University of Trieste
Department of Physics

Ph.D. School of Physics

Measurement of the associated production of a Z boson and b quarks in proton-proton collisions at $\sqrt{s} = 8$ TeV with the CMS experiment at LHC

XXVI Cycle

SSD: FIS/01

Supervisor
Dr. Giuseppe Della Ricca
Co-Supervisor
Dr. Fabio Cossutti
Director of the School
Prof. Paolo Camerini

Candidate
Vieri Candelise

Academic Year 2012/2013

The pursuit of truth and beauty
is a sphere of activity in which we
are permitted to remain children
all our lives.
Albert Einstein

Contents

Introduction

The Standard Model of elementary particles is one of the greatest achievements in the history of science. Thanks to its continuous experimental confirmations, obtained in particles accelerators experiments over the decades, our understanding in the structure of the most intimate nature of matter and universe has reached a staggering level of detail. The milestone discovery of the Higgs boson made by the ATLAS and CMS experiments in 2012 consecrated the importance and physics potentials of the Large Hadron Collider (LHC) at CERN, completing the picture of the fundamental particles and forces of the Standard Model, projecting particle physics into a new era. Within the scientific program of the LHC experiments, the "re-discovery" of the Standard Model particles in high energy collisions and the measurement of their basic properties are a crucial part for both the understanding of the interactions between quarks in the framework of quantum chromodynamics (QCD) and the measurements of the properties of bosons in electroweak interactions. The associated production of a Z boson in association with b quarks process is predicted by the Standard Model and holds unique features. The measurement of the properties of the $Z + b$ final state in proton-proton interactions shows a vast physics phenomenology, and offers the opportunity to investigate different sectors of the Standard Model. Starting from the test of the predictions in perturbative QCD of the b quark content inside the proton, and its production mechanism through the hard scattering, the study of the associated production with a Z boson provides an important test for the study of the background processes of several Higgs boson and supersymmetric particles decays, and may give information on eventual fourth generations of fermions as well. Furthermore, through the study of the kinematics of the $Z + b$ final state, the presence of potential new particles can be investigated, beside the particular model under consideration. Another key feature of this process is the possibility of measuring the polarization asymmetry of the $Z+b$ system, which remains today the last Standard Model observable to have some tension between the experimental measurements and the theoretical predictions, and, if such discrepancy would be confirmed, it could provide a starting point in the new physics searches at LHC. In the first Chapter of this thesis, a brief theoretical introduction of the Standard Model is presented, describing the electroweak interaction and the bosons masses through the Higgs mechanism, the strong interactions sector and the QCD basic principles, ending with a summary of the main open questions of the model and the possible extensions ideated to solve them. In Chapter 2 the $Z + b$ final state is described in details, describing the main QCD models for the production mechanism in proton collisions and how the calculations are implemented in several simulations. The main concepts of the event generations at LHC is presented together with a description of the most important generators used in this analysis. In the second half of the Chapter the physics phenomenology of the $Z + b$ process is

presented for both Standard Model and New Physics scenarios involving this signature in the final state. Chapter 3 offers a detailed description of the experimental apparatus employed for the measurements made in this thesis. The Large Hadron Collider is introduced, and the Compact Muon Experiment is described showing the various sub-detectors structure. In Chapter 4 the reconstruction algorithms employed in CMS in order to identify and build the physics objects (electrons, muons, jets, missing energy) needed for the measurements made in this thesis are presented, with particular emphasis on their performances. After a general description of the Particle Flow algorithm, used to combine all the CMS informations coming from the different sub-detector measurements, dedicated Sections are presented to describe the single objects reconstruction techniques. Chapter 5 is dedicated to the event selection, including all the informations on the data samples and trigger requirements, physics variables cuts introduced to discriminate the $Z + b$ signature, background evaluation. Special Sections are dedicated to the $t\bar{t}$ background extraction using a data-driven method and on the b, c, light flavoured quark fractions estimation in the selected sample. The event topology is studied presenting the most relevant kinematic distributions of the $Z + b$ final state at detector level. In Chapter 6, after the description of the unfolding technique used to deconvolve the measured distributions from the detector effects, the systematic error sources are evaluated showing for each of them the strategy adopted to quote the final contribution. Finally, the differential cross section measurements of the $Z + b$ and $Z +$ jets processes as a function of several observables of the event are presented together with the relative differential cross section ratios. The results obtained for the dielectron dimuon and combined channels are then presented with the comparison between unfolded data and several QCD theoretical predictions.

Chapter 1

The Standard Model of Particle Physics

Matter and antimatter in the Universe are the result of continuous interactions of elementary particles over time. The Standard Model is the quantum theory that describes particles and fields in terms of the dynamics and unification of three of the four fundamental forces of Nature: the electromagnetic force, the weak nuclear force and the strong nuclear force. Within the development of quantum mechanics, since the first decades of the 20th century, fundamental physics has grown and built up piece by piece over extended and always more and more precise experimental confirmations: from the discovery of the first antiparticle, the positron, that confirmed the Dirac theory in 1932, through the discovery of the W and Z particles at CERN in 1983 by Rubbia and Van der Meer, consolidating the idea of electroweak unification by Weinberg, Salam and Glashow, until the discovery of the top quark in 1995 at Tevatron and the Higgs boson at the LHC today. For the predictive power and the incredible amount of experimental confirmations, the Standard Model of elementary particles is beyond any doubt one of the most successful scientific theories in human history. In this Chapter, I will briefly introduce the theoretical framework of the Standard Model, starting from the Electroweak Unification and the Brout-Englert-Higgs model and continuing with Quantum Chromodynamics. In the end I will touch on the still unsolved main questions of the theory, for which this model cannot be fully considered a satisfactory description of nature, yet.

1.1 Electroweak Interactions

The electroweak interaction is the unified description of the electromagnetism and weak nuclear force that arise at a fixed energy scale, of around 100 GeV, called *electroweak scale*. The theoretical framework of this unified force in the context of field theory was introduced and developed by Glashow, Weinberg and Salam in the 70's of the last century [1, 2, 3], starting from the quantum description of electromagnetism (QED, quantum electrodynamics), and the V-A theory of the weak force. The main idea is to construct a Yang-Mills theory that includes in the same symmetry the photon and the vector bosons, already introduced in the V-A description of the weak interactions.

1.1.1 The Electroweak Unification

The first step towards the unification is to build up a Lagrangian in which the internal symmetries are the result of the Noether theorem applied to the weak and electromagnetic currents:

$$J_\mu^{weak} = \bar{\ell}\gamma_\mu(1-\gamma_5)\nu_\ell, \quad J_\mu^{e.m.} = \bar{\ell}\gamma_\mu\ell,$$

where γ_μ, γ_5 are the 4-dimensional Dirac matrix, and the quarks, leptons and neutrinos fields are gathered together into the Dirac spinors:

$$\psi = \begin{pmatrix} \nu_\ell \\ \ell \end{pmatrix}_{\ell=e,\mu,\tau}, \quad \begin{pmatrix} U \\ D \end{pmatrix}_{D=d,s,b}^{U=u,c,t}.$$

The transformations that operate under this doublet's components are the *internal,* related to the Lorentz invariance, and the *chiral* related to the γ_5 Dirac matrix

$$\begin{cases} \psi \to e^{i\alpha\tau/2}\ \psi \\ \psi \to e^{i\frac{\beta\tau\gamma_5}{2}}\ \psi \end{cases},$$

where α, β are real matrices and τ are the Pauli matrices. This set of transformations generates an Algebra, and the minimal gauge group possible is $SU(2)_L \times U(1)_Y$, where L describes the left-chiral component of the ψ field and Y is the weak hypercharge, a combination of the neutral current by $SU(2)$ and the electromagnetic current by $U(1)$. The gauge field associated to the $SU(2)_L \times U(1)_Y$ defines the electroweak Lagrangian

$$\mathcal{L}_{EW} = -\frac{1}{4}W_{\mu\nu}W^{\mu\nu} - -\frac{1}{4}B_{\mu\nu}B^{\mu\nu} + \bar{\psi}i\gamma_\mu D^\mu\psi,$$

where the interaction between W, B particles, the Yang-Mills gauge field of the weak isospin and weak hypercharge, are described in the first part of the formula, while the second part is the Dirac Lagrangian of left-handed leptons and quarks doublets. The D_μ term is the covariant derivative that denote the couplings as $g, g^{'}$ defined as

$$D_\mu\psi = \left[\partial_\mu + igW_\mu\frac{\tau}{2} - \frac{ig'}{2}B_\mu\right]\psi.$$

The description of matter through *three* generations of quarks and leptons, with electroweak interactions between their quantum fields, evolved naturally from the Cabibbo quark mixing theory into the Cabibbo-Kobayashi-Maskawa matrix, that describes the quark mixing induced by weak interaction in an exactly three-generation of particles vision of the Standard Model.

The Cabibbo-Kobayashi-Maskawa Matrix

Up to the heavy quarks discovery, the Cabibbo theory [4] guaranteed the universality of weak interaction through a rotation mechanism of the u and d type quark by an angle θ_C (the Cabibbo angle). The Cabibbo's 2×2 real matrix allows to calculate the amplitude of charged current process involving quarks, and adjusted the discrepancy seen in the measurements of many decay branching fractions, such as the ratio between the amplitude of muon decay $\mu \to e\nu_e\bar{\nu_\mu}$ (proportional to 1) and neutron

decay $n \to pe\nu_e$ (proportional to $\cos^2\theta_C$). The introduction of the Cabibbo angle cured the inconsistency between data and theory, and it was found out that

$$\theta_C \sim 13.4^{\circ}.$$

Following the discovery of the *charm* quark in 1974, and later of the *beauty* quark in 1977, together with the discovery of the CP violation in the K^0 system, a natural extension of the Cabibbo idea that includes the new three-famdily version of the weak interactions theory was formulated by Kobayashi and Maskawa[5], taking into account the presence of three new quarks and their mixing and amplitudes, and the CP violation in heavy flavours by inserting a complex phase δ in a 3×3 matrix. A representation of the Cabibbo-Kobayashi-Maskawa matrix can be written as follows:

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

The matrix describes the probability for a transition between a quark q to a quark $q^{'}$, proportional to $|V_{qq'}|^2$. For example, the amplitude of the *top* quark decay $t \to Wb$ is calculated from the theory and it is equal to 1 according to the fact that comes directly from $|V_{tb}|^2$. It is found that the value of this quantity is experimentally very close to one by studying the Wb final state. The explicit expression of the matrix, exploiting three Cabibbo angles $\theta_{1,2,3}$ is then

$$\begin{pmatrix} c\theta_1 & -s\theta_1 c\theta_3 & s\theta_1 c\theta_3 \\ s\theta_1 c\theta_2 & c\theta_1 c\theta_2 c\theta_3 - s\theta_2 s\theta_3 e^{-i\delta} & c\theta_1 c\theta_2 s\theta_3 + s\theta_2 c\theta_3 e^{i\delta} \\ s\theta_1 s\theta_2 & c\theta_1 s\theta_2 c\theta_3 + c\theta_2 s\theta_3 e^{i\delta} & c\theta s\theta_2 s\theta_3 - c\theta_2 c\theta_3 e^{i\delta} \end{pmatrix}$$

where $c\theta = \cos\theta$ and $s\theta = \sin\theta$. An important feature of the CKM matrix from the experimental point of view came from the unitarity condition of the matrix:

$$\sum_k V_{ik} V_{jk}^* = 0.$$

Since there are six flavours, the condition can be depicted as six different triangles in the complex plane with the same area, and each side of these triangles (called *unitarity triangles*) has length VV^*. The angles of these triangles are related to the CP-violating phases of quark fields, and are accessible to dedicated experiments. The most precise measurements of the unitarity triangles are made by BELLE, BaBar, and LHCb [6].

1.1.2 Electroweak Symmetry Breaking

In the theory described so far, all the particles are massless. This fact is in great contradiction with the experimental evidence that fermions and vector bosons are massive, and thus some kind of new mechanism must be introduced in order to make the whole model consistent with the experimental observations. One of the easiest ways to build a mass term and assign it to fermions and vector bosons was firstly proposed by Higgs, Englert and Brout [7, 8] with the introduction of a scalar field in the electroweak Lagrangian, through the spontaneous breaking of the $SU(2)_L \times U(1)_Y$ symmetry group. A simple and direct way to perform the symmetry

breaking of electroweak interactions comes from the Higgs Mechanism, inspired by the quantum behaviour of superconductivity. A complex scalar field of $SU(2)_L$ with isospin equal to one in the form

$$\frac{1}{\sqrt{2}} \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$$

and included in the electroweak Lagrangian

$$\mathcal{L} = \mathcal{L}_{EW} + D_\mu \phi D^\mu \phi - V(\phi).$$

The Higgs potential $V(\phi)$ is a quartic function of the field

$$V(\phi) = -\mu^2 \phi\phi^\dagger - \lambda(\phi\phi^\dagger)^2.$$

For $\mu^2 > 0$, the potential is a concave function with an absolute minimum in the origin, bringing to a stable quantum vacuum expectation value for the field ϕ equal to 0. When $\mu^2 < 0$ on the other hand, a spontaneous symmetry breaking arises for the degeneration of minima around the quartic valley of $V(\phi)$ (Fig. 1.1.1). In this case, the quantum vacuum expectation value is found to be

$$< 0|\phi|0 >= v,$$

the configuration of minimum energy is not symmetric and the symmetry is spontaneously broken. When this mechanism is absorbed inside the electroweak Lagrangian, the breakage of the symmetry is made by the scalar field, leaving the electromagnetic current conserved in the scheme

$$SU(2)_L \times U(1)_Y \rightarrow U(1)_{em}.$$

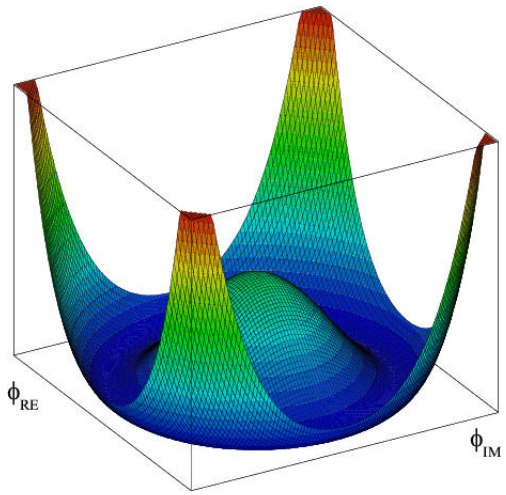

Figure 1.1.1: Three-dimensional Higgs potential as a function of the complex scalar field ϕ components.

Now the covariant derivative originates mass terms to vectorial fields and fermionic fields. If Z_μ and A_μ are the massive and electromagnetic fields that diagonalize the mass matrix, which comes up from the interaction between the ϕ field and $\mathcal{L}_{EW}$, it turns out that the physics fields are obtained by a rotation of the Yang-Mills fields W_μ, B_μ by a mixing angle θ_W, known as Weinberg angle

$$\begin{cases} Z_\mu = W_\mu^3 \cos\theta_W - B_\mu \sin\theta_W \\ A_\mu = W_\mu^3 \sin\theta_W - B_\mu \cos\theta_W \end{cases},$$

where Z_μ is a vectorial neutral field. In the end, the particles associated to the complete Lagrangian are two charged and one neutral vector boson, related to the three components of W_μ, the $W^\pm$ boson and the Z boson, the photon γ related to A_μ , and the Higgs boson H related to the scalar ϕ. All the masses are linked to the mixing electroweak angle θ_W by the relation

$$\frac{M_W^2}{M_Z^2} = \cos^2\theta_W.$$

This is a key parameter of the Standard Model, and it was measured with unprecedent precision by the LEP experiments through the couplings of the Z boson to fermions [9] :

$$\sin^2\theta_W = 0.23120 \pm 0.00015.$$

The Higgs mass only depends on the expectation value of the vacuum v

$$M_H = \sqrt{\lambda v^2}.$$

The first experimental evidence for the Standard Model was the discovery of the weak neutral current, associated to the Z_μ field. The concept of a neutral current is one of the most important breakthroughs in particle physics, and the first experimental evidence was found in the *Gargamelle* bubble chamber [10], exploiting the interactions of high energy neutrinos in matter, through the reaction

$$\nu_e + N \rightarrow e^- + X.$$

1.1.3 The Higgs Boson

For more than 50 years, the Higgs particle predicted by the Standard Model was one of the main puzzles in particle physics. First of all, the discovery of such boson was fundamental in order to validate the Standard Model and to make it a realistic theory. Moreover, beside the masses, the presence of the Higgs is crucial to avoid many divergences in processes like the t-channel WW scattering. Without the Higgs, this process would violate the unitariety of the S-matrix giving a cross section that diverges with energy [11]. In some way the situation is similar to the Fermi theory of weak interactions , where the W boson plays the role of a massive corrector in the nuclear β decay of the neutron. For decades the Higgs was hunted in colliders experiment, first in the e^+e^- LEP collider at CERN (at a center of mass energy around the Z mass pole in the first phase, LEPI, and increased up to a maximum of 210 GeV in the phase II), that put a first limit of 114 GeV on its mass at 95% confidence level [12]. Then the hunt passed to Fermilab experiments in $p\bar{p}$ collision at $\sqrt{s} = 1.96$ TeV, where many parts of the mass spectra of the Higgs were excluded [13]. One of the main challenges in the Higgs searches was the impossibility to make a theoretical prediction for its mass through indirect searches by evaluating its quantum loop corrections, as it was made in the past for the *top* quark before its discovery at Fermilab. The reason is that, while for the *top* the quantum corrections to the W, Z propagator are proportional to m_t^2, in the Higgs case the dependences is proportional to $\log(M_H/M_t)$ and thus the sensitivity is largely worst.

1.1.4 The Discovery

The striking answer in the search of the Higgs boson came from CERN on 4th July 2012 at the LHC experiments, when the discovery of a neutral scalar boson in pp collision at a center of mass energy of 7 and 8 TeV was announced. The two multipurpose experiments designed to measure the decay products of the Higgs, CMS and ATLAS, found a particle compatible with the Standard Model Higgs boson

within a significance greater than 5σ. The main contribution to the discovery were the $H \to \gamma\gamma$ and the $H \to ZZ \to 4\ell^{\pm}$ channels, employing the extraordinary accuracy reached by the two experiment in precise measurements of leptons and photons energy and momentum. In Fig.1.1.2 the $\gamma\gamma$ invariant mass measured with the CMS experiment in 2012 shows the peak around the candidate Higgs boson mass around 125.5 GeV. The final result, a milestone in physics history, established a new benchmark in our understanding of the Universe [14] [15] (ATLAS+CMS combined result):

$$M_H = 125 \pm 0.2(\text{stat.}) \pm 0.7(\text{syst.}) \text{ GeV}.$$

Many other measurements on the Higgs have been performed by ATLAS and CMS, including couplings to bosons and fermions, the spin and the CP-parity. All the results are consistent between the two experiments and consistent with the prediction of the Standard Model for a Higgs particle.

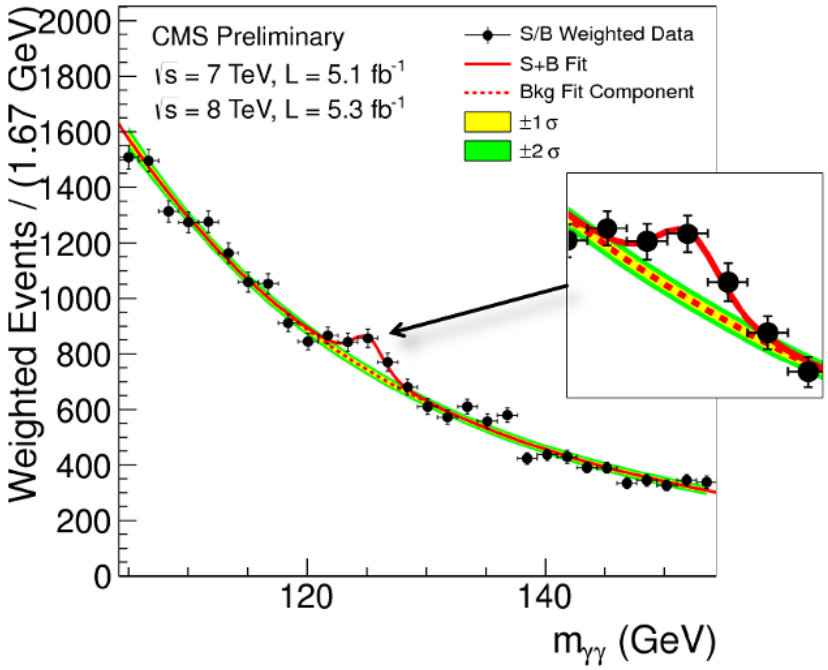

Figure 1.1.2: $H \to \gamma\gamma$: the di-photon invariant mass spectrum.

1.2 Quantum Chromodynamics

In the first half of the 50's of the last century, with the relentless growth in the number of particles discovered in bubble chamber experiments, which interacted with matter producing cascades of decays, the so-called *hadrons* invaded the scenario of experimental physics at the time. It seemed, therefore, that such a vast "zoo" of particles could not be a fundamental representation of matter. The idea of classifying hadrons through three basic quantities, the quarks, was first proposed in 1961 by Gell-Mann and Ne'eman, in what was called the *eightfold way*

[16] i.e. the description of particles observed through the properties of the symmetry group $SU(3)$. In the plain hypercharge-isospin, the *eightfold way* perfectly reproduced the spectrum of all the particles observed, for example: two doublets (n, p); (Ξ^0, Ξ^-), a triplet $(\Sigma^+\Sigma^-, \Sigma^0)$, and a singlet (Λ). The success of the model was the prediction of two particles experimentally discovered only a few years after the introduction of the theory: the η and Ω^-. The next step was to identify the fundamental representation of $SU(3)$ with a $3 + \bar{3}$ (from which all combinations of observed hadrons would emerge) as a set of $3 + 3$ fundamental particles called quarks[1] (and antiquarks), fermions with fractional charge in three flavours: up, down, strange. Baryons consist of three quarks, while mesons consist of quark pairs. A new quantum number that marks out strong interactions, called *colour*, was introduced in order to describe the behaviour of the Δ^{++} particle, a state made of three identical up-type quarks: $|uuu>$ that would violate the Pauli principle. The experimental test for the existence of colour occurs in the measurement of the cross section ratio $R = \sigma(e^+e^- \to \mu^+\mu^-)/\sigma(e^+e^- \to q\bar{q})$, that is directly proportional to the inverse of the number of colours $1/N_C$. In the representation of quarks in three different colours (conventionally green, blue, red), the Δ^{++} would be no more just the product of three identical fermions but three quantum mechanically speaking different particles. Over the years, the discovery of neutral currents and the *GIM* [2] mechanism underlines the needs to introduce other generations of quarks to explain certain cross sections in weak neutral current induced processes. Further confirmations of the *GIM* mechanism predictions were established by the discovery of the *beauty* and *top* quarks.

1.2.1 The QCD Lagrangian

Quantum Chromodynamics (QCD) is the gauge theory of strong interactions, describing the dynamics of coloured quarks and gluons. The QCD is derived in a Yang-Mills theory from the invariance for local symmetry in the non-Abelian colour group, $SU(3)$, where quarks are in its fundamental representation. The Lagrangian is written as

$$\mathcal{L} = -\frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a + \sum_q \bar{\psi}_{q,a}(i\gamma^\mu\partial_\mu\delta_{ab} - g_s\gamma^\mu t^C_{ab}A^C_\mu - m_q\delta_{ab})\psi_{q,b},$$

where the quark fields holding a colour index a are represented by the spinor ψ (a ranges from 1 to the total number of colours $N = 3$). A_μ field is the gluon field, and $C = 1, ..., N_C - 1$ are the 8 different types of gluon, that belong, as predicted by the $SU(3)$ group, to its adjoint representation. The 3×3 matrices t^C_{ab} are the $SU(3)$ generators defined as describing the fact that a quark-gluon interaction rotates the quark's colour in the $SU(3)$ space. The field tensor is

$$G^A_{\mu\nu} = \partial_\mu A^A_\nu - \partial_\nu A^A_\mu - g_S f_{ABC}A^B_\mu A^C_\nu.$$

[1] The Gell-Man's word "quark" comes from a James Joyce's quotes by his book "Finnegans Wake" (1939): *"Three quarks for Muster Mark!"*.

[2] The Glashow-Illiopoulus-Maiani mechanism [17] consists in the introduction of a heavy (~1.5 GeV) quark to accomodate the divergent amplitude in kaons decay such as $K_L \to \mu^+\mu^-$. The first evidence of a meson state made of this new quark (called *charm*), the J/ψ, was discovered at SLAC by Richter and at BNL by Ting through a narrow peak in the e^+e^- invariant mass distribution around 3 GeV.

Finally $g_S = \alpha_S/4\pi$ is the QCD coupling constant. Together with the quark masses m_q, the strong coupling is the fundamental parameter of QCD.

1.2.2 Asymptotic Freedom and Quark Confinement

In QCD there are two peculiar aspects, largely tested through years of experiments. The *confinement* is the phenomenon that explains why quarks cannot be ever detected or isolated singularly. As a consequence of this principle, quarks cannot be observed directly as free particles. After a quark production, at very short distances, the quantum nature of QCD predicts the generation of virtual quark-antiquark (or gluons) pairs, produced with decreasing energy while increasing distance. In the Lund model, they couple together through colour strings forming observable uncoloured baryons and mesons. It is used to refer to the two extreme conditions as "parton level" (short distances, intense strong force), in which nothing can be observed, and "hadron level" (large distances and strong force approaching to zero), where particles made of quarks can be observed and detected. The scenario of strong force ideally equal to zero when distances are close to zero is called *asymptotic freedom*: in this situation quarks are essentially free [18]. In the regime of asymptotic freedom, the perturbation theory is a very satisfactory description of QCD physics observables, giving precise predictions about what can be tested in collider experiments. This approach is called perturbative-QCD, or pQCD. The opposite case, of large distances and low momentum transferred, where it is no more possible to compute QCD in a perturbative approach, is still an ongoing and under study effort. In this framework QCD path integrals are computed in a discrete set of space-time points (lattice-QCD). A test of these properties of QCD can be performed in the context of measurements involving the dependence of α_S with energy.

