVD method is the one used as default while the Bayesian method is performed only as a cross check. In both the cases the choice of a value for the regularization parameter (k) is an important step in the unfolding procedure. For the Bayesian approach, the regularization is achieved by stopping iterations before reaching the “true” inverse. The regularization parameter is just the number of iterations. In principle, this has to be tuned according to the sample statistics and binning. In practice, the results are fairly insensitive to the precise setting used and four iterations are usually sufficient. In the case of the SVD method the value of the regularization parameter can be chosen in the range $[2, N_{bins}]$. The regularization needs to be tuned according to the distribution, binning, and sample statistics in order minimize the bias due to the choice of the Monte Carlo truth sample (which dominates at small

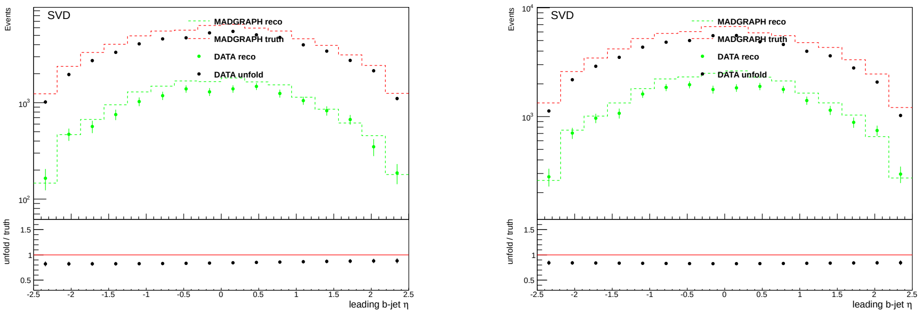
k) and avoiding results too dominated by statistical fluctuations (which grow at large k). In order to choose a value for the regularization parameter in the SVD procedure, the closure test has been studied for each distribution. The value of the regularization parameter that allows to obtain the most reliable closure test in each distribution (both in the Z+1b and Z+2b cases) is chosen. The ratio between the unfolded results and generator level distributions in the closure test, using the chosen parameter, is indeed the closest to the unity and the oscillating contributions are reduced. The results of the unfolding procedure applying the SVD method are presented in fig. 6.5 - 6.14.



(a) Unfolded distribution of the transverse momentum of the b tagged jet in the sample Z+1b jet where the Z boson is identified trough its deacay mode into electrons.

(b) Unfolded distribution of the transverse momentum of the b tagged jet in the sample Z+1b jet where the Z boson is identified trough its deacay mode into muons.

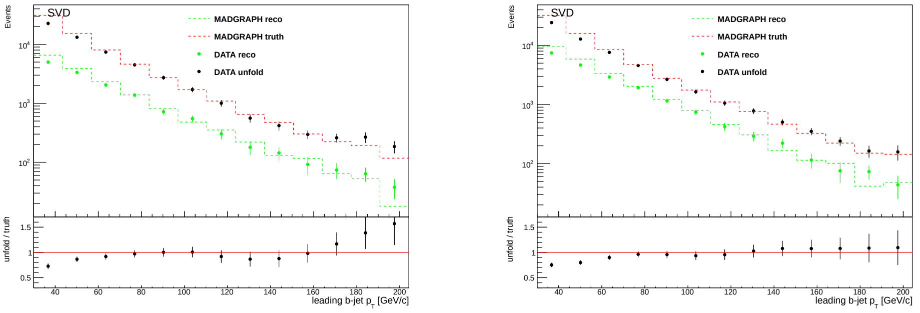
Figure 6.5: Unfolding results for the leading b jet momentum in electrons and muons final state.



(a) Unfolded distribution of the pseudorapidity of the b tagged jet in the sample Z+1b jet where the Z boson is identified trough its deacay mode into electrons.

(b) Unfolded distribution of the pseudorapidity of the b tagged jet in the sample Z+1b jet where the Z boson is identified trough its deacy mode into muons.

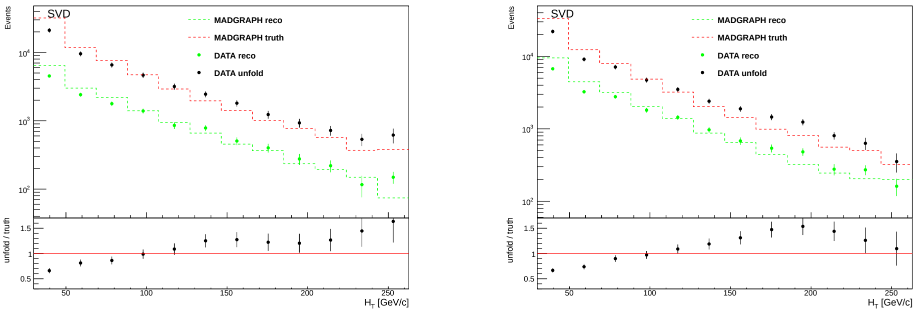
Figure 6.6: Unfolding results for the leading b jet pseudorapidity in electrons and muons final state.



(a) Unfolded distribution of the transverse momentum of the Z boson in the sample Z+1b jet where the Z is identified trough its deacay mode into electrons.

(b) Unfolded distribution of the transverse momentum of the Z boson in the sample Z+1b jet where the Z is identified trough its deacay mode into muons.

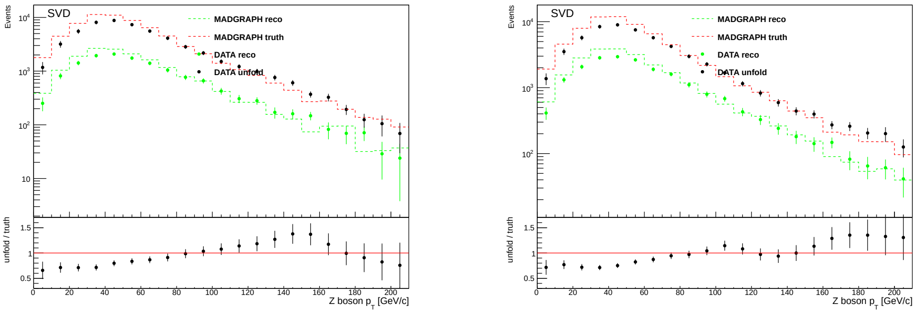
Figure 6.7: Unfolding results for the transverse momentum of the Z boson in electrons and muons final state.



(a) Unfolded distribution of the H_T variable in the sample Z+1b jet where the Z is identified trough its deacay mode into electrons.

(b) Unfolded distribution of the H_T variable in the sample Z+1b jet where the Z is identified trough its deacay mode into muons.

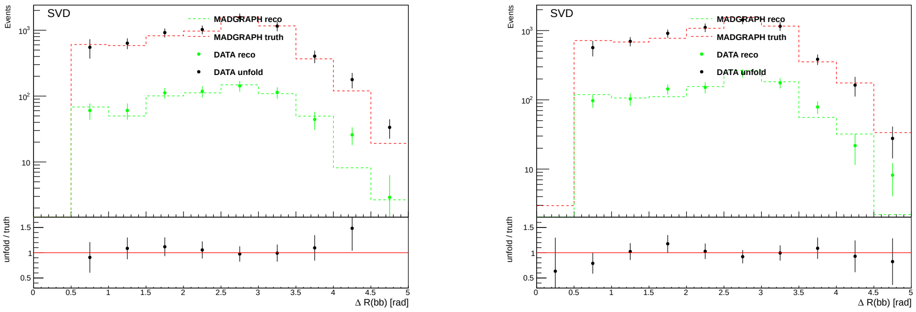
Figure 6.8: Unfolding results for the H_T variable in electrons and muons final state.



(a) Unfolded distribution of the $\Delta\phi$ variable between the Z boson and the leading b jet in the sample Z+1b jet where the Z is identified trough its deacay mode into electrons.

(b) Unfolded distribution of the $\Delta\phi$ variable between the Z boson and the leading b jet in the sample Z+1b jet where the Z is identified trough its deacay mode into muons.

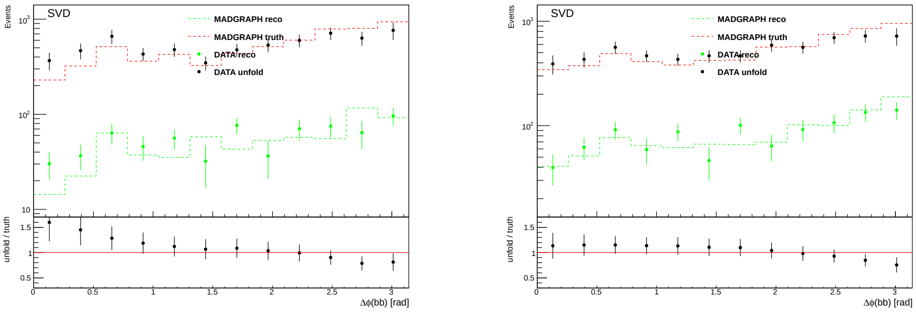
Figure 6.9: Unfolding results for the $\Delta\phi$ variable between the Z boson and the leading b jet in electrons and muons final state.



(a) Unfolded distribution of the ΔR between the two b tagged jets in the sample Z+2b where the Z is identified trough its deacay mode into electrons.

(b) Unfolded distribution of the ΔR between the two b tagged jets in the sample Z+2b where the Z is identified trough its deacay mode into muons.

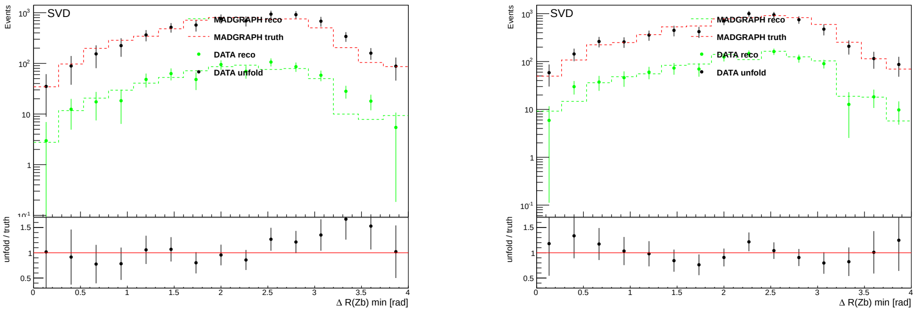
Figure 6.10: Unfolding results for the variable ΔR between the two b tagged jets in electrons and muons final state.



(a) Unfolded distribution of the $\Delta\phi$ between the two b tagged jets in the sample Z+2b where the Z is identified trough its deacay mode into electrons.

(b) Unfolded distribution of the $\Delta\phi$ between the two b tagged jets in the sample Z+2b where the Z is identified trough its deacay mode into muons.

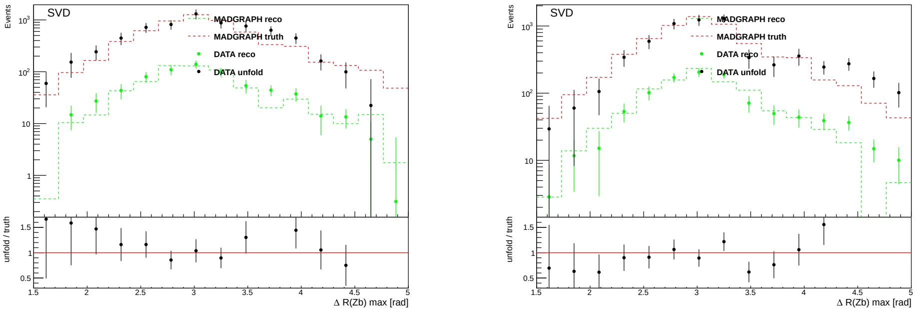
Figure 6.11: Unfolding results for the variable $\Delta\phi$ between the two b tagged jets in electrons and muons final state.



(a) Unfolded distribution of the ΔR between the Z boson and the closest b tagged jet in the sample Z+2b where the Z is identified trough its deacay mode into electrons.

(b) Unfolded distribution of the ΔR between the Z boson and the closest b tagged jet in the sample Z+2b jet where the Z is identified trough its deacay mode into muons.

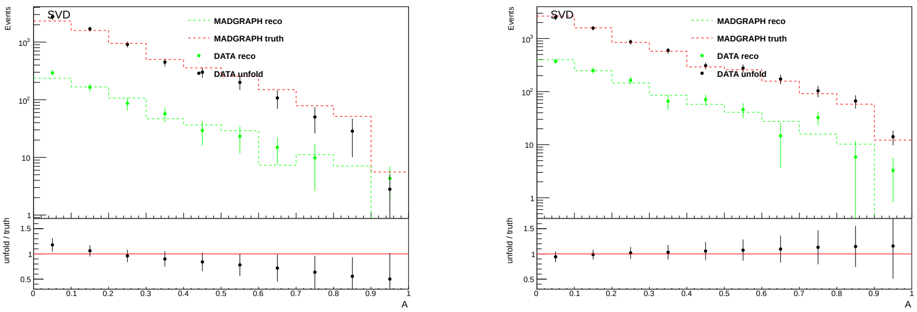
Figure 6.12: Unfolding results for the variable ΔR between the Z boson and the closest b tagged jet in electrons and muons final state.



(a) Unfolded distribution of the ΔR between the Z boson and the further b tagged jet in the sample Z+2b where the Z is identified trough its deacay mode into electrons.

(b) Unfolded distribution of the ΔR between the Z boson and the further b tagged jet in the sample Z+2b jet where the Z is identified trough its deacay mode into muons.

Figure 6.13: Unfolding results for the variable ΔR between the Z boson and the further b tagged jet in electrons and muons final state.



(a) Unfolded distribution of the asymmetry between the Z boson direction and the b tagged jet direction in the sample Z+2b where the Z is identified trough its deacay mode into electrons.

(b) Unfolded distribution of the asymmetry between the Z boson direction and the b tagged jet direction in the sample Z+2b jet where the Z is identified trough its deacay mode into muons.

Figure 6.14: Unfolding results for the of the asymmetry between the Z boson direction and the b tagged jet direction variable in electrons and muons final state.

6.2 Statistical and Systematic uncertainties

This section deals with the possible sources of uncertainties associated to this measurement. The main contributions to the total systematic uncertainty come from:

- jet energy correction and resolution
- pileup
- luminosity
- background estimate
- efficiency
- unfolding

The sources of statistical uncertainty related to the Z+b cross section measurement are:

- statistical uncertainty associated to the datasets used for data and Monte Carlo contributions
- statistical errors on the elements of the response matrix built in the unfolding procedure
- top-antitop background estimation method
- b purity extraction procedure

each of them will be described in the following subsections. The total contributions of the systematic uncertainties is then added in quadrature to the statistical error in the final cross section measurement.

Jet energy correction and resolution

The main uncertainty in jet reconstruction comes from the uncertainty on the jet energy corrections (JEC) and the jet energy resolution (JER).

- jet energy corrections (JEC).
  As explained in the Chapter 4, the jet energy measured in the detector is in general different from the corresponding particle jet energy. The main cause for this energy mismatch is the non uniform and non linear response of the CMS calorimeters to the jet showers. Moreover, electronic noise and pileup particles originating from the additional proton-proton interactions can introduce biases and smear the reconstructed jet energy. All these effects are taken into account applying a set of corrections to the raw jet four-momentum vector p^{raw} in order to obtain the corrected

jet four-momentum p^{corr}. To evaluate the systematic uncertainty induced by this correction, all the reconstructed jet transverse momenta are varied as:

$$p^{up} = p^{corr} \cdot (1 + \sigma_{JEC}), \tag{6.16}$$

$$p^{down} = p^{corr} \cdot (1 - \sigma_{JEC}), \tag{6.17}$$

where the σ_{JEC} is the addition in quadrature of all the uncertainties relative to the jet correction factors.

- jet energy resolution (JER).
  Experimental results show that jet energy resolution (JER) in data is worse than in the simulation therefore the jet energy in the Monte Carlo needs to be smeared to better describe the data. For this reason another set of corrections has to be applied to the reconstructed jet momentum in simulated events. In order to estimate the uncertainty due to this correction, two additional sets of correction factors are used and the impact on the measured distribution is evaluated. These new sets have been obtained by systematically shifting the measured factors down and up of one standard deviation.

The systematic effect both for the JEC and JER correction is estimated by comparing the distributions obtained using the standard corrections and their variation.

Pileup

The Monte Carlo reweighting procedure adopted in order to reproduce the expected number of the pileup interactions correctly, carries out some systematic uncertainties that must be propagated through the analysis [139]. The pileup model employed has several sources of systematic error:

- uncertainty in the number of interactions in data.
  It is estimated from the measured luminosity in each bunch crossing multiplied by an average total inelastic cross section. Two sources of error, that are also the dominant contributions, come up: the luminosity uncertainty, and the uncertainty in the total inelastic cross section. Combining the errors on the luminosity and cross section measurements, the estimated number of interactions has a total uncertainty of about 3.9% in 2012.

- differences between the PYTHIA model of pileup and what is observed in the data. There could be for example, differences in multiplicity, momentum and energy spectra, angular distribution. An uncertainty in order to take into account these effects is needed.

- errors associated with the beam size along the z axis. The bunch length changes by approximately 1 cm during the course of each fill. It increases as the fill progresses, which implies that, as the luminosity decreases, pileup interaction are easier to distinguish from the primary interaction. None of these effects are in the simulation.