1.3 Running of α_S

The functional dependance of the coupling constant with the energy scale μ_R, the renormalization scale, is given by the Renormalization Group Equations for α_S

$$\beta(\alpha_S) = \mu_R^2 \frac{\mathrm{d}\alpha_S}{\mathrm{d}\mu_R^2} = -(b_0\alpha_S^2 + b_1\alpha_S^3 + b_2\alpha_S^4 + ...),$$

where the scale μ_R represents the energy scale where QCD is evaluated. For energies close to the typical scale of the process of interest, $\mu_R^2 \simeq Q^2$, the meaning of $\alpha_S(\mu_F)$ is the strength of the strong interaction between particles under study. The coefficients of the expansion b_n are the $n-$loop coefficients of the β function, depending on the number of flavours n_f. For example the first order of the expansion is

$$b_0 = \frac{33 - 2n_f}{12\pi}.$$

Of particular importance is the minus sign in the $\beta(\alpha_S)$ expansion, expressing the meaning of the asymptotic freedom: the strong coupling becomes weaker for processes with high Q^2. For instance, in the range of $10^2 - 10^3$ GeV, the value of α_S is close to 0.1. Neglecting all the b_n coefficients but b_0, an exact solution of the

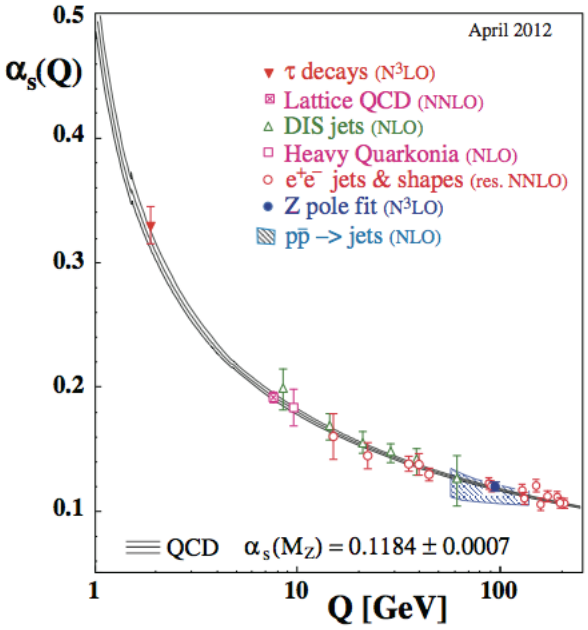

Figure 1.3.1: The strong coupling constant running over the momentum transferred Q of the process in different experimental measurements.

differential equation for the beta-function

$$\alpha_S(\mu_R^2) = \frac{1}{b_0 \log(\frac{\mu_R^2}{\Lambda_{QCD}^2})},$$

where Λ_{QCD} is the perturbative cut-off over the renormalization's integrals. The meaning of this cut-off is the validity of the perturbative regime approximation, beyond which the integrals would diverge. For many experimental studies, the strong coupling is evaluated at a fixed energy scale, tipically of the order of the electroweak scale, $\mu_R \simeq M_Z$ (see Fig.1.3.1).

1.3.1 Hadronic Collisions

The hadron-hadron scattering amplitude is one of the more important results gathered from pQCD. Particularly important is of course the case of proton-proton collisions (pp), essentially the starting point of the LHC physics. The pp cross section can be seen as the convolution of the *hard* qq cross section $\hat{\sigma}_{ij\to X}$ inside the proton, with the parton distribution functions $f_1(x_1, Q^2)$, $f_2(x_2, Q^2)$ (PDF), integrated over the Bjorken variables x_1, x_2, defined as the fraction of the 4-momentum carried out by the two partons inside the proton during the collision

$$\sigma(pp \to X) = \sum_{q,g=0}^{n} \alpha_S^n(\mu_R^2) \sum_{ij} \int \mathrm{d}x_1 \mathrm{d}x_2 f_1(x_1\mu_F^2) f_2(x_1\mu_F^2) \cdot \hat{\sigma}_{ij\to X}^{(n)}(s; x_1, x_2, \mu_R^2, \mu_F^2)$$

where the n index runs over the perturbative order and s is the squared center of mass energy of the collision and μ_F is the scale momentum. The tree-level process, where no emission of gluons or quarks holds, is called "Leading Order" (LO) and

takes place when $n = 0$. Further orders are called Next to Leading Order (NLO, $n = 1$), Next to Next to Leading Order (NNLO, $n = 2$) and so on. As it can be seen in the formula, the PDFs play a fundamental role in the description of the scattering, and it is very important to have several experimental tests accessible for their values. Perturbative QCD cannot predicts PDFs, since they contain also the low energy (non-perturbative) information about the scattering. As a consequence, PDFs distributions are extracted from data, through fit techniques to the most sensitive observables. The energy scale evolution of the PDFs can be calculated in pQCD solving the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equation [19, 20, 21]

$$\frac{\partial f(x,Q^2)}{\partial \log Q^2} = \frac{\alpha_S}{4\pi}\int_x^1 \frac{\mathrm{d}z}{z}\{P_{aa'}(z,\alpha_S)f(\frac{x}{z},Q^2)\}.$$

The Altarelli-Parisi splitting function $P_{aa'}$ represents the probability to have the trasformation of parton with momentum x to another with momentum z, $a(x) \to a'(z)$, by means of one or more quarks or gluons emission. Evaluating $P_{aa'}$ in terms of α_S powers, it's possible to calculate the amplitude for processes involving quark or gluon emissions at a fixed order:

$$P_{aa'}(z,\alpha_S) = P_{aa'}(x)\ |_{LO} + \frac{\alpha_S}{2\pi}P_{aa'}(z)\ |_{NLO} + (\frac{\alpha_S}{2\pi})^2 P_{aa'}(z)\ |_{NNLO}.$$

The most important result came from the deep inelastic $e+p$ scattering experiments at HERA [22] and neutrino-nucleus ($\nu + N$) at NuTeV [23], where the structure functions $F_q = xf_q$ are measured (see plot in Fig. 1.3.2) and PDFs are extracted by means of a fit. At hadron colliders like LHC, many processes and observables may be useful to measure and constrain PDFs in certain ranges if x, such as the Drell-Yan process W/Z + jets as well as the general properties of W/Z boson decays (for example the W charge asymmetry, or the Z boson momentum distribution).

1.3.2 Quarks and Gluons Radiation

The perturbation order n expresses the different phenomenologies of a certain QCD process. When $n > 0$ the QCD amplitude is the result of the sum of the tree-level diagram, plus all the possible Feynman diagrams coming out from the emission of n real or virtual quarks and gluons. If a gluon g is emitted in a direction that is approximately parallel to the direction of emission of a quark q and its momentum is close to zero, a simplification on the matrix element can be made:

$$\lim_{\theta\to 0}\lim_{E_n\to 0}|\mathcal{M}_n(p_1,...,\ p_n)|^2\mathrm{d}\Phi = |\mathcal{M}_{n-1}(p_1,...,\ p_{n-1})|^2\mathrm{d}\Phi\frac{\alpha_s C_i}{\pi}\cdot\frac{d\theta_i^2}{\theta_i}\cdot\frac{dE_n}{E_n}.$$

These limits are known as *collinear limit* ($\theta \to 0$) and *soft limit* ($E \to 0$) , and since they produce as a result divergent integrals, they are of great importance for QCD, especially because they may change the event topology (Fig. 1.3.3).

1.3.3 Next to Leading Order Calculations

Currently, NLO available calculations included in Monte Carlo simulations cover a wide range of physics processes, starting with two particles annihilation to a maximum of five final state objects, $N \leq 6$. The frontier of NLO calculation at LHC

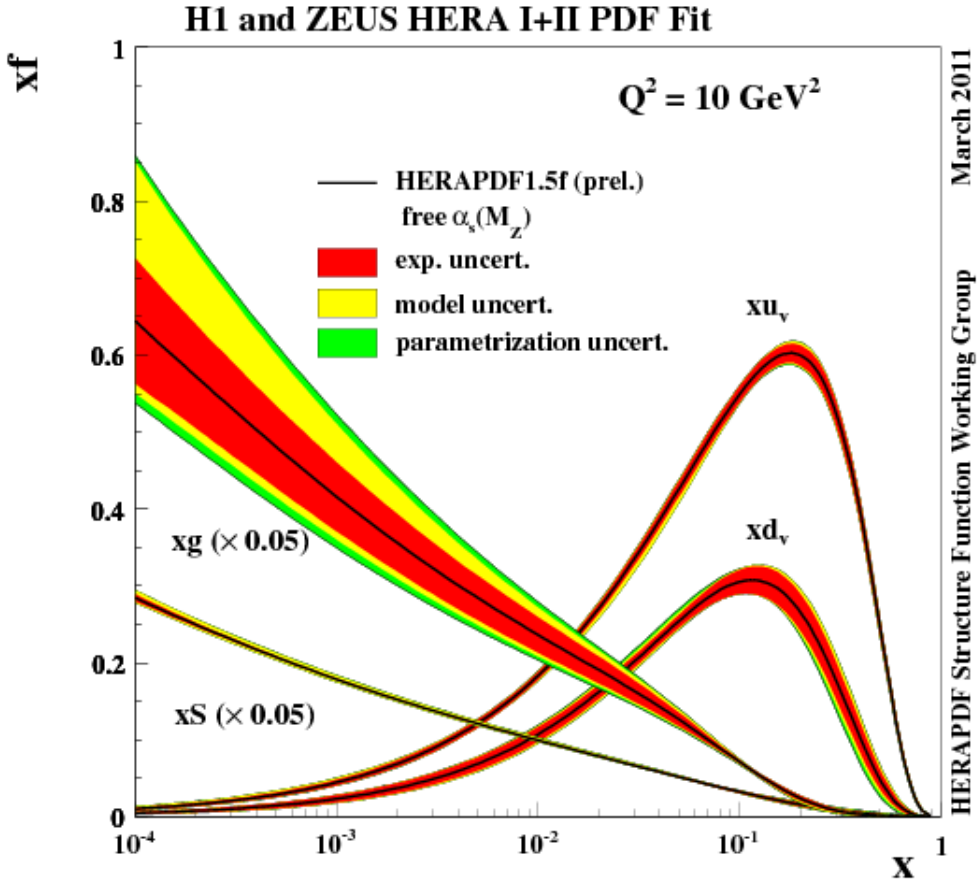

Figure 1.3.2: Proton structure functions measured in 2011 at HERA as a function of x for up, down, strange quarks and gluons.

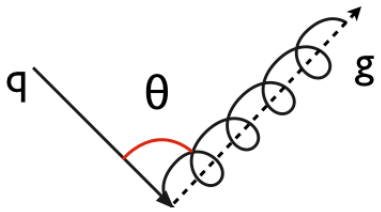

Figure 1.3.3: Leading order emission in the quark line of a gluon with a θ angle. Collinear limit arise when the emission comes up with very small θ angles.

includes very important Standard Model processes like W/Z + 4-jets, $t\bar{t}b\bar{b}$, $b\bar{b}b\bar{b}$, WW + 2-jets [24, 25, 26]. It is already possible to extend the calculation beyond the NLO in few processes, requiring the resummation of all the Feynman diagrams at the α_S^2 order, the Next to Next to Leading Order (NNLO). From the computational point of view, it is extremely challenging to compute the exponentially growth of all the diagrams involving more and more loops. The most relevant results available that are fully inclusive of the NNLO are W/Z, H, $\gamma\gamma$ in pp collisions [27, 28, 29]. A more detailed description of event generators techniques and how they are connected to the $Z + b$ particular case is left to the next Chapter. In the $n = 1$ case, the Next To Leading Order (NLO), the process is

$$2 \rightarrow N$$

with N final state objects, and $N \leq 5$, depending on the physics under consideration. Given an observable K in a certain physics case, its NLO prediction is the sum over all the Feynman diagrams of the LO order plus the interference between the N tree level particles with the N 1-loop amplitudes:

$$\sigma_K^{NLO} = \sigma^{LO} + \alpha_S^{n-1}(\mu_R^2) \int \mathrm{d}\Phi_{n+1} |\mathcal{M}_{n+1}^2(p_1, ..., p_{n+1})| K_{n+1} +$$

$$+\alpha_S^{n-1}(\mu_R^2) \int \mathrm{d}\Phi_n 2Re(\mathcal{M}_n(p_1, ..., p_n)\mathcal{M}^*_{n;1-loop}(p_1, ..., p_n)) \cdot K_n.$$

NLO contributions appear as gluon loops to be summed in different Feynman diagrams (Fig. 1.3.4). One important issue to be noticed is that the amplitudes of the loop diagrams are described by infinite integrals, due to the infrared and collinear divergencies described before. However, many observables K are not affected by this problem if they are "infrared and collinear safe", i.e. they respect the following rules:

$$\lim_{p_\alpha \rightarrow 0} K_{n+1}(p_1, ..., , p_\alpha, ..., , p_n) = K_n(p_1, ..., , p_n)$$

In the specific physics case of this thesis work, several observables under analysis are not infrared and collinear safe. A detailed discussion is postponed to the next Chapter. To calculate the effects of NLO corrections, the action of adding to the diagram under consideration all the contributions proportional to α_S is needed. There are two types of diagrams:

- real emission: gluon radiation $qq \rightarrow X + (n)g$ as it is shown in Fig. 1.3.5(a)

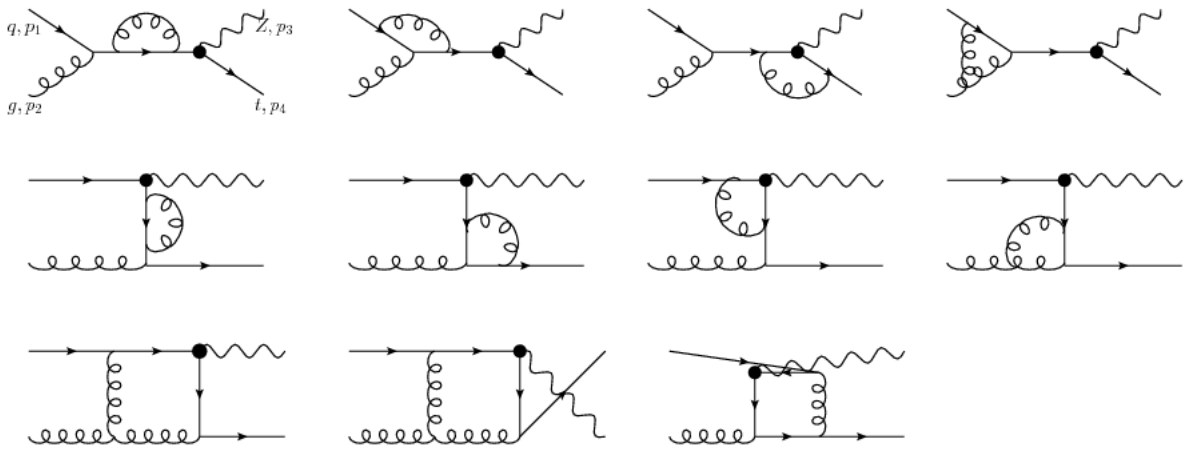

Figure 1.3.4: Example of NLO diagrams loop in single heavy quark production in association with a Z boson.

- virtual emission: loop corrections that involves emission and re-absorbption of one or more gluons (Fig. 1.3.5 (b)).

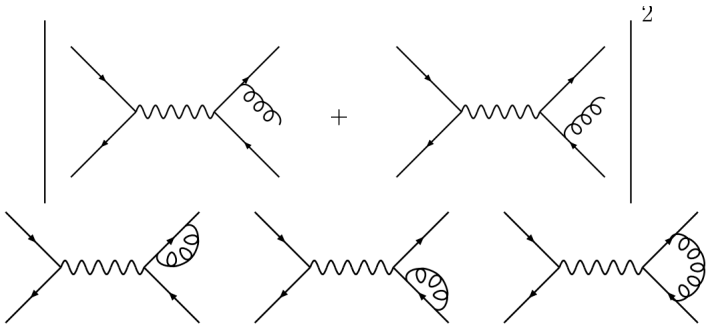

Figure 1.3.5: Real emission sums (top) and virtual emission loops at NLO (bottom).

1.3.4 Hadronization and Jet Clustering

Due to the quark confinement phenomenon, gluons and quarks produced in proton-proton collisions cannot be detected as free particles since they carry out a colour charge that must be conserved between the initial and the final state of the scattering. Hereafter the initial quark production, the strong potential decreases as the square of distance, and a single particle starts to tend to a stable configuration of uncoloured composite object recombining themselves according to the QCD rules by the emission of gluons initial state radiation. Final objects obtained by this process are mesons and baryons, detectable by their interaction in matter. This *hadronization* happens in the same direction of the initial parton that originates the chain, and the subsequent emission of particles around this direction, makes a cone of quarks and gluons known as *jet*. The bulk of a jet is a mixture of particles, mostly light charged mesons (π, K), photons form π^0 decays, a small fraction of baryons $(p, n, \Lambda,)$ neutrinos and leptons coming from the semileptonic decays of b and c quarks $(b \to c + l + \nu)$. Most of the energy produced in the hadronic shower can be measured by means of an hadronic calorimeter. The full reconstruction of the

complex jet can be done by many different algorithms. One of the most performing vastly used today in collider experiments is the anti-k_T algorithm [30].

The $anti - k_T$ algorithm

Clustering algorithms are essential when we want to analyze events involving quarks and gluons in the final state, because they provide an estimation of the particle energy and momentum that originates the jet, by taking into account the information of all the particles inside the jet cone. The anti-k_T algorithm starting point is the calculation of the distance of the two particles i, j :

$$d_{ij} = \min(1/k_T^2(i), 1/k_T^2(j)) \cdot \frac{\Delta R_{ij}^2}{R^2},$$

where R is a radius parameter that characterizes the method (for example we refer to anti-k_T-05 if it is $R = 0.5$) and ΔR is the distance between the two particles in the $(\eta - \phi)$ plane:

$$\Delta R = \sqrt{|\eta_i - \eta_j|^2 + |\phi_i - \phi_j|^2}$$

In this framework, ϕ is the azimuthal angle, $0 \leq \phi \leq 2\pi$ and $\eta = -\log((\tan\theta/2))$ is the pseudorapidity, proportional to the polar angle θ [3]. The transverse momentum of particles is denoted by $k_T = |k| \sin\theta$. The ith particle has a distance from the beamline B given by:

$$d_{iB} = 1/k_{Ti}^2$$

and the minimum between d_{iB} and d_{ij} is calculated. If this minimum is d_{iB}, the anti-k_T removes the particle i form a list of candidate and call the object "jet". If the minimum is d_{ij}, the momenta k of the particle i,j are then summed constructing a new particle, and then the algorithm is iterated again until only jets are left. Since no particle at $d > R$ can be counted, the shape of the clusterized jet is a cone around the original particle direction. A simulated jet reconstructed with the anti-k_T algorithm is shown in Fig. 1.3.6. An important feature of this method is that it takes into account for the infrared and collinear limits ensuring reliable prediction at all orders in QCD.

1.4 Beyond the Standard Model

Over the last decades, the predictions of the Standard Model got several validations through the discovery of every single boson predicted, the measurement of parameters and couplings, and finally the Higgs discovery within generations of collider experiments. The robustness of the model, in terms of experimental predictivity, allows to believe in a unified vision of three fundamental interactions of the Universe today. However, the Standard Model still has several obscure points, and it is unable to describe some of the experimental observations concerning the fundamental elements in the description of matter. In this Section the main unresolved issues of the Standard Model will be shown, together with the main theoretical attempts designed to solve them.

[3] The pseudorapidity is the optimal radial coordinate in hadron collider experiments, because of the PDFs. It is defined as the limit of the rapidity $y = \frac{E+p_z}{E_z-p_z}$ in the ultra relativistic case where the particle velocity is close to the speed of light.

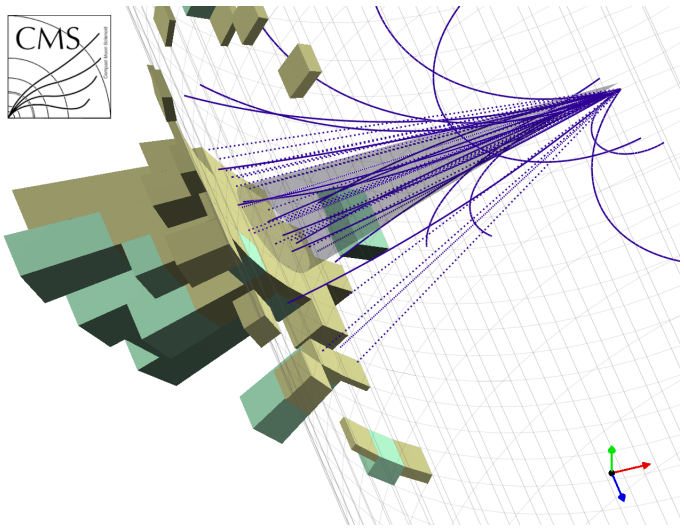

Figure 1.3.6: A graphical representation of a simulated jet clustered with the anti-k_T algorithm in the CMS experiment, and measured as a release of energy in the hadronic calorimeter.

The Hierarchy Problem

The Hierarchy Problem is one of the biggest dilemma in the whole theory of the Standard Model, causing an instability of the entire theoretical system. The problem can be faced on two apparently different fronts: On the one hand, the crucial point is derived from the observation of a huge discrepancy (of the order of 10^{32}) between electroweak force and gravity: this fact is reflected in the mass difference between the W, Z bosons that define the scale of electroweak interactions (order of 10^2 GeV), and the Planck mass, that defines the scale beyond which gravity must be described by quantum mechanics ($M_P = \sqrt{\frac{\hbar c}{G_N}}$, order of 10^{19} GeV). On the other hand, vector bosons acquire mass by means of the Higgs boson, whose vacuum instability produces corrections to its very own mass of the order of the Planck energy, making the whole Standard Model completely unrealistic and immeasurable. To correct the Higgs mass, these loop effects must be canceled by *something* in the renormalization process. The most accredited solution to this problem remains today to employ theories involving *supersymmetry*[4], which will exactly cancel out the differences by the introduction of a huge counter-contribution induced by *s-type top* quarks (*stops*) [32]. Extensive searches have been done in the first phase of data taking of LHC, showing no evidence of supersymmetry in a wide range of parameters and the masses of the s-particles.

Dark Matter

Gravitational effects in large scale structures of the Universe, especially in galaxies and clusters of galaxies, shown by calculations based on the rotational dynamics of their visible mass, bring to the surprising discovery [33] that the Universe consists

[4]Supersymmetry (SUSY) is an extension of the Standard Model where an extra symmetry is introduced by adminting the existence of SM particles partners with integer$\longleftrightarrow$fractional spin inversion ($s-particles$). The minimal extension possible is the $MSSM$ (Minimal Supersymmetric Standard Model [31]).

mainly of a different type of matter, not interacting with the Standard Model fields called *dark matter*. According to the latest measurements, the amount of dark matter accounts for 26.8% of the total matter of the Universe, and only the remaining 5% is made by ordinary matter. The searches from the point of view of experimental and theoretical dark matter hunting is one of the major efforts in particle physics. There are two fronts: direct searches, in cosmic rays by underground experiments, and searches at hadron colliders, where dark matter would be produced in pairs of neutral particles that may belong to different models. In the case of supersymmetry, dark matter is incarnated by *neutralinos* ($\tilde{\chi}^0$), often called Weakly Interacting Massive Particles or WIMPS, in processes such as

$$pp \rightarrow X \rightarrow \tilde{\chi}^0\tilde{\chi}^0$$

To date no evidence of dark matter has been found in any experiment [34].

Matter/Antimatter asymmetry

In the baryogenesis epoch after the Big Bang, where baryonic particles and antiparticles were produced for the first time, the content of matter and antimatter was perfectly equal. What we see today, instead, is a Universe dominated by matter, where antiparticles appear only in cosmic rays and cosmic phenomena, or produced by humans inside accelerators. Moreover, the antimatter annihilates with matter for which large structures of antimatter would not be permitted. The mechanism which makes, at a certain moment of baryogenesis, matter and antimatter symmetry broken in favor of the matter is still unknown [35]. In particle physics, the natural connection would appear by the CP symmetry violation in mesons containing b, c flavoured quarks,where the symmetry that turns particle into antiparticle (charge conjugation C) together with parity (P) does not provide the same representation of particles before the operation of the product of the CP symmetry. The CP violation is studied today at LHC by the rare decays of B mesons, and particularly by measure of the Cabibbo-Kobayashi-Maskawa matrix elements that provide the size of the effect of the CP violation in the Standard Model.

Neutrino Masses

In the Standard Model, neutrinos are described as massless particles, and the Higgs mechanism does not provide for a Higgs boson coupling to neutrinos. In 1998, the SuperKamiokande experiment measures for the first time the oscillation of neutrino flavours $\nu_e \longleftrightarrow \nu_\mu$, as predicted by B.Pontecorvo in 1957 [36]. This is possible only if neutrinos are massive. By combining the information coming from cosmology (either supernova neutrinos and solar ones) and oscillation experiments like OPERA and SuperKamiokande, we find out that neutrino masses are different for the three lepton families and are of the order of a few eV [37]. The neutrino mass is explained in various extensions of the Standard Model, for example in the case of Majorana/Dirac representation, in the study of right-handed extended symmetry group, and by the so-called *see-saw* mechanism [38].

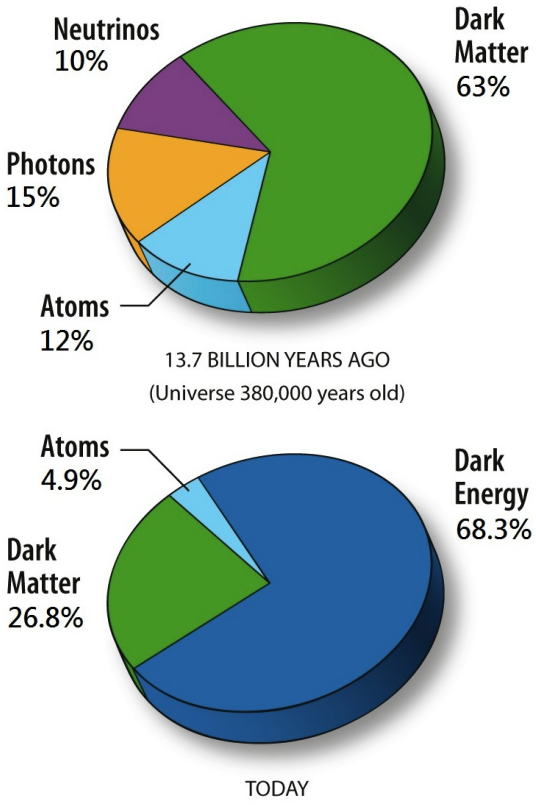

Figure 1.4.1: Relative fractions of matter and dark matter in our Universe at the Big Bang time (top) and today (bottom).

Chapter 2

Phenomenology of $Z + b$ events

The associated production of vector bosons (V) and jets in hadronic collisions is a benchmark for the understanding of the perturbative QCD sector of the Standard Model. Processes involving W, Z bosons, together with the emission of one or more quarks or gluons are a probe for the PDFs choices at the collider energy, and give information about the hard scattering structure and the strong energy scale used to model the simulation. Measuring the kinematic observables in $V + jets$ topology at collider energy can be a powerful tool which allows to unveil the QCD structure of the Monte Carlo simulation, and tune the different constants and parameters chosen to describe the process. From many point of views, it is of great importance to study the particular case of the *beauty* quarks production together with a Z boson, described in the Standard Model by the interaction Lagrangian

$$\mathcal{L} = \frac{e}{\sin^2\theta_W \cos^2\theta_W}\bar{b}\left[\gamma^\mu(g_L^b + g_L^b)\cdot\frac{1-\gamma_5}{2} + (g_R^b + g_R^b)\cdot\frac{1+\gamma_5}{2}\right]b\cdot Z_\mu,$$

that describes the $Zb\bar{b}$ vertex through the $g_{L,R}^b$ tree-level left, right couplings. In this Chapter I will discuss the relevance of the $Z + b$ final state in pp collision through the reaction

$$pp \to Z + b,$$

at the LHC, starting with the description of the Standard Model predictions (Sec. 2.1), at the different pQCD orders, and the interpretations and approximations used to build up the Monte Carlo simulations (Sec. 2.2). In Section 2.3 I will focus my attention on the review of the most important $Z + b$ final state signatures that come up from the dynamics of new fields and particles in scenarios of physics beyond the Standard Model. Since this channel is a significant background (very often a dominant one) for many important searches at the LHC, a measurement of the Zb production might be a strong tool to improve significantly the understanding and the precision in the description of such processes. Finally I will briefly discuss about a very important feature of this channel, that is the possibility to measure the couplings of the Z, b particles and thus the effective electroweak Weinberg angle, exploiting the *polarization asymmetry* of the Zb system. This quantity still remains the only Standard Model observable with a sizeble statistical tension in electroweak precision tests (around 3σ) and in the electroweak global fit, even including as a fixed parameter the latest Higgs mass measurement [39]. This puzzle emerged through the experimental tests at LEP, and it might be interpreted as an important window in new physics searches.