A variation of ±5% in the number of interactions is then applied to cover all the uncertainties described due to pileup modeling.

Luminosity

As the integrated luminosity enters in the expression of the differential cross sections measured in this analysis, the luminosity uncertainty directly affects these measurement. The measurement of the LHC luminosity delivered to CMS in the 2012 proton-proton physics run has been measured using the cluster counting technique from the silicon pixel detector [140]. The overall uncertainty on the integrated luminosity is estimated to be 2.6% for 2012.

Background estimate

A systematic uncertainty is related to the the background subtraction procedure because of the uncertainties in the cross sections used to rescale the simulated background samples. Before the background subtraction from data, each sample is rescaled by a factor f_{norm}:

$$f_{norm} = \frac{L \cdot \sigma_{bkg}}{N_{gen}}, \tag{6.18}$$

where L is the integrated luminosity, σ_{bkg} is the theoretical cross section of a given background process and N_{gen} is the number of events generated. The theoretical cross section for each process is shifted by a 10% of their central value.

The $t\bar{t}$ background has been evaluated using a data driven approach as explained in the Sec. 5.4.1. The contribution of the uncertainty due to top background estimate is given by the errors on the normalization, extracted from the fit to the invariant mass distribution (M_{e^+,e^-}) with the $e\mu$ method.

b purity

A not negligible contribution to the backgrounds comes from Z+jets events where charm and light/gluon jets are misidentified as b jets. In order to extract the different quark components a fit to the Secondary Vertex mass distribution has been done using the templates built in the Monte Carlo. The systematic error due to this method is obtained varying the normalization factors of a a 10% of their central values. The systematic uncertainty is the evaluated comparing the cross sections obtained with the central values and those obtained after the variation.

Efficiency

Systematic uncertainties associated to the scale factors applied to Monte Carlo samples are considered. Several sources of uncertainties are due to the Tag and Probe method applied in order to extract the efficiency in data and Monte Carlo. In particular two major categories should be considered. The first category includes uncertainties related to the specific choices made in the definition of the tag and probe procedure such as the invariant mass (M_{ee}) window, the choice of the tag selection criteria or the choice of Monte Carlo. Sources coming from the methodology, such as background estimation and subtraction, dependence on pileup scenario and hadronic activity in the event, belong to the second category. However an important advantage in using the scale factors, defined as the ratio between the efficiency measured in data and the one estimated in Monte Carlo, is the reduction of several common systematic uncertainties as possible correlations between tag and probe or hadronic activity in the event. The main source of systematic uncertainties that still remains applying the scale factors instead of the efficiencies in data and Monte Carlo samples is due to the background estimation in the data sample used in the Tag and Probe method. The efficiencies in Monte Carlo are calculated from simple counting procedure done on signal sample while a fit is used for data and this is the reason why the uncertainty in data arising from background parametrization and shape used in fit do not cancel during evaluation of the scale factors. The scale factors measurement is performed in the 2D-bins of electron transverse energy and pseudorapidity. For each bin the scale factors are evaluated using a different background shape and the maximum shift of central values is taken as associated systematic:

$$\Delta\rho = (\mid \rho_{cent} - \rho_{var} \mid)_{max}, \tag{6.19}$$

where ρ_{cent} is the central result and ρ_{var} is the result obtained by varying the background parametrization. The systematic uncertainty related to the scale factor application is estimated 1.5% in the dielectron final state and 2% in the muons final state.

The b tagging efficiency is measured in data using several methods applied to multijet events and $t\bar{t}$ events. The efficiency measured in data is compared to the one measured in the simulation; scale factors $SF_b = \varepsilon_b^{data}/\varepsilon_b^{MC}$ are then calculated and used to rescale the Monte Carlo sample. Several systematic uncertainties affect the b tagging efficiency measurement:

- pileup: the measured b tagging efficiency depends on the number of proton-proton events superimposed to the primary interaction of interest. In order to match the data better, each Monte Carlo event is reweighted to agree with the average observed pileup. The data and Monte Carlo samples, used in the efficiency estimation, are split into two parts, corresponding to events with a number of reconstructed primary vertices below and above the average value. Each subsample is then used to extract the scale factors $SF_b = \varepsilon_b^{data}/\varepsilon_b^{MC}$. The systematic uncertainty is computed as the difference between the two SF_b values.

- muon p_T cut: for b tagging efficiency estimation, dedicated QCD samples with the requirement of a reconstructed muon have been produced. The central value of the b tagging efficiency is extracted from data with muon transverse momentum (p_T^{μ}) greater then 5 GeV. Another sample of events is selected when p_T^{μ} is increased to quantify how this cut affects the b tagging efficiency. The scale factors are calculated in the two subsamples and the maximum deviation observed from the central value is taken as a systematic uncertainty.
- gluon splitting: QCD events may have a larger fraction of gluon splitting into a $b\bar{b}$ pair than what is generated in the simulation. A Monte Carlo sample where the number of events with gluon splitting was artificially changed has been generated. Results obtained with this modified gluon-splitting Monte Carlo sample are then compared to the ones with the default sample. The observed deviation is quoted as a systematic uncertainty.

The b tagging scale factors give a contribution of 3% of the total cross section to the systematic uncertainty in the Z+1b case and up to 6% in the Z+2b case.

Unfolding

One of the biggest contributions to the total systematic uncertainties comes from the unfolding procedure. In order to estimate it, a new weighted response matrix is built. The weight used to rescale each element of the matrix is given by the ratio between the pre-unfolded data and the reconstructed Monte Carlo. The difference between the results obtained with this new response matrix and the standard one is taken as systematic error. Another source of systematic error due to the unfolding procedure is related to the statistic errors on the elements of the response matrix.

6.3 Combination

The selection applied to identify the Z boson is based on the reconstruction and selection of electrons and muons that are candidate decay products of the Z. In the previous chapter the details of this selection has been provided and also several distributions shown separately for the sample where a Z boson is selected using its decay mode into electrons and for the other sample where the Z boson has been identified through its decay into muons. The differential cross sections presented in this chapter have been obtained, instead, combining electron and muon channels after checking the consistency between results achieved for electron and muon final state. To combine measurement of the same observable individual uncertainties and their correlations are taken into account. The systematic uncertainties in the combined cross sections are calculated by considering the eventual correlation between the electrons final state and the muon final state. The combination is done bin by bin considering the uncertainties fully correlated or uncorrelated between the muon and electron samples, while the correlation between the bins of the distribution is not treated. The uncertainties related to the following sources have been treated as fully correlated:

- the jet energy correction and jet energy resolution
- the procedure adopted to reweight the Monte Carlo in order to reproduce the pileup in the simulated sample
- the unfolding procedure
- the estimation of background contributions from the Monte Carlo
- the integrated luminosity estimate

The above uncertainties are fully correlated because they have been estimated with the variation of a certain parameter in the same way in the electron and muon distribution. The study of the effects of these variation in the muon and the electron distribution provides two independent estimates of the same quantity. So for these uncertainties an average value is taken as global effect. The sources that induce uncertainties that can be considered uncorrelated are:

- the application of scale factors to Monte Carlo samples
- the b purity estimation procedure
- the $t\bar{t}$ background extraction
- the available statistics in the data and Monte Carlo samples

The value of the $i-th$ bin of the combined cross section is obtained by the weighted average of the values of the $i-th$ bin in the electron and muon distribution. The weight is given by the quadrature sum of all the contributions of uncorrelated uncertainties. The error of the weighted average is given by its variance:

$$\sqrt{\frac{1}{\frac{1}{\sigma_{ele}^2}+\frac{1}{\sigma_{muon}^2}}}. \tag{6.20}$$

Tables 6.13-6.2 show the systematic error contributions (in percent) for the differential cross section measurement as a function of the leading b tagged jet pseudorapidity for the combined channels in the sample Z+1b jet and Z+2b jets respectively. A breakdown of all the statistics and systematics error sources together with the relative total error expressed in percentage on the other observables measured in the analysis is given in Appendix 6.6. The dominant source of uncertainty is the unfolding whose contribution ranges from 2% to over 45% depending on the distribution considered, both in the Z+1b and Z+2b cases.

bin	**data stat**	**bkg stat**	**eff syst**	**jec syst**	**jer syst**	**pu syst**	**bkg syst**	**ttbar stat**	**bfit stat**	**btag syst**	**unfold stat**	**unfold syst**	**lumi syst**	**total stat**	**total syst**	**total error**
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	1.0	2.5	1.4	1.8	0.2	0.2	0.2	0.2	4.2	3.0	1.0	2.6	2.6	4.3	5.9	7.3
2	1.0	2.4	1.3	1.8	0.2	0.2	0.2	0.2	4.2	3.0	1.0	2.5	2.6	4.3	5.8	7.2
3	1.0	2.3	1.4	1.7	0.2	0.2	0.2	0.2	4.2	3.0	0.9	2.3	2.6	4.3	5.7	7.1
4	0.9	2.2	1.4	1.7	0.2	0.2	0.2	0.2	4.2	3.0	0.9	2.1	2.6	4.3	5.5	7.0
5	0.9	2.1	1.3	1.6	0.2	0.2	0.2	0.2	4.1	3.0	0.8	1.9	2.6	4.2	5.4	6.8
6	0.8	2.0	1.3	1.6	0.2	0.2	0.2	0.2	4.1	3.0	0.8	1.6	2.6	4.2	5.2	6.7
7	0.8	1.9	1.3	1.5	0.2	0.2	0.2	0.2	4.1	3.0	0.7	1.4	2.6	4.2	5.1	6.6
8	0.8	1.8	1.3	1.5	0.3	0.2	0.2	0.2	4.1	3.0	0.7	1.3	2.6	4.2	5.1	6.6
9	0.8	1.8	1.3	1.5	0.3	0.2	0.2	0.2	4.1	3.0	0.7	1.2	2.6	4.2	5.0	6.5
10	0.8	1.9	1.4	1.5	0.4	0.2	0.2	0.2	4.1	3.0	0.7	0.9	2.6	4.2	5.0	6.5
11	0.8	2.0	1.4	1.6	0.4	0.2	0.2	0.2	4.1	3.0	0.7	0.5	2.6	4.2	5.0	6.5
12	0.9	2.1	1.3	1.6	0.5	0.3	0.2	0.2	4.1	3.0	0.7	0.2	2.6	4.2	5.0	6.6
13	0.9	2.2	1.3	1.6	0.5	0.3	0.2	0.2	4.2	3.0	0.8	0.4	2.6	4.3	5.1	6.6
14	1.0	2.3	1.3	1.7	0.5	0.3	0.2	0.2	4.2	3.0	0.8	0.8	2.6	4.3	5.2	6.8
15	1.0	2.4	1.3	1.8	0.6	0.3	0.2	0.2	4.2	3.0	0.8	1.1	2.6	4.3	5.4	6.9
16	1.0	2.5	1.4	1.8	0.6	0.3	0.2	0.2	4.2	3.0	0.9	1.3	2.6	4.3	5.5	7.0
17	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 6.1: Statistical and systematic error contributions (in percent) for the differential cross section measurement as a function of the leading b tagged jet pseudorapidity in the sample Z+1b jet for the combined channels.

bin	**data stat**	**bkg stat**	**eff syst**	**jec syst**	**jer syst**	**pu syst**	**bkg syst**	**ttbar stat**	**bfit stat**	**btag syst**	**unfold stat**	**unfold syst**	**lumi syst**	**total stat**	**total syst**	**total error**
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	4.8	5.6	1.3	5.9	1.6	0.6	0.6	2.6	0.0	6.0	5.3	2.7	2.6	5.5	12.2	13.4
2	4.6	5.4	1.3	5.8	1.5	0.7	0.6	2.6	0.0	6.0	5.2	2.6	2.6	5.3	12.0	13.1
3	4.4	5.2	1.4	6.1	1.5	0.8	0.6	2.6	0.0	6.1	5.0	2.6	2.6	5.1	12.1	13.1
4	4.2	5.0	1.3	6.0	1.3	0.7	0.6	2.6	0.0	6.0	4.9	2.8	2.6	5.0	11.8	12.8
5	4.0	4.7	1.3	6.2	1.2	0.8	0.6	2.6	0.0	6.0	4.7	2.9	2.6	4.8	11.8	12.7
6	3.8	4.5	1.3	6.3	1.0	0.8	0.5	2.7	0.0	6.0	4.5	3.0	2.6	4.6	11.6	12.5
7	3.6	4.3	1.3	6.4	0.8	0.8	0.5	2.7	0.0	6.0	4.4	3.0	2.6	4.5	11.6	12.4
8	3.5	4.2	1.3	6.6	0.9	0.8	0.5	2.7	0.0	6.0	4.3	2.9	2.6	4.4	11.6	12.4
9	3.5	4.2	1.3	6.8	1.0	0.8	0.5	2.8	0.0	6.1	4.3	2.8	2.6	4.5	11.7	12.5
10	3.6	4.3	1.3	6.9	1.2	0.8	0.5	2.8	0.0	6.0	4.3	2.7	2.6	4.6	11.8	12.6
11	3.8	4.5	1.3	7.0	1.5	0.8	0.5	2.8	0.0	6.0	4.5	2.4	2.6	4.7	11.9	12.8
12	4.1	4.8	1.3	7.1	1.7	0.9	0.5	2.8	0.0	6.0	4.6	2.3	2.6	4.9	12.2	13.1
13	4.3	5.0	1.3	7.2	1.9	0.9	0.5	2.8	0.0	6.0	4.8	2.0	2.6	5.1	12.4	13.4
14	4.5	5.3	1.3	7.3	2.1	1.0	0.5	2.8	0.0	6.0	5.0	1.8	2.6	5.3	12.7	13.7
15	4.7	5.5	1.3	7.4	2.2	0.9	0.5	2.8	0.0	6.0	5.2	1.6	2.6	5.5	12.8	13.9
16	4.9	5.8	1.4	7.9	2.7	1.2	0.6	2.8	0.0	6.7	5.4	1.7	2.9	5.7	13.9	15.0
17	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 6.2: Statistical and systematic error contributions (in percent) for the differential cross section measurement as a function of the leading b tagged jet pseudorapidity in the sample Z+2b jet for the combined channels.

6.4 Generator level analysis

The generator level analysis should mimic as closely as possible the analysis done on the real data. That means the same selection requirements have to be applied also in this analysis and that physics effects which affects measurements, as the final state radiation (FSR), have to be considered and properly treated at generator level. Experimentally, depending on the reconstruction and the detector limitations, one may not be able to identify the real photon emitted close to the lepton. During the object reconstruction phase these photons are inevitably merged into the lepton. In order to allow valid comparisons, the theory and the experiment should use the same definitions of the lepton object. Among the generated final state particles a pair of leptons (electrons or muons) with opposite sign and highest transverse momentum have to be selected. In order to take into account FSR effects it is necessary to consider additionally the photons around the lepton. One should define a cone around the lepton which should be used for the "dressing", adding explicitly photons to the lepton. In such way the effective contribution from FSR photons is considered in the estimation of the lepton four momentum. These leptons are called *dressed* leptons and they should be the closest to the real objects. The final generator lepton objects considered in the analysis are so composed of a "bare" lepton and all the photons inside a cone of $\Delta R = 0.1$ and they are required to have:

- transverse momentum greater than 20 GeV
- the eta variable has to be in the range [-2.4,2.4]
- a minimum distance of 0.5 in ΔR between the jet and the two leptons identified as decay products of the Z boson

For the lepton pair passing this selection, the invariant mass is calculated requiring it to be in the allowed Z mass window, which has been chosen to be 71 GeV $< M_{l^+l^-} <$ 111 GeV. The jets are clusterized using the anti-kt algorithm with a clusterization parameter $\Delta R = 0.5$. Before the clusterization some particles are removed from the sample of final state particles. The leptons and the FSR photons selected are eliminated from the sample of final state particles to avoid to count them as jets and also the neutrinos are not considered in the construction of the jets. The jets must have a transverse momentum greater than 30 GeV and be inside the acceptance defined by the range in η of $[-2.4, 2.4]$. After the jet selection an algorithm is used to identify those that are originated by the hadronization of a b quark. A jet is tagged as a b jet if a B hadron is present in the history of one of the particles that constitute it. The B hadron found has to be close to the jet inside a cone of $\Delta R = 0.5$.

6.4.1 RIVET

RIVET (Robust Independent Validation of Experiment and Theory) [141] is a C++ class library that produces simulated distributions which can be directly compared to measured data for Monte Carlo validation and tuning studies. It can also be used without reference data to compare two or more generators to each other for tune comparison. The final purpose is to have a RIVET analysis relative to each physics paper published, storing the relative experimental data in a file format adopted by RIVET. This last step allows for an automated generation of the bin-edges of a histogram, starting from the one available with the real data. Furthermore, it allows for an easy comparison between data and Monte Carlo, with supplied scripts. The RIVET library includes all the tools needed to calculate physical observables and it also contains a set of analysis already implemented from LEP, HERA, RHIC, KEK-B, Tevatron Runs I and II, LHC and more. This means that all the given analyses can still be run independently from the details of different experiments. RIVET operates on the HepMC [142] event record that is a standardized event format for Monte Carlo generators and that makes RIVET generator-independent. As for the generator level analysis described above, the RIVET analysis should mimic as closely as possible the analysis done on the real data. The steps followed to build the RIVET analysis for the present work are the same explained in the previous section.

6.5 Cross section measurement

In this chapter the results of the measurement of the differential cross section for the production of a Z boson with exactly one b jet or at least two b jets are shown. The differential cross section $\frac{d\sigma}{dx}$ (with x one of the measured observables) is given by the following equation:

$$\frac{d\sigma}{dx} = \frac{N_{sel} - N_{bkg}}{\varepsilon \cdot \int L dt}, \tag{6.21}$$

where N_{sel} is the number of the events passing the selection requirements as described in the previous chapter, N_{bkg} is the estimated number of different background contributions, ε is an efficiency term and $\int L dt$ is the integrated luminosity. The different Monte Carlo background samples have to be rescaled taking into account the number of generated events and the theoretical cross section calculations (eq. 6.18). The $t\bar{t}$ background is normalized using the factor (c_t) extracted by fitting the sidebands of the dielectron or dimuon invariant mass distribution using the sidebands obtained in the $e\mu$ sample. The N_{bkg} term comprises also the $Z+$ jets events where the jets coming from the hadronization of c or light quarks have been wrongly tagged as b jets. These samples are corrected for the scale factors extracted from the Secondary Vertex

Mass fitting procedure before being subtracted. In the case of the associated production of a Z boson and exactly one b tagged jet also migrations from the Z+2b sample have to be considered. So those events with more than 1 b tagged jet have to be rescaled and reduced. Migrations do not affect the Z+2b jet that is an almost pure sample after the selection applied. The Monte Carlo samples are also corrected for the scale factors $SF = \varepsilon^{data}/\varepsilon^{MC}$ in order to take into account the difference between the efficiencies measured in data and those predicted by the simulation. The application of the scale factors to Monte Carlo samples makes possible the background subtraction as the data and Monte Carlo samples are consistent. In fig. 6.15 - 6.31 the differential cross sections for the associated production of the Z boson with exactly or at least two b tagged jets as function of several observables are shown. The plots show the results obtained for the combined channels electrons plus muons with the statistical uncertainty. The measurements are deconvolved from the detector effects using the unfolding technique as described in Sec. 6.1 and are compared with different theoretical calculations made by MadGraph 4FS and MadGraph 5FS event generators, both interfaced with the parton shower PYTHIA6 and POWHEG plus PYTHIA6.

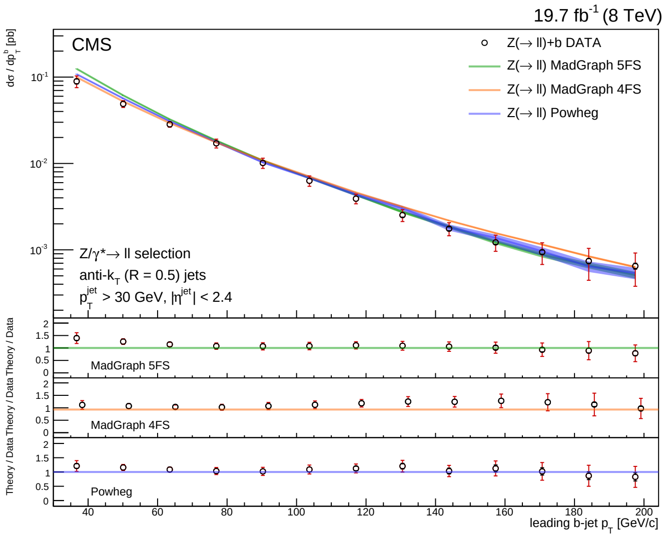


Figure 6.15: Unfolded differential Z+1b jet cross section as a function of the leading b jet transverse momentum for electrons and muons final state, compared with the different theoretical expectations: MadGraph 4Fs (orange band), MadGraph 4Fs (green band) and POWHEG (blue band).

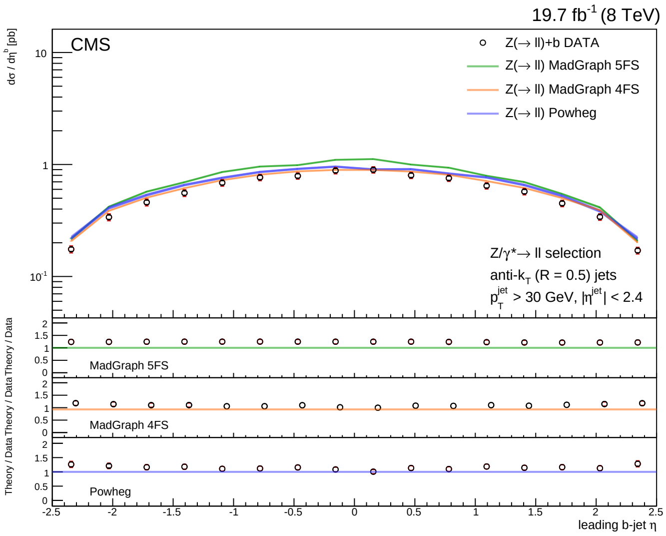


Figure 6.16: Unfolded differential $Z{+}1b$ jet cross section as a function of the leading b jet pseudorapidity for electrons and muons final state, compared with the different theoretical expectations: MadGraph 4Fs (orange band), MadGraph 4Fs (green band) and POWHEG (blue band).

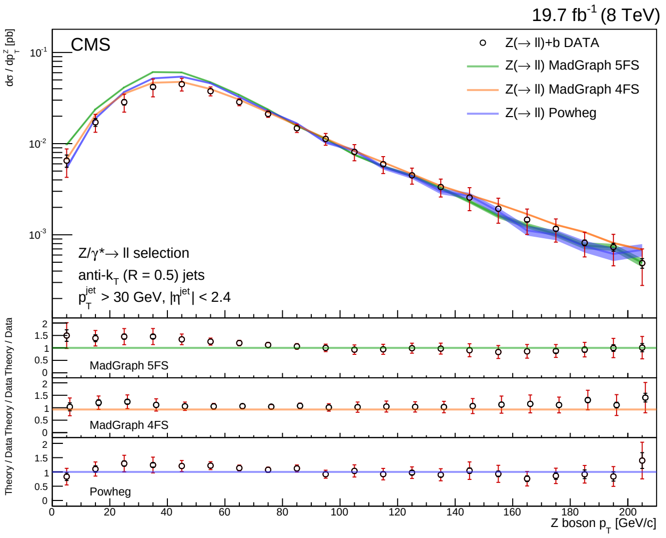


Figure 6.17: Unfolded differential $Z{+}1b$ jet cross section as a function of the Z boson transverse momentum for electrons and muons final state, compared with the different theoretical expectations: MadGraph 4Fs (orange band), MadGraph 4Fs (green band) and POWHEG (blue band).

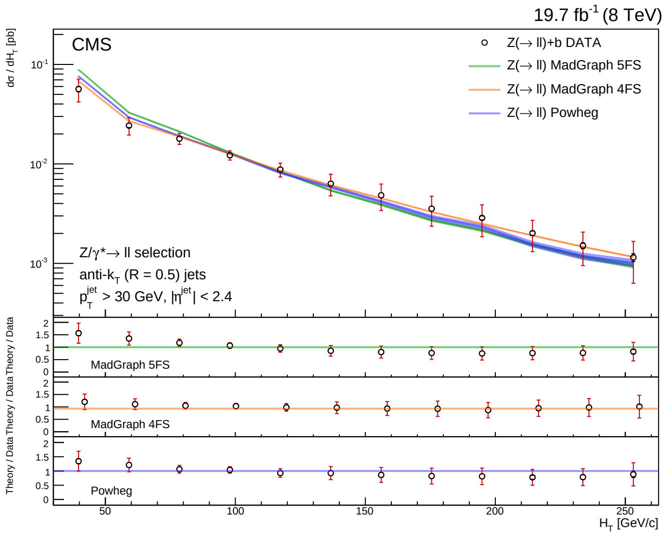


Figure 6.18: Unfolded differential $Z{+}1b$ jet cross section as a function of the H_T variable for electrons and muons final state, compared with the different theoretical expectations: MadGraph 4Fs (orange band), MadGraph 4Fs (green band) and POWHEG (blue band).

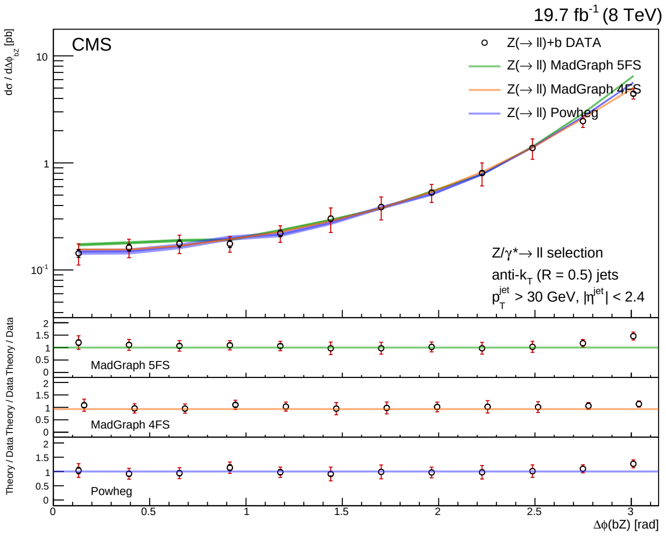


Figure 6.19: Unfolded differential $Z{+}1b$ jet cross section as a function of the $\Delta\phi$ between the Z boson and the leading b tagged jet for electrons and muons final state, compared with the different theoretical expectations: MadGraph 4Fs (orange band), MadGraph 4Fs (green band) and POWHEG (blue band).

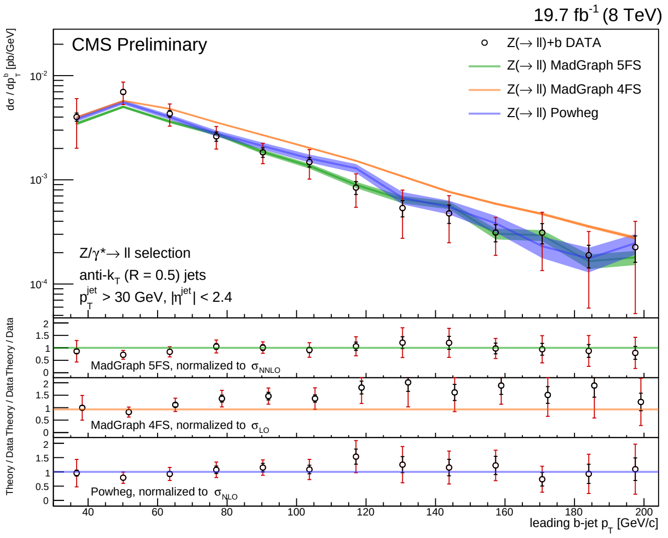


Figure 6.20: Unfolded differential $Z{+}2b$ jet cross section as a function of the leading b jet transverse momentum for electrons and muons final state, compared with the different theoretical expectations: MadGraph 4Fs (orange band), MadGraph 4Fs (green band) and POWHEG (blue band).

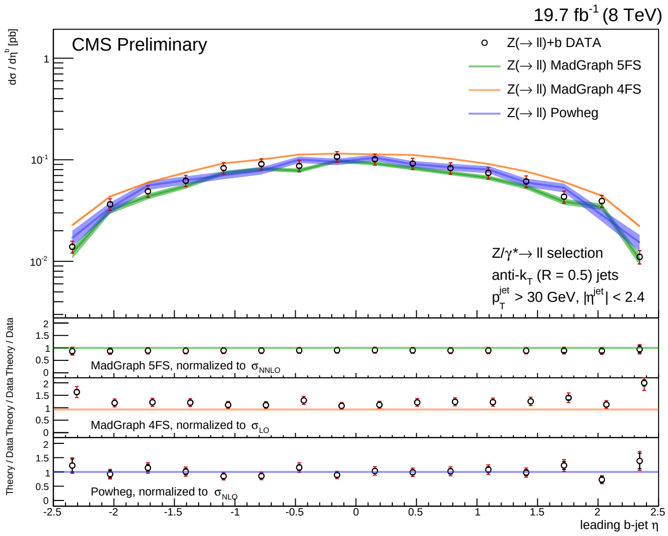


Figure 6.21: Unfolded differential $Z{+}2b$ jet cross section as a function of the leading b jet pseudorapidity for electrons and muons final state, compared with the different theoretical expectations: MadGraph 4Fs (orange band), MadGraph 4Fs (green band) and POWHEG (blue band).

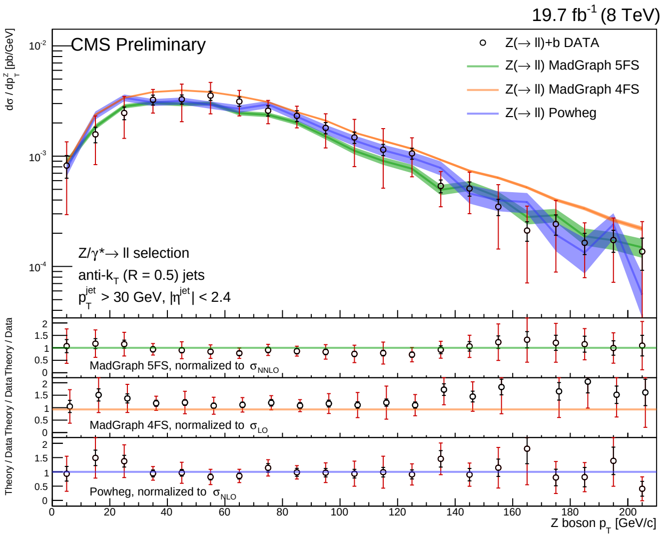


Figure 6.22: Unfolded differential Z+2b jet cross section as a function of the Z boson transverse momentum for electrons and muons final state, compared with the different theoretical expectations: MadGraph 4Fs (orange band), MadGraph 4Fs (green band) and POWHEG (blue band).

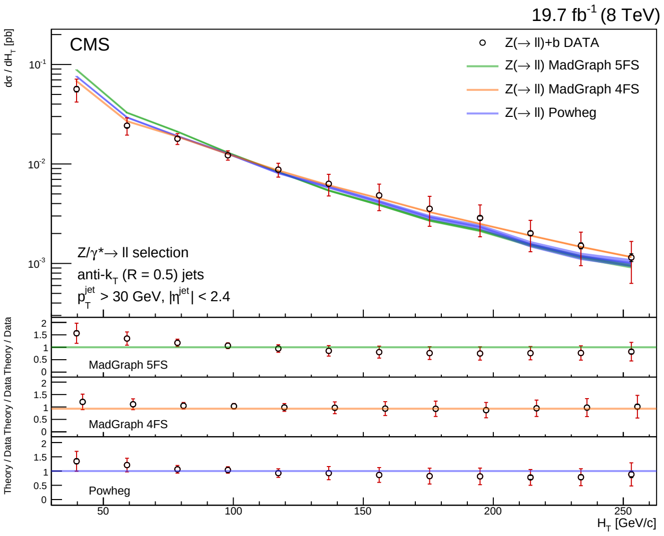


Figure 6.23: Unfolded differential Z+1b jet cross section as a function of the H_T variable for electrons and muons final state, compared with the different theoretical expectations: MadGraph 4Fs (orange band), MadGraph 4Fs (green band) and POWHEG (blue band).

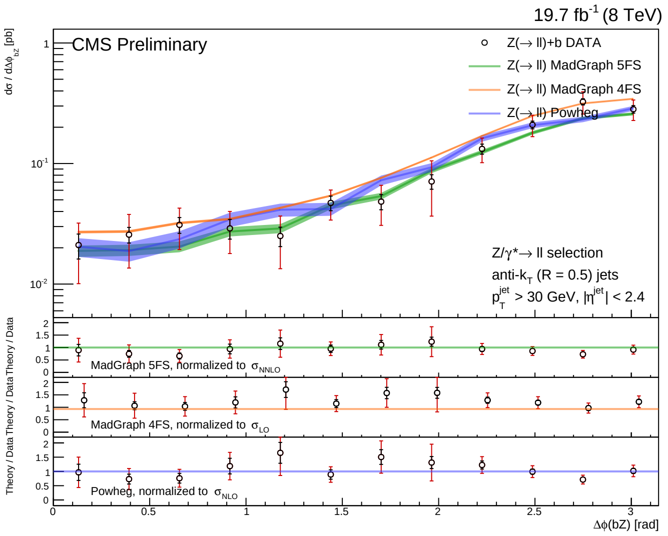


Figure 6.24: Unfolded differential Z+2b jet cross section as a function of the $\Delta\phi$ between the Z boson and the leading b tagged jet for electrons and muons final state, compared with the different theoretical expectations: MadGraph 4Fs (orange band), MadGraph 4Fs (green band) and POWHEG (blue band).