2.1 QCD Interpretation of the $pp \rightarrow Z + b$ Process

The interpretation of the production dynamics of an electroweak Z boson in association with b quarks follows essentially two indipendent approaches in perturbative QCD. Depending on the treatment and description of the b quark characteristics, it is possible to perform the matrix element calculation of the process using a *massive* approach, called 4-flavour scheme (4FS), or a *massless* one, for which we refer to the 5-flavour scheme (5FS).

2.1.1 The 4-Flavour Scheme

The 4-flavour scheme [40] is the interpretation of the $Z + b$ final state in the framework of an effective QCD theory, with $n_f = 4$ quark flavours involved in the computation of the running of the strong coupling constant α_S. The evolution dynamics is operated by the DGLAP equation, and the b quark is fully decoupled and treated as an indipendent quantity. The b in this approach is considered as a massive particle, and its mass is set to be of the order of magnitude of the hard scale of the process. Since the b quark mass, m_b, is much larger than the proton mass, the b quarks are always produced in pairs, and the tree level Feynman diagrams for the Zb production in pp collision comes out basically from the following processes (Fig. 2.1.1):

- $q + \bar{q} \rightarrow Z + b + \bar{b}$
- $g + g \rightarrow Z + b + \bar{b}$

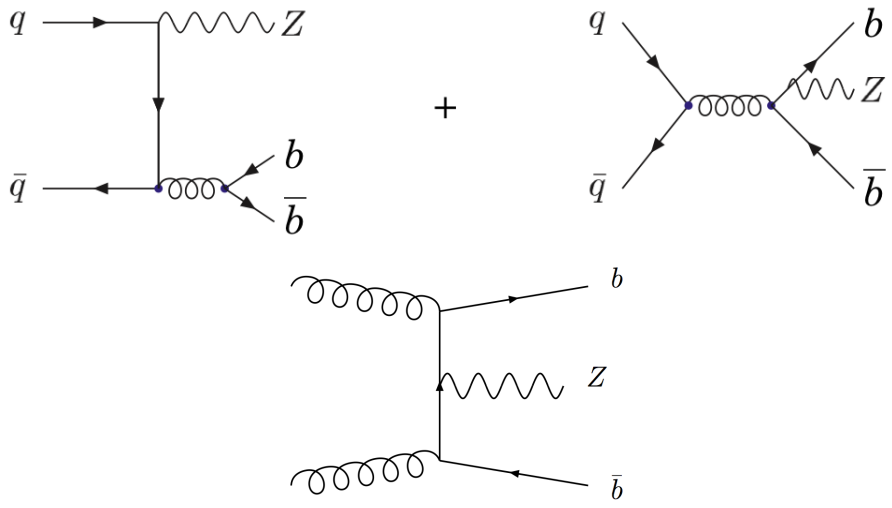

Figure 2.1.1: Top: Example of the tree level $qq \rightarrow Zb\bar{b}$ diagrams in the 4FS. Bottom: The $Zb\bar{b}$ production through gluons anihilation $gg \rightarrow Zb\bar{b}$ diagrams in the 4FS.

In this scheme, a b quark is emitted collinearly to the beam line, and so a $Z + b$ final state arise. The PDF of the b quark is set to zero, and its effect is taken into account by the *gluon splitting* contribution $g \rightarrow b\bar{b}$, and the non-zero b quark mass acts as a regulator for the *gluon splitting* collinear divergence. More generally, this

class of diagrams is also called Fixed Flavour scheme (FFS) in QCD, and they are valid at any order in perturbation theory under the condition that the heavy quark mass m_Q is about the same order of magnitude of the process under consideration. When the scale becomes large with respect to the m_Q, the scheme breaks up and the theory gives no more reliable results.

2.1.2 The 5-Flavour Scheme

Another class of QCD calculation scheme is the so called *Variable Flavour Scheme* (VFS). In this case the computation of the matrix element is done considering the quark masses as very small contributions, or even equal to zero. When the b quark mass is set to zero in the $Z+b$ calculation, the sub-class of the VFS is called 5-flavour scheme [41], indicating the fact that this time $n_f = 5$. The main difference with respect to the 4-flavour scheme is that now the gluon splitting contribution is included within a b parton density function. It is important to notice that gluon splitting may come either from an initial state radiation or a final state radiation. This effect has typically the form of a large contribution in the cross section calculation, like

$$\sigma_{Zb} \sim \alpha_S^n \cdot \log^n(\frac{M_Z^2}{m_b^2}),$$

where n is the perturbation order, and these *large logarithms* are resummed into a b PDF (for the initial state) or a fragmentation function (for the final state) using the DGLAP equations. Essentially, the 5FS scheme represents the process

$$b + g \rightarrow Z + b.$$

The leading order diagram for this process is depicted in Fig. 2.1.2.

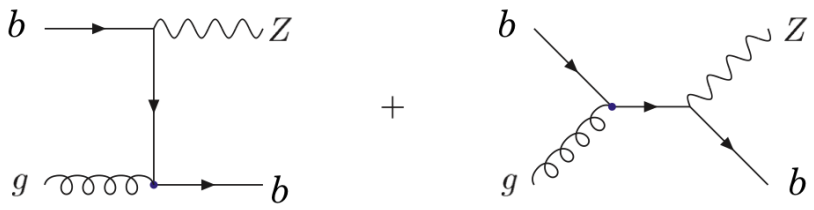

Figure 2.1.2: Leading order 5FS diagram for $gb \rightarrow Zb$.

The main obvious advantage of the 5FS is the simplification introduced by "transforming" a $2 \rightarrow 3$ ($pp \rightarrow Zb\bar{b}$) process into a $2 \rightarrow 2$ ($pp \rightarrow Zb$) process and thus an important simplification in the calculation at higher orders in pQCD. Within this interpretation of the process, the b quark is treated just like the other quarks inside the protons, i.e. it contributes to the total wave function of the proton in the initial state. A direct comparison of the 4FS and the 5FS, showing the advantages and disadvantages that emerge in different physics situation at LHC, is described in the next Section.

2.1.3 Effects of the Different Interpretations

The main difference between the two approaches, as introduced in the previous Section, is the simplification of the calculation in the 5-flavour scheme, i.e. in the assumption $m_b = 0$. It is found that, at leading order, the number of external legs in the diagrams sum is largely reduced with respect to the 4-flavour scheme. Another semplification is the reduction of the magnitude of the scale, and the absence of many effects induced by the heavy quark masses that in the 5FS only appear at high orders, and can be added as external effects. Moreover, probably the most sizeble effect is the resummation of the large logarithm arising from the gluon splitting inside the b-PDF. Anyway, in the end, this is of course an approximation: only the 4FS can take into account the full kinematics of the final state with all the quarks masses effects carried out in the QCD calculations. The 4FS is much more complicated from a computational point of view, but gives the full kinematic and brings to a more complete understanding and physically correct vision of the NLO. However, the definition of the two schemes is made in a way such as they are exactly identical when all orders in pQCD are accounted for. The differences come up at fixed orders: the different ordering in the expansions gives slightly different results, and so different predictions. Many studies have been carried out in order to compare the two schemes at a theoretical level for LHC energies, and the latest calculations suggest as a final result that the large logarithms resummed inside the b-PDFs give a contribution that is not as big as expected. The reason why this happen can be found in the kinematics of events under consideration. For each Q^2/m_b^2 terms, a phace space suppression factor mitigates it at all orders, and so the perturbative expansion of the 4FS remains reliable. The 4FS appears to be more adeguate to describe exclusive and differential observables, while the 5FS gives smaller uncertainties on inclusive observables. Detailed calculations can be found in dedicated studies, see for instance an introduction in Reference [42].

2.1.4 Overview of Available Calculations

Many calculations are available at NLO performed in the 4FS and in the 5FS for processes containing a b quark. In Table 2.1, a list of the currently available NLO processes for the two schemes is presented, while detailed calculation can be found in Reference [43]. No calculation at NNLO currently exist for the 4FS, while for the 5FS there are just three processes available: $pp \to Hb\bar{b}$, $pp \to Zb\bar{b}$, $pp \to Ab\bar{b}$ [44] .

4FS, $m_b \neq 0$, $pp \to$	5FS, $m_b = 0$, $pp \to$
$b\bar{b} + X$	$tj + X$
$b\bar{b}b\bar{b} + X$	$tW^{\pm} + X$
$t\bar{t}b\bar{b} + X$	$H(b\bar{b}) + X$
$tbj + X$	$W^{\pm}bj + X$
$Hb\bar{b} + X$	$Zbj + X$
$W^{\pm}b\bar{b} + X$	$W^{\pm}b + X$
$Zb\bar{b} + X$	$Zb + X$

Table 2.1: NLO available calculations at $\sqrt{s} = 7$ TeV for the $pp \to b + X$ process [43] [44].

Event Generators

Different techniques are employed in order to provide the $Z + b$ final state kinematic phenomenology at the LHC center of mass energy. Theoretical calculations are performed inside Monte Carlo *events generators* (for a complete description see for instance Ref. [45]), which give samples of events needed to make a comparison to data and to perform comparisons and tests between different models. It is important to notice that this comparison is possible only after taking into account the detector effects in data, using statistical techniques to correct and bring data back to the particle level. The chain of processes described by probability distributions calculated in quantum field theory is produced by sampling these distributions through pseudo-random event generation, using the Monte Carlo integration. The generators follow the whole chain of the physics process, starting from the proton collision, described as a collision of partons, then the hard scattering, resonances production (for example an intermediate vector boson), initial and final state radiation of photons, quark and gluons, eventual minor effects and multiple parton interactions. At this point the hadronization happens and the fragmentation is calculated in order to characterizes hadrons and their subsequent decays. There are many different event generators with different characteristics and some of them are essentially dedicate to particular events descriptions. In the following, event generators for $Z+b$ physics in the different schemes are presented. The discussion of the data to theory predictions comparison and details of the unfolding procedure are left to Chapter 6.

MadGraph

MadGraph [46] is an event generator dedicated to the matrix element calculation of a certain process. It does not completely simulate the events, and thus it needs an interfaced program to perform the parton shower after the computing of the matrix element. For this purpose, the most used interface is Pythia [47], and so the whole simulated process appears to be generated by the MadGraph+Pythia configuration. After defining the starting process that is needed to be simulated at parton level, i.e. $q\bar{q} \to Zb$, given the parameters that define the reaction, MadGraph calculates the Born amplitude of the $2 \to n$ process, eventually by adding jets to the final state. After the generation, the event is stored in a dedicated format [48], and then it is ready to be read by Pythia to calculate the final state and carry out the parton shower, clustering partons into jets, hadronization and decays. The interface between the matrix element calculation and the parton shower is a delicate task, and a complete description of the problem can be found in Reference [49]. MadGraph is a possible choice to generate samples of the tree-level $Z + b$ matrix element calculation. Currently, both the 4FS and the 5FS are available, used and tested, using the MadGraph+Pythia system. The comparison of the MadGraph prediction for the two flavour schemes with 8 TeV LHC data are performed, to our knowledge, for the first time in this thesis work (Chapter 5).

*a*MC@NLO

The automatic event generator at next to leading order, aMC@NLO [50], allows to calculate *any* amplitudes at NLO accuracy, and includes the matching to the parton shower to produce events ready for the hadronization. This is a matrix element-only

generator, and the basic idea is to use the MC@NLO approach for the calculations [51]. The advantage is a very fast events generator, computing automatically all the needed quantities for the process under consideration. This is the only prediction at next-to leading order for the $Z + b$ case, and only the massive 4F scheme actually is available.

SHERPA

A different choice for the generation of b quarks in association with vector bosons can be made using the SHERPA event generator [52]. This program provides a full descriptions of the pp collision. Unlike MADGRAPH, SHERPA relies on its own parton shower technique, and accounts for the merging with the matrix element itself. For this reason it may be considered as a *standalone* program for event generation. It is structured in a very modular way, easy to customize with respect to the different characteristic of the processes parameters. Nevertheless, the matrix element calculation is computed with SHERPA with a maximum of 8 partons in the final state. On the other hand, the accuracy is again at leading order. Both 4FS and 5FS are available in SHERPA, so this will be another generation considered in terms of the data comparison in the next Chapters.

POWHEG

POWHEG [53] is another approach to calculate NLO matrix elements matched with the parton shower. The same approach is implemented in several codes. For instance, the POWHEG BOX [54] starts from the Born level diagram $B(\Phi_B)$ in its phace space Φ_B, calculates the amplitude, the colour and the spin correlations, the colour structure and the finite part of the real part of the virtual correction $R(\Phi_R)$ to the process under consideration. Given these informations, it automatically builds the soft and collinear radiations using dedicated Sudakov form factors Δ,

$$\mathrm{d}\sigma = \bar{B}(\Phi_B)\mathrm{d}\Phi_B \left[\Delta(p_T^{min}) + \frac{R(\Phi_R)}{B(\Phi_B)}\Delta(k_T(\Phi_R))\mathrm{d}\Phi\right],$$

where k_T is the momentum of the emitted particle. The NLO contributions are kept inside the $\bar{B} = B + \left[V + \int \mathrm{d}\Phi R(\Phi)\right]$ term, considering the virtual V and the real R part of it. After the generation, the POWHEG output needs to be interfaced to a chosen showering program, for example PYTHIA6. The $Z + b$ process inside POWHEG is calculated only in the 5F scheme.

2.2 Physics of $Z + b$ final state

The physics of the $Z+b$ system at the LHC covers a wide range of different interesting phenomena, either belonging to a Standard Model nature or a *New Physics* nature. At the LHC, the search of new particles and forces is one of the main tasks of the experiments purposes. From a large variety of theoretical interpretations, New

Physics is expected to appear at the LHC energy, and it is needed to account for the Standard Model open issues discussed in Chapter 1, first of all the Higgs vacuum instability. Many theories beyond the Standard Model involve b quarks and Z bosons in the final state, so it is of extreme importance to have a very accurate description of the Standard Model Zb process in order to model as precisely as possible this background in searches. In this Section, the most relevant processes involving a $Z + b$ signature are presented for both the Standard Model physics and the Beyond the Standard Model physics.

2.3 The W mass measurement

The mass of the W boson, m_W, is a key parameter of the Standard Model. A precise measurement of this quantity significantly improves our understanding of the electroweak symmetry breaking, in the light of the Higgs boson discovery, and can be used in the global electroweak fit. The comparison of its directly-measured mass with the Standard Model prediction is a powerful test of the model. The value of the m_W has been previously measured in $p\bar{p}$ collisions at the Tevatron [55] and in e^+e^- annihilations at LEP [56]. The W mass measurement at the LHC involves additional experimental challenges with respect to the previous measurements. The high pile-up environment of the high luminosity LHC runs will produce much more energy deposits uncorrelated to the W boson giving important differences in the uncertainties related to the resolution. Moreover, the production mechanism of the W boson is different between the Tevatron and the LHC. In the framework of a measurement of the W mass in the $W \to \ell\nu$ decay channel, the knowledge of the W momentum is important while considering the transverse mass of the boson, a function of the lepton p_T through the relation

$$m_T = \sqrt{2p_T^{\ell} E_T^{miss}(1 - \cos\,\Delta\phi^{\ell\nu})},$$

where p_T^{ℓ} is the lepton momentum, E_T^{miss} is the missing transverse energy related to the neutrinos and $\Delta\phi^{\ell\nu}$ is the azimuthal angular difference between the lepton and the neutrino directions. Since the W boson p_T is mostly affected by the p_T^{ℓ} measurement, that is sensitive to detector and resolution effects, a possible solution to improve the precision of the measurement is to study the better-known Z boson momentum measurement. Following this approach, an important issue to be considered is the difference in the production mechanism between the W and the Z, related for instance to the contribution given by the process

$$b\bar{b} \to Z + b\bar{b},$$

which could be not negligible. The different b quark PDF definition in this particular case, i.e. the choice of the flavour scheme in the QCD calculations inside generators, can enlight differences, leading to a theoretical systematic uncertainty on the boson momentum parametrization that can be important and difficult to evaluate from a theoretical point of view. Furthermore, the use of a different b quark PDF definition can in principle have an effect on the Z boson momentum shape. For these reasons, a systematic measurement of the $Z + b$ production in the 4FS and 5FS can give important informations on the study of the theoretical uncertainties related to the Z boson parametrization and thus in the W mass measurement at LHC.

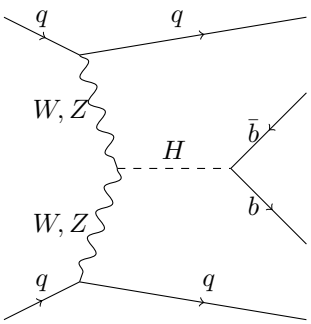

Figure 2.4.1: Higgs boson production in vector boson fusion at LHC decaying into b quarks pairs .

2.4 The Standard Model $H \rightarrow b\bar{b}$ decay

According to the electroweak symmetry breaking, the Standard Model Higgs boson couples with the other particles with a strength that is proportional to the decaying particles masses. For this reason, the most frequent decay mode for the 125 GeV Higgs is the $H \rightarrow b\bar{b}$ channel, which has the highest cross section. On the other hand, this channel is very challenging from an experimental point of view in a high luminosity hadronic collider. In such colliders, like the LHC, the $b\bar{b}$ final state is hidden by all the QCD background ruins survived after the hard collision. Nevertheless, very strong efforts to perform this measurement have been made, and finally the contribution of this channel has been added to the final combination in the Higgs discovery [14]. The associated production of a Z boson and heavy flavour quarks is one of the main backgrounds in the Standard Model Higgs studies in the low mass region. The $Z+b$ final state with a b quark missing in the reconstruction of the final state can mimic the Higgs boson signature in the associated production with a Z, $pp \rightarrow Z^* \rightarrow ZH \rightarrow \bar{b}b\bar{\ell}\ell$, and in the vector boson fusion process $pp \rightarrow ZZ \rightarrow H\bar{b}b$ (Fig. 2.4.1).

2.4.1 Supersymmetric Higgses

Supersymmetric extensions of the Standard Model always involve the presence of more than one Higgs doublet. For example, in the MSSM more than one Higgs boson $(h^{\pm}, A, H)$ interacts with SM particles. In this framework, *SUSY* Higgses might have a large enhancement in the strength of couplings to the b quarks, produced in association with a Higgs boson and decaying to muons or taus pairs [57]. The 5FS scheme in particular has the same exact signature as the Z boson case as it is shown in Fig. 2.4.2, so the $bg \rightarrow Zb$ for this channel is the main background. The strength of the coupling is proportional to the $\tan\beta$ parameter, namely the ratio between the supersymmetry vacuum expectation value and the Standard Model one. The previously cited enhancement in the h production would be proportional to $(\tan\beta)^2$, so this channel can be a promising way to search for new particles at LHC. Many searches have been already performed but no supersymmetric Higgs signal has been found yet [58].

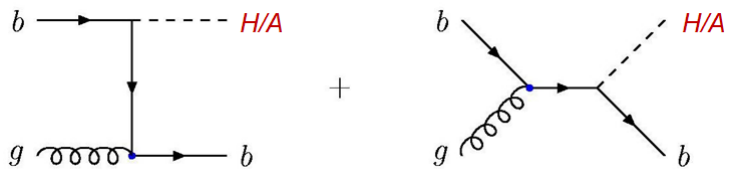

Figure 2.4.2: Supesymmetric Higgs bosons H/A produced in association with a b quark .

2.4.2 4th Generations of Heavy Quarks

Many models that attempt to extend the number of generations of quark and leptons have been formulated in the last decades, adding a new doublet of heavy quarks $b't'$ [59]. These 4th generation quarks are heavier than the Standard Model ones and they couple to vector bosons and decay into standard b, t quarks. Of particular relevance for the study discussed until now is the b' decay

$$b' \rightarrow Zb,$$

which clearly mimics the signature of the Standard Model bZ production (Fig. 2.4.3). Searches of such new particles have been performed by ATLAS and CMS [60] [61], excluding their existence by analyzing the shape of many kinematics observables, enhanced by the eventual b' production, such as the invariant mass distribution of the bZ system (which may have a peak around the new particle mass). Still some models that involve these new quarks are not fully excluded by experiments (vector-like T, B, heavy gluons [62]), so it is of great importance to dominate the Zb process in order to maximize the discovery power.

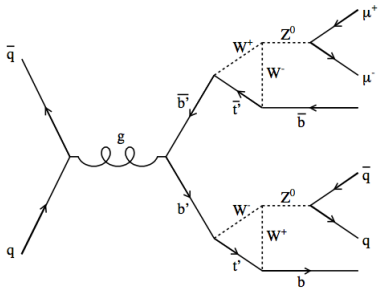

Figure 2.4.3: 4th generation of b quarks (b') decay chain in $q\bar{q}$ annihilations .

2.5 The Zb Polarization Asymmetry

Electroweak precision tests carried out by LEP experiments at CERN in electron-positron annihilations [63], provide an outstanding set of experimental confirmation of every part of the Standard Model. The four experiments located at LEP took data by the electron-positron collisions near the Z pole in the first phase, and then over the threshold of WW production in the second phase. In the framework of the Z boson properties measurements, the measurement of the axial and vector

coupling of the Z to fermions, namely g_A and g_V, is strongly important and is indirectly accessible by evaluating the forward-backward asymmetry of the fermion f production

$$A_{FB}^f = \frac{\sigma_F^f - \sigma_B^f}{\sigma_F^f + \sigma_B^f},$$

defined as the ratio between the forward and backward cross section,

$$\sigma_F = \int_0^1 \frac{\mathrm{d}N}{\mathrm{d}\cos\theta}\mathrm{d}\cos\theta \ , \ \sigma_B = \int_{-1}^0 \frac{\mathrm{d}N}{\mathrm{d}\cos\theta}\mathrm{d}\cos\theta,$$

with respect to the polar angle θ, defined as the beam line angle. It turns out that exploiting the cross section formula, this quantity (easily measurable at colliders by events counting) is related to the Z boson coupling to fermions at its pole:

$$A_{FB}^f = \frac{3}{4}\mathcal{A}_e \cdot \frac{g_A g_V}{g_A^2 + g_V^2} = \frac{3}{4}\mathcal{A}_e \cdot \mathcal{A}_f.$$

The quantity $\mathcal{A}_f$ is mostly sensitive to the b quarks and thus to their coupling to the Z through the effective Weinberg angle $\sin^2\theta_W^{(b)}$. Measurement techniques employed by the LEP experiments have been carried out exploiting different strategies: lepton tagging, D meson from the B hadron decay reconstruction or evaluating the *jet charge*. Indipendent measurements were also performed in the same period by the SLC collider using polarized electron beams. The final LEP result for A_{FB}^b was presented by combining all the four experiments, giving an asymmetry of

$$A_{FB}^b = 0.0992 \pm 0.0016$$

showing a discrepancy within the Standard Model prediction at 2σ [64]. Recently, with the discovery of the Higgs boson, an update of the electroweak global fit was performed, using all the most precise measurements ever made for each Standard Model parameter, and by inserting the value of the new boson mass measured in 2012 (left as a free parameter before the discovery) [39]. The result of the global fit shows that Standard Model predictions spectacularly improve the most important variables precision, such as the W mass and the t mass, decreasing their total errors of tens of MeV, but enlarges the discrepancy of the b asymmetry to 2.9 σ (Fig. 2.5.1). It is clear that this is the key Standard Model parameter to be measured at LHC after the Higgs boson discovery and the measurements of its properties. Many strategies on how to measure the asymmetry in pp collisions are under investigation, and the $Z + b$ precision understanding is the central ingredient in order to perform such measurement. A possible strategy for an LHC measurement for the polarization asymmetry is proposed by the study summarized in Reference [65]. According to this proposal, the angle between the Z boson direction and the b quark direction, in the rest frame of the Z boson, can be employed to discriminate the different polarization state (helicity amplitudes) of the Zb system to be used as input for the asymmetry calculation. If this tension will remain, or even get bigger, it will be the first indirect sign of physics beyond the Standard Model predictions. It is interesting thus to spoil the different theories that explain such divergencies with new phenomena.

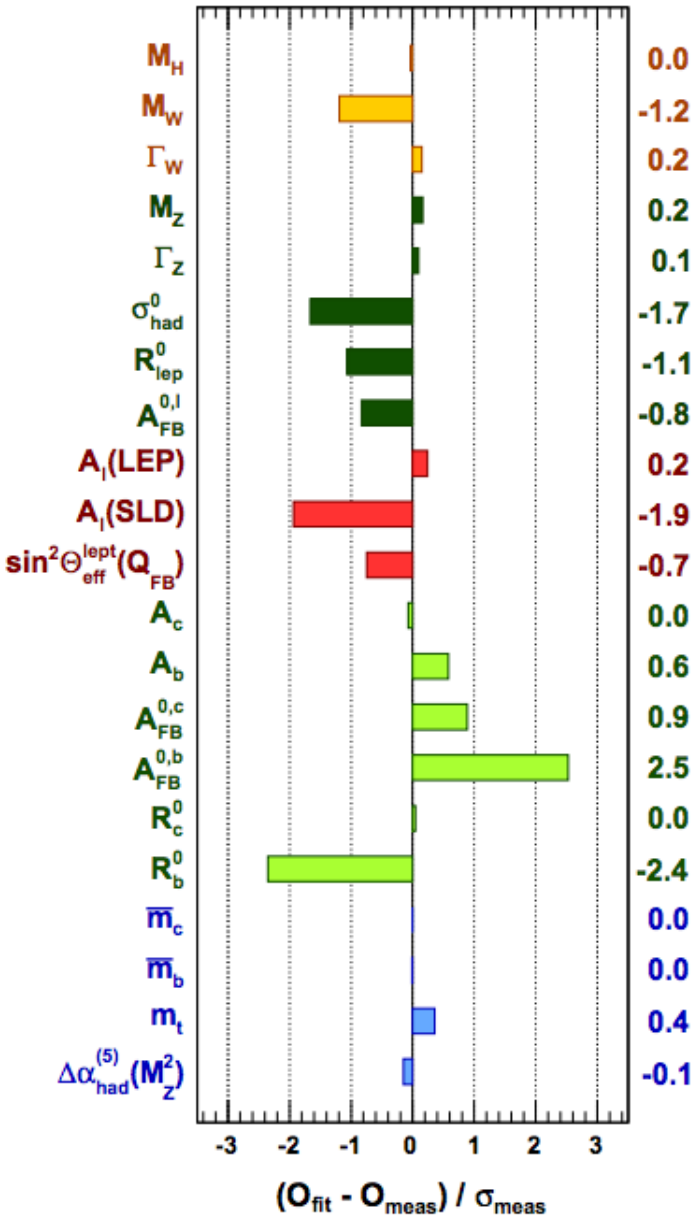

Figure 2.5.1: The Electroweak Global Fit result for the key Standard Model parameters (with $M_H = 125$GeV). On the horizontal axis, the difference between the value extracted from the fit and the corresponding measured one is shown in terms of deviation from the theoretical expectations. As it can be seen, the most impressive disagreement comes from the the $A^{0,b}_{FB}$ asymmetry results..