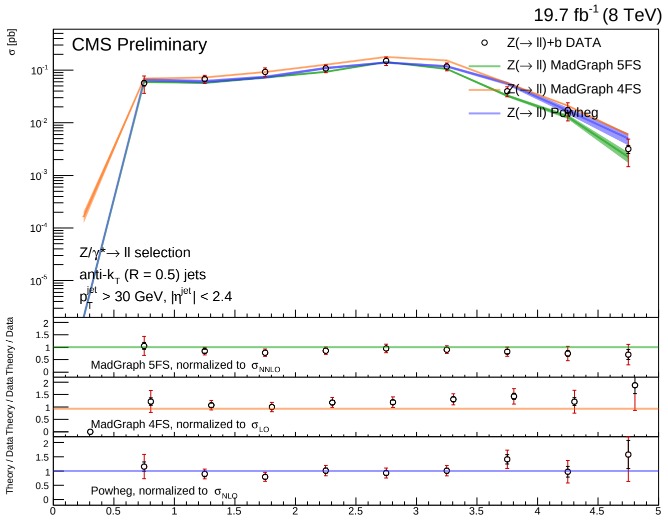


Figure 6.25: Unfolded differential Z+2b jet cross section as a function of the ΔR between the two b tagged jets for electrons and muons final state, compared with the different theoretical expectations: MadGraph 4Fs (orange band), MadGraph 4Fs (green band) and POWHEG (blue band).

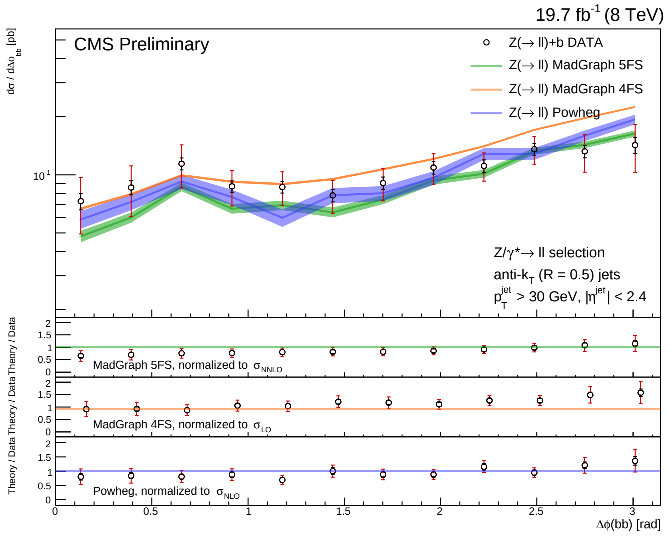


Figure 6.26: Unfolded differential $Z+2b$ jet cross section as a function of the $\Delta\phi$ between the two b tagged jets for electrons and muons final state, compared with the different theoretical expectations: MadGraph 4Fs (orange band), MadGraph 4Fs (green band) and POWHEG (blue band).

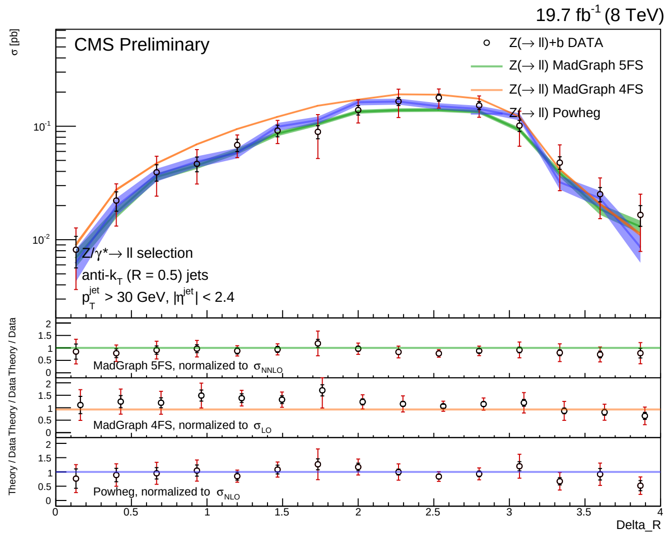


Figure 6.27: Unfolded differential $Z+2b$ jet cross section as a function of the ΔR between the Z boson and the closest b tagged jets for electrons and muons final state, compared with the different theoretical expectations: MadGraph 4Fs (orange band), MadGraph 4Fs (green band) and POWHEG (blue band).

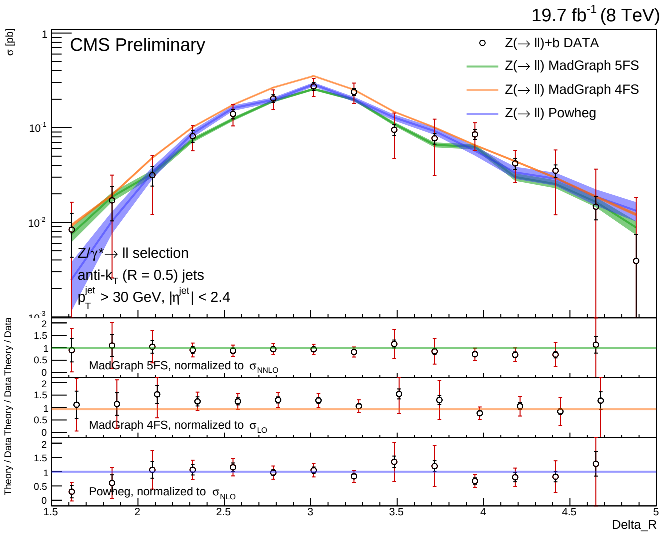


Figure 6.28: Unfolded differential Z+2b jet cross section as a function of the ΔR between the Z boson and the further b tagged jets for electrons and muons final state, compared with the different theoretical expectations: MadGraph 4Fs (orange band), MadGraph 4Fs (green band) and POWHEG (blue band).

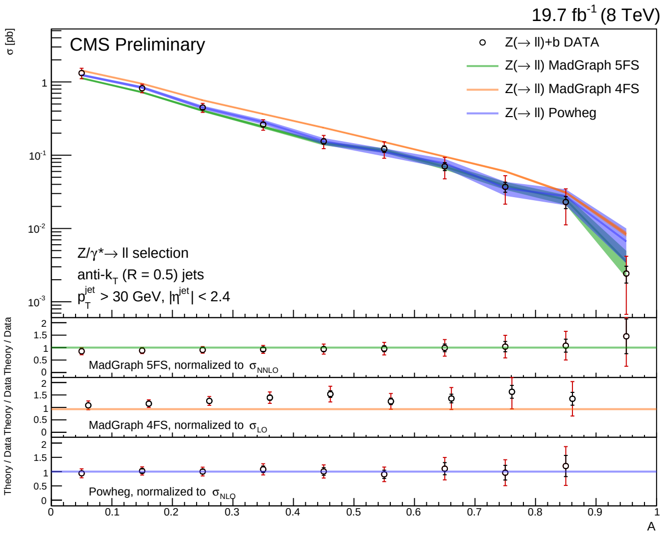


Figure 6.29: Unfolded differential Z+2b jet cross section as a function of the asymmetry between the Z boson direction and the b tagged jet direction for electrons and muons final state, compared with the different theoretical expectations: MadGraph 4Fs (orange band), MadGraph 4Fs (green band) and POWHEG (blue band).

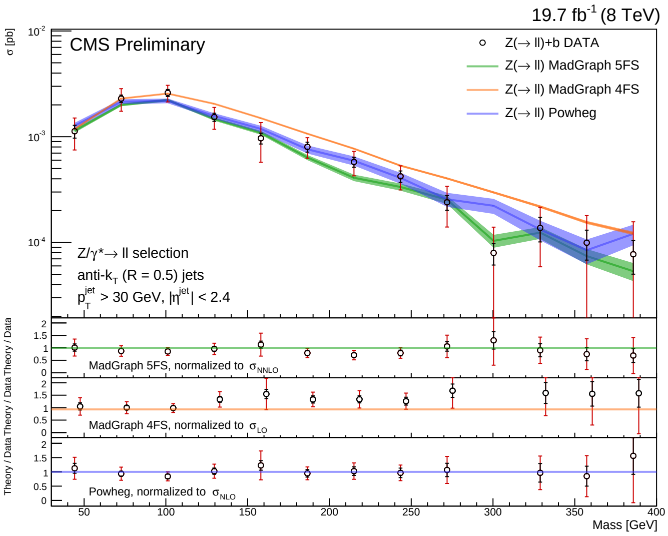


Figure 6.30: Unfolded differential Z+2b jet cross section as a function of the invariant mass between the two b tagged jets for electrons and muons final state, compared with the different theoretical expectations: MadGraph 4Fs (orange band), MadGraph 4Fs (green band) and POWHEG (blue band).

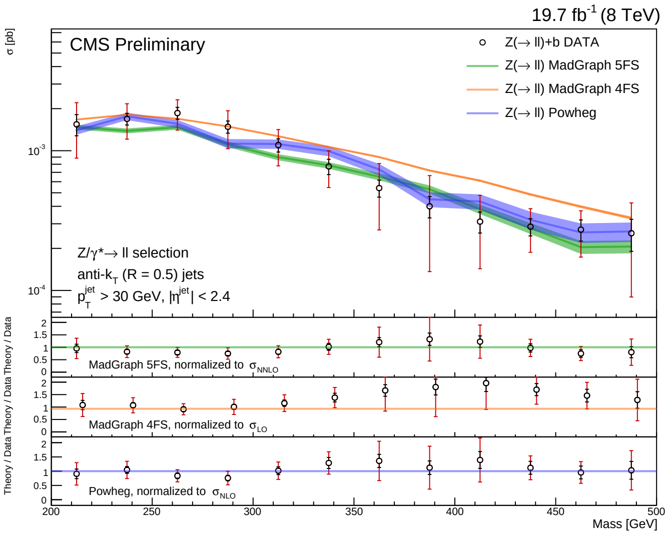


Figure 6.31: Unfolded differential Z+2b jet cross section as a function of the invariant mass of the system with the Z boson the two b tagged jets for electrons and muons final state, compared with the different theoretical expectations: MadGraph 4Fs (orange band), MadGraph 4Fs (green band) and POWHEG (blue band).

6.6 Interpretation of the results

In the previous section, the differential cross sections for the associated production of a Z boson and exactly one or at least two b jets, measured at 8 TeV, as function of different observables have been presented. The results shown, corrected for the detector effects by mean of the SVD unfolding, have been obtained combining the electron and the muon channels taking into account correlations between the systematic uncertainties. All the results have been compared to theoretical expectations using a dedicated RIVET analysis. Theoretical predictions at leading order in QCD are computed with the MadGraph event generator in the four-flavour (4FS) and five-flavour (5FS) schemes and with POWHEG that provides a next-to-leading order matrix element calculation limited to the emission of one jet of any flavour. Both MadGraph and POWHEG event generators have been interfaced with Pythia 6 tune Z2*. The parton density functions adopted for MadGraph plus Pythia 6 is CTEQ6L1 for the 5FS and MSTW2008 for the 4FS, while the CT10 is used inside POWHEG plus PYTHIA 6. The theoretical predictions have been rescaled to the available cross section calculation: next-to-next leading order cross section (MadGraph 5FS case) or to the leading order cross section (MadGraph 4FS) or to the next-to leading order cross section (POWHEG).

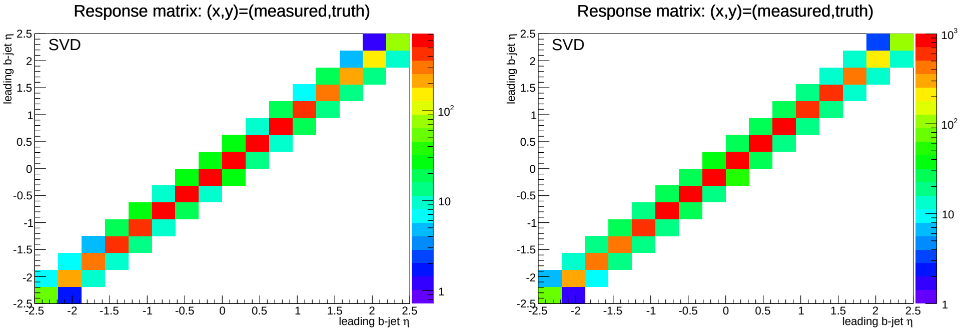
The comparison between the measured differential cross section for the associated production of a Z boson and exactly one b jet and the theoretical predictions provided both by MadGraph plus PYTHIA 6 and POWHEG is found in good agreement within the measurement systematic and statistical uncertainties. A difference (order 30%) is observed between data and theoretical expectation by MadGraph at low transverse momentum of the leading b jet and of the Z boson. This discrepancy would probably be reduced with a next-to leading order (NLO) matrix element calculation with MadGraph. Differences of the order of 10% can be seen between the predictions done with the 4FS and 5FS approaches especially in the b jet transverse momentum and Z boson p_T where a small discrepancy in the comparison with data is found for the 4FS in the low momentum region. In the Z+2b jets final state the MadGraph 5FS provides a better description than the MadGraph 4FS in all the kinematic variables. In particular the 5FS seems to reproduce more precisely the high transverse momentum region of the leading b jet and of the Z boson. Moreover an overestimation of the 4FS prediction to data can be noted in the differential cross section as function of the pseudorapidity especially in the forward region. In order to better exploit the different approaches in the theoretical calculations understanding the impact of the b quark mass other kinematic distributions have been studied in the case of the Z+2b jet final state. The differential cross section as function of angular variables and the invariant mass of the system formed by the b jets and the Z boson have been shown. Some evidence for disagreement between 4FS predictions and data are also visible in these cases. The theoretical uncertainties coming from the renormalization scale and those related to the different PDF sets are not yet considered in these final results and their estimate and inclusion represent the next step of this analysis.

Conclusions

The measurement of the differential cross sections as function of different observables for the associated production of a Z boson and b jets have been presented. This thesis is based on data collected with the CMS detector in 2012. The data sample analysed corresponds to proton-proton collision events at a center-of-mass energy of 8 TeV and to an integrated luminosity L = 19.78 fb^{-1} in the $Z \to e^+e^-$ selection and L = 19.75 fb^{-1} in the $Z \to \mu^+\mu^-$ selection. The production of b quarks in association with a vector boson Z is a challenging topic both under a theoretical and experimental point of view. It is interesting to study both the process with a Z boson and exactly 1 b jet (Z+1b) and also that with a Z boson associated with at least 2 b jets (Z+2b). The associated production of the Z boson and exactly one b jet is of interest for precision tests of pQCD. Since predictions use different approaches to calculate this production mechanism (massive versus massless b-quarks, 4/5-flavor approaches, LO versus NLO), the study of this process provides information to distinguish between these different approaches. Moreover, the Z +1b jet process can be of interest to study the b content of the proton. The production of a Z boson in association with at least two b jets is especially interesting in the context of searches. The Z+2b represents the main background for processes with the same final state like the Standard Model Higgs decay $H \to Z(ll)Z(bb)$ and the Higgs Strahlung ($qq \to ZH(bb)$) or for search of new particles as SUSY Higgs or fourth generation heavy quarks. Specific requirements on some variables have been applied in order to discriminate the Z + b signature. Then, different background contributions have been evaluated using Monte Carlo simulations, except in the case of $t\bar{t}$ background whose extraction has been performed using a data driven approach. The b jets fraction, defined as the number of events after the selection that truly contains one or at least two b tagged jets, is an important variable in this analysis. For this reason a particular attention has been dedicated to estimate the contamination of different quark flavours extracted simultaneously in the Z + 1b and Z + 2b samples, in order to take into account possible inefficiencies in the tagging of jets and so migrations of events from one sample to the other. In order to compare the measured data with the theoretical predictions, the measured observables have been corrected to deconvolve them from the detector effects with the unfolding method. Different sources of systematic uncertainties and their effects have been explored. Finally the differential cross sections measured have been compared with QCD theoretical predictions with several implementations: leading order calculation made by MadGraph 5FS and MadGraph 4FS, next-to-leading order calculation provided by POWHEG (next-to leading order calculation limited to the emission of one jet). The comparison between the measured differential cross section for the associated production of a Z boson and exactly one b jet and the theoretical predictions provided both by MadGraph and POWHEG plus PYTHIA 6 (tune Z2*) is found in good agreement accounting for the measurement uncertainties except for some discrepancies (order 30%) observed in the comparison using the theoretical expectation by MadGraph at low transverse momentum of the leading b jet and of the Z boson. This discrepancy would probably be reduced with a next-to leading order matrix element calculation with MadGraph. For the $Z{+}2b$ jets differential cross sections, all prediction provide a reasonable description of the data within the large experimental uncertainties. In this case the MadGraph 5FS seems to provide a better description than the MadGraph 4FS and POWHEG in all the kinematic variables. This is probably due to the fact that the next-to leading order matrix element calculation of POWHEG is limited to the emission of only one jet. In these comparisons the theoretical uncertainties coming from the renormalization scale and those related to the different PDF sets are not considered so an important step for this analysis could be the estimation of these contributions. It could be very interesting to compare these results with theoretical predictions obtained using next-to leading order matrix element provided by MadGraph four-flavor and five-flavor schemes.

Appendix A

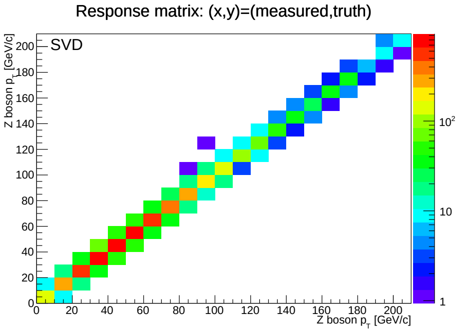
Unfolding response matrices



(a) Response matrix for the leading b jet pseudorapidity variable in the electron final state with the Z+1b selection.