2.5.1 The *Beautiful Mirrors* Models

An intriguing way to explain the asymmetry in the b quark production measured at LEP is to assume that some new mechanism would enhance the Zb coupling in an extended version of the Standard Model. The class of theories designed for this purpose is called *Beautiful Mirrors* Models [66]. The principal idea is the introduction of a doublet of left/right vector-like quarks in SU(2), ω and χ, called *mirror quarks*

$$\Psi_{L/R} = \begin{pmatrix} \omega_{L/R} \\ \chi_{L,R} \end{pmatrix}^{-1/3}_{-4/3}.$$

These new quarks are able to mix together with the standard b quarks producing an increase of the couplings that will lead to the measured A^b_{FB} value. In order to be consistent with the electroweak precision tests measurements, these quarks must interact only with the third generation, and assuming that, every Standard Model results is kept safely. Also the symmetry breaking of SU(2)xU(1) will not enter in this extension, so the mirror quarks masses are generated by the Higgs term of the form $y_{L/R}\bar{\psi}\phi b + h.c.$. After the diagonalization of the mass matrix for the new quarks, the connection to the b quark asymmetry will be enlighted by the *new* Zb vertex calculation, leading to a correction of the left/right coupling of the b g_A, g_V by a shift δ

$$\delta g_{L/R} = \frac{g}{2\cos\theta_W}(1 - K_{L/R})$$

where K is a combination of the $b-\omega$ and $b-\chi$ mixing angle, and the ω, χ masses can be obtained by the standard Yukawa couplings. In particular, the right component's sine of this mixing, $\sin^2\theta_R$, is primarily responsable of the A^b_{FB} shift.

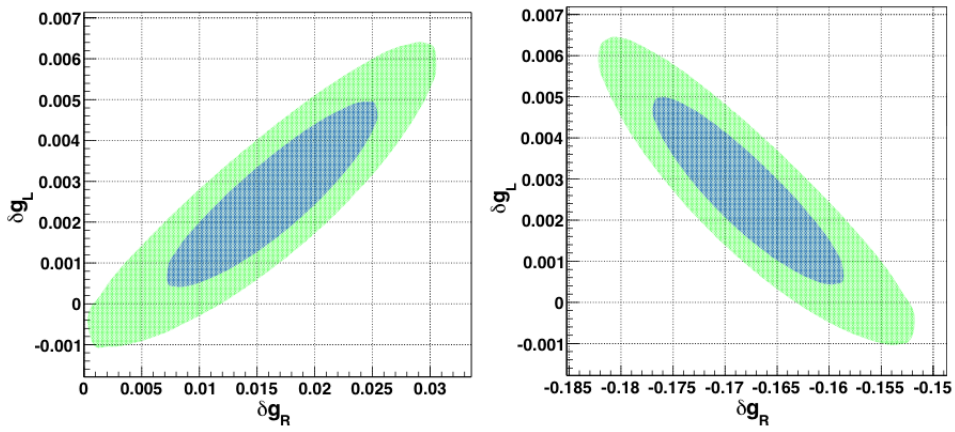

Figure 2.5.2: The couplings parameter space. The blue (green) shaded regions correspond respectively to 1σ (2σ) agreement with the best fit shifts δ.

At the LHC the mirror quarks can be produced in QCD by pairs, and through electroweak interaction singly. In the assumption of no mixing between $\chi - \omega$, the ω mass eigenstates, $b_{1,2}$, can decay into Z or H bosons only through flavour-changing neutral current processes

$$b_{1,2} \to Z(H)b,$$

with a branching ratio of 52%(Z) and 48%(H) for a Higgs mass of 125 GeV and $b_{1,2}$ mass of 500 GeV. For masses up to 1 TeV, one expects to have 10 signal events

of $b_{1,2} \to Zb$ with 300 fb^{-1} at $\sqrt{s} = 14$ TeV over 3 background events. Within this assumption, it is possible to measure the $\sin^2\theta_R$ at 20% level [67]. Using the modified couplings δg inside the A_{FB}^b expression, the discrepancy (if this is a physical one) would be eventually cured (Fig. 2.5.2).

Chapter 3

The LHC and the CMS experiment at CERN

In this Chapter I will describe the Large Hadron Collider (LHC) that makes proton beams interact in the underground laboratories at CERN, providing the 8 TeV collision data used in this thesis. The accelerator principles and characteristics will be described in the first part of the Chapter. The results of the proton-proton collisions are measured by the sub-detectors of the Compact Muon Solenoid (CMS) which is described in details in the second half of the Chapter. Finally, a brief description of the trigger structure will be discussed.

3.1 Large Hadron Collider

The LHC [68] is the most powerful and large particles collider ever built. It is located in the underground of the French-Swiss national border near Geneva (Fig. 3.1.1). The LHC was built in place of the previous electron-positron accelerator LEP, in operation until the end of 2000, and its circumference occupies its 27 km underground tunnel. It is designed to accelerate proton[1] bunches at 99,9999991% the speed of light with an ultra cold superconducting magnet system reaching design center of mass energy of 14 TeV. The intensity of the LHC proton beams is described by the machine luminosity:

$$L = \frac{k \cdot f \cdot N_p^2}{4\pi\rho^2},$$

where k is the number of proton bunches, f is the revolution frequency of the bunches inside the ring, N_p is the number of protons and ρ is the mean quadratic radius of protons spatial distribution in a plain that is orthogonal to the beam direction. The luminosity is related to the cross section σ of a given process by the relation

$$\frac{\mathrm{d}N}{\mathrm{d}t} = \sigma \cdot L,$$

where N is the number of events. The main quantity used to express the statistics recorded by an LHC experiment is the integrated luminosity $\int L dt$ expressed in barn^{-1} [2]. The design instantaneous luminosity of the LHC is $10^{34}\mathrm{cm}^{-2}\mathrm{s}^{-1}$. The

[1] The LHC also accelerates heavy ions in a dedicated physics program, but this specific case is not relevant for the purposes of this thesis.

[2] 1 barn = $10^{-28}\mathrm{m}^2$

Figure 3.1.1: The aerial view of CERN, in the border between France and Swiss where the 27 km tunnel of LHC is located in the underground, following the yellow line.

LHC accelerated protons at a center of mass energy of 7 TeV in 2011 and 8 TeV in 2012, and it is planned to reach the design energy and luminosity in the next years. Before reaching the point where the collision takes place, proton beams experience a long series of subsequent accelerations. First, protons are accelerated in the LINAC and reach an energy of 50 MeV, then they raise their energy to 1.4 GeV by passing through the Proton Synchrotron (PS) and finally, before entering in the LHC, they enter inside the Super-Proton Synchrotron (SPS) that injects them to the LHC at 450 GeV. Once they reach the the last stage of the acceleration process, protons are further accelerated to the main energy (8 TeV) through 400 MHz radiofrequency cavities, that increase the proton energy of 0.5 MeV per turn. Two separated cavities are employed with two magnetic fields in order to accelerate in the opposite direction the two same-charged beams. The LHC is equipped with 1232 superconducting 14.2 m long Niobium-Titanium dipole magnets, cooled at a temperature of 1.9 K by means of super-fluid Helium, that creates a bending magnetic field of about 8.3 T. The magnets are placed in eight curved sections around the LHC perimeter. The 4 TeV beams are made of up to 2808 bunches with a designed number of protons of 10^{11} each, and they have a spread in the z direction of about 15μm. The high luminosity of the LHC is possible thanks to the high frequency of the bunch crossings coming to a maximum value of 40 MHz, corresponding to a collision every 25 ns. In 2012, the LHC worked with a bunch crossing rate of 50 ns. There are four interaction points where the collisions are recovered by four different experiments. The two multi-purpose experiments, ATLAS [69] and CMS [70], and two dedicated experiments, ALICE [71] and LHCb [72], respectively designed for heavy ions physics and flavour physics.

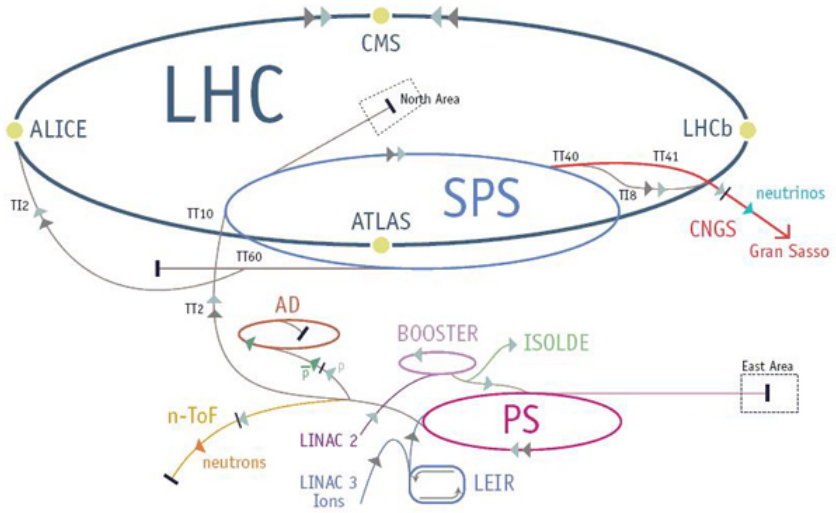

Figure 3.1.2: The LHC injection chain.

3.2 Compact Muon Solenoid Experiment

The Compact Muon Solenoid (CMS) experiment is one of the two general-purpose detectors installed at the LHC. The general idea behind the project is to deeply exploit the discovery potential that comes from the energy and luminosity performances of the LHC, and thus the physics measurements accessible with CMS range from the Higgs boson mass, the Standard Model fundamental parameters, perturbative QCD, together with the extremely vast possibilities of the New Physics theories and their eventual new particles associated. Dealing with so many different measurements means that the experiment must be able to measure with the highest precision the energy and momentum of electrons, photons, muons, hadronic and electromagnetic showers and missing energy. In order to cover such a variety of physics phenomena, many characteristics are required for CMS:

- excellent inner tracking system able to measure particles momentum from low to high values
- optimal resolution in electromagnetic showers measurements for high energy electrons and photons
- hermetic hadronic calorimeters for a good reconstruction of the missing energy and jets
- highly performing muon system and strong magnet field to the best momentum measurement for high energy muons.

The CMS experiment fully satisfies these requirements with a series of dedicated concentric sub-detectors, inserted in its cylindric shell, symmetric around the beam direction, with a radius of 7.5 m, a length of 22 m, and a weight of 12 500 tons. In Fig. 3.2.1, an insight of CMS shows the various sub-detectors composing the experiment: from the inner part of the cylinder, corresponding to the nearest distance to the beam line and moving outwards, it shows an inner silicon tracker system, a crystal electromagnetic calorimeter, an hadronic calorimeter, and the superconducting

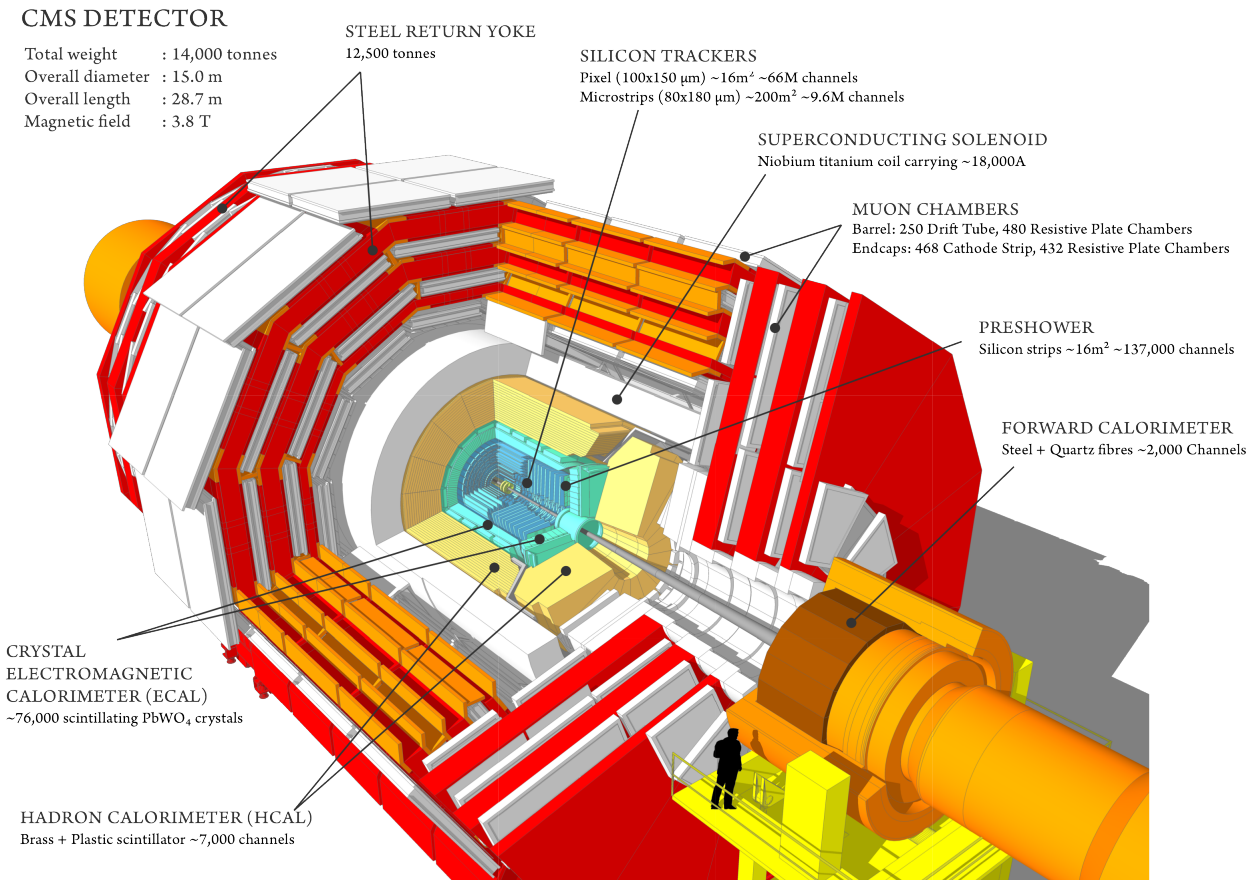

Figure 3.2.1: Insight of the Compact Muon Solenoid, showing all the different sub-detectors around the beam line and forming outward the cylindrical shape layers of the experiment.

solenoidal magnet with its return yoke, where the muon drift chambers are installed. Given the geometry of the experiment, two sections can be identified: the *barrel*, corresponding to the central part of the cylinder, and the *endcaps*, corresponding to the two hermetic discs closing the central body at their two extremities.

3.2.1 Coordinates System in CMS

The cartesian coordinates system in CMS has the origin coincident to the interaction point, in the symmetry center of the cylinder. The x axis points towards the center of the LHC ring, the y axis points outwards to the surface of the cylinder, and the z axis is set along the beam axis. A natural frame for a cylindrical symmetry is identified by the three variables (r,η,ϕ) with r defined as the radial distance from the beam line, ϕ as the azimuthal angle with respect to the z axis, and

$$\eta = \text{-ln}\left[\text{tg}\left(\frac{\theta}{2}\right)\right]$$

is the *pseudorapidity*, with θ defined as the polar angle with respect to the z axis. The pseudorapidity plays a fundamental role in hadron colliders. It can be derived as the ultra-relativistic limit of the *rapidity*:

$$y = \frac{1}{2}\ln\left(\frac{E + p_L}{E - p_L}\right)$$

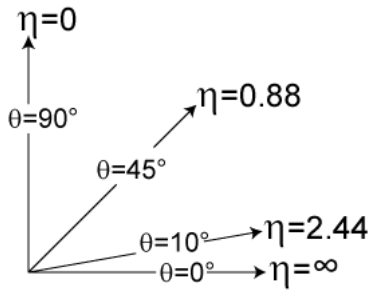

Figure 3.2.2: The pseudorapidity variation as a function of the polar angle θ.

where p_L is the longitudinal component of the particle momentum and E is its energy. The rapidity difference Δy is invariant under Lorentz boost transformation along the axis direction, and for this reason it is appropriate to describe the relativistic kinematics of the high energy particles produced in the pp collisions. The pseudorapidity has an asymptotic behaviour at $\theta = 0$ and is zero for $\theta = \pi/2$ (Fig. 5.6.3). In the protons collision, the total four-momentum conservation requires 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 a priori, since it would require the full knowledge of the PDFs. For this reason in order to be able to apply the four momentum conservation, only the transverse motion is considered, since the sum of the transverse momenta of the proton constituents is zero. As a consequence, in order to fully describe the particles kinematics, the radial and angular informations coming from the η, ϕ coordinates are associated to the transverse kinematic variables:

$$p_T = p \cdot \sin\,\theta$$

$$E_T = E \cdot \sin\,\theta$$

3.2.2 Inner Tracker System

The inner tracker [73] is the closest detector with respect to the beamline, and it is designed to reconstruct charged particle tracks and measure with very high efficiency their momentum. It covers the pseudorapidity range of $|\eta| < 2.4$ and is able to reconstruct electron and muon tracks with a 95% efficiency. The challenge of this detector is to face off two problems: the material budget constraints and the ability to reach the high performances in track reconstruction in the very high pile-up (and radiation) environment that characterizes the LHC high luminosity program. To deal with these characteristics, it was decided to build, for the first time in high energy physics, a tracker that is completely based on silicon materials. There are three zones of the CMS tracker system with different characteristics, identified by the distance to the interaction point r:

- $r \sim 10$ cm : the tracker is fully covered by pixel detectors of size 100×150 μm^2. The occupancy of the detector is 10^{-4} per pixel per bunch crossing. This zone, the closest to the interaction vertex, is the highest particle flux one, estimated to be about 10^7particles/s.

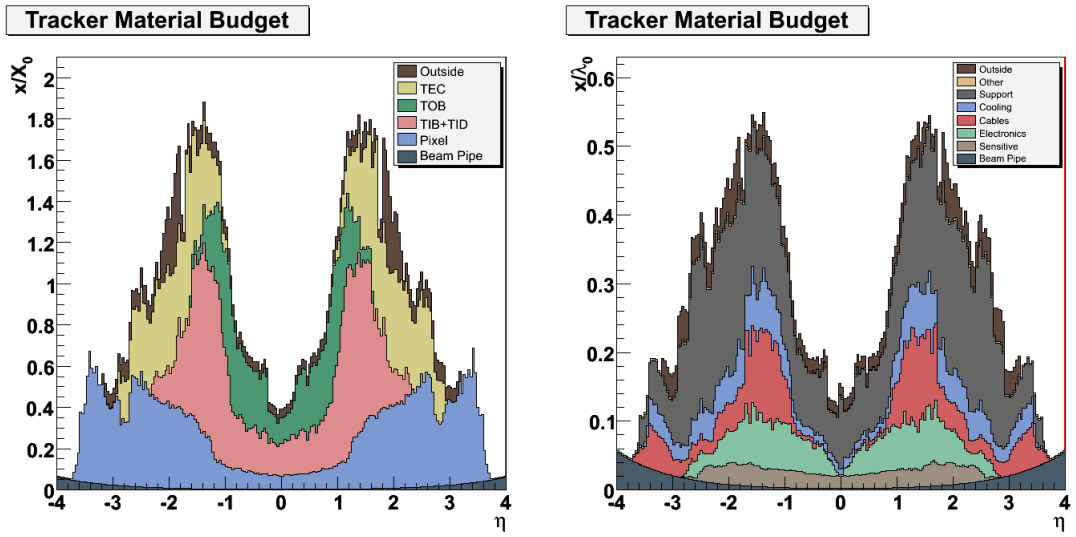

Figure 3.2.3: Radiation length (on the left) and interaction length (on the right) of the tracker as a function of η.

- $20 < r < 55$ cm : in this zone the detector is made of silicon microstrips of 10 cm $\times$ 80 μm
- $r > 55$ cm : the detector is made of larger-pitch silicon microstrips with a size of 25 cm $\times$ 180 μm

The pixel detector (66 million of pixels) is composed by three cylindrical layers in the barrel region, placed respectively at $r = 4.7$, 7.3 and 10.2 cm, and two closing end-discs, at $r = 6$ to 15 cm. In this configuration each charged particle releases at least two hits in the pixels. The whole pixel detector is covered by a shield of microstrips detectors (9.6 million), schematically divided into two barrel sections. In the inner barrel ($20 < r < 55$ cm) there are four layers of microstrips, while in the outer barrel ($r < 110$ cm) there are six layers. Furthermore, three additional disc layers have been inserted between the inner barrel and the endcaps. The endcaps are composed by nine layers of discs, which are double sided for the first, the second and the fifth one. The pixel and microstrip detectors composing the CMS tracker are shown in Fig. 3.2.4-3.2.4. The tracker depth is evaluated as a function of η with simulation studies in terms of two quantities: the radiation length X/X_0[3] and the interaction length λ/λ_0[4]. The distribution of these two quantities, enlightening the several contributions coming from the different parts of the tracker is presented in Fig. 3.2.3.

3.2.3 Electromagnetic Calorimeter

The performances in electrons and photons energy measurements is of prime importance for the CMS physics program. A highly performing electromagnetic calorimeter (ECAL) is necessary to substantially increase the physics potentials in photon and electron final states, for example in the golden - diphoton channel of the Higgs

[3] X_0 It is the mean distance over which a high-energy electron loses all but 1/e of its energy by bremsstrahlung and 7/9 of the mean free path for pair production by a high-energy photon.

[4] λ_0 is defined as the mean free path of hadrons before interacting with the tracker material

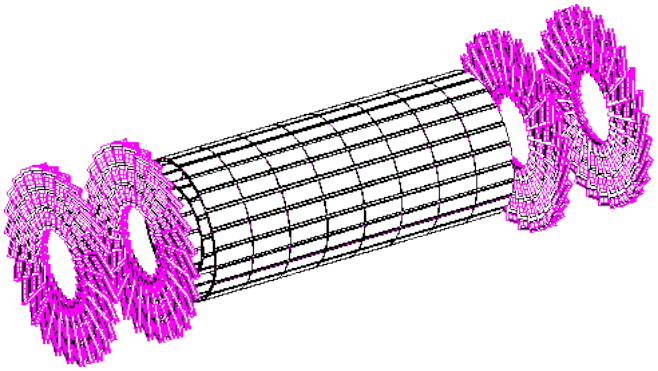

Figure 3.2.4: Schematic view of the pixel detector.

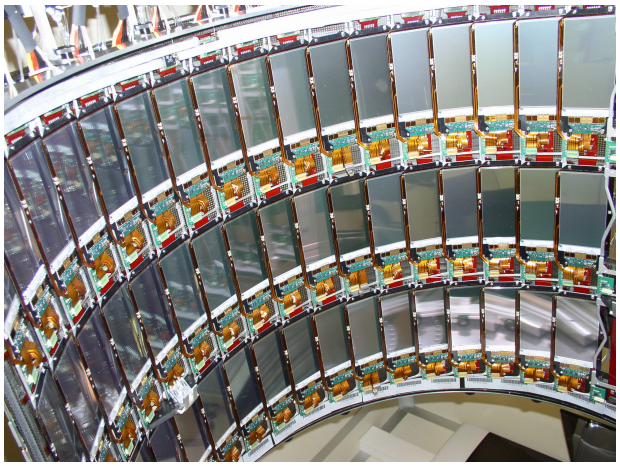

Figure 3.2.5: Layer 5 section of the silicon microstrips detector in the inner barrel region.

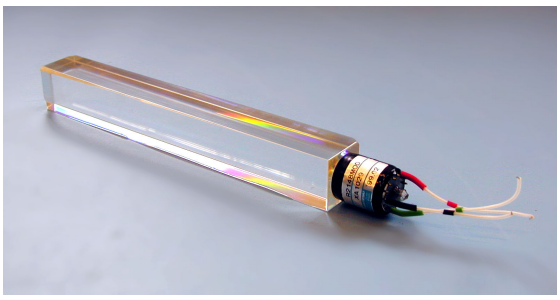

Figure 3.2.6: An example of a $PbWO_4$ crystal composing the ECAL calorimeter.

decay $H \rightarrow \gamma\gamma$. The CMS ECAL [74] (Fig.3.2.7) is a hermetic and homogeneous calorimeter made of about 75000 lead tungstate ($PbWO_4$) scintillating crystals divided into a barrel region ($|\eta| < 1.479$) and two discs (endcaps, up to $\eta = 3$). The choice of the $PbWO_4$ crystals comes from several factors. One of the most important features of this inorganic crystal is the very short radiation length X_0, estimated [5] to be about 0.89 cm, which permits to build a compact detector suitable to the CMS magnet (Subsec. 3.2.6). The lateral shower containment is efficiently ensured by the small Molière radius $R_M = 2.2$ cm, and the light emission is fast enough to work with the highest LHC bunch crossing rate of 25 ns: the emission time is about 5 ns for the primary scintillation and 15 ns for the secondary one. Finally, the radiation-hard is fair enough to sustain years of high luminosity work conditions and can be corrected using monitoring systems. In order to amplify the light yield, ECAL takes advantage of silicon avalanche photodiodes (APD) in the barrel, and vacuum phototriodes (VPT) in the endcaps. The barrel, made of 18 supermodules, has a granularity of 360 crystals in the ϕ direction and 170 along η. Crystals have a quasi-projective geometry and they are 23 cm long (Fig. 3.2.6) with 25.8 X_0. In the endcaps, the crystals (22 cm and 24.7 X_0) are gathered together in 5x5 units called *supercrystals*. In front of the endcaps, in the region $1.653 < |\eta| < 2.6$ a preshower detector is inserted, made of two layers sampling calorimeter filled with lead and silicon. The CMS ECAL calorimeter energy resolution, given by the expression

$$\frac{\sigma_E}{E} = \frac{S}{\sqrt{E}} \oplus \frac{N}{E} \oplus C,$$

has been measured by means of test beams, giving:

- $S = 2.8\%$ (stochastic term),
- $N = 124$ MeV (noise term),
- $C = 0.3\%$ (constant term).

3.2.4 Hadronic Calorimeter

The hadronic calorimeter [75] (HCAL) is designed to measure the energy and momentum of hadronic particles, and to give also a measurement of the missing energy of the events. The requirements for this detector are a good energy resolution, high

[5] $X_0 = \frac{716.4 \cdot A}{Z(Z+1)\log(287/\sqrt{Z})}\text{g}\cdot\text{cm}^{-2}$, where A is the mass number and Z is the atom number.

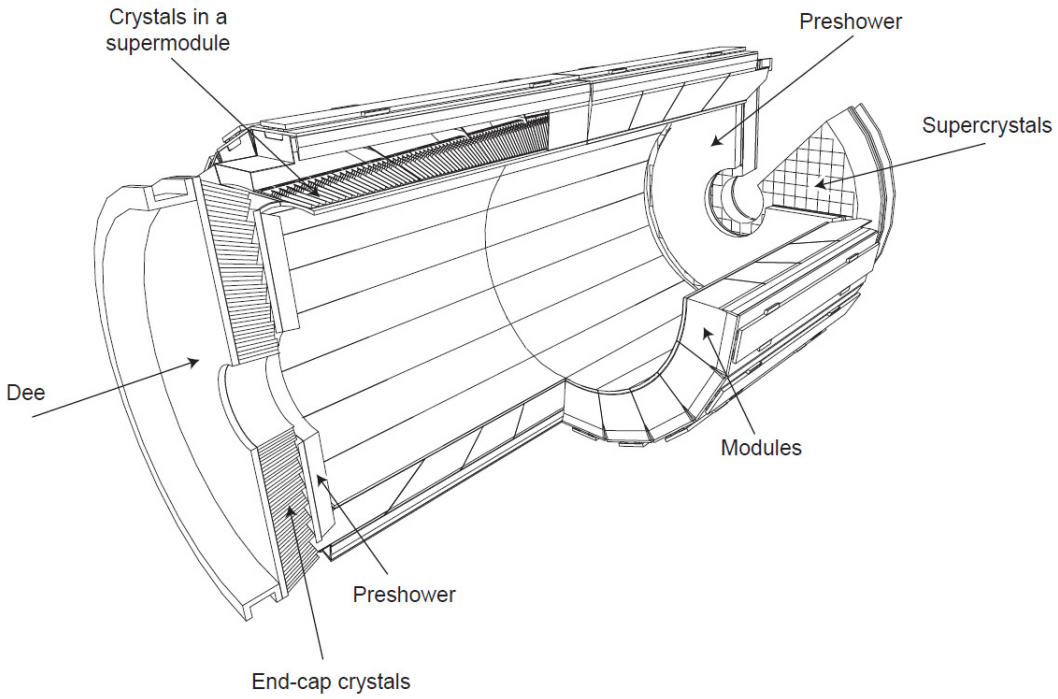

Figure 3.2.7: The ECAL calorimeter (section) showing the different modules in the barrel, the endcaps and the preshower.

hermeticity and transverse granularity. The CMS HCAL is divided into two parts: the central hadronic calorimeter, covering the range $|\eta| < 1.3$ in the barrel and $1.3 < |\eta| < 3$ in the endcaps, and the forward calorimeter (HF), placed outside the magnet yoke, 11 m away from the interaction point, on both sides ($3 < |\eta| < 5$). The HCAL is a sampling calorimeter, made of an absorber of brass alternate with plastic scintillators as active materials, for a transverse granularity of $\Delta\phi \times \Delta\eta = 0.087 \times 0.087$. The HF, designed to sustain very high particle multiplicities and radiations, is structured with a sampling strategy based on iron and quartz fibers. The HF fibers have two different lengths in order to isolate the electromagnetic component of the hadronic shower, and its granularity is $\Delta\phi \times \Delta\eta = 0.175 \times 0.175$. The HCAL energy resolution, for a π^0 of a given energy E, is

$$\frac{\sigma_E}{E} = \frac{100\%}{\sqrt{E}} \oplus 8\%$$

with a total thickness of 10 interaction lengths λ_i. In order to contain the most energetic showers, an extra layer of active material was added behind the solenoid, which increases the total effective thickness by about 3λ and improves by 10% the energy resolution for 300 GeV pions. A schematic view of HCAL is given in Fig.3.2.8.

3.2.5 Magnet System

The CMS magnet field [76] is produced by a relatively small solenoid, designed to give, within the compact size of the detectors, a strong bending power that starts

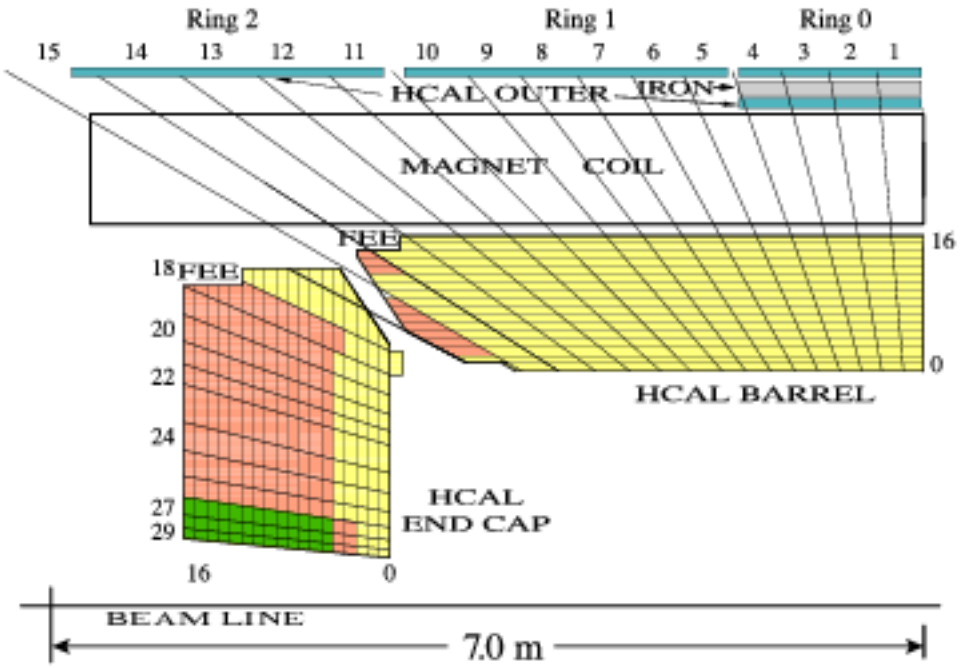

Figure 3.2.8: Side view of HCAL.

from the collision point, essential to reach high resolution in the measurement of the muon momentum[6]. The magnet is a 13 m long superconductive coil with a diameter of about 6 m, made of Niobium-Titanium. The strength of the field around CMS is uniform, and at its center reach 3.8 T with an associated current of 18kA and a total energy of 2.4 GJ. The flux of the magnetic fields returns by a saturated iron yoke that holds mechanically the whole structure of the experiment.

3.2.6 Muon Spectrometer

For their minimum energy loss inside matter, muons are the only particles that can pass through all the CMS sub-detectors without being absorbed. For this reason, the CMS muon system [77] is placed outside the magnet coil, with a detector in the barrel ($|\eta| < 1.2$) and a different one in the endcaps ($1.2 < |\eta| < 2.4$). The global muon system consists of four layers of stations embedded into the iron of the magnet return yoke within a 1.5 T magnetic field. The entire system, both in the barrel (with six stations) and in the endcaps (with four stations), is equipped with resistive plate chambers detectors (RPC), for a total of 612 chambers. RPC detectors are gas chambers installed in parallel with very high time response (about 3 ns). RPC are essential to deal with the high speed request of the trigger system (Subsec. 3.2.7), able to identify muons with high efficiency. The barrel region is made of 195000 drift tubes (DT), placed in twelve planes. In the endcap, the muon system is made

[6]For a charged particle in a magnetic field, the momentum is related to the field B and the relative curvature R by the formula (expressed in GeV)

$$p = 0.3 \cdot B \cdot R.$$

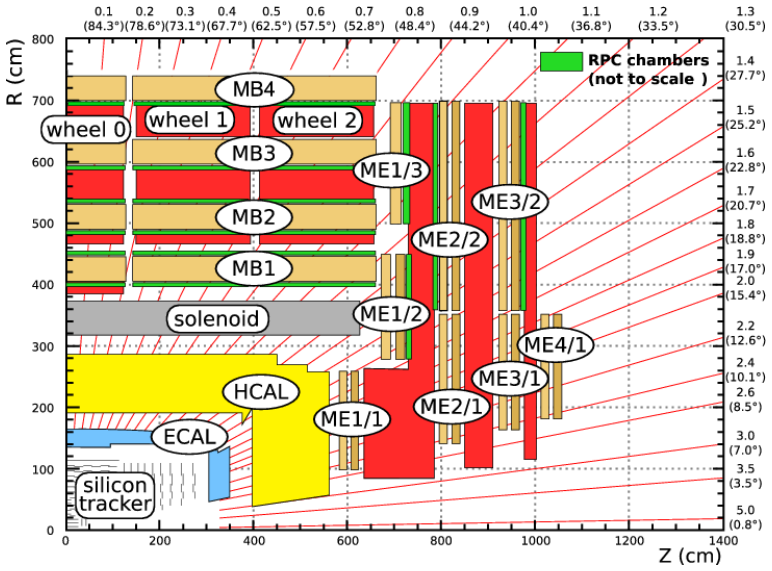

Figure 3.2.9: Transverse view of the CMS muon spectrometer detectors. The RPC chambers (in green) are shown together with muon chambers in the barrel (MB) and in the endcaps (ME)

of cathode strip chambers (CSC[7]) divided into six modules. A schematic view of the CMS muon system is shown in Fig.3.2.9-3.2.10.

3.2.7 Trigger System

One of the great technologic challenges of the LHC experiments is to deal with a huge amount of data to be stored in a limited storage capability. At the nominal LHC luminosity, the expected rate of events is about 10^9 Hz, with an average size for a raw event of about 1 MB. Given that rate, it is clearly not possible to record all the proton collision informations into a physical support, which would have a capability of around 100 Hz. However, the most of interactions are *soft*, i.e. low momentum particles, that are not interesting for the physics purposes of CMS. For this reason it is necessary to filter this information and record only the physical interesting part of the collisions, and thus lowering the rate enough but still keeping all the physics potentials. The CMS trigger system is designed to reduct the event rate down to 10^7 Hz by means of a two level scheme: the Level-1 trigger (L1), mainly hardware-based and the High Level Trigger (HLT), that exploits also software systems. A detailed scheme of the CMS trigger system is presented in Fig. 3.2.11.

Level-1 Trigger

The first level of the CMS trigger system [78] (L1) is designed to reduce the rate of events down to 50-100 kHz. The L1 takes a decision about keeping or discarding data from a particular bunch crossing within 3.2 μs; if the L1 accepts the event, the data are moved to be processed by the HLT. Since the available time of the decision is so short, instead of using the whole CMS information, the L1 trigger takes advantage of the fast response coming from the calorimeter and muon systems

[7]CSC detectors are multi-wire proportional chambers with a stripe-segmented chatode plane

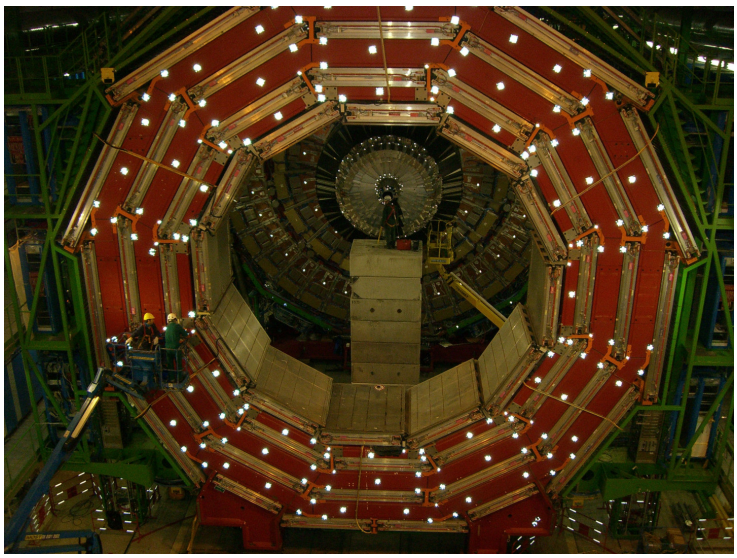

Figure 3.2.10: Picture of the muon chambers embedded in the iron return yoke of the CMS magnet.

only. The information coming from the Calorimetric Trigger and the Muon Trigger are then passed to the Global Trigger that takes the final decision. The Calorimetric trigger is based on the information of the *trigger towers,* a 5X5 matrix of ECAL crystals, grouped in 4X4 squares. The Calorimetric Trigger is able to distinguish classes of candidates: electrons, photons, jets and tau jets and missing energy by the energy deposits analysis. The Muon Trigger runs in parallel taking information from the muon system. The final information is then merged and passed to the Global Trigger, that analyzes the characteristics of the single object before passing the information to the HLT.

High Level Trigger

The role of the High Level Trigger [79] (HLT) is to reduce the rate down to 100 Hz. The HLT is based on sophisticated and dedicated software tools called *HLT Paths* that are designed to discriminate specific topologies of events. The HLT exploits a system that connects informations coming from all the sub-detectors of CMS and controlled by an event manager system. The High Level Triggers have access to all the information used in L1 stored locally in the L1 trigger crates. Consequently, the HLT can make further combinations on the digital list of objects transmitted from L1, since much information is not available on the time scale of the L1 trigger decision. The tracking information and the full granularity calorimeter information is taken at the HLT stage, and the combination of all the CMS informations is taken to make the final decision about weather keeping or discarding the given event. The High Level Triggers are designed to achieve a rejection factor of 10^3, writing up 100 events per second to the mass storage. The last stage of HLT processing is the reconstruction and event filtering with the aim of making datasets of different physics processes on easily accessed media.

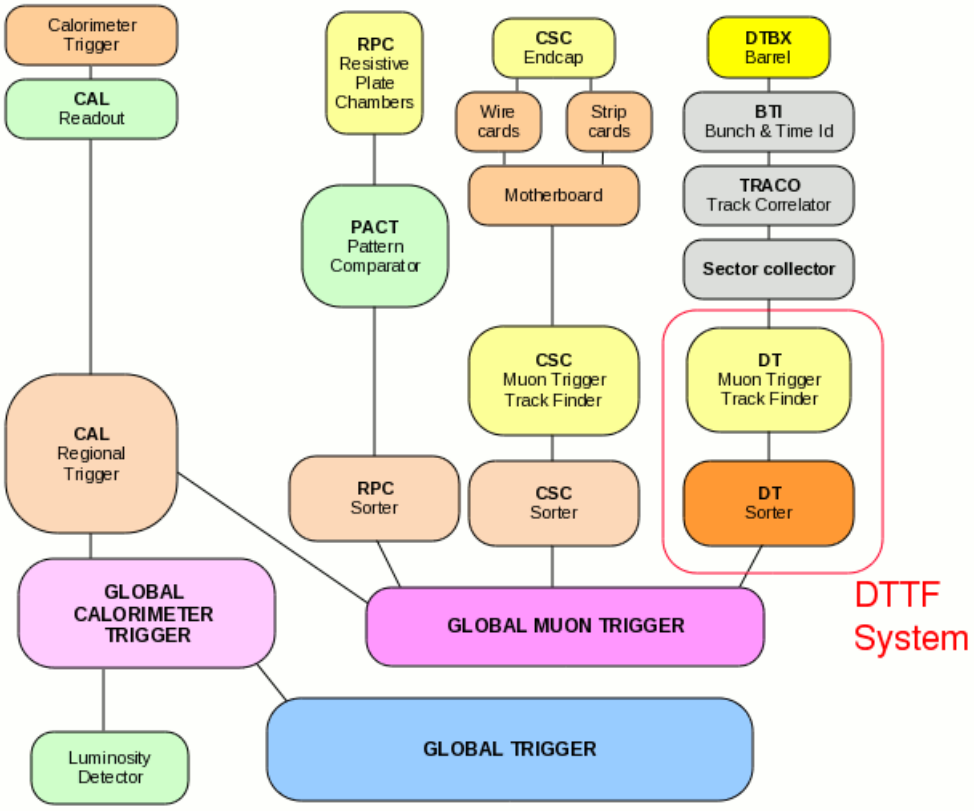

Figure 3.2.11: Detailed scheme of the CMS trigger showing all the steps needed to filter the interesting data from the entire pp collision information.

Chapter 4

Physics Objects Reconstruction

In order to characterize the $Z+b$ events, a detailed knowledge of the physics objects used in the analysis is very important. This Chapter is dedicated to the electrons, muons, jets and b-jets reconstruction with the CMS experiment, and the description of the techniques adopted in order to measure and associate the energy measurements to the different particles. In the first part of the Chapter (Sec. 4.1) the Particle Flow algorithm will be presented. This is the most used and performing algorithm for object reconstruction in CMS, and thus all the particles involved in the analysis will be selected after the Particle Flow reconstruction. In the second part of the Chapter (Sec. 4.2) the reconstruction of leptons, jets and b-jets will be discussed, and the selection criteria to optimize the event description to be used in the analysis (Chapter 5) will be described. Finally, a discussion on the efficiency evaluation for leptons and jets will be presented.

4.1 The Particle Flow Algorithm

Ideated to give a full description of the physics objects reconstructed in the CMS experiment, the Particle Flow algorithm [80, 81] is a standard tool used in almost all the physics analyses of CMS. The main feature of the algorithm is that it takes advantage of all the sub-detectors of CMS by analyzing the information on the different position and energy measurements (made by calorimeters, tracking, muon chambers (Fig. 4.1.1)), and combines them in order to characterize the single particle, giving a more precise result with respect to a single detector measurement. Within the Particle Flow it is possible to discriminate, for each event, leptons, jets, missing energy (i.e. neutrinos) using all the information coming from CMS. It is possible to determine tracks, isolation variables, quark flavours discriminators in order to have a comprehensive description of the event under consideration. The starting point of the algorithm is to take the calorimeter clusters or the track measurements, then the information taken by the other sub-detectors is combined to build the physics object.

4.1.1 Track Reconstruction

Charged particle trajectories are measured by the CMS tracking system. A single track is defined as the sum of all the hits inside the silicon and pixel detectors that provide a measurement of the path followed by a charged particle coming from

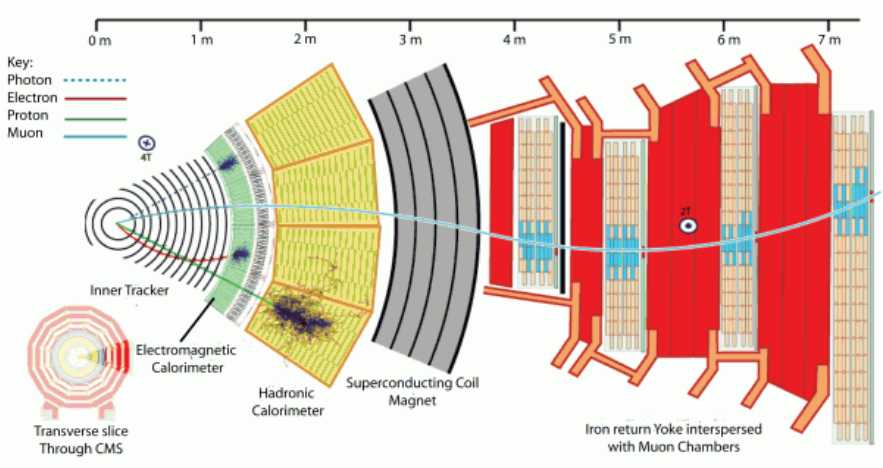

Figure 4.1.1: Transverse slide view of the CMS sub-detectors energy measurement for different particles. The Particle Flow algorithm combines all the information given by each detector to build the physics object.

the pp collision. The reconstruction starts with the collection of these hits, and then a single *seed* is generated if it satisfies quality requirements on the vertex, pixel and silicon detectors measurements. Then this seed is given as input to a Kalman Filter, an iterative fit technique that starting from a seed, estimates the statistical compatibility of the hit with the track (taking into account the energy loss in the material and the magnetic field) and finally extracts the first estimation of its momentum (p_T) and pseudorapidity (η). After this recognition a series of quality cuts are applied in order to optimize the track identification efficiency. For this purpose, an iterative algorithm has been developed to ensure high quality in the track selection [82]. After this selection the fake rate is estimated to be at the order of percent, while the tracking efficiency is greater than 90% as it can be seen in Fig. 4.1.2. Finally the track is given to the PF algorithm as input.

4.1.2 Calorimetric Clusters Reconstruction

The calorimetric information is essential to measure and identify electromagnetic particles (electrons and photons) through their energy loss in matter as well as charged and neutral hadrons. The basic idea is the clustering of the energy deposits released by the different interacting particles inside the ECAL and HCAL calorimeters and in the ES. The clustering algorithms (see Sec. 4.2) run separately in the endcap region and in the barrel region of these detectors. The clusters are constructed if they have a minimum energy threshold, and in this case they are labeled as *cluster seeds.* When a cluster seed has at least one calorimetric cell in common with another calorimetric one, they are gathered together if they have an energy that is found to be 2σ away from the measured calorimeter noise energy in ECAL (80 to 300 MeV) and HCAL (800 MeV). These new objects are then called Particle Flow seeds. At this point a linking algorithm starts to interconnect the informations from the various measurements in order to identify the different particles.

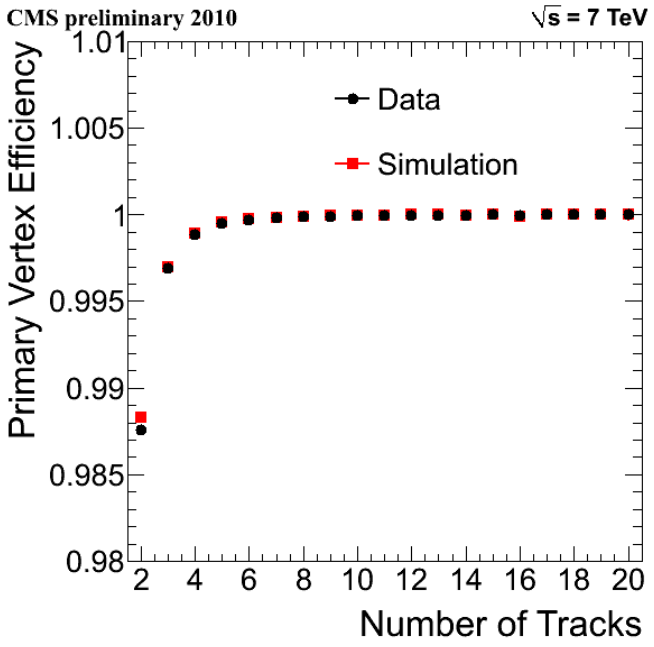

Figure 4.1.2: Tracking efficiency measured at 7 TeV in CMS compared to simulation.

4.1.3 The Linking Algorithm

The heart of the PF algorithm is the connection between the different basic objects (track seeds, ECAL/HCAL clusters) aiming to provide, as an output, the reconstruction of the single particle under consideration. All the possibilities are taken into account based on the specific particle type, for example ECAL to track, or ECAL to HCAL, by linking the information. This procedure leads to new reconstructed objects called *blocks*:

- The extrapolation of a track from an ECAL or HCAL cluster to the last hit in the tracker has a link whose quality is defined as the precision of the extrapolation, i.e. the precision of the point in the ECAL cluster where the track is taken and back-extrapolated to the corresponding tracker hit position. An example of this procedure is given in Fig. 4.1.3.

- The ECAL to HCAL or ECAL to ES clusters link is done if the cluster position in the more granular calorimeter (ES or ECAL) is within the cluster boundaries in the less granular calorimeter (ECAL or HCAL).

- The charged particle track to a muon track link is realized when the χ^2 of the fit between the two tracks gives a result that is below a predefined threshold.

Bremsstrahlung photons emitted by electrons are also taken into account by the PF algorithm. For each track, starting from the intersection points with the tracker layers, tangents are extrapolated to ECAL. If clusters are found by the intersection of a track tangent with the calorimeter, they are linked to the first track and identified as bremsstrahlung photons. The link quality corresponds to the ΔR between the cluster position and the extrapolated point of the tangent in ECAL.

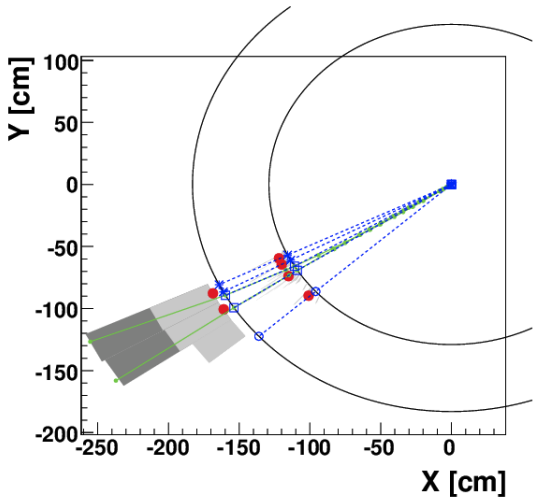

Figure 4.1.3: CMS xy view of the HCAL seed cluster back-extrapolated to corresponding tracks in a $K^0_L \to \pi^+\pi^-\pi^0$ event. The cluster positions are represented by dots, the simulated particles by dashed lines, and the position of their impact on the calorimeter surfaces by various open markers.

4.2 Objects Reconstruction

Particle Flow Electrons

The main characteristics of an electron candidate are a short track associated with significant bremsstrahlung photon emission inside the tracker material. Since electrons may have a great bending in the 3.8 T axial field generated by the solenoid surrounding the tracker and the calorimeters, the emission is widely spread-out in the ϕ direction. However this is dependent on the electron p_T, directly proportional to the magnetic field bending power. Two algorithms have been developed:

- **ECAL Driven Reconstruction**: In the barrel, the algorithm searches for single crystals which gained a total energy above the threshold of 1 GeV, and sorts them in decreasing energy. This crystals are called *seeds*. Around each seed, it opens in the ϕ-direction a 5-crystal wide strip, reaching up to ± 17 crystals, and adds to the row containing all the seeds all the 5-crystal rows which have a total energy larger than 0.1 GeV. In the endcaps the algorithm first organizes crystals in 5×5 matrices, and groups the subset of them which lie within an azimuthal distance of 0.3 rad. The final set obtained through this procedure is called *supercluster,* and it is used to seed the reconstruction of the electron track, done with a fitting procedure (called *Gaussian Sum Filter*). This reconstruction based on an ECAL-driven super-clustering strategy is the best choice for high p_T isolated electrons.

- **Tracker Driven Reconstruction:** Low p_T electrons are not isolated, and their photons emission inside the material has shorter curvature and appears at high azimuthal angles. These electrons come, for example, from the semileptonic charm and beauty decays and from low mass resonances ($J/\psi, \Upsilon$) and so a big contamination make the cluster measurement too much polluted to be

reliable. One of the great features of the PF reconstruction is the possibility to measure these low energy electrons as well, by extrapolating a straight line tangent from the electron track to the ECAL calorimeter from each tracker layer, around which most of the tracker material is concentrated. The electron identification and the momentum reconstruction are performed by combining the track and supercluster observables.

Particle Flow Muons

Before the PF reconstruction, CMS takes advantage of three different strategies to build up a reconstructed muon: the association between track hits measured in the muon spectrometer chambers and in the tracker, the tracker-only measurement or the spectrometer-only measurement. A global fit to the information coming from all the three methods combined statistically give rise to a *global muon.* The PF algorithm takes as input a global muon and, if the combined momentum is within 3σ of the tracker-only reconstruction, the corresponding track is removed from the *block*.

Charged Hadrons

Charged hadron candidates are reconstructed by charged tracks linked to any number of calorimeter (ECAL or HCAL) clusters, and which are not identified as electrons.

Neutral Hadrons and Photons

If the calibrated energy of the closest ECAL and HCAL clusters linked to a track is significantly larger than the total associated charged particle momentum, neutral particles might be present. If this energy is found to be larger than the total ECAL energy, a photon is created with this ECAL energy and a neutral hadron is created with the remaining part of the energy. The remaining ECAL and HCAL clusters, either originally not linked to any track or for which the link was disabled, are identified as photons and neutral hadrons, respectively.