(b) Response matrix for the leading b jet pseudorapidity variable in the muon final state with the Z+1b selection.

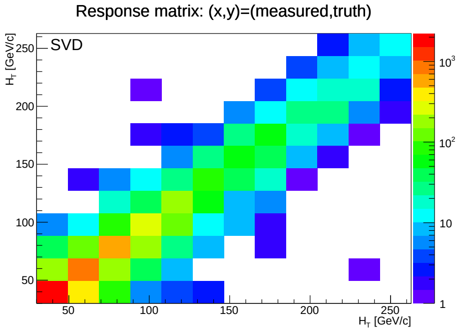
Figure 6.32: Response matrix for the leading b jet pseudorapidity variable with the Z+1b selection.



(a) Response matrix for the Z boson transverse momentum in the electron final state with the Z+1b selection.

(b) Response matrix for the Z boson transverse momentum in the muon final state with the Z+1b selection

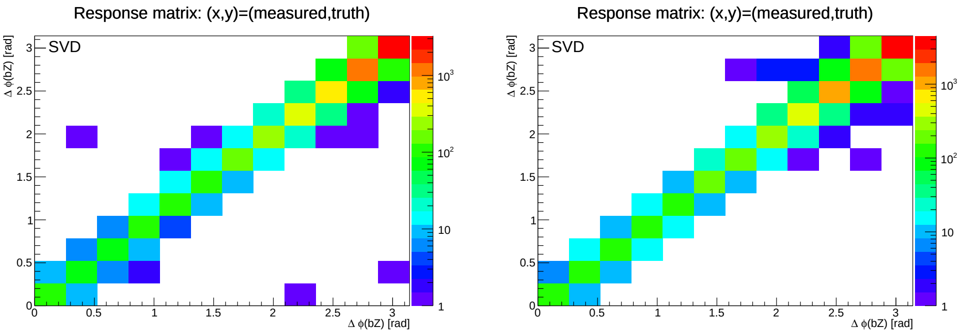
Figure 6.33: Response matrix for the Z boson transverse momentum with the Z+1b selection.



(a) Response matrix for the H_T variable with the Z+1b selection in the electron final state.

(b) Response matrix for the H_T variable with the Z+1b selection in the muon final state.

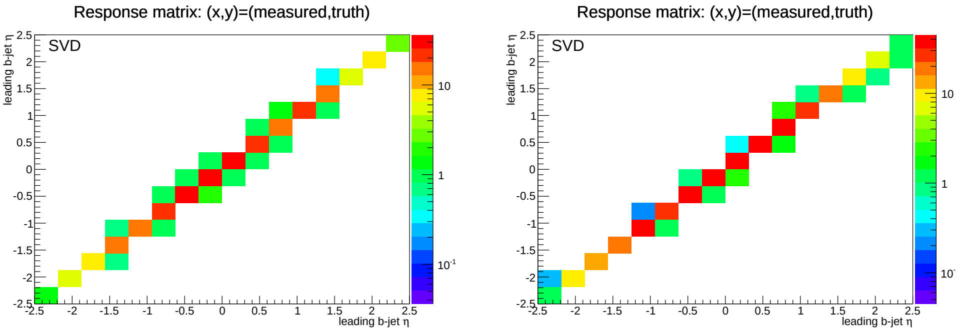
Figure 6.34: Response matrix for the H_T variable with the Z+1b selection.



(a) Response matrix for the $\Delta\Phi(Zb)$ variable with the Z+1b selection in the electron final state.

(b) Response matrix for the $\Delta\Phi(Zb)$ variable with the Z+1b selection in the muon final state.

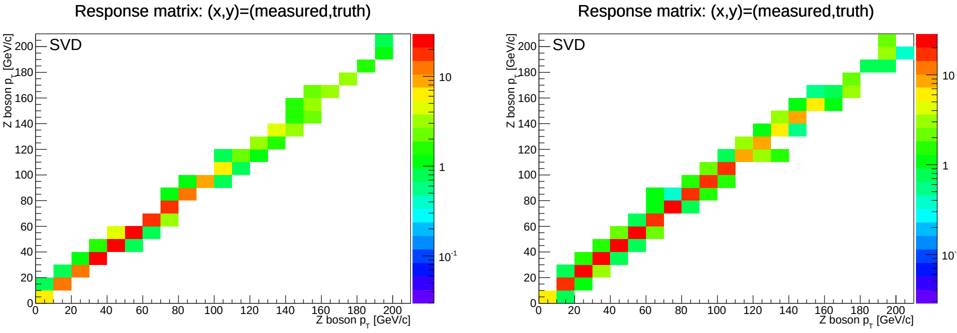
Figure 6.35: Response matrix for the $\Delta\Phi(Zb)$ variable with the Z+1b selection.



(a) Response matrix for the leading b jet pseudorapidity variable in the electron final state with the Z+2b selection.

(b) Response matrix for the leading b jet pseudorapidity variable in the muon final state with the Z+2b selection.

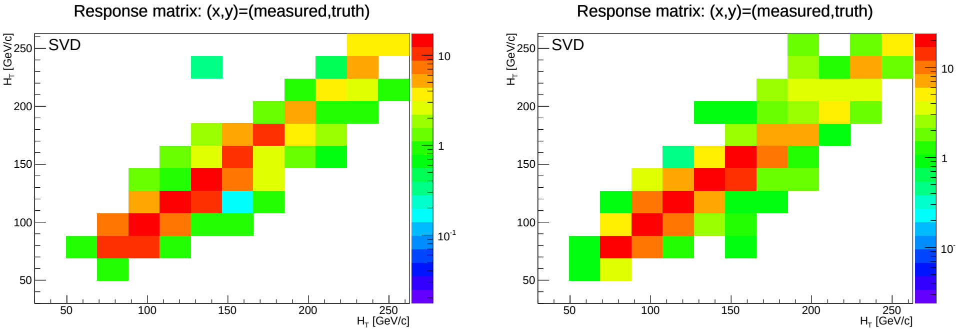
Figure 6.36: Response matrix for the leading b jet pseudorapidity variable with the Z+2b selection.



(a) Response matrix for the Z boson transverse momentum in the electron final state with the Z+2b selection.

(b) Response matrix for the Z boson transverse momentum in the muon final state with the Z+2b selection

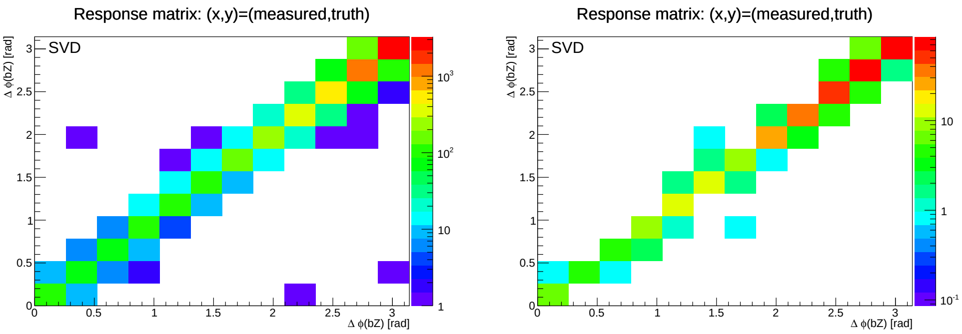
Figure 6.37: Response matrix for the Z boson transverse momentum with the Z+2b selection.



(a) Response matrix for the H_T variable with the Z+2b selection in the electron final state.

(b) Response matrix for the H_T variable with the Z+2b selection in the muon final state.

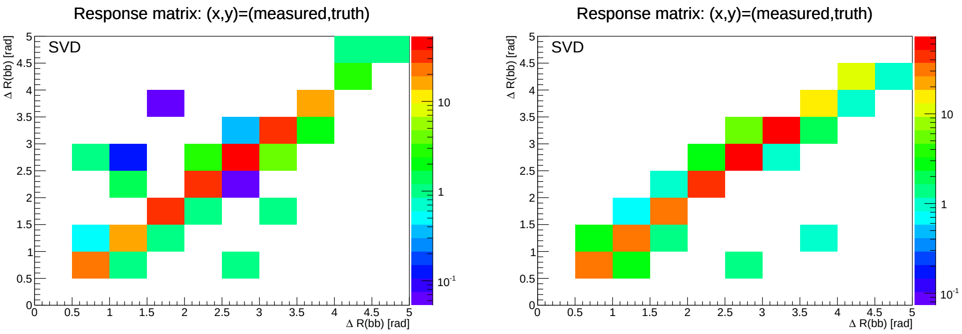
Figure 6.38: Response matrix for the H_T variable with the Z+2b selection.



(a) Response matrix for the $\Delta\Phi(Zb)$ variable with the Z+2b selection in the electron final state.

(b) Response matrix for the $\Delta\Phi(Zb)$ variable with the Z+2b selection in the muon final state.

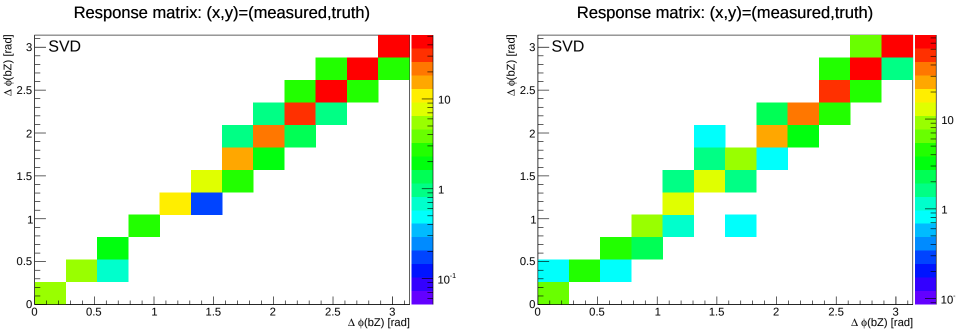
Figure 6.39: Response matrix for the $\Delta\Phi(Zb)$ variable with the Z+2b selection.



(a) Response matrix for the $\Delta R(bb)$ variable with the Z+2b selection in the electron final state.

(b) Response matrix for the $\Delta R(bb)$ variable with the Z+2b selection in the muon final state.

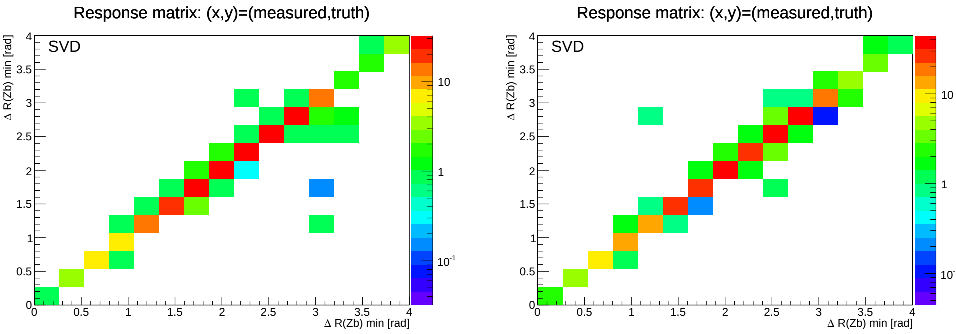
Figure 6.40: Response matrix for the $\Delta R(bb)$ variable with the Z+2b selection.



(a) Response matrix for the $\Delta\phi(bb)$ variable with the Z+2b selection in the electron final state.

(b) Response matrix for the $\Delta\phi(bb)$ variable with the Z+2b selection in the muon final state.

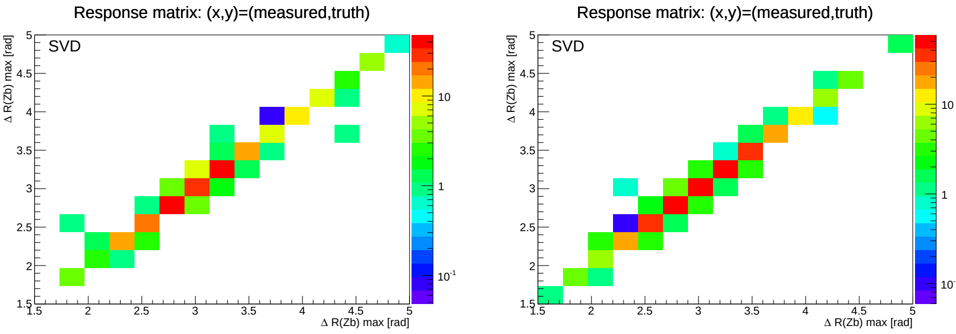
Figure 6.41: Response matrix for the $\Delta\phi(bb)$ variable with the Z+2b selection.



(a) Response matrix for the $\Delta R(Zb)^{min}$ variable with the Z+2b selection in the electron final state.

(b) Response matrix for the $\Delta R(Zb)^{min}$ variable with the Z+2b selection in the muon final state.

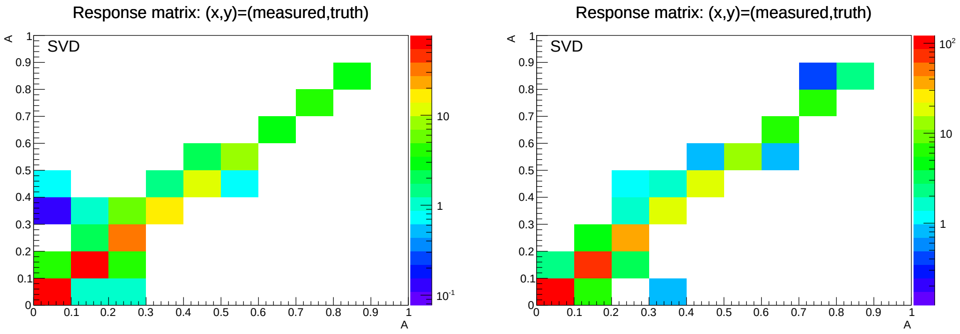
Figure 6.42: Response matrix for the $\Delta R(Zb)^{min}$ variable with the Z+2b selection.



(a) Response matrix for the $\Delta R(Zb)^{max}$ variable with the Z+2b selection in the electron final state.

(b) Response matrix for the $\Delta R(Zb)^{max}$ variable with the Z+2b selection in the muon final state.

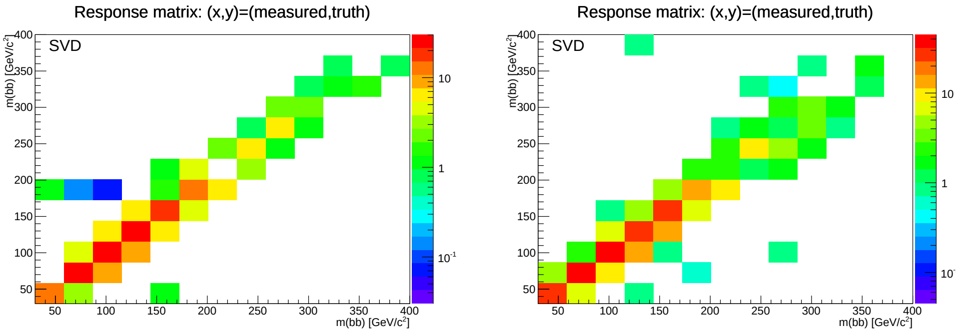
Figure 6.43: Response matrix for the $\Delta R(Zb)^{max}$ variable with the Z+2b selection.



(a) Response matrix for the asymmetry A_{Zbb} variable with the Z+2b selection in the electron final state.

(b) Response matrix for the asymmetry A_{Zbb} variable with the Z+2b selection in the muon final state.

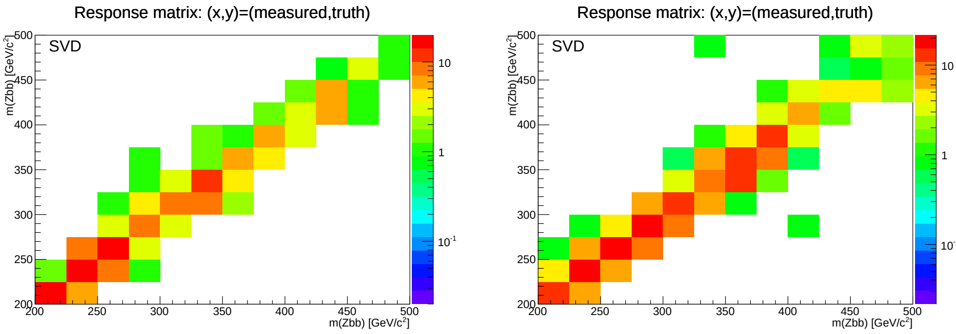
Figure 6.44: Response matrix for the asymmetry A_{Zbb} variable with the Z+2b selection.



(a) Response matrix for the invariant mass variable between the two b jets with the Z+2b selection in the electron final state.

(b) Response matrix for the invariant mass variable between the two b jets with the Z+2b selection in the muon final state.

Figure 6.45: Response matrix for the invariant mass variable between the two b jets with the Z+2b selection.



(a) Response matrix for the invariant mass variable between the two b jets and the Z boson with the Z+2b selection in the electron final state.

(b) Response matrix for the invariant mass variable between the two b jets and the Z boson with the Z+2b selection in the muon final state.