Jet Energy Corrections

Particle Flow jets (PF jets) are simply constructed by applying the chosen algorithm taking as input the PF candidates list of the event within the cone of clusterization. In this analysis, the algorithm employed for the jets clusterization is the anti-k_T (described in details in Subsec.1.2.4), that iterates the calculation of particles momenta in the list inside a cone of $\Delta R = 0.5$. The results, as it can be seen in Fig. 4.2.1, is that the PF reconstruction is much more precise than other approaches [80], like calorimetric reconstruction for instance, especially in the low p_T regions. For several physics aspects, corrections to the jet energy must be applied in order to obtain a uniform response of the energy with respect to the η and p_T of the jet. The CMS strategy for these corrections, called Jet Energy Corrections (JEC), is to factorize the different contributions (*Levels*) coming from the different physics underlying the jet reconstruction process [83]. Three of these are considered mandatory when requiring a reconstructed jet:

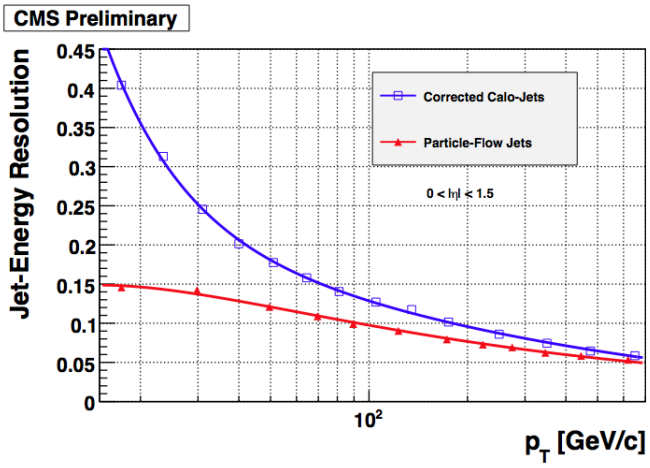

Figure 4.2.1: Jet resolution as a function of the jet p_T for the PF reconstruction (red curve) and the calorimetric reconstruction (blue curve).

- **Level 1 (LV1 - offset)** : is the pile-up correction, taking into account the effect of overlapping collisions
- **Level 2 (LV2 - relative)** : corrects for the non-uniformity of the jet energy with respect to the CMS sub-detectors (jet η)
- **Level 3 (LV3 - absolute)** : corrects for the non-uniformity of the jet response as a function of the jet p_T
- **Residual (RES)** : accounts for the residual differences between data and Monte Carlo.

The factorized approach consists in the sequential application of the three levels LV1+LV2+LV3 in order to correct the jet p_T.

$$p_T^{corr} = [C_{LV1}(p_T) \cdot C_{LV2}(p_T, \eta) \cdot C_{LV3}(p_T)]\, p_T^{uncorr}$$

The correction values are derived from Monte Carlo simulation and then applied on the data. The RES correction is applied if significant discrepancies are found in the comparison. The LV1 correction is calculated by subtracting the additional energy which contaminates the inside of a jet cone by secondary proton-proton collisions (pile-up). This is done at the clusterization time, using the k_T algorithm, by adding soft particles in the event increasing the mean energy and taking as a variable the mean of this energy distribution per unit of area (called ρ, jet energy density). This variable is then multiplied by an effective jet *area* and the resulting quantity is then subtracted from the jet energy, correcting it for both the pile-up contribution and the underlying events one (left side of Fig. 4.2.2). This method is known as ρ - FastJet subtraction [84]. The LV2 correction is evaluated by measuring the jet response across its pseudorapidity range. The response at the center of the barrel is taken as reference. The correction is evaluated in dijet events, in which one jet is required to be in the barrel ($|\eta| < 1.3$) and the response of the second (probe) jet is studied as a function of the pseudorapidity of the probe. Residual corrections (RES) are then applied in order to correct for the residual discrepancies between data and

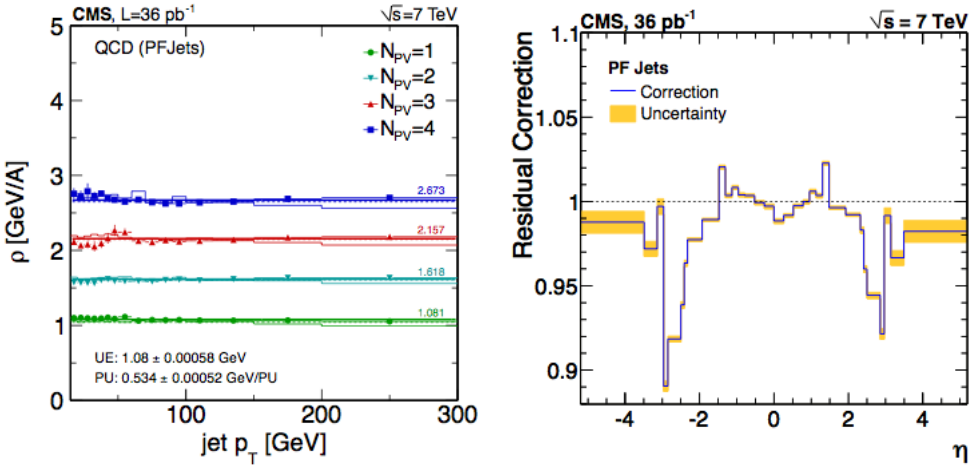

Figure 4.2.2: Left: The energy density as a function of the jet p_T showing the contribution of pile-up and underlying events (LV1 correction) for PF jets in QCD events. Right: Residual correction as a function of the jet η.

Monte Carlo (right side of Fig. 4.2.2). Other minor corrections exist, but they are considered not mandatory in the jet reconstruction and depend on the particular analysis under consideration:

- **Level 4 (LV4 - electromagnetic fraction)** : makes the jet response uniform with respect to the electromagnetic energy fraction
- **Level 5 (LV5 - jet flavour correction)** : corrects for the jet flavor dependence
- **Level 7 (LV7 - parton jet correction)** : corrects the jet p_T by bringing it back to the originating parton p_T (average)
- **Level2Level3Residual (LV2LV3Res - residual calibration)**: accounts for the residual energy calibration discrepancies in data.

Jet Energy Resolution

The PF jets transverse momentum resolution is a fundamental quantity needed in any final state regarding jets. It represents the precision of the jet used in the analysis, and thus gives the systematic error associated to the reconstruction with respect to the *true* generator-level jet. The precision of the knowledge of the jet momentum is essential to understand the process under consideration under many aspects, from the QCD prediction to the unfolding and the missing transverse energy understanding. After correcting the PF reconstructed jet by the application of the JEC, the p_T resolution is evaluated by two techniques, taking advantage of two particular topology of events :

- Dijet Asymmetry, in events containing two jets
- Jet Balance Method, in γ + jets events

The dijet asymmetry method exploits momentum conservation in the transverse plane of the dijet system and it permits to extract the jet resolution using both the RMS of the characterizing kinematic distributions and fitting a resolution function REF. The γ+jet balance method exploits the balance in the transverse plane between the photon and the recoiling jet and uses the photon momentum. The p_T^{γ} distribution provides information on the p_T^{jet} resolution in a given p_T^{γ} range. Results on the jet momentum resolution measured by CMS with 7 TeV data, with the dijet and γ+jets methods are presented in Fig. 4.2.3 as a function of the jet momentum [85].

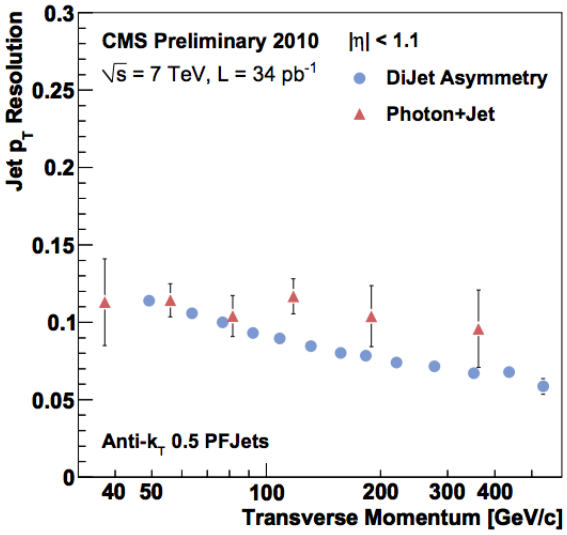

Figure 4.2.3: The jet energy resolution measured in 7 TeV CMS data with the γ + jets method (red triangles) and the dijet method (blue points).

Missing Transverse Energy

The Missing Transverse Energy ($\not\!\!E_T$ or MET) is defined as the imbalance in the plane transverse to the colliding proton beams. This imbalance may arise for different reasons:

- particles escaping detection by CMS, like neutrinos or other weakly interacting particles to be discovered[1]
- detector effects
- physics not from the collision: beam halo, cosmics, pile-up
- every other possible sources.

The $\not\!\!E_T$ is defined as the magnitude of the vector

$$\text{MET} = -\sum_i E_x^i - \sum_i E_y^i,$$

[1]For this reason, the $\not\!\!E_T$ is considered one of the most powerful discriminant in the search of new physics processes al LHC.

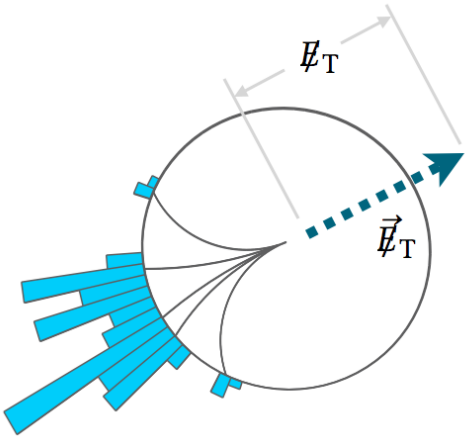

Figure 4.2.4: Scheme of the direction and magnitude of the missing transverse energy in a simulated CMS event.

where E_j^i is the energy of the i input object projected along the j-axis. It can be useful also to define ratio between the missing energy and its resolution, called *MET significance*:

$$\text{MET}^{sign} = \frac{\text{MET}}{\sigma(\text{MET})}.$$

The Missing Transverse Energy is evaluated by considering all the reconstructed particles present in the PF list, and so it refers to a PF MET. As for the jets, also missing energy needs to be corrected. The strategy of the sequential correction levels (called *types*), is similar to the jet energy correction, but unlike the JEC in which corrections are factors, in the case of MET corrections, they are defined as additive terms. Many contributions can be individuated [86], so the corrected value will be

$$\text{MET}^{corr} = \text{MET}^{raw} + \Delta,$$

where Δ represents the various types of correction:

$$\Delta = C_{T0} + C_{T1} + C_{T2},$$

- **Type-I**

The type one correction to the PFMET, C_{T1}, is the contribution from the PF objects which are clustered as jets whose energies are above a threshold and which are corrected for the LV1 JEC. This correction is the propagation of the LV2LV3 residual correction to the PFMET, and can be calculated as

$$C_{T1} = -\sum_j (p_{Tj}^{LV123} - p_{Tj}^{LV1}),$$

where p_{Tj}^{LVk} is the transverse momentum of the jet j after applying the $k-$th JEC level.

- **Type-II**

This correction accounts to the contribution from the offset of jet energy, which was subtracted from the uncorrected energy by LV1 JEC. It can be defined as

$$C_{T2} = (C^{uncl} - 1) \left\{ -\sum_j (p_{Tj}^{uncorr} - p_{Tj}^{LV1}) - \sum_j p_{Tj}^{uncorr} - \sum_{i \not\subset j} p_{Ti} \right\},$$

where the index i runs over all the unclustered particles.

- **Type-0**

The PFMET Type-0 corrections remove an estimate of neutral pile-up contributions as well as identified charged pile-up contributions from reconstructed PFMET. Pile-up interactions have little true MET because they produce few invisible particles, e.g., neutrinos from Kaon decays. Therefore, if we were able to measure all visible particles precisely and accurately, pile-up interactions would not much degrade the MET reconstruction. However, in practice, because our measurement of visible particles is not perfect, the MET reconstruction degrades as the number of the pile-up interactions increases. The Type-0 correction is built starting from the assumption that the sum of the neutral component of the pile-up particles momenta is proportional to the sum of the charged component by means of a factor R^0:

$$\sum_{i \in \text{neuPU}} p_T^i = -R^0 \sum_{i \in \text{chPU}} p_T^i.$$

The correction term is calculated by

$$C_{T0} = (1 - R^0) \sum_{i \in \text{chPU}} p_T^i.$$

The detailed calculations of the Type-0 term can be found in dedicated studies [87].

4.2.1 Reconstruction of $Z \rightarrow e^+e^-$

Dielectrons pairs coming from a Z boson are selected if they have opposite charge and a momentum of at least 20 GeV inside the geometrical range of $|\eta| \leq 1.442$ and $1.566 \leq |\eta| \leq 2.4$, reconstructed with the Particle Flow algorithm. The selected pair is taken only when its invariant mass lies inside the Z boson mass window $M_{e^+e^-} \in [71; 111]$ GeV. Electrons must satisfy a set of selection cuts aimed to guarantee the quality of the candidates by minimizing the fake rate and maximizing the identification efficiency. The cuts used in the electron identification are the following:

- $\Delta\phi$: the azimuthal difference between the track position extrapolated at the calorimeter surface (with parameters computed at the vertex) and the ECAL supercluster energy barycenter;
- $\Delta\eta$: the pseudorapidity difference between the track position extrapolated at the calorimeter surface (with parameters computed at the vertex) and the ECAL supercluster energy barycenter;
- $\sigma_{i\eta i\eta}$: the second moment of the ECAL cluster energy distribution along the η direction;

- H/E : the ratio between the energy deposit recorded in the HCAL tower directly behind the ECAL supercluster seed, and the ECAL supercluster energy;
- d_0 : transverse distance with respect to the primary vertex
- d_z : the distance of closest approach in the transverse plane between the electron track and the z coordinate of the primary vertex.

Cuts on these variables are applied in order to have a working point of about 80% in the electron selection (*medium ID*), and are categorized in two categories depending on the geometrical region under consideration (barrel, endcap). Moreover, two cuts are applied to reject electrons coming from photon conversion inside material layers of the CMS tracker, and at least one missing hit is allowed in the electron track measurement. Finally, to reject electrons coming from the semileptonic decay of b, c quarks inside jets, electrons must be isolated. Isolation is defined as the sum of all the detector energy measurements (reconstructed by the Particle Flow algorithm) in the CMS sub detectors, electromagnetic energy deposit in ECAL, hadronic deposit in HCAL and all tracks, inside a fixed ΔR cone in the $\eta - \phi$ plane around electron candidate, optimized to 0.3:

$$I_e = \frac{1}{p_T^e} \left[\sum_{trk} E_T + \sum_{ECAL} E_T + \sum_{HCAL} E_T \right]_{\Delta R \leq 0.3}$$

In order to be isolated, an electron candidate must pass the requirement $I_e < 0.15$. A full list of the identification cuts described in this section and used to define the $Z \to e^+e^-$ process is presented in Table 4.1.

Variable	**Barrel**	**Endcup**
$\Delta\phi$	< 0.004	< 0.007
$\Delta\eta$	< 0.06	< 0.03
$\sigma_{i\eta i\eta}$	< 0.01	< 0.03
H/E	< 0.12	< 0.10
d_0	< 0.02	< 0.02
d_z	< 0.1	< 0.1

Table 4.1: Identification cuts in the electrons selection.

4.2.2 Reconstruction of $Z \to \mu^+\mu^-$

Particle Flow muons are selected in pairs with opposite charge in the Z mass range $M_{\mu^+\mu^-} \in [71; 111]$ with at least $p_T > 30$ GeV and $|\eta| < 2.5$. A set of variable cuts defines a *tight muon* for a candidate that is reconstructed together with the CMS muon chambers and tracker *(global muon)* :

- $\chi^2/d.o.f.$ of the global track;
- Matched Hits: number of hits reconstructed in the tracker;
- Matched Segments: number of matched segments in the muon stations;
- d_{xy} : track transverse impact parameter;

- d_z : longitudinal impact parameter.

The list of the global muon cuts is presented in Table 4.2. Muons must be isolated, defining the isolation variable

$$I_\mu = \frac{1}{p_T^\mu}\left[\sum_{trk} p_T + \sum_{ECAL} E_T + \sum_{HCAL} E_T\right]_{\Delta R \leq 0.3}$$

and imposing $I_\mu < 0.2$. Since the measurement of the muon momentum is affected by the reconstruction capability and the limited knowledge of the physical configuration of the detector, an extra correction is needed. The momentum scale correction is extrapolated using processes that are very well under control, such as J/ψ and Z resonances decays. This method is called *MuScle Fit* correction [88].

Variable	**Cut**
$\chi^2/n.d.f.$	< 10
Matched Hits	> 10
Matched Segments	> 2
d_{xy}	< 0.02 cm
d_z	< 1 cm

Table 4.2: Tight selection cuts for Global Muon candidates.

4.2.3 Lepton Efficiency Evaluation

In order to obtain a cross section measurement, a crucial point is the evaluation of each contribution to the total efficiency, defined as the product of the single efficiency coming from the different physics objects selection. In the $Z + b$ final state, apart from the b sector efficiency) several contributions to the total efficiency can be identified regarding the Z boson reconstruction:

- ϵ_{RECO} : efficiency of reconstructing leptons;
- ϵ_{HLT} : efficiency of the high level trigger selection;
- ϵ_{SEL} : selection efficiency referring to the cuts applied in the selection;

The three components of the above mentioned list are applied in sequence,

$$\epsilon_l = \epsilon_{RECO} \cdot \epsilon_{HLT} \cdot \epsilon_{SEL},$$

and refer to the Z boson reconstruction and decay efficiency. The ϵ_l values are computed by means of the *Tag&Probe* method [88] in data and Monte Carlo, aiming at the extraction of the data to Monte Carlo *scale factors* for each component:

$$\mathrm{SF}_j(p_T, \eta) = \frac{\epsilon^j_{DATA}(p_T, \eta)}{\epsilon^j_{MC}(p_T, \eta)},$$

where the efficiency for the j-th component of the total efficiency ϵ_l is evaluated in data and Monte Carlo as a function of the momentum and pseudorapidity of the lepton used in the *Tag&Probe* method.

4.2.4 The *Tag&Probe* Method

Tag&Probe (T&P) is a data driven technique used to calculate efficiencies. It exploits the properties of a well known resonance, such as the J/ψ or, in this particular case, the Z boson decay product. The two electrons or muons coming from the Z decay are selected with asymmetric criteria. One of them, called the *tag* lepton, has very tight selection criteria and very low fake rate. The other one, called *probe* lepton, has looser criteria that depend on the specific efficiency calculation in consideration. The passing *probe* has tighter criteria than the probe, but not tighter than the *tag*. Efficiency of the *probe* is defined as the number of passing probes divided by the total number of probes. defining the probe efficiciency as

$$\epsilon_{probe} = \frac{N_{passing\ probes}}{N_{total\ probes}},$$

in the T&P the formula takes into account passing (PP) and failing (FP) probes combination

$$\epsilon_{probe} = \frac{N_{PP}}{N_{FP} + N_{PP}}.$$

To extract the signal, a fit to the invariant mass distribution of the electron and muon pairs is made using a Cruijff function[2] for the passing and the failing candidates. This function is found to be the best choice in order to describe the asymmetric tails in the Z shape due to the bremsstrahlung and the final state radiation, since it takes into account the lepton radiative tails in the exponential expression of the function. The background is modeled by a negative exponential function. Finally, the fit procedure makes use of an extended unbinned maximum likelihood fit. Rather than correcting the cross section for the total efficiency (the correction will be taken into account by the unfolding procedure (described in Chapter 6)), the differences seen between Monte Carlo and data for each efficiency component are restored by an event by event correction with the appropriate scale factors. In Fig. 4.2.5 the electrons *MediumID* (described in Sec.3.2) efficiency is shown for the 8 TeV recorded in 2012, evaluated with the T&P method as a function of the electron momentum for several electron pseudorapidity ranges, together with the relative scale factors. Muon *TightID* efficiencies as a function of the muon p_T in several η regions and as a function of the number of vertices are shown in Fig. 4.2.6. For both, the comparison with the simulation is made with a Drell-Yan $Z \to e^+e^-$, $Z \to \mu^+\mu^-$ sample.

4.2.5 Reconstruction of *b* quarks

The identification of jets originating from b quarks hadronization is a crucial point in the whole LHC physics program, because of the wide phenomenology of interesting phenomena involving b jets in the final state. Moreover, this is a central point in the $Z + b$ studies discussed in the next Chapters. For these reasons, it is of primary importance to have an efficient algorithm able to discriminate the b quarks from the other non-b jet background processes.

[2]The Crujiff function is a centered Gaussian with different left-right resolutions and non-Gaussian tails: $f(x) = \exp((x-m)^2/2(\sigma^2_{L,R} + \alpha_{L,R}(x-m)^2))$.

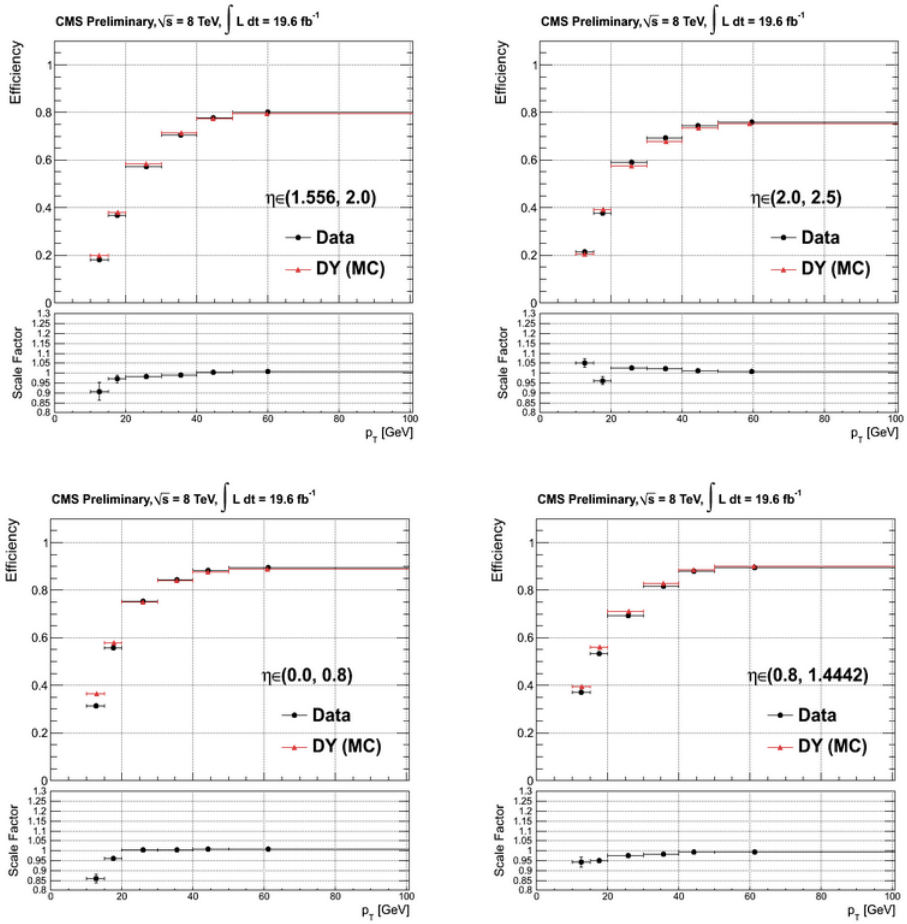

Figure 4.2.5: Electron MediumID efficiency with the Tag&Probe method in $Z \rightarrow e^+e^-$ events in different electron η regions as a function of the electron momentum.

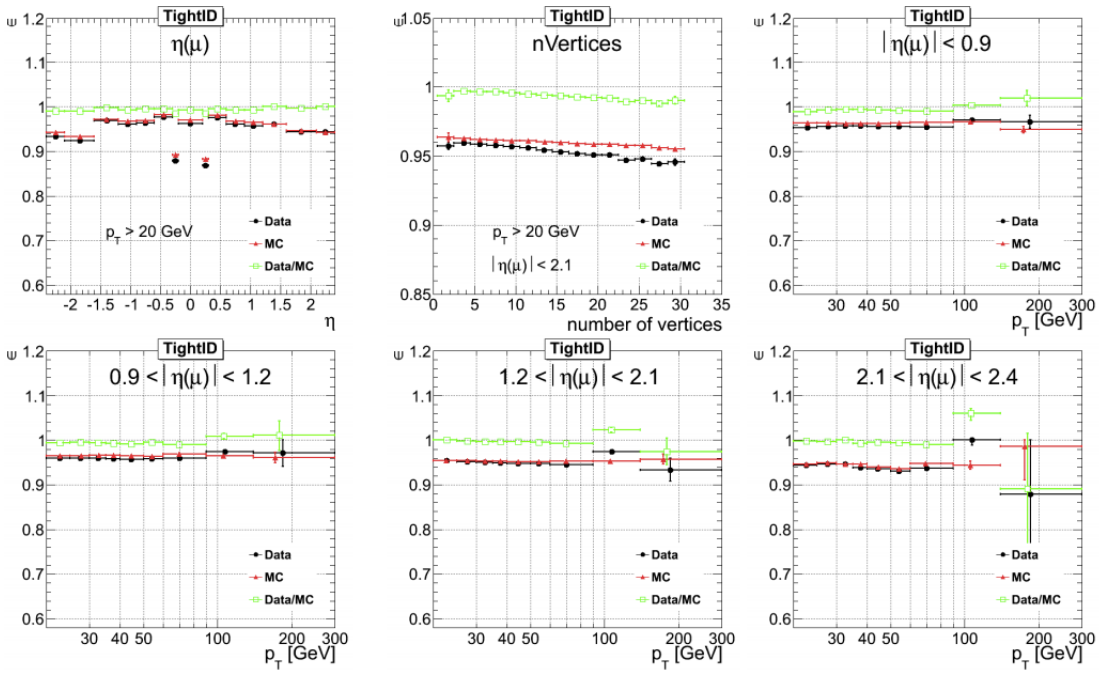

Figure 4.2.6: Muons TightID efficiency with the Tag&Probe method in $Z \rightarrow \mu^+\mu^-$ events in different electron η regions as a function of the electron momentum.

The b-tagging

The identification of the b-jets can be made exploiting the properties of the B hadrons production and weak decay [6]. There are several unique properties of the B hadrons ($B^{\pm}_{u(c)} = |b\bar{u}(\bar{c}) >$, $|\bar{b}u(c) >, B^0_{d(s)} = |b\bar{d}(\bar{s}), \bar{b}d(s) >$) family. Because of their large mass (about 5 GeV, the most massive quarks of the Standard Model that have enough lifetime to hadronize and produce bounded states) and their coupling to the charm, b quarks-made hadrons lifetime is relatively long. Through the weak interactions theory, using the B total decay width Γ_B, that the mean lifetime is ($\tau_B = \frac{\hbar}{\Gamma_B}$):

$$\tau_B \propto G_F |V_{cb}|^2 m_b^5.$$

Using the current measured values for the Fermi constant G_F (from the muon decay rate) and V_{cb} (from the b-factories experiments), for ultra relativistic B mesons, the lifetime is found to be around 1.5 ps, equivalent to an average decay length of $\lambda = (\beta\gamma)c\tau = 450\mu$m. This distance corresponds to a flight distance that can be observed in the LHC experiments tracker systems. After this distances, the B hadron decays, thus producing an observable secondary vertex inside the tracker, displaced from the primary vertex with all the subsequent corresponding tracks (Fig. 4.2.7). In addition, the high B mass has as adjunctive consequence an high charged tracks multiplicity, estimated to be around 5 per B decay. Another characteristic of the b topology comes from the B semileptonic branching ratio: around 10.7% of the cases an electron or a muon is produced inside the b-jet. Most of the b-jet energy is carried out by the B hadron present inside the jet itself, because of the big value of the b quark parton density function.

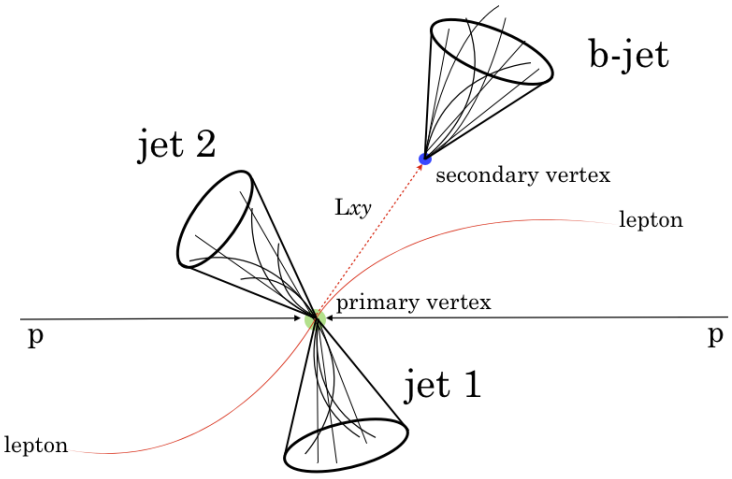

Figure 4.2.7: Secondary vertex production scheme.

Several algorithms have been developed in order to perform the b-tagging. Historically two big categories can be distinguished. One option is to exploit the lepton identification, measuring electrons and muons track's impact parameter from the semileptonic B decay. The impact parameter of a track with respect to the primary vertex can be used to distinguish the decay products of a b hadron from prompt tracks. Examples of algorithms of this kind 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. Another possibility is to build an algorithm based on the secondary vertex reconstruction. The presence of a secondary vertex and its associated kinematic variables can be used to discriminate between b and non-b jets. An example is the *Simple Secondary Vertex* (SSV) that uses the significance of the flight distance as the discriminating variable. A complete description of these algorithms can be found in [89]. A more sophisticated and complex approach is to combine with advanced statistical techniques all the information coming from the secondary vertex measurement and the track-based measurement. This approach leads to a more involved b-tagging algorithm called *Combined Secondary Vertex* (CSV).

Secondary Vertex Tagging

The Combined Secondary Vertex (CSV) [90] exploits the combination of several kinematic informations, starting with the reconstruction of a secondary vertex in the event, to build a discriminating variable (CSV discriminator). The topology of the event is first characterized by the discrimination of all the tracks that are not originated by the primary vertex, reconstructing the secondary vertex by means of the *Trimmed Kalmann Vertex Finder* [91]. Basically, a series of kinematic cuts are set to reconstruct the secondary vertex. The distance between the primary vertex and the secondary vertex in the transverse plane must be in the distance range of 100 μm - 2.5 cm, and must satisfy the condition

$$\frac{L_{xy}}{\sigma_{L_{xy}}} > 3,$$

where $\sigma_{L_{xy}}$ is the error associated to the distance L_{xy} in the transverse plane. Then the invariant mass of the charged particles associated to the vertex is calculated and a cut of $M_{charged} < 6.5$ GeV is applied. A special cut is dedicated to the rejection of the K_S^0 . In order to reject K_S^0 decays, vertices with two oppositely charged tracks are rejected if their mass is within a window of 50 MeV around the nominal K_S^0 mass. After this chain the algorithm discriminates three categories:

- **RecoVertex** : Secondary vertex candidate is reconstructed. More than one secondary vertex can be reconstructed.
- **PseudoVertex** : In the case of no reconstructed secondary vertex candidate found, *PseudoVertex* is created using charged particle tracks not compatible with the primary vertex, having a signed transverse impact parameter significance greater than two, if at least two such tracks are present in the jet.
- **NoVertex** : It is the case when neither *PseudoVertex* or secondary vertex candidate are found and the tagging relies only on the tracks measurement.

Variables to be used as imput for the CSV algorithm are defined for the different categories. For the first category, when **RecoVertex** are found, many observables, such as the invariant mass of all the secondary vertex's coming tracks (Secondary Vertex Mass) or their rapidities energy and multiplicity, are constructed to discriminate b-jets. In the case of **PseudoVertex**, most of these variables still have some discriminating power but since there is no attempt to fit the geometrical position of the *PseudoVertex*, the significance of the flight distance is not used for this category, whereas the other variables enter into the list of input variables. For the **NoVertex**, case, no additional variables are defined. The last step to build the CSV discriminator is the combination of the variables described above. In order to have a single variable as an output of the combination of the different kinematic observables, a likelihood ratio technique is employed. The key point is that these variables behave differently for b quarks, c quarks and light flavours (u, d, s, g) . For this reason the Likelihood function allows to discriminate the different contributions. The function is defined as

$$\mathcal{L}^{\mathrm{b,c,light}} = f^{\mathrm{b,c,light}}(\alpha) \cdot \prod_{i=1}^{N} f_{\alpha}^{\mathrm{b,c,light}}(x_i)$$

where $\alpha =$ RecoVertex, PseudoVertex, NoVertex, x_i are the input variables, $f_{BG}(c)$ and $f_{BG}(\mathrm{light})$ are the expected a-priori probabilities for the c and light content in non-b jets $f^{\mathrm{b,c,light}}(\alpha)$ is the probability for flavour $b, c, light$ to fall into category α and $f^{\mathrm{b,c,light}}(x)$ is the probability density function for variable x for category α and flavour $b, c, light$. The CSV discriminator $\mathrm{d_{CSV}}$ then is computed by

$$\mathrm{d_{CSV}} = f_{BG}(c) \cdot \frac{\mathcal{L}^{\mathrm{b}}}{\mathcal{L}^{\mathrm{b}} + \mathcal{L}^{\mathrm{c}}} + f_{BG}(\mathrm{light}) \cdot \frac{\mathcal{L}^{\mathrm{b}}}{\mathcal{L}^{\mathrm{c}} + \mathcal{L}^{\mathrm{light}}}$$

The probability density functions are extracted from a statistically independent sample of simulated QCD events, and depend on the transverse jet energy and pseudorapidity. The distribution of the CSV discriminator for the three categories is presented in Fig. 4.2.8, obtained from Monte Carlo QCD sample. In order to

select b-jets, a cut on the CSV discriminator is set to 0.89 (tight working point). For the estimation of the CSV discriminator in the $Z + b$ selection, see Chapter 5.

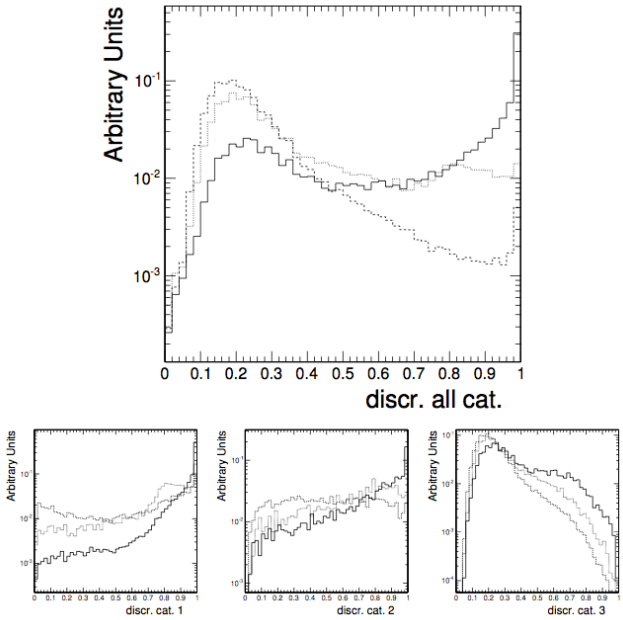

Figure 4.2.8: The CSV discriminator distributions obtained from a Monte Carlo QCD sample for the three different categories separetely (bottom) and for the combination (top). The three curves represent the light, bottom and charm component of the total CSV distribution.

Track Impact Parameters Tagging

Another important strategy to tag b quarks, tested and used already in LEP experiments especially for the $Z \to b\bar{b}$ measurements [92], is to exploit the information coming from the three-dimensional impact parameter. The advantage of this technique is that it can rely on the excellent performances of the pixel detector on the z axis. By means of the sign of the impact parameter, a prompt track can be discriminated from a track coming from a B hadron decay: positive impact parameter has the same sign as the scalar product of the vector pointing from the primary vertex to the point of closest approach with the jet direction. A variable used as a discriminator is the impact parameter significance, defined as the the ratio between the impact parameter and its uncertainty. As any track based observable, the impact parameter strongly depends on the p_T, η of the track. Detailed studies can be found in [82]. Two $b-$tagging discriminator can be built from the impact parameter calculation. The *Jet Probability* (JPB) algorithm [93] uses an estimate of the likelihood that all tracks associated to the jet come from the primary vertex:

$$P_{jet} = \Pi \sum_{k=0}^{N-1} \frac{-\ln(\Pi)^k}{k!},$$

with $\Pi = \prod_{i=1}^{N} \max\{P_i, 0.005\}$, N is the number of tracks, and P_i is the estimated probability for track i to come from the primary vertex. By weighting the tracks with higher impact parameter (up the fourth with highest value) another tagger can

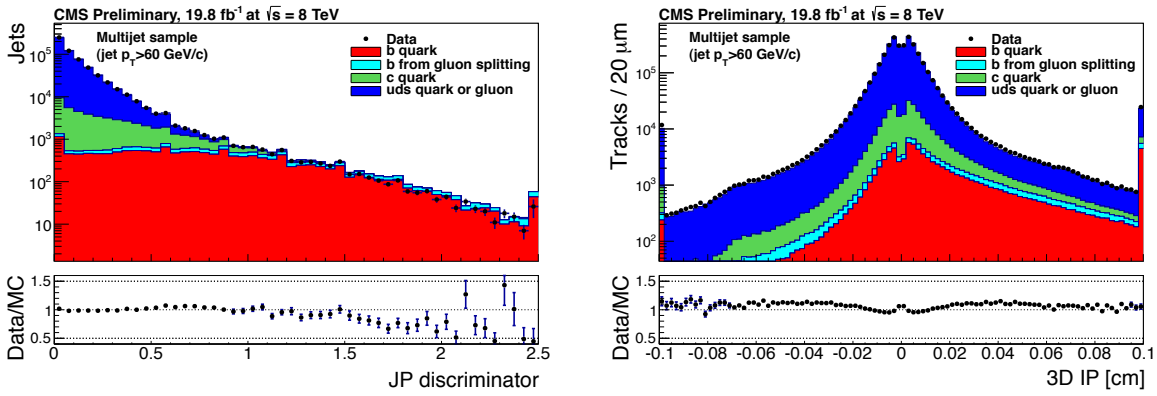

Figure 4.2.9: Jet Probability discriminator (left) and 3D impact parameter distribution (right) measured with 8 TeV collision data in 2012 by CMS. Different flavours composition are presented in different colours showing the tagging power of the tagger.

be defined by the same definition of the JP, called *Jet B Probability* (BJP). Shape of the three dimensional impact parameter and the JPB discriminator calculated with 8 TeV collision data are presented in Fig.4.2.9, where the relative fraction of b quarks, c quarks and light flavours are enlightened in different colours.

The Tagging Efficienc

An example of the CSV algorithm performances is described in Fig. 4.2.10, comparing to the Monte Carlo efficiency evaluated for the other algorithms described above [89]. The light flavour and c-jet misidentification probabilities as a function of the b-jet efficiency for PF jets with $p_T > 60$ GeV in multijets events show the above mentioned better performances of the combined algorithm. The b- tagging efficiency is defined as

$$\epsilon_b = \frac{f_b^{tag} \cdot N_{data}^{tag}}{f_b^{tag} \cdot N_{data}^{tag} + f_b^{untag} \cdot N_{data}^{untag}},$$

where f_b is the fraction of b jets extracted from data and N are the total yield of tagged and untagged jets. The ϵ_b values can be calculated with several methods, basically depending on the b-jet p_T region under consideration. For example, the *PtRel* and the *System8* methods provide precise measurements for the lower part of the spectrum, while the *IP3D* and the *LT* methods have been designed for high jet p_T. In the case of the CSV tagger, the b-tagging efficiency evaluated with all the above mentioned methods, measured in data and Monte Carlo, is reported in Fig.4.2.11 in terms of the relative scale factors coefficients $\mathrm{SF}_b = \frac{\epsilon_b^{data}}{\epsilon_b^{MC}}$ measured in a multijets sample with 8 TeV data of the 2012 CMS data taking as a function of the jet p_T. Also the combination of all the different methods is shown. The application of the $b-$tagging scale factors depends on the number of tagged $b-$jets with respect to the total number of selected jets. In the $Z + (\geq 1b)$ case, where at least one of the n jets is tagged, the event is weighted by a factor

$$w(\geq 1|n) = 1 - w(0|n),$$

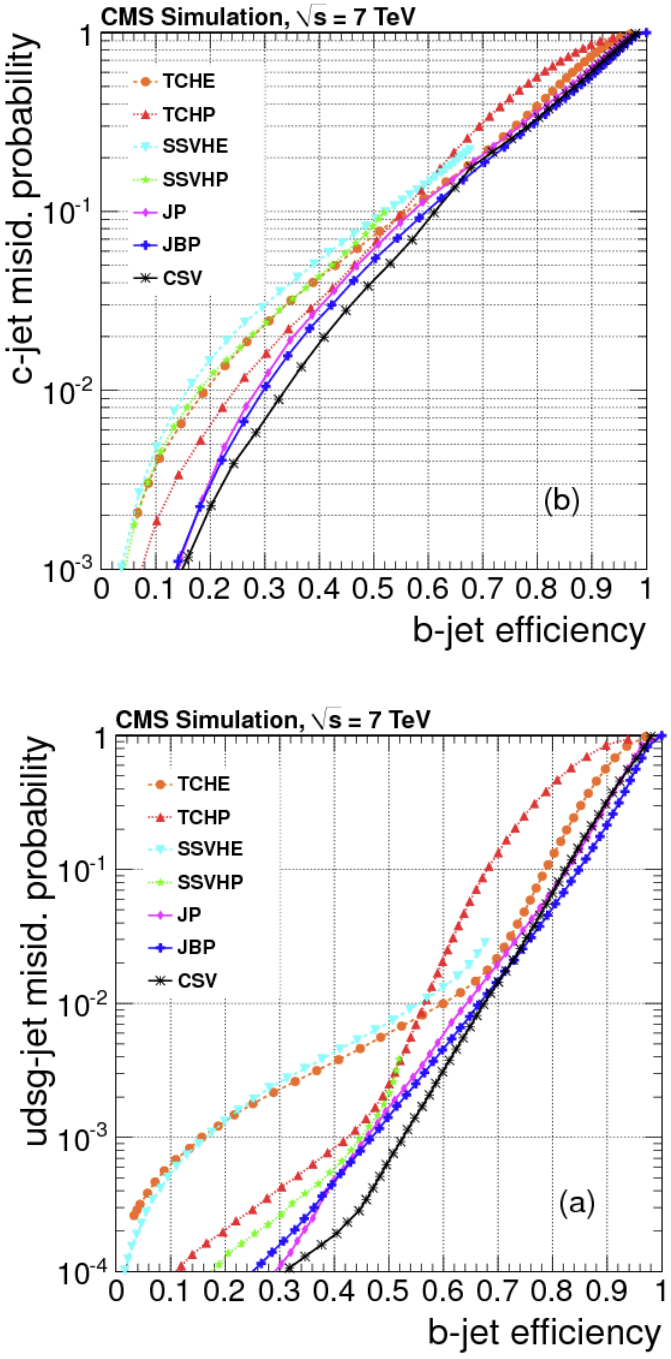

Figure 4.2.10: The mis-identification probability separately evaluated for light ($udsg$) flavours (top) and c (bottom) as a function of the b jet efficiency, for the different taggers described in this Section (coloured lines).

where

$$w(0|n) = \prod_{i=1}^{n}(1 - \mathrm{SF}_i),$$

and SF_i is the $i-$th scale factor. In the case of two $b-$tagged jets, $Z + (\geq 2b)$, the event weight is defined as

$$w(\geq 2|n) = 1 - w(0|n) - w(1|n)$$

$$w(1|n) = \sum_{j=1}^{n}\left[\prod_{i=1, i\neq j}^{n}(1 - \mathrm{SF}_i)\right]\mathrm{SF}_j.$$

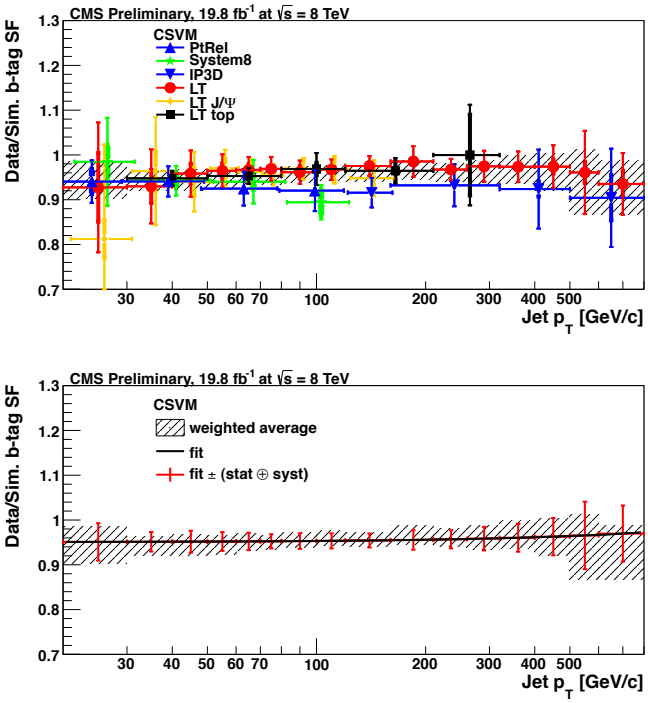

Figure 4.2.11: The Combined Secondary Vertex Scale Factors calculated with different techniques employed to evaluate the b-tagging efficiency (coloured lines) as a function of the jet p_T with 8 TeV data.

Chapter 5

Analysis of $Z + b$ Events

The aim of this work is to perform the measurement of the differential cross sections for events with a Z boson and at least one b quark in the final state

$$pp \rightarrow Z(\rightarrow \ell^+\ell^-) + b + X,$$

with $\ell = e, \mu$, as a function of a set of kinematic observables which are most sensitive to the structure of the initial hard process. In this Chapter I will explain the experimental procedure and the strategy adopted in order to identify and unveil the $Z + b$ final state from the LHC pp collision at 8 TeV. I will start with the description of the data samples and the statistics employed for the measurement, then I will pass through the selection criteria used to select muons, electrons and jets characterizing the $Z+$jets events topology. Particular attention is then dedicated to the pile-up problem, i.e. the description of events with many extra vertices due to soft interactions and underlying events in the pp collisions. Finally, the techniques employed to identify and select jets coming from the hadronization of a b quark ($b-$jets) are presented, together with the description of the method used to extract the b purity. In section 5 a description of the background events that mimic our signal, and the strategy adopted to evaluate each single contribution are discussed, with dedicated attention on the dominant $t\bar{t}$ component. The final results, showing the distributions of the kinematic observables for the $Z+$jets and $Z+b$ event topology after the selection chain, are presented. I will refer to the results of this Chapter as "detector level" results, focusing on the fact that no theoretical comparison or unfolding is done at this stage.

5.1 Luminosity, Datasets and Trigger

5.1.1 Electrons and Muons Datasets

The total integrated luminosity exploited in this analysis is based on the full statistics recorded in the CMS experiment in 2012 pp collisions data taking periods: 19.8 fb^{-1} in the $Z \rightarrow e^+e^-$ selection and 19.7 fb^{-1} in the $Z \rightarrow \mu^+\mu^-$ selection, pre-selected and stored into two primary datasets with a dilepton or dimuon pair trigger selection, called `Double Electron` and `Double Muon`. The different sub-periods that generate the full sample, according to the different timescale and technical conditions of the machine, are described in Table 5.1-5.2.

Run	Number of Events	Integrated luminosity $[\text{fb}^{-1}]$
2012 A	12964286	0.889
2012 B	23571931	4.429
2012 C	33843769	7.152
2012 D	34526899	7.318
full 2012	104906885	19.78

Table 5.1: Number of events and relative integrated luminosity for the partial runs of the total 2012 CMS statistics employed for the `Double Electron` primary dataset.

Run	Number of Events	Integrated luminosity $[\text{fb}^{-1}]$
2012 A	6432930	0.889
2012 B	29282993	4.429
2012 C	36820243	7.152
2012 D	38006513	7.318
full 2012	110542679	19.78

Table 5.2: Number of events and relative integrated luminosity for the partial runs of the total 2012 CMS statistics employed for the `Double Muon` primary dataset.

5.1.2 Trigger Selection

Candidate electrons from the `Double Electron` dataset are requested to be matched with one of the two specific high level trigger objects, selected with asymmetric transverse momentum cuts:

- `HLT_Ele17_CaloIdL_CaloIsoVL_Ele8_CaloId_CaloIsoVL`
- `HLT_Ele8_CaloIdT_CaloIsoVL_Ele17CaloId_CaloIsoVL`

These triggers require at least one electron to have a transverse momentum of at least 17 GeV and at least another electron with $p_T > 8$ GeV and both electrons must pass some minimal calorimetric identification cuts to reduce the huge background of QCD-induced electrons. Muons must be matched with the HLT trigger object after the `Double Muon` preselection:

- `HLT_Mu17_Mu8`

Again an asymmetric selection on the two objects is made, requiring at least a muon with $p_T > 17$ GeV and at least another muon with $p_T > 8$ GeV. Selected electron and muon pairs, triggered with the HLT criteria described above, are then passed through a series of identification, isolation and reconstruction cuts in order to give high momentum isolated objects coming from a Z boson decay. Finally, for top background studies (Subsec. 5.4.2), electrons and muons are required to pass a dedicated $e\mu$ high level trigger menu, defined with similar criteria in the momentum threshold as the dielectron and dimuon case:

- `HLT_Mu17_Ele8_CaloIdT_CaloIsoVL_TrkIdVL_TrkIsoVL`.

Process	**Generator**	**Order**	**Cross Section** [pb]
Z/γ^*	FEWZ [96]	NNLO, $m_b = 0$	3503.71
$Z/\gamma^* + b\bar{b}$	MCFM [97]	LO, $m_b \neq 0$	76.75
$t\bar{t}$	MCFM [98]	NLO	225.197
$W +$ jets	MCFM [99]	NLO	31200
WW	MCFM [99]	NLO	54.838
ZZ	MCFM [99]	NLO	8.059
WZ	FEWZ [96]	NLO	33.21
$W +$ jets	FEWZ [96]	NNLO	31200

Table 5.3: Theoretical predictions for the Standard Model cross sections used in this analysis.

Process	**Generator**	**Number of Events**
Z/γ^*	MadGraph5F	30459503
$Z/\gamma^* + b\bar{b}$	MadGraph4F	7551580
Z/γ^*	sherpa	43915740
Z/γ^*	powheg ee	3297045
Z/γ^*	powheg $\mu\mu$	3283740
$t\bar{t}$	MadGraph	6923750
$W +$ jets	MadGraph	57709905
WW	pythia	10000431
ZZ	pythia	9799908
WZ	pythia	10000283
$W +$ jets	MadGraph	57709905

Table 5.4: Signal and background simulated samples, for the different MC generators choices, with the total number of generated events.

5.1.3 Monte Carlo Samples

The results shown in the following sections are always obtained with proton-proton collision data, are always compared with simulated events using a full Monte Carlo simulation. Simulated signal events, modeled by Drell-Yan $Z/\gamma^* +$ jets final state, are produced by the matrix element calculation of MadGraph [94] and powheg [53] interfaced with Pythia6 (Tune Z2*) [47] for the hadronization. The full reconstructed MadGraph simulation is used as default for the main data to MC agreement study, as well as the background processes. Two different versions of MadGraph are used, exploiting the different descriptions of the b quark production in the hard scattering (4-flavour scheme and 5-flavour scheme, described in detail in Subsec. 2.1.1-2.1.2). The detailed CMS detector and simulation is currently based on the geant4 [95] program, describing electromagnetic and hadronic interactions of particles with the CMS detector material. It is also the main tool for modeling the full CMS detector and geometry simulation. The theoretical calculations for the standard model cross sections in proton-proton collisions at 8 TeV inside the generators are made using the following parameters as input for the calculation:

- $m_Z = 91.1876$ GeV
- $m_W = 80.398$ GeV
- $\Gamma_W = 2.141$ GeV
- PDF : CTEQ6M
- $m_t = 172.5$ GeV
- $m_b = 4.8$ GeV
- $m_c = 1.27$ GeV
- $\alpha_{\mathrm{QED}} = 0.00729735$

In Table 5.3 the complete list of simulated processes for signal and backgrounds employed in this analysis is presented, with the respective theoretical cross section at the given order in pQCD, the generators used for the cross section calculations and details of the production characteristics. The number of events obtained after the $Z + b$ selection for signal and background processes simulated in the analysis with different generators used are presented in Table 5.4.

5.2 Events Selection

Events are required to have a Z boson reconstructed through its decay products (electron and muon pairs), in association with at least one hadronic jet to form the $Z +$ jets sample. Physics objects are always reconstructed using the Particle Flow algorithm, as described in Chapter 4. Dielectrons pairs coming from a Z boson are selected if they are matched with the corresponding trigger object, have opposite charge and a transverse momentum of at least 20 GeV inside the geometrical range of $|\eta| \leq 1.442$ and $1.566 \leq |\eta| \leq 2.4$. The selected pair is taken only if its invariant mass lies inside the Z boson mass window $M_{e^+e^-} \in [71; 111]$ GeV. Electrons must satisfy a set of selection cuts aimed to guarantee the quality of the candidate by minimizing the fake rate and maximizing the identification efficiency (*MediumID*, see Subsec. 4.2.3). Dimuon pairs from a Z boson are selected if they are matched with the corresponding trigger object, have opposite charge and a transverse momentum of at least 20 GeV inside the geometrical range of $|\eta| \leq 2.4$. The selected pair is taken only if its invariant mass lies inside the Z boson mass window $M_{\mu^+\mu^-} \in [71; 111]$ GeV. Muons must satisfy the *tight* selection criteria (see Subsec. 4.2.2). Furthermore, electrons and muons must be isolated in order to reject other leptons coming from the semileptonic decay of heavy flavours (Subsec. 4.2.1-4.2.2). Jets are clustered by the anti-k_T algorithm with a cone size of $\Delta R = 0.5$. To suppress jets that are produced by minimum bias and underlying events, a cut on the jet momentum is applied, requiring $p_T > 30$ GeV. Furthermore, only jets inside the tracker acceptance are used, requiring $|\eta| < 2.5$. Leptons coming from different processes inside jets are removed from the clusterization, and several cuts on the energy fraction composing the jets are applied (Sec. 4.2) in order to ensure a better rejection of energy contamination from neutrinos, photons and leptons (Table 5.5). In data and Monte Carlo simulation, the jet energy is corrected by the subsequent application of the first three levels of jet energy corrections, LV1+LV2+LV3 (Sec. 4.2). Only in data, the residual correction LV2LV3Res is also applied. The b jets are tagged using the Combined Secondary Vertex algorithm with a tight working point: CSV > 0.895 (see Subsec. 4.2.5 for a complete description). Finally, events are selected if they contain a Z boson reconstructed decaying into e^+e^- or $\mu^+\mu^-$ and at least one b - tagged jet in order to construct the $Z + b$ final sample.

Jet Energy Composition	Cut
Neutral Hadronic Energy Fraction	< 0.99
Charged Hadronic Energy Fraction	> 0
Neutral Electromagnetic Fraction	< 0.99
Charged Electromagnetic Fraction	> 0.9
Charged Multiplicity	> 0
Number of Constituents	> 1

Table 5.5: Jet energy composition cuts.