Figure 6.46: Response matrix for the invariant mass variable between the two b jets and the Z boson with the Z+2b selection.

Appendix B

Statistical and systematic uncertainties

bin	**data stat**	**bkg stat**	**eff syst**	**jec syst**	**jer syst**	**pu syst**	**bkg syst**	**ttbar stat**	**bfit stat**	**btag syst**	**unfold stat**	**unfold syst**	**lumi syst**	**total stat**	**total syst**	**total error**
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	1.7	4.2	1.4	1.7	1.1	0.3	0.1	0.0	6.5	3.0	1.5	12.4	2.6	6.7	14.0	15.5
2	1.6	3.7	1.3	1.2	1.3	0.5	0.2	0.2	3.5	3.0	1.6	5.4	2.6	3.9	8.1	9.0
3	1.7	3.8	1.3	1.6	1.1	0.6	0.2	0.3	2.9	3.0	1.7	2.2	2.6	3.3	6.6	7.4
4	1.9	4.1	1.3	3.5	0.6	0.8	0.2	0.4	2.6	3.0	2.1	8.6	2.6	3.3	11.3	11.7
5	2.2	4.9	1.3	2.7	1.1	0.5	0.2	0.5	2.4	3.0	2.8	10.7	2.6	3.3	13.1	13.5
6	2.7	5.9	1.3	1.9	1.5	0.3	0.1	0.5	2.5	3.0	3.0	10.4	2.6	3.7	13.2	13.7
7	3.3	7.2	1.4	2.4	0.9	1.0	0.1	0.6	3.0	3.0	3.7	7.4	2.6	4.5	12.0	12.8
8	3.8	8.5	1.3	3.6	1.3	1.7	0.1	0.6	3.4	3.0	4.1	9.9	2.6	5.2	14.9	15.8
9	4.2	9.2	1.4	3.5	2.0	1.7	0.1	0.5	3.1	3.0	4.8	11.2	2.6	5.2	16.4	17.2
10	4.5	9.7	1.3	4.2	1.8	0.8	0.1	0.5	2.4	3.0	5.7	16.1	2.6	5.1	20.6	21.2
11	5.2	10.9	1.3	5.5	3.0	1.0	0.2	0.4	1.5	3.0	7.2	23.1	2.7	5.5	27.6	28.1
12	6.8	14.1	1.3	6.8	6.7	1.8	0.3	0.3	1.2	3.3	9.7	34.0	2.8	6.9	39.5	40.1
13	7.9	16.4	1.2	6.2	7.7	1.8	0.2	0.3	0.7	3.0	11.5	33.9	2.5	7.9	40.9	41.6
14	8.6	17.8	1.2	6.2	8.7	1.9	0.2	0.3	0.7	3.0	12.5	35.4	2.6	8.6	43.2	44.0

Table 6.3: Systematic error contributions (in percent) for the differential cross section measurement as a function of the leading b tagged jet transverse momentum in the sample Z+1b jet for the combined channels.

bin	**data stat**	**bkg stat**	**eff syst**	**jec syst**	**jer syst**	**pu syst**	**bkg syst**	**ttbar stat**	**bfit stat**	**btag syst**	**unfold stat**	**unfold syst**	**lumi syst**	**total stat**	**total syst**	**total error**
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	6.1	14.7	1.4	15.0	8.0	6.1	0.1	0.2	14.0	3.0	4.0	19.5	2.6	15.3	30.9	34.4
2	3.3	8.1	1.4	12.1	5.3	2.8	0.1	0.1	9.1	3.0	2.4	11.4	2.6	9.7	20.0	22.2
3	2.5	6.1	1.3	9.7	2.8	0.7	0.1	0.1	6.5	3.0	2.1	16.9	2.6	6.9	21.1	22.2
4	2.0	5.0	1.3	4.9	0.3	0.7	0.1	0.1	5.1	3.0	1.9	19.7	2.6	5.5	21.4	22.1
5	1.7	4.3	1.3	1.5	1.6	1.3	0.1	0.1	4.0	3.0	1.5	13.7	2.6	4.3	15.3	15.9
6	1.7	4.2	1.3	1.1	0.8	0.8	0.1	0.2	3.5	3.0	1.7	8.4	2.6	3.9	10.5	11.2
7	1.8	4.4	1.3	1.7	0.9	0.5	0.1	0.3	3.2	3.0	2.0	4.6	2.6	3.7	8.1	8.9
8	2.0	4.6	1.4	1.8	0.9	0.3	0.2	0.3	2.8	3.0	2.0	2.9	2.6	3.5	7.5	8.2
9	2.2	5.0	1.3	1.7	0.6	0.6	0.2	0.4	2.8	3.0	2.3	6.5	2.6	3.6	9.7	10.3
10	2.3	5.2	1.3	1.8	0.5	0.9	0.3	0.4	2.4	3.0	2.9	12.2	2.6	3.4	14.4	14.8
11	2.5	5.5	1.3	1.9	0.8	1.0	0.3	0.4	2.3	3.0	2.9	18.2	2.6	3.4	19.8	20.1
12	2.9	6.3	1.3	2.5	1.0	0.4	0.2	0.4	2.5	3.0	3.1	18.8	2.6	3.8	20.7	21.0
13	3.3	7.3	1.3	2.7	1.3	0.8	0.3	0.3	2.6	3.1	3.9	16.8	2.7	4.2	19.5	20.0
14	3.6	8.1	1.3	2.6	1.2	1.6	0.3	0.3	2.5	3.1	4.6	19.0	2.7	4.4	21.8	22.2
15	3.8	8.5	1.3	2.5	0.9	2.2	0.4	0.2	2.2	3.1	5.1	25.5	2.7	4.4	27.9	28.3
16	3.9	8.7	1.2	2.2	0.5	1.4	0.4	0.2	1.8	3.0	5.4	28.0	2.6	4.3	30.2	30.6
17	4.4	9.7	1.4	1.9	0.4	1.0	0.4	0.2	1.8	3.1	6.2	28.0	2.7	4.8	30.6	31.0
18	5.1	11.4	1.4	2.0	0.8	0.6	0.5	0.1	2.2	2.9	7.6	23.8	2.5	5.5	27.9	28.5
19	6.6	15.0	1.3	2.6	0.9	2.2	0.6	0.1	3.0	3.1	9.2	22.6	2.6	7.2	29.2	30.1
20	9.1	21.1	1.3	3.3	1.3	4.2	0.7	0.1	4.0	3.1	11.4	26.6	2.7	9.9	36.5	37.8
21	11.3	26.1	1.4	3.6	1.6	5.9	0.7	0.0	4.4	3.0	13.6	27.9	2.7	12.1	41.4	43.1
22	12.8	29.8	1.4	4.0	2.0	6.8	0.7	0.0	4.9	3.1	15.0	30.7	2.7	13.7	46.2	48.2

Table 6.4: Statistical and systematic error contributions (in percent) for the differential cross section measurement as a function of the Z boson transverse momentum in the sample Z+1b jet for the combined channels.

bin	**data stat**	**bkg stat**	**eff syst**	**jec syst**	**jer syst**	**pu syst**	**bkg syst**	**ttbar stat**	**bfit stat**	**btag syst**	**unfold stat**	**unfold syst**	**lumi syst**	**total stat**	**total syst**	**total error**
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	1.7	4.4	1.4	2.7	1.1	0.3	0.1	0.1	6.1	3.0	1.3	24.2	2.6	6.3	25.1	25.9
2	2.1	5.1	1.3	1.1	1.0	0.9	0.2	0.1	3.8	3.0	2.1	17.8	2.6	4.4	19.2	19.7
3	2.2	5.2	1.3	5.0	0.9	0.2	0.4	0.1	3.4	3.0	2.3	8.1	2.6	4.1	11.8	12.5
4	2.4	5.4	1.3	5.9	2.2	0.3	0.3	0.3	3.2	3.0	2.5	2.3	2.6	4.0	9.9	10.7
5	2.6	5.6	1.4	4.7	1.5	0.4	0.3	0.4	3.0	3.0	2.8	12.5	2.6	4.0	15.4	15.9
6	2.7	5.9	1.3	4.1	1.5	0.8	0.3	0.6	2.6	3.0	3.0	22.5	2.6	3.8	24.3	24.6
7	3.1	6.4	1.3	6.0	1.6	1.4	0.3	0.7	2.6	3.0	3.5	27.5	2.6	4.0	29.5	29.8
8	3.4	7.0	1.4	7.5	1.6	2.0	0.3	0.8	2.6	3.0	4.1	30.8	2.6	4.3	33.1	33.3
9	3.6	7.4	1.4	7.9	1.6	1.8	0.3	0.9	2.6	3.0	4.5	32.6	2.6	4.5	35.0	35.3
10	4.3	8.5	1.4	6.3	3.0	0.8	0.2	1.0	2.5	3.0	5.3	31.9	2.6	5.1	34.4	34.8
11	6.2	12.5	1.2	4.4	4.5	0.8	0.2	1.1	2.7	3.0	7.5	32.4	2.6	6.9	36.3	37.0
12	8.5	17.6	1.3	3.9	7.1	2.2	0.1	1.0	2.3	3.3	10.8	37.9	2.9	8.9	44.2	45.1
13	9.6	19.9	1.2	3.5	7.2	2.8	0.1	1.1	2.4	3.1	12.1	34.2	2.7	9.9	42.5	43.6

Table 6.5: Statistical and systematic error contributions (in percent) for the differential cross section measurement as a function of the H_T variable in the sample Z+1b jet for the combined channels.

bin	**data stat**	**bkg stat**	**eff syst**	**jec syst**	**jer syst**	**pu syst**	**bkg syst**	**ttbar stat**	**bfit stat**	**btag syst**	**unfold stat**	**unfold syst**	**lumi syst**	**total stat**	**total syst**	**total error**
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	6.2	13.8	1.3	5.8	2.3	1.6	0.4	0.7	5.8	3.0	6.2	11.9	2.6	8.5	20.8	22.4
2	4.3	9.5	1.3	5.1	2.8	1.2	0.2	0.6	6.0	3.1	3.6	13.1	2.7	7.4	18.2	19.6
3	4.3	9.4	1.4	5.1	3.1	1.3	0.1	0.6	6.3	3.0	3.9	12.8	2.6	7.6	17.9	19.5
4	4.2	9.2	1.5	4.3	0.9	0.5	0.2	0.6	6.3	2.9	4.2	9.0	2.5	7.6	14.9	16.7
5	3.9	8.6	1.4	4.0	2.0	0.7	0.3	0.6	6.0	2.9	3.8	11.5	2.5	7.2	16.1	17.6
6	3.3	7.2	1.3	4.4	1.9	1.4	0.4	0.6	4.6	3.0	3.1	23.0	2.6	5.8	25.1	25.8
7	3.1	6.6	1.3	4.1	1.0	1.2	0.3	0.6	4.4	3.0	3.2	21.7	2.6	5.4	23.7	24.3
8	2.8	6.2	1.4	3.6	1.4	0.4	0.2	0.5	4.2	2.9	3.5	16.1	2.5	5.1	18.5	19.2
9	2.3	5.1	1.4	3.4	1.1	0.8	0.3	0.4	3.5	3.0	2.4	22.5	2.6	4.2	23.9	24.3
10	1.9	4.4	1.3	2.9	0.9	1.1	0.2	0.2	3.4	3.0	2.0	20.0	2.6	3.9	21.2	21.6
11	1.6	3.8	1.3	2.5	0.6	0.5	0.2	0.1	3.6	3.0	1.7	10.4	2.6	3.9	12.2	12.8
12	1.2	3.1	1.3	0.5	0.6	0.5	0.1	0.1	4.4	3.0	1.2	8.3	2.6	4.6	9.9	10.9
13	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 6.6: Statistical and systematic error contributions (in percent) for the differential cross section measurement as a function of the $\Delta\phi$ between the Z boson and the b tagged jet in the sample Z+1b jet for the combined channels.

bin	**data stat**	**bkg stat**	**eff syst**	**jec syst**	**jer syst**	**pu syst**	**bkg syst**	**ttbar stat**	**bfit stat**	**btag syst**	**unfold stat**	**unfold syst**	**lumi syst**	**total stat**	**total syst**	**total error**
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	12.2	15.9	1.4	12.7	4.9	1.3	0.3	1.5	0.0	6.0	14.5	21.4	2.6	12.3	34.0	36.1
3	7.6	9.1	1.3	6.4	1.8	1.1	0.7	2.4	0.0	6.0	8.4	1.5	2.6	7.9	15.7	17.6
4	6.8	7.7	1.3	6.5	0.9	0.9	0.7	2.8	0.0	6.0	7.7	9.6	2.6	7.4	17.2	18.8
5	6.9	7.8	1.3	7.3	1.7	1.6	0.7	3.5	0.0	6.0	7.6	1.8	2.6	7.7	15.0	16.9
6	6.6	7.8	1.3	6.6	2.1	1.8	0.5	3.3	0.0	6.1	7.3	8.0	2.6	7.4	16.6	18.1
7	7.0	8.3	1.3	6.9	1.6	1.8	0.4	2.7	0.0	6.0	6.5	4.7	2.6	7.5	15.2	17.0
8	8.5	10.5	1.3	8.7	2.8	2.3	0.3	1.5	0.0	6.0	11.4	3.7	2.6	8.6	19.7	21.5
9	12.9	15.6	1.2	13.3	5.3	2.2	0.1	1.0	0.0	6.0	20.7	18.3	2.6	13.0	35.5	37.8
10	17.6	20.7	1.2	18.0	6.8	2.3	0.1	0.7	0.0	5.8	29.0	30.3	2.5	17.6	51.1	54.0
11	20.6	25.7	1.2	24.0	7.5	2.2	0.0	0.6	0.0	6.8	36.3	44.0	3.0	20.6	67.9	70.9

Table 6.7: Statistical and systematic error contributions (in percent) for the differential cross section measurement as a function of the ΔR between the two b tagged jets in the sample Z+2b jet for the combined channels.

bin	**data stat**	**bkg stat**	**eff syst**	**jec syst**	**jer syst**	**pu syst**	**bkg syst**	**ttbar stat**	**bfit stat**	**btag syst**	**unfold stat**	**unfold syst**	**lumi syst**	**total stat**	**total syst**	**total error**
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	9.7	11.1	1.3	11.2	3.4	3.1	0.1	1.6	0.0	6.0	11.6	22.3	2.6	9.9	30.8	32.4
2	8.2	9.5	1.4	10.3	2.4	2.3	0.1	1.6	0.0	6.1	10.1	21.0	2.6	8.3	28.2	29.4
3	6.9	8.1	1.4	9.1	1.4	1.0	0.2	1.7	0.0	6.2	8.6	17.6	2.7	7.1	24.2	25.2
4	6.4	7.5	1.3	6.7	1.2	0.5	0.3	1.9	0.0	6.0	7.6	13.7	2.6	6.7	19.8	20.9
5	6.6	7.7	1.4	6.1	1.6	1.1	0.5	2.3	0.0	6.1	7.3	12.0	2.6	7.0	18.6	19.8
6	6.9	7.9	1.3	5.8	2.0	1.7	0.6	2.7	0.0	6.0	8.1	9.9	2.6	7.4	17.7	19.2
7	6.7	7.7	1.4	6.2	1.7	2.1	0.7	3.0	0.0	6.0	7.2	10.8	2.6	7.3	17.8	19.3
8	6.7	7.6	1.3	7.4	1.5	2.0	0.7	3.4	0.0	6.0	6.9	7.5	2.6	7.5	16.4	18.0
9	6.5	7.5	1.4	8.1	1.1	1.6	0.7	3.5	0.0	6.0	6.2	3.7	2.6	7.4	15.0	16.7
10	6.2	7.3	1.4	8.1	0.9	2.1	0.7	3.4	0.0	6.0	5.8	3.8	2.6	7.0	14.7	16.3
11	6.7	8.3	1.4	6.6	1.3	2.4	0.7	3.3	0.0	5.9	7.2	14.6	2.6	7.4	20.7	22.0
12	8.8	11.0	1.3	5.3	2.2	2.1	0.6	2.9	0.0	6.0	9.2	20.2	2.6	9.3	26.4	28.0
13	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 6.8: Statistical and systematic error contributions (in percent) for the differential cross section measurement as a function of the $\Delta\phi$ between the two b tagged jets in the sample Z+2b jet for the combined channels.

bin	**data stat**	**bkg stat**	**eff syst**	**jec syst**	**jer syst**	**pu syst**	**bkg syst**	**ttbar stat**	**bfit stat**	**btag syst**	**unfold stat**	**unfold syst**	**lumi syst**	**total stat**	**total syst**	**total error**
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	30.7	34.2	1.2	5.5	1.6	1.5	1.2	3.9	0.0	6.5	27.1	11.1	2.8	30.9	46.0	55.4
2	19.0	21.0	1.4	7.0	1.5	1.5	0.6	3.6	0.0	6.0	16.8	20.8	2.6	19.4	35.4	40.4
3	15.8	17.7	1.4	9.8	2.7	3.3	0.4	4.5	0.0	6.0	14.2	23.0	2.6	16.5	34.7	38.4
4	13.9	16.0	1.4	7.4	1.5	3.4	0.2	5.1	0.0	5.9	13.5	18.8	2.5	14.8	30.0	33.5
5	11.5	13.0	1.3	7.2	1.9	0.9	0.3	5.0	0.0	6.0	9.2	1.0	2.6	12.5	18.9	22.7
6	10.8	12.3	1.2	7.7	2.6	1.8	0.5	4.7	0.0	6.0	8.7	7.4	2.6	11.8	19.8	23.1
7	12.0	14.3	1.3	7.6	2.1	2.7	0.9	5.4	0.0	6.3	10.5	33.7	2.7	13.2	39.7	41.8
8	9.0	10.5	1.3	4.1	4.3	2.3	0.6	3.5	0.0	6.0	8.0	13.3	2.6	9.6	20.9	23.0
9	7.7	9.3	1.4	6.2	1.7	1.1	0.7	2.5	0.0	5.8	8.7	21.7	2.5	8.1	26.8	28.0
10	7.1	8.6	1.3	8.7	1.9	1.2	0.5	1.4	0.0	6.0	8.3	6.5	2.6	7.3	17.6	19.1
11	8.0	10.0	1.4	9.0	1.6	1.0	0.5	1.3	0.0	6.2	10.4	6.9	2.7	8.1	19.7	21.3
12	11.2	14.4	1.4	12.7	6.2	1.9	0.6	0.5	0.0	6.5	12.8	20.4	2.8	11.2	32.3	34.2
13	12.6	16.3	1.3	10.3	8.5	3.7	0.5	0.3	0.0	6.5	16.8	30.6	2.8	12.6	41.6	43.4
14	14.5	18.8	1.3	7.8	7.6	3.6	0.3	0.1	0.0	6.2	23.0	16.2	2.7	14.5	36.4	39.2
15	20.7	27.0	1.4	10.6	7.6	4.1	0.1	0.3	0.0	6.0	34.3	13.3	2.6	20.7	48.1	52.4
16	24.7	32.2	1.6	11.5	9.2	4.1	0.1	0.3	0.0	5.6	41.4	30.6	2.4	24.7	62.9	67.6

Table 6.9: Statistical and systematic error contributions (in percent) for the differential cross section measurement as a function of the ΔR between the Z boson and the closest b tagged jet in the sample Z+2b jet for the combined channels.

bin	**data stat**	**bkg stat**	**eff syst**	**jec syst**	**jer syst**	**pu syst**	**bkg syst**	**ttbar stat**	**bfit stat**	**btag syst**	**unfold stat**	**unfold syst**	**lumi syst**	**total stat**	**total syst**	**total error**
0	72.5	81.1	1.6	18.1	1.4	33.0	0.4	16.3	0.0	18.3	48.8	109.9	7.9	74.3	151.2	168.5
1	47.8	51.8	1.3	7.7	0.9	9.2	0.2	10.5	0.0	6.9	43.4	42.7	3.0	48.9	81.2	94.8
2	38.2	41.1	1.4	7.9	0.5	6.9	0.5	9.7	0.0	8.1	31.3	52.9	3.5	39.4	75.2	84.9
3	22.1	24.1	1.3	7.7	0.8	0.5	0.9	7.4	0.0	7.2	20.9	46.0	3.1	23.3	57.1	61.6
4	13.6	14.7	1.4	9.0	1.5	2.6	0.9	5.5	0.0	6.3	14.0	11.4	2.7	14.6	26.1	29.9
5	10.7	12.0	1.3	7.2	0.3	0.5	0.7	3.9	0.0	6.1	12.5	10.7	2.6	11.4	22.6	25.3
6	8.8	10.4	1.3	8.1	1.0	1.6	0.8	3.2	0.0	6.1	10.1	11.1	2.6	9.3	21.2	23.1
7	9.2	10.8	1.4	5.3	5.0	3.6	0.4	3.0	0.0	6.0	9.8	7.2	2.6	9.7	19.4	21.7
8	7.8	9.3	1.4	6.9	2.1	0.6	0.3	1.6	0.0	5.9	9.0	15.9	2.6	8.0	22.7	24.1
9	12.8	16.9	1.3	9.8	10.0	5.2	0.7	1.5	0.0	7.5	15.3	39.2	3.3	12.9	48.5	50.2
10	13.1	16.0	1.3	9.1	10.0	3.0	0.9	1.3	0.0	7.5	12.9	51.9	3.3	13.1	58.1	59.6
11	12.1	14.4	1.3	7.2	7.6	1.0	0.4	0.9	0.0	6.2	17.3	15.9	2.7	12.2	30.3	32.7
12	13.3	15.8	1.2	10.6	6.3	3.8	0.1	0.8	0.0	6.3	18.2	20.6	2.7	13.3	35.0	37.4
13	15.1	18.9	1.3	13.2	8.4	1.4	0.2	0.2	0.0	6.9	23.9	53.6	3.0	15.1	64.0	65.8
14	27.5	46.7	1.6	23.1	7.9	8.1	0.3	0.8	0.0	8.4	34.7	132.1	3.6	27.5	146.9	149.5

Table 6.10: Statistical and systematic error contributions (in percent) for the differential cross section measurement as a function of the ΔR between the Z boson and the further b tagged jet in the sample Z+2b jet for the combined channels.

bin	**data stat**	**bkg stat**	**eff syst**	**jec syst**	**jer syst**	**pu syst**	**bkg syst**	**ttbar stat**	**bfit stat**	**btag syst**	**unfold stat**	**unfold syst**	**lumi syst**	**total stat**	**total syst**	**total error**
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	4.9	5.9	1.3	6.5	1.1	1.0	0.6	2.1	0.0	6.0	5.6	9.0	2.6	5.3	15.4	16.3
2	4.3	5.1	1.3	6.9	1.1	0.7	0.6	2.5	0.0	6.0	4.4	3.3	2.6	5.0	12.3	13.3
3	5.0	5.9	1.3	7.0	1.6	0.8	0.5	3.1	0.0	6.0	4.8	2.4	2.6	5.9	12.7	14.0
4	6.2	7.4	1.3	6.8	2.2	0.7	0.5	3.6	0.0	6.0	6.0	5.4	2.6	7.2	14.7	16.4
5	7.6	8.9	1.4	6.7	2.2	1.3	0.4	3.8	0.0	6.0	7.3	10.9	2.6	8.5	18.7	20.5
6	9.4	11.3	1.3	7.2	2.3	2.1	0.3	4.0	0.0	5.9	9.7	14.5	2.6	10.2	23.2	25.3
7	11.8	14.5	1.3	8.0	2.3	2.9	0.3	4.2	0.0	6.0	13.2	19.7	2.6	12.5	29.9	32.4
8	15.0	18.7	1.4	8.8	2.2	3.6	0.4	4.2	0.0	6.3	17.2	26.9	2.7	15.6	38.9	41.9
9	18.2	22.7	1.4	8.9	2.3	4.2	0.4	4.2	0.0	6.3	21.4	33.9	2.7	18.7	47.7	51.2
10	25.3	32.8	1.3	16.1	3.0	6.5	0.6	5.4	0.0	10.6	28.6	47.0	4.6	25.8	67.4	72.2
11	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table 6.11: Statistical and systematic error contributions (in percent) for the differential cross section measurement as a function of the asymmetry between the Z boson direction and the b tagged jet direction in the sample Z+2b jet for the combined channels.

bin	**data stat**	**bkg stat**	**eff syst**	**jec syst**	**jer syst**	**pu syst**	**bkg syst**	**ttbar stat**	**bfit stat**	**btag syst**	**unfold stat**	**unfold syst**	**lumi syst**	**total stat**	**total syst**	**total error**
0	21.8	27.5	1.3	17.4	9.4	4.0	0.0	0.0	0.0	6.7	21.9	24.1	2.7	21.8	47.7	52.5
1	13.8	17.7	1.4	12.4	5.1	1.7	0.1	0.7	0.0	6.0	13.6	14.3	2.6	13.8	30.5	33.5
2	7.9	9.5	1.2	8.1	3.4	1.6	1.4	1.4	0.0	6.2	7.8	15.3	2.7	8.0	22.7	24.0
3	7.4	8.6	1.3	6.6	3.2	1.2	1.3	2.1	0.0	6.0	8.4	4.2	2.6	7.7	16.2	17.9
4	8.3	9.9	1.3	9.8	2.9	2.3	0.3	3.0	0.0	6.0	9.2	11.3	2.6	8.9	21.6	23.3
5	10.8	12.7	1.3	6.6	7.4	2.8	0.1	4.6	0.0	6.0	14.1	31.6	2.6	11.8	38.9	40.6
6	9.9	11.0	1.3	7.2	2.9	2.1	0.1	4.2	0.0	6.0	10.4	5.4	2.6	10.8	19.2	22.0
7	10.1	11.1	1.2	6.9	2.3	1.4	0.1	3.5	0.0	6.0	10.7	15.2	2.6	10.7	23.8	26.1
8	11.9	13.4	1.3	8.5	3.6	3.1	0.1	3.4	0.0	6.0	11.9	6.6	2.6	12.4	22.5	25.6
9	15.4	18.1	1.2	14.2	8.3	6.0	0.2	3.9	0.0	6.0	15.8	23.6	2.7	15.8	38.5	41.7
10	22.5	25.2	1.1	24.5	18.9	7.5	0.6	5.4	0.0	6.9	25.8	52.8	3.1	23.1	71.8	75.4
11	25.8	27.7	1.0	10.2	6.9	3.6	0.0	4.9	0.0	5.8	24.8	31.3	2.6	26.3	50.7	57.1
12	31.3	34.6	1.2	26.7	2.5	3.5	0.0	4.1	0.0	7.0	28.6	51.8	3.0	31.5	74.1	80.5
13	34.9	39.4	1.4	40.8	2.6	3.2	0.0	3.9	0.0	9.1	33.6	69.9	3.9	35.2	96.6	102.8
14	30.6	33.3	1.2	30.9	2.7	2.7	0.4	3.7	0.0	6.1	30.4	46.0	2.5	30.9	71.9	78.2

Table 6.12: Statistical and systematic error contributions (in percent) for the differential cross section measurement as a function of the invariant mass of the two b tagged jets in the sample Z+2b jet for the combined channels.

bin	**data stat**	**bkg stat**	**eff syst**	**jec syst**	**jer syst**	**pu syst**	**bkg syst**	**ttbar stat**	**bfit stat**	**btag syst**	**unfold stat**	**unfold syst**	**lumi syst**	**total stat**	**total syst**	**total error**
0	25.4	30.4	1.2	20.3	10.3	5.9	1.3	0.8	0.0	5.6	29.3	38.1	2.4	25.4	61.8	66.8
1	17.1	20.6	1.2	16.5	8.6	3.8	1.0	1.6	0.0	5.9	19.8	17.5	2.5	17.2	39.1	42.7
2	9.2	10.5	1.3	11.7	7.8	0.3	0.6	2.8	0.0	6.0	9.6	16.4	2.6	9.6	26.7	28.4
3	9.2	10.3	1.4	9.8	2.5	1.6	0.5	3.5	0.0	5.9	10.0	12.2	2.6	9.8	22.4	24.5
4	9.3	10.5	1.3	13.2	3.6	1.4	0.5	3.8	0.0	6.3	10.4	19.0	2.7	10.0	28.6	30.3
5	10.1	11.7	1.3	7.5	4.7	1.3	0.5	4.1	0.0	6.3	10.8	18.5	2.7	10.9	26.9	29.0
6	12.0	14.2	1.5	3.4	3.2	2.1	0.5	3.8	0.0	5.8	12.2	16.8	2.5	12.6	26.5	29.3
7	13.6	16.7	1.5	8.2	3.3	1.1	0.4	3.1	0.0	5.8	14.0	41.1	2.5	13.9	47.8	49.8
8	17.2	21.2	1.4	14.7	4.1	1.4	0.2	2.9	0.0	5.7	14.8	55.6	2.5	17.4	63.6	65.9
9	17.0	21.3	1.3	15.6	9.3	5.1	0.2	2.6	0.0	6.4	14.8	39.2	2.7	17.2	51.2	54.0
10	14.0	17.5	1.3	10.1	9.3	4.2	0.3	1.9	0.0	6.8	16.5	12.1	3.0	14.1	31.4	34.4
11	17.2	20.3	1.3	5.3	7.2	1.7	0.7	2.1	0.0	6.1	19.9	9.5	2.6	17.3	32.0	36.4
12	25.7	30.6	1.4	14.0	11.1	1.6	1.4	2.4	0.0	8.0	30.4	36.0	3.5	25.8	59.7	65.0
13	28.3	33.6	1.3	14.5	10.7	1.7	1.2	2.3	0.0	6.7	33.4	33.3	2.9	28.4	61.1	67.4

Table 6.13: Statistical and systematic error contributions (in percent) for the differential cross section measurement as a function of the invariant mass of the system with the Z boson the two b tagged jets in the sample Z+2b jet for the combined channels.

Bibliography

[1] S. L. Glashow. Partial Symmetries of Weak Interactions. *Nucl. Phys.*, 22:579–588, 1961.

[2] Abdus Salam. Weak and Electromagnetic Interactions. *Conf.Proc.*, C680519:367–577, 1968.

[3] S. Weinberg. A Model of Leptons. *Phys.Rev.Lett..*, 19:1264–1266, 1967.

[4] The ALEPH, DELPHI, L3, OPAL, SLD Collaborations, the LEP Electroweak Working Group, the SLD Electroweak, and Heavy Flavour Groups. *Phys. Rep.*, 427:257, 2006.

[5] The ALEPH, DELPHI, L3, OPAL Collaborations, and the LEP Electroweak Working Group. Electroweak Measurements in Electron-Positron Collisions at W-Boson-Pair Energies at LEP. *Phys. Rep.*, 532:119, 2013.

[6] J. Beringer et al. Review of Particle Physics. *Phys. Rev. D*, 86:010001, 2012.

[7] P. W. Higgs. Broken symmetries, massless particles and gauge fields. *Phys.Rev.Lett.*, 12(2):132–133, 1964.

[8] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Phys.Rev.Lett.*, 13:321–323, 1964.

[9] ATLAS Collaboration. Observation of a new particle in the search for the standard model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1–29, 2012.

[10] CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys.Lett.B716*, page 30–61, 2012.

[11] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson. Experimental Test of Parity Conservation in Beta Decay. *Phys.Rev.*, 105:1413–1414, 1957.

[12] E. Fermi. Trends to a Theory of beta Radiation. *Nuovo Cim.*, 11:1–19, 1934.

[13] The ALEPH, DELPHI, L3, OPAL, SLD Collaborations, the LEP Electroweak Working Group, the SLD Electroweak, and Heavy Flavour Groups. Precision Electroweak Measurements on the Z Resonance. *Phys. Rept.*, 427:257, 2006.

[14] Makoto Kobayashi and Toshihide Maskawa. CP Violation in the Renormalizable Theory of Weak Interaction. *Prog.Theor.Phys.*, 49:652–657, 1973.

[15] N. Cabibbo. Unitary Symmetry and Leptonic Decays. *Phys. Rev. Lett.*, 10:531, 1963.

[16] R. Barate et al. Search for the standard model Higgs boson at LEP. *Phys.Lett.*, B565:61–75, 2003.

[17] T. Aaltonen et al. Higgs boson studies at the Tevatron. *Phys. Rev. D*, 88:052014, 2013.

[18] Vardan Khachatryan et al. Observation of the diphoton decay of the Higgs boson and measurement of its properties.

[19] Serguei Chatrchyan et al. Measurement of the properties of a Higgs boson in the four-lepton final state.

[20] Serguei Chatrchyan et al. Search for the standard model Higgs boson produced in association with a W or a Z boson and decaying to bottom quarks.

[21] Serguei Chatrchyan et al. Measurement of Higgs boson production and properties in the WW decay channel with leptonic final states.

[22] Serguei Chatrchyan et al. Evidence for the 125 GeV Higgs boson decaying to a pair of τ leptons.

[23] M. Gell-Mann. A Schematic Model of Baryons and Mesons. *Phys.Lett.*, 8:214–215, 1964.

[24] M. Y. Han and Y. Nambu. Three Triplet Model with Double SU(3) Symmetry. *Phys.Rev.*, 139:B1006–B1010, 1965.

[25] H. Fritzsch, M. Gell-Mann, and H. Leutwyler. Advantages of the Color Octet Gluon Picture. *Phys.Lett.*, B47:365–368, 1973.

[26] D. J. Gross and F. Wilczek. Asymptotically Free Gauge Theories. *Phys. Rev.*, D8:3633–3652, 1973.

[27] S. Weinberg. Nonabelian Gauge Theories of the Strong Interactions. *Phys.Rev.Lett.*, 31:494–497, 1973.

[28] Siegfried Bethke. Experimental Tests of Asymptotic Freedom. *Prog. Part. Nucl. Phys.*, 58:351–386, 2007.

[29] D.J. Gross and Frank Wilczek. Ultraviolet Behavior of Nonabelian Gauge Theories. *Phys.Rev.Lett.*, 30:1343–1346, 1973.

[30] J. Smit. Introduction to quantum field on a lattice, A robust mate. *Cambridege Lect. Notes Phys*, 15:1, 2002.

[31] R. K. Ellis H. Georgi M. Machacek H. D. Politzer and G. G. Ross. Factorization and the Parton Model in QCD. *Phys.Lett.*, B78:281, 1978.

[32] M.L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A.D. Polosa. ALPGEN, a generator for hard multipar- ton processes in hadronic collisions. *JHEP*, 0307:001, 2003.

[33] F. Caravaglios, M.L. Mangano, M. Moretti, and R. Pittau. A New approach to multijet calculations in hadron collisions. *Nucl.Phys.*, B539:215–232, 1999.

[34] R.D. Ball, V. Bertone, S. Carrazza, C.S. Deans, and L. Del Debbio et al. Parton distributions with LHC data. *Nucl.Phys.*, B867:244–289, 2013.

[35] CTEQ Collaboration. The Coordinated Theoretical-Experimental Project on QCD.

[36] A. D. Martin, R. G. Roberts, W. J. Stirling, and R.S. Thorne. Physical gluons and high E_T jets. *Phys.Lett. B*, 604:61, 2004.

[37] H1 and ZEUS Collaboration. Inclusive diffraction at HERA. *Nucl.Phys.Proc.Suppl..*, 117:403–407, 2003.

[38] CMS Collaboration. Measurement of associated W + charm production in pp collisions at 7 TeV. *JHEP*, 02:013, 2014.

[39] ATLAS Collaboration. Measurement of the production of a W boson in association with a charm quark in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *JHEP*, 1405:068, 2014.

[40] G. Altarelli and G. Parisi. Asymptotic Freedom in Parton Language. *Phys.Lett.*, B126:298, 1977.

[41] Y. L. Dokshitzer. Calculation of the Structure Functions for Deep Inelastic Scattering and e^+e^- Annihilation by Perturbation Theory in Quantum Chromodynamics. *Sov.Phys.JETP*, 46:641–653, 1977.

[42] V.N. Gribov and L.N. Lipatov. Deep inelastic $e-p$ scattering in perturbation theory. *Sov.J.Nucl.Phys.*, 15:438–450, 1972.

[43] PDF constraints and extraction of the strong coupling constant from the inclusive jet cross section at 7 TeV. Technical Report arXiv:1410.6765. CMS-SMP-12-028. CERN-PH-EP-2014-238, CERN, Geneva, 2013. Comments: Submitted to Eur. Phys. J. C.

[44] Super-Kamiokande Collaboration. Evidence for oscillation of atmospheric neutrinos. *Phys.Rev.Lett.*, 81:1562–1567, 1998.

[45] SNO Collaboration. Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory. *Phys.Rev.Lett.*, 89, 2002.

[46] KamLAND Collaboration. First results from KamLAND: Evidence for reactor anti-neutrino disappearance. *Phys.Rev.Lett.*, 90, 2003.

[47] OPERA Collaboration. Search for $\nu_\mu \to \nu_e$ oscillations with the OPERA experiment in the CNGS beam. *JHEP*, 1307:004, 2013.

[48] OPERA Collaboration. New results on $\nu_\mu \to \nu_\tau$ appearance with the OPERA experiment in the CNGS beam. *JHEP*, 1311:036, 2013.

[49] UA2 Collaboration. Evidence for $Z0$ to e^+e^- at the CERN anti-p p Collider. *Phys.Lett.*, B129:130–140, 1983.

[50] CDF Collaboration. Measurements of inclusive W and Z cross sections in p anti-p collisions at s**(1/2) = 1.96-TeV. *J.Phys.*, G34:2457–2544, 2007.

[51] D0 Collaboration. Measurement of W and Z boson production cross sections. *Phys.Rev.*, D60:052003, 1999.

[52] D0 Collaboration. Measurement of the shape of the boson transverse momentum distribution in $p\bar{p} \to Z \setminus \gamma^* e^+e^- + X$ events produced at $\sqrt{s}$=1.96 TeV. *Phys.Rev.Lett.*, 100:102002, 2008.

[53] CDF Collaboration. Measurement of inclusive jet cross-sections in $Z/\gamma^* \to e^+e^-$ + jets production in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. *Phys.Rev.Lett.*, 100:102001, 2008.

[54] T. Aaltonen et al. Measurement of cross sections for b jet production in events with a Z boson in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. *Phys.Rev.*, D79:052008, 2009.

[55] D0 Collaboration. Measurement of the ratio of differential cross sections $\sigma(p\bar{p} \to Z + bjet)/\sigma(p\bar{p} \to Z + jet)$ in $p\bar{p}$ collision at $\sqrt{s} = 1.96$ TeV. *Phys.Rev.*, D87(9):092010, 2013.

[56] Matthias Schott and Monica Dunford. Review of single vector boson production in pp collisions at $\sqrt(7) =$ TeV. *Eur.Phys.J.*, C74:2916, 2014.

[57] F. Febres Cordero, L. Reina, and D. Wackeroth. W- and Z-boson production with a massive bottom-quark pair at the Large Hadron Collider. *Phys. Rev.*, D80:034015, 2009.

[58] J. Campbell, R. K. Ellis, F. Maltoni, and S. Willenbrock. Associated Production of a Z Boson and a Single Heavy-Quark Jet. *Phys. Rev. D*, 69:074021, 2004.

[59] J. M. Campbell and R. K. Ellis. Radiative corrections to Z b anti-b production. *Phys. Rev.*, D62:114012, 2000.

[60] R. Frederix et. al. W and Z/$gamma^*$ boson production in association with a bottom-antibottom pair. *JHEP*, 09:061, 2011.

[61] Nicolas Greiner, Alberto Guffanti, Thomas Reiter, and Jurgen Reuter. NLO QCD corrections to the production of two bottom-antibottom pairs at the LHC. *Phys. Rev. Lett.*, 107:102002, 2011.

[62] F. Maltoni, G. Ridolfi, and M. Ubiali. b-initiated processes at the LHC: a reappraisal. *JHEP*, 1207:022, 2012.

[63] John Campbell and R. K. Ellis. Next-to-leading order corrections to W + 2jet and Z+ 2jet production at hadron colliders. *Phys. Rev. D*, 65:113007, 2002.

[64] CMS Collaboration. Measurement of the production cross sections for a Z boson and one or more b jets in pp collisions at $\sqrt{s} = 7$ TeV. *arXiv hep-ex/*, 1402.1521, 2014.

[65] ATLAS Collaboration. Measurement of differential production cross-sections for a Z boson in association with b-jets in 7 TeV proton-proton collisions with the ATLAS detector. *Journal of High Energy Physics*, 10(10):141–49, 2014.

[66] J. M. Campbell and R. K. Ellis. *Phys. Rev. D*, 62:114012, 2000.

[67] R. K. Ellis J. M. Campbell and D. L. Rainwater. Next-to-leading order QCD predictions for $W + 2$ jet and $Z + 2$ jet production at the CERN LHC. *Phys. Rev. D*, 68:094021, 2003.

[68] F.Febres Cordero, L.Reina, and D.Wackeroth. NLO QCD corrections to W boson production with a massive b-quark jet pair at the Fermilab Tevatron $p\overline{p}$ collider. *Phys. Rev. D*, 74:034007, 2006.

[69] F.Febres Cordero, L.Reina, and D.Wackeroth. NLO QCD corrections to $Zb\bar{b}$ production with massive bottom quarks at the Fermilab Tevatron. *Phys. Rev. D*, 78:074014, 2008.

[70] F.Febres Cordero, L.Reina, and D.Wackeroth. W and Z boson production with a massive bottom-quark pair at the Large Hadron Collider. *Phys. Rev. D*, 80:034015, 2009.

[71] S. Hoche J. Alwall and F. Krauss et al. Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions. *Eur. Phys. J.*, C53:473, 2008.

[72] S. Catani, F. Krauss, and R. Kuhn et al. QCD matrix elements + parton showers. *JHEP*, 0111:063, 2001.

[73] T. Sjostrand and M. van Zijl. A multiple-interaction model for the event structure in hadron collisions. *Phys. Rev.*, D36:2019, 1987.

[74] A. Buckley et al. General-purpose event generators for LHC physics. *Phys.Rept.*, 504:145–233, 2011.

[75] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 0605:026, 2006.

[76] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand. Parton Fragmentation and String Dynamics. *Phys.Rept.*, 97:31–145, 1983.

[77] Paolo Bartalini and Livio Fano. Multiple partonic interactions at the LHC. Proceedings, 1st International Workshop, MPI 08, Perugia, Italy, October 27-31, 2008. 2010.

[78] R. Field. Physics at the Tevatron. *Acta Phys.Polon*, B39:2611–2672, 2008.

[79] R. Field. Early LHC underlying event data—Findings and surprises. *Hadron Collider Physics Symposium*, 2010.

[80] J. Pumplin et al. New generation of parton distributions with uncertainties from global QCD analysis. *JHEP*, 07:012, 2002.

[81] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer. MadGraph 5: Going Beyond. *JHEP*, 1106:128, 2011.

[82] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re. A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX. *JHEP*, 1006:043, 2010.

[83] P. Nason. A new method for combining NLO QCD with shower Monte Carlo algorithms. *JHEP*, 11, 2004.

[84] Guido Altarelli and Martin W. Grunewald. Precision electroweak tests of the standard model. *Phys.Rept.*, 403-404:189–201, 2004.

[85] ALEPH, DELPHI, L3, OPAL, SLD, LEP Electroweak Working Group, SLD Electroweak Group, and SLD Heavy Flavour Group Collaborations. *Phys. Rept.*, 427:257, 2006.

[86] M. Beccaria, G. Marcorini, G. Panizzo, and C. Verzegnassi. New Physics signals from measurable polarization asymmetries at LHC. *Phys.Lett.*, B730:149–154, 2014.

[87] S. Heinemeyer et al. Handbook of LHC Higgs cross sections: 3. Higgs properties. *CERN Report CERN-2013-004*, 2013.

[88] W. S. Hou. Source of CP Violation for the Baryon Asymmetry of the Universe. *Chin. J. Phys.*, 47:134, 2009.

[89] Georges Aad et al. Search for down-type fourth generation quarks with the ATLAS detector in events with one lepton and hadronically decaying W bosons. *Phys.Rev.Lett.*, 109:032001, 2012.

[90] Lyndon Evans and Philip Bryant. LHC Machine. *JINST*, 3:S08001, 2008.

[91] M. Benedikt, P. Collier, V. Mertens, J. Poole, and K. Schindl. LHC Design Report, Vol. 3: The LHC Injector Chain. *CERN-2004-003-V-3*, 2004.

[92] K. Aamodt et al. The ALICE experiment at the CERN LHC. *JINST*, 3:S08002, 2008.

[93] G. Aad et al. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.

[94] S. Chatrchyan et al. The CMS experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

[95] Jr. Alves A. Augusto et al. The LHCb Detector at the LHC. *JINST*, 3:S08005, 2008.

[96] CMS Collaboration. CMS tracker design report: Technical Design Report. *Technical Design Report CMS*, 1997.

[97] CMS Collaboration. The CMS electromagnetic calorimeter project: Technical Design Report. . *Technical Design Report CMS*, 1997.

[98] P. Adzic et al. Radiation hardness qualification of PbWO4 scintillation crystals for the CMS Electromagnetic Calorimeter. *JINST*, 5:P03010, 2010.

[99] P. Adzic et al. Energy resolution of the barrel of the CMS Electromagnetic Calorimeter. *JINST*, 2:P04004, 2007.

[100] CMS Collaboration. Intercalibration of the barrel electromagnetic calorimeter of the CMS experiment at start-up. *JINST*, 3:P10007, 2008.

[101] CMS Collaboration. The CMS hadron calorimeter project: Technical Design Report. *Technical Design Report CMS*, 1997.

[102] CMS Collaboration. The CMS muon project: Technical Design Report. *Technical Design Report CMS*, 1997.

[103] CMS Collaboration. CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems. *Technical Design Report CMS*, 1997.

[104] Sergio Cittolin, Attila Racz, and Paris Sphicas. MS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project. *Technical Design Report CMS*, 1997.

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

[106] CMS Collaboration. Particle-flow commissioning with muons and electrons from J/ψ and W events at 7 TeV. *CMS-PAS-PFT-10-003*, 2010.

[107] W. Adam, R. Frühwirth, A. Strandlie, and T. Todorov. Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC. *Journal of Physics G: Nuclear and Particle Physics*, 31:9, 2005.

[108] CMS Collaboration. Identification of b-quark jets with the CMS experiment. *JINST 8*, page P04013, 2013.

[109] P. Billoir. Progressive track recognition with a Kalman like fitting procedure. *Comput. Phys. Commun.*, 57:390, 1989.

[110] P. Billoir and S. Qian. Simultaneous pattern recognition and track fitting by the Kalman filtering method. *Nucl. Instrum. Meth. A*, 294:219, 1990.

[111] CMS Collaboration. Track Reconstruction in the CMS tracker. *CMS-PAS*, TRK-09-001, 2009.

[112] E. Meschi et al. Electron Reconstruction in the CMS Electromagnetic Calorimeter. *CMS Note*, 034, 2001.

[113] M. Anderson, A. Askew, A.F. Barfuss, D. Evans, F. Ferri, K. Kaadze, Y. Maravin, P. Meridiani, and C. Seez. Review of clustering algorithms and energy corrections in ECAL. *IN-2010*, 008, 2010.

[114] CMS Collaboration. Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at $\sqrt{s} = 7$ TeV. *JINST 8*, P09009, 2013.

[115] P. Adzic et al. Reconstruction of the signal amplitude of the CMS electromagnetic calorimeter. *Eur. Phys. J. C*, page 46, 2006.

[116] W. Adam et al. Reconstruction of electrons with the Gaussian-Sum Filter in the CMS tracker at the LCH. *Journal of Physics G: Nuclear and Particle Physics*, 31(9), 2005.

[117] CMS Collaboration. Performance of CMS muon reconstruction in cosmic-ray events. *JINST 5*, T03022, 2010.

[118] CMS Collaboration. CMS Physics Technical Design Report, volume I: Detector performance and software. *TDR CERN-LHCC-2006-001*, CMS-TDR-008-1, 2006.

[119] S. Bolognesi, M.A. Borgia, R. Castello, C. Mariotti, M. De Mattia, and T. Dorigo. Calibration of track momentum using dimuon resonances in CMS. *CMS-AN-2010/059*, 2010.

[120] CMS Collaboration. Jet Plus Tracks Algorithm for Calorimeter Jet Energy Corrections in CMS. *CMS PAS*, JME-09-002, 2009.

[121] Gavin P. Salam and Gregory Soyez. A practical seedless infrared-safe cone jet algorithm. *JHEP*, 05:086, 2007.

[122] S. Catani, Yu.L. Dokshitzer, M.H. Seymour, and B.R. Webber. Longitudinally-invariant Kt-clustering algorithms for hadron-hadron collisions. *Nuclear Physics B*, 406:187–224, 1993.

[123] G. Soyez. The SISCone and anti-kt jet algorithm. *DIS2008*, 2008.

[124] M. Cacciari, G. Salam, and G. Soyez. The anti-kt jet clustering algorithm. *JHEP*, 0804:063, 2008.

[125] Dokshitzer and Yuri L. et al. Better jet clustering algorithms. *JHEP*, 9708:001, 1997.

[126] M. Cacciari and G. P. Salam. Pileup subtraction using jet areas. *Phys. Lett. B*, 659:119, 2007.

[127] M. Cacciari, G. P. Salam, and G. Soyez. The Catchment Area of Jets. *JHEP*, 04:005, 2007.

[128] UA2 Collaboration. Measurement of Production and Properties of Jets at the CERN anti-p p Collider. *Z. Phys. C*, 20:117, 1983.

[129] The CMS Collaboration. Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS. *CMS JME-10-011*, 2011.

[130] W.Waltenberger, R.Fruwirth, and P.Vanlaer. Adaptive Vertex Fitting. *J. Phys. G: Nucl. Part. Phys.*, 34:N343, 2007.

[131] CMS Collaboration. Performance of b tagging at $\sqrt(s)$ =8 TeV in multijet, $t\bar{t}$ and boosted topology events. (CMS-PAS-BTV-13-001), 2013.

[132] CMS Collaboration. Public CMS Data Quality Information.

[133] S. Agostinelli et al. Geant 4 simulation toolkit. *Nucl. Inst. Meth. A*, 506:250–303, 2003.

[134] Christian Weiser. A combined secondary vertex based b-tagging algorithm in CMS. *Technical Report CMS-NOTE*, 014, 2006.

[135] Andreas Hocker and Vakhtang Kartvelishvili. SVD approach to data unfolding. *Nucl. Instrum. Meth.*, A372:469–481, 1996.

[136] G. D'Agostini. A Multidimensional unfolding method based on Bayes' theorem. *Nucl.Instrum.Meth.*, A362:487–498, 1995.

[137] T. Adye. Unfolding algorithms and tests using RooUnfold. *arXiv*, page 1105.1160, 2011.

[138] C.E.Lawson and R.J.Hanson. Solving Least Square Problems. *Prentice-Hall Inc.*, 1974.

[139] `http://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupSystematicErrors`.

[140] The CMS Collaboration. CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update. *Technical Report CMS-PAS-LUM-13-001*, 2013.

[141] A. Buckley, J. Butterworth, L. Lonnblad, H. Hoeth, J. Monk, H. Schulz, J. Eike von Seggern, F. Siegert, and L. Sonnenschein. Rivet user manual v. 1.6.0. *arXiv:1003.0694v6.*

[142] `http://lcgapp.cern.ch/project/simu/HepMC/`.