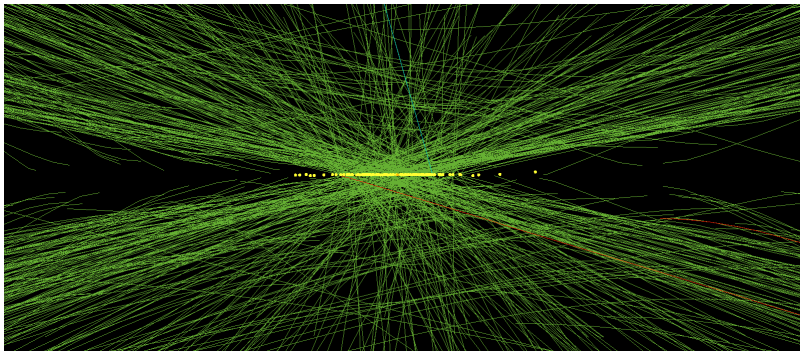

Figure 5.3.1: An example of 78 reconstructed vertices in the CMS tracker, in 2012 data taking, run 198609.

5.3 Pile-Up Modeling

One of the most important problems to face off at very high luminosity hadronic collisions is the presence of multiple pp collisions taking place either in the same bunch crossing (*in-time* pile-up events) or due to events in different crossings *(out-of-time* pile-up events). The effect of such extra events, not coming from the primary vertex, manifests as extra energy deposits inside calorimeters and extra particles (mostly with low momentum), polluting the specific process under study. Increasing luminosity, pile-up becomes even more important and it is crucial to precisely estimate it to ensure an uncontaminated measurement of the specific physics observable under consideration. In 2012 data taking, at the LHC energy, $\sqrt{s} = 8$ TeV, the average number of pile-up vertices is measured to be about 21. The highest number of reconstructed vertices in the CMS tracker until today was recorded in 2012 data taking, showing the impressive scenario of 78 events (Fig. 5.3.1). In order for the Monte Carlo to match with the observed pile-up distribution in data, a reweighting procedure has been used. The reweighting technique is based on the estimation of the minimum bias cross section together with the instantaneous luminosity for each bunch crossing in order to produce a set of weights, defined by the ratio between the observed number of pile-up events in data and in simulation in the ith bin:

$$w_{\mathrm{PU}}(i) = \frac{N_{\mathrm{DATA}}(i)}{N_{\mathrm{MC}}(i)}$$

Analyzing the Monte Carlo samples, each event is thus reweighted by $w_{\mathrm{PU}}(i)$. The pile-up events distribution is calculated by the convolution of the bunch crossing instantaneous luminosity and the total inelastic pp cross section.

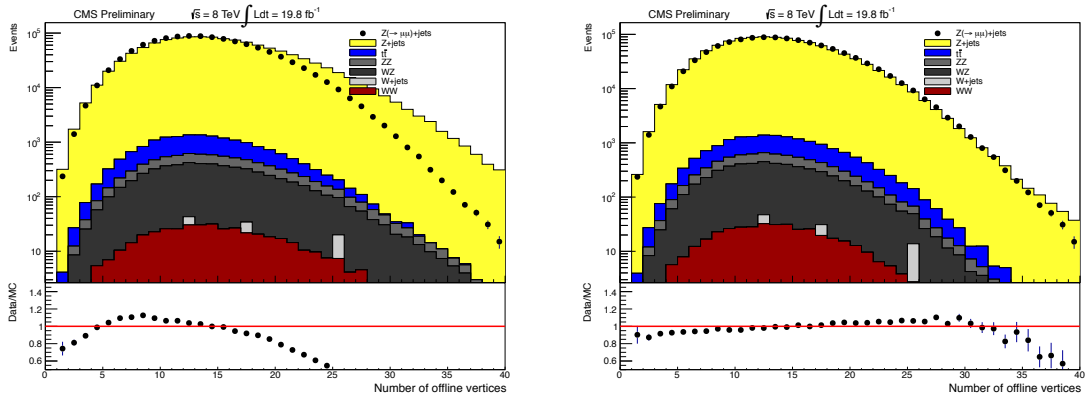

Figure 5.3.2: Number of reconstructed vertices before the reweighting (left plot) and after the reweighting (right plot) for $Z+$jets selection with $\mu^+\mu^-$ final state .

5.3.1 Monte Carlo Reweighting Tests

In order to test the Monte Carlo reweighting procedure, and to make sure that simulation takes into account this effect according to the pile-up seen in data for each event, the number of reconstructed vertices after the electrons and muons selection described above has been chosen as a test. In Fig. 5.3.2 the distribution of the reconstructed vertices in $Z \to e^+e^-$, $Z \to \mu^+\mu^-$ events with at least one associated jet is shown, comparing data to MadGraph inclusive Drell-Yan plus jets Monte Carlo with the different background contributions included. As it can be seen, the reweighting procedure fixes the discrepancy between simulation and data, giving a reasonable comparison after the correction. The agreement is assessed by the ratio between data and Monte Carlo in the bottom part of the plot, showing a very good agreement in the comparison.

5.3.2 Isolation correction

Isolation for electrons and muons is calculated as a sum of energy inside a cone (as described in Subsec. 4.2.1 - 4.2.2). When high pile-up occurs, charged particles from vertices other than the primary vertex can be found inside the cone and contribute to the isolation value of the muon or electron candidate under consideration. It is also possible that neutral energy, coming from neutral hadrons, with no tracks reconstructed in the tracker and so impossible to assign to vertices, enters in the isolation cone. For such cases, an extra correction to the isolation value is applied, called *delta beta* correction:

$$\Delta\beta = I_{charged\ hadron} + \max\{0, I_\gamma + I_{neutral\ hadron} - \frac{1}{2}\frac{I_{PU-charged}}{p_T^\mu}\}.$$

where the $I_{charged,neutral}$ terms represent the electromagnetic, hadronic, neutral and pile-up originated contribution to the isolation, evaluated in $Z \to \mu^+\mu^-$ and $Z \to e^+e^-$ events. A requirement for the corrected $\Delta\beta$ isolation for muons and electrons isolation of $\Delta\beta < 0.2$ is applied.

5.4 Background Estimation

Several Standard Model processes may come up with a very similar topology to the $Z + b$ final state. The b jets may be produced from the c quark decay, from the t quark decay, or in multijet production in association with W, Z bosons. Diboson production (WW, ZZ, WZ) in association with jets is also an important class of processes that can generate a $Z + b$ final state, as well as QCD interactions. A brief description of the physics background processes evaluation is described in this Section. A complete list of the simulated number of events and their cross section can be found in Table 5.3.

5.4.1 Vector Bosons Production

Dibosons (W, Z) production may mimic the $Z + b$ final state by missing the reconstruction of one or more leptons with the emission of at least one jet. They can be summarized in the following processes:

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

Such backgrounds are evaluated from Monte Carlo simulations. A smaller source of background contamination can be the production of a single $W^\pm$ in association with jets, where a secondary lepton is wrongly identified. The presence of τ lepton pairs, produced from a Z/γ^* in association with jets may also be considered as fake signal, since it may lead to two electrons or muons plus missing energy due to the neutrinos emission:

- $Z/\gamma^* \rightarrow \tau^+\tau^- + \text{jets} \rightarrow 2\ell^\pm + 4\nu + \text{jets}$.

Other backgrounds are the general QCD production of multijet might mimic the signal topology, and the single top production in the $s-$channel and in the $t-$channel. From Monte Carlo studies, these contributions are found to be negligible, as it can be seen in the distributions (Sec. 5.6).

5.4.2 Top Background Estimation

The top-antitop ($t\bar{t}$) production is the dominant background in the $Z + b$ analysis since top quarks always produce bottom quarks, in association with two leptons from the W decays in the about 10% of the cases following the decay chain :

$$t\bar{t} \rightarrow W^+W^-b\bar{b} \rightarrow \ell^+\ell^-\nu\bar{\nu}b\bar{b}.$$

In order to study this background, the same selection used for $Z(\rightarrow ee, \mu\mu) + b$ sample is applied to a $(e\mu) + b$ sample, with leptons triggered by a specific $e\mu$ HLT path and dataset (Subsec. 5.1.2). Events are selected if the $e\mu$ invariant mass lies

in the mass range [50; 200] GeV. In addition, an extra cut on the missing transverse energy requiring $\not\!\!E_T > 100$ GeV is applied in order to select events in a region of the phase space that is substantially dominated by this background. The number of $e\mu$ events is expected to be twice the number of $ee, \mu\mu$ events:

$$N_{e\mu} = c_t \cdot N_{\ell\ell}, \qquad \ell\ell = ee, \mu\mu$$

$$\text{with} \qquad c_t \sim 2.$$

When comparing the $e\mu$ results obtained in data with the $t\bar{t}$ Monte Carlo results obtained in the standard ($\ell\ell$) selection, some distributions appear to be shifted, as it can be seen for example in the Z boson momentum distribution (Fig. 5.4.1). The Monte Carlo appears not to be fully reliable in the description of the $t\bar{t}$ background component in the $Z+b$ analysis. For this reason it has been decided to estimate the $t\bar{t}$ background shape directly from the $e\mu$ sample in data. The normalization of the $e\mu$ sample is obtained by fitting the $\ell\ell$ invariant mass sidebands in data with the $e\mu$ invariant mass sidebands template. The results of the fit used to extract the c_t normalization coefficient are presented in Fig. 5.4.2 for the dielectron and dimuon final state. To check the consistency of the fit, the c_t coefficient has been extracted also by fitting the missing transverse energy templates of the $e\mu$ shape in data to the missing energy distribution obtained in the $\ell\ell$ sample (Fig. 5.4.3). Results obtained are in good agreement between the two fits. The full summary of the fit results obtained in electrons and muons final state for the inclusive Z + jets and for the $Z+b$ process are reported in Table 5.6. In the following, both the normalization and the shape of the $t\bar{t}$ background contribution will be taken from the $e\mu$ data sample after subtracting all the other background contributions evaluated from the Monte Carlo predictions.

Channel	$e\mu$ **Invariant Mass Fit**	$\not\!\!E_T$ **Fit**	$\not\!\!E_T$ **Significance Fit**
$Z \to ee + b$	0.435 ± 0.010	0.473 ± 0.014	0.433 ± 0.009
$Z \to \mu\mu + b$	0.560 ± 0.011	0.576 ± 0.012	0.562 ± 0.011
$Z \to ee + \mathrm{j}$	0.454 ± 0.008	0.441 ± 0.010	0.446 ± 0.011
$Z \to \mu\mu + \mathrm{j}$	0.578 ± 0.010	0.596 ± 0.017	0.556 ± 0.013

Table 5.6: The c_t coefficient extracted by fitting different distributions for the Z+jets and $Z + b$ in muons and electrons final state.

5.5 Estimation of the *b* Fraction

The b-jets fraction f_b is defined as the number of events after the $Z + b$ selection that truly contains at least one b-tagged jet. By means of the f_b estimation, the contamination of different quark flavours can be extracted in addition, so the inclusive sample will be divided in its b- , c-, light−flavour components. Purity is a crucial parameter for this analysis since the final cross section for the $Z + b$ final state depends on the b-purity, and so the Monte Carlo simulation has to be corrected for the real fraction of events containing b quarks actually present in data.

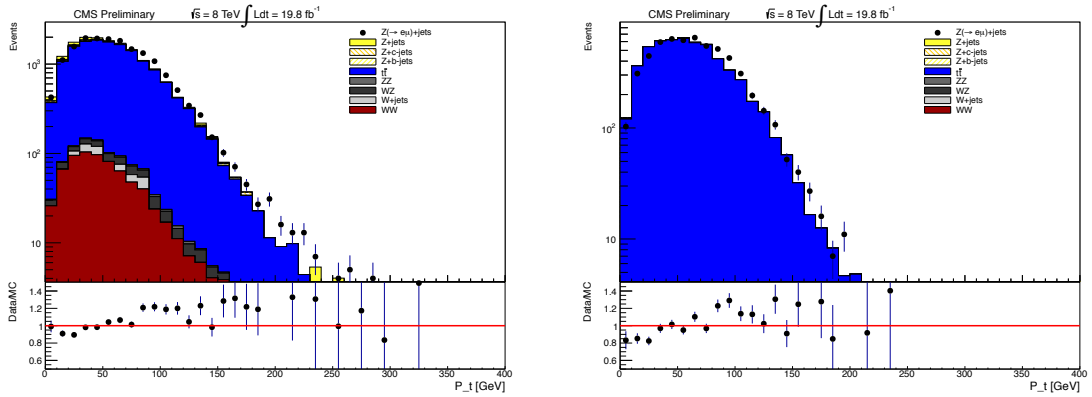

Figure 5.4.1: Electron-muon pairs transverse momentum in the $Z+$jets (left) and $Z+b$ (right) sample after the $e\mu$ selection. The plots show the different shapes between the corrected $t\bar{t}$ Monte Carlo sample (blue area)and the $e\mu$ data (black points). In the ratio plot a very similar trend of the shift can be seen for both distributions.

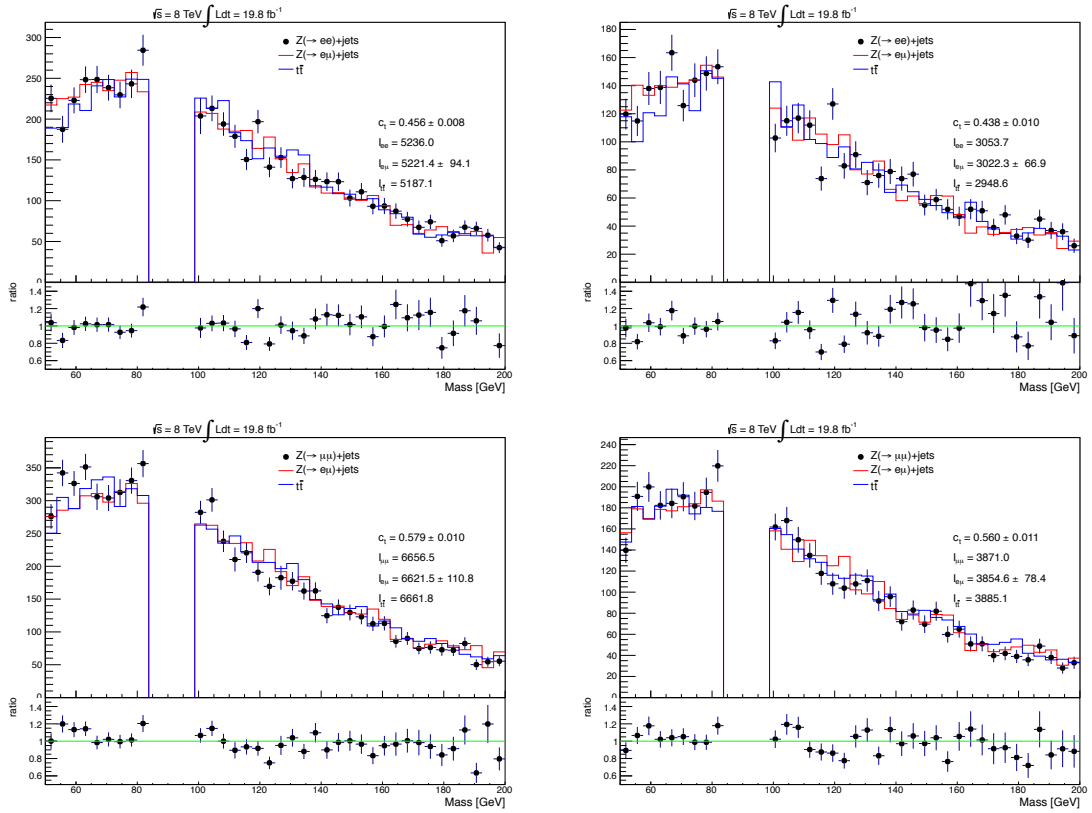

Figure 5.4.2: The $Z \to e^+e^-$ (top) and $Z \to \mu^+\mu^-$ (bottom) invariant mass sidebands $M_{ee/\mu\mu} < 85$ GeV and $M_{ee/\mu\mu} > 100$ GeV fitted to extract the c_t coefficient, for the inclusive $Z+$jets (left) and the $Z+b$ selection. Black dots represent data, blue line represents the $t\bar{t}$ Monte Carlo and red line represents the $e\mu$ sidebands shapes. Also the total integral of the three distributions (ee,$e\mu$,$t\bar{t}$) is shown in the plot.

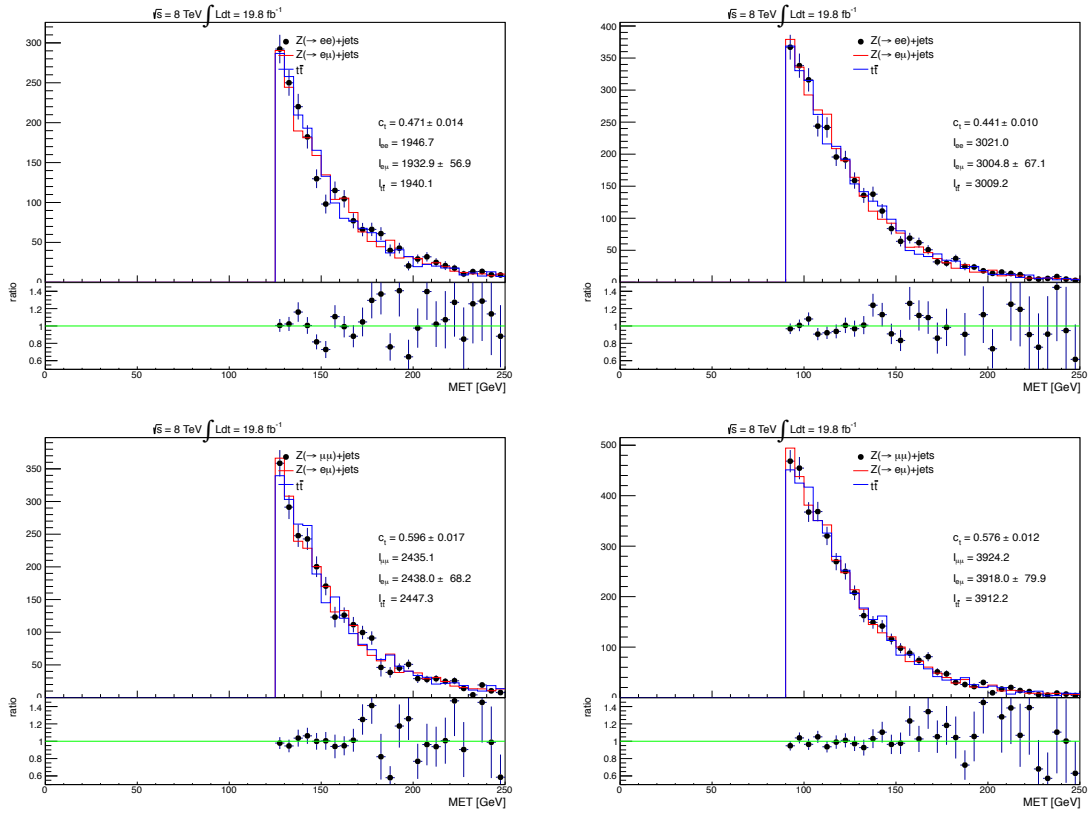

Figure 5.4.3: The missing transverse energy distributions for Z + jets (left) and $Z + b$ (right) fitted to extract the c_t coefficient. Black dots represent data, blue line represents the $t\bar{t}$ Monte Carlo and red line represents the $e\mu$ sidebands shapes. Also the total integral of the three cases (ee,$e\mu$,$t\bar{t}$) is shown in the plot.

Distribution	b **fraction %**	c **fraction %**	*light* **fraction %**
Jet Probability B (ee)	59.0±3.9	30.6±6.1	10.4±3.5
Jet Probability B ($\mu\mu$)	60.0±3	29.5±3.4	10.7±1.6
B-Jet Probability (ee)	62.7±2.8	30.7±3.3	6.6±1.5
B-Jet Probability ($\mu\mu$)	63.1±2.4	29.7±2.6	7.2±1.1
Secondary Vertex Mass (ee)	70.3±7.8	20.9±7.7	8.8±10.3
Secondary Vertex Mass ($\mu\mu$)	66.9±2,8	21.7±5.1	11.1±2.3

Table 5.7: Fraction of beauty, charm and light (up, down, strange and gluons) quark jets in $Z + b$ events, extracted by fitting the Jet Probability Discriminator and the Secondary Vertex Mass Distribution for dielectron and dimuon final state.

In this analysis the purity is estimated by the Monte Carlo simulation using the generator information on the different flavour composition of each jet. Two distributions have been chosen in past analyses [100, 101] in order to fit the different shape of the different contributions. One variable chosen to fit the b-jets purity is the Secondary Vertex Mass distribution (SVTX Mass). This variable is defined as the invariant mass of all the tracks coming from a reconstructed secondary vertex, assuming the π^0 mass for each. The reason why this variable is so effective in the flavour separation derives from the properties of the invariant mass of the hadrons originated by the secondary vertex. Tracks coming from the B meson generally are originated by vertices with bigger masses than the ones coming from a D meson, and of course this one has a higher invariant mass value than light flavours. Another way to extract the b-purity is to exploit the properties of the Jet Probability Discriminators (JPB and BJP, see Subsec. 4.5.1). The fit procedure starts after the request to have at least one b-tagged jet (using the tight CSV working point) associated with a Z boson decaying in muons or electrons in the final state. The diboson and top background is subtracted, then the templates of the $b, c,$ light fraction of the JPB Discriminator distribution, estimated from simulation, are fitted to the data. The JPB Discriminator distribution is presented in Fig. 5.5.1 showing the different flavours composition. As a check for the consistency of the fit procedure, the SVTX Mass is presented in Fig. 5.5.2 together with the BJP discriminators, with the additional request of at least one reconstructed secondary vertex, in Fig. 5.5.3. The two fits, both in the dimuon and dielectron final state, show compatible results for the b-purity extracted by the fit. All the results are given in Table 5.7. In order to better understand the physics of the b-fraction inside the $Z + b$ events, the correction factor c_b containing the information on the fraction of true b in the event, extracted from the fit, has been studied as a function of the most important kinematic variables defining the topology of the event. The dependency of the c_b as a function of the most energetic (leading) b-jet in the event, its pseudorapidity, and the Z boson transverse momentum in muons and electrons final state has been studied, and no evident functional trend is observed in each of these variables (Fig. 5.5.4).

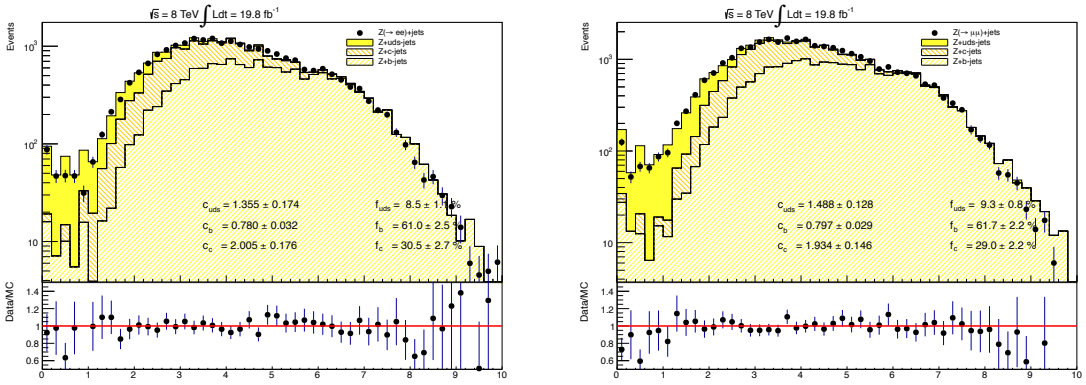

Figure 5.5.1: Jet Probability distribution after the $Z+b$ selection with a Z decaying into electrons (left) and muons (right). The fraction of events containing b quarks is shown with yellow stripes while c fraction is represented by orange stripes. Results of the fit to the different templates are depicted in the plot as $f_{b,c,light}$ for the different flavour percentage inside the sample and with the corresponding scaling factor $c_{b,c,light}$.

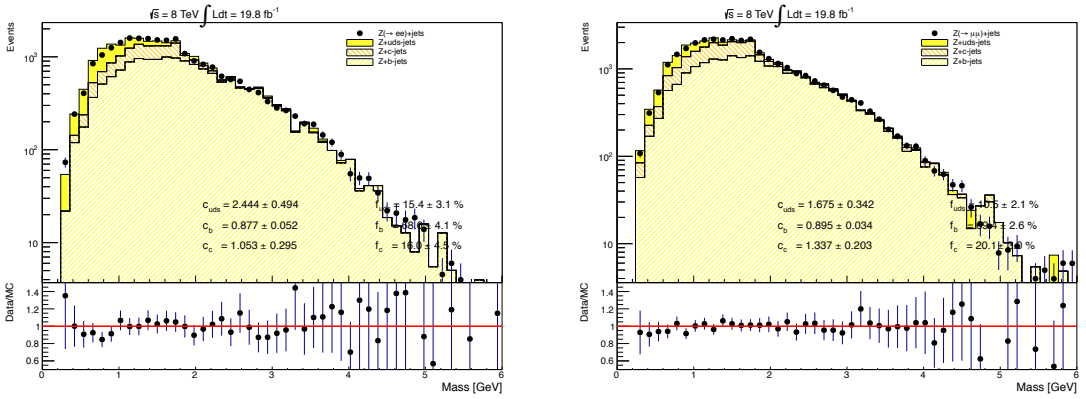

Figure 5.5.2: Secondary Vertex Mass distribution after the $Z+b$ selection with a Z decaying into electrons (left) and muons (right). The fraction of events containing b quarks is shown with yellow stripes while c fraction is represented by orange stripes. Results of the fit to the different templates are depicted in the plot as $f_{b,c,light}$ for the different flavour percentage inside the sample and with the corresponding scaling factor $c_{b,c,light}$.

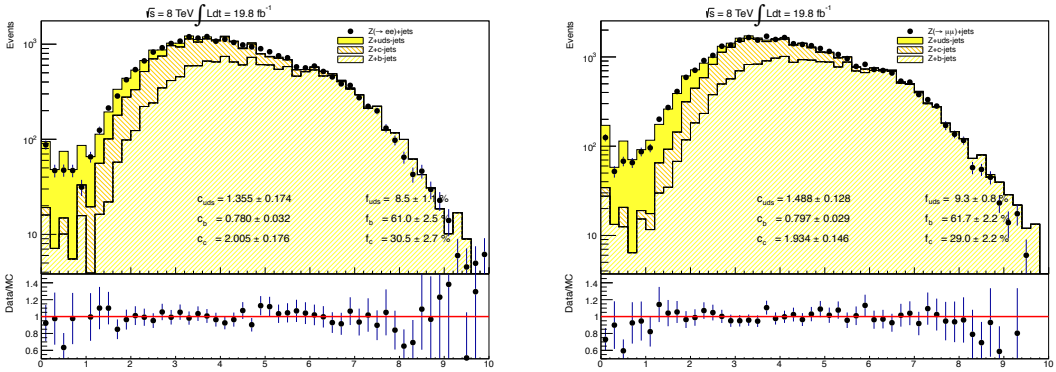

Figure 5.5.3: The B jet probability discriminator (BJP) distribution after the $Z+b$ selection with a Z decaying into electrons (left) and muons (right). The fraction of events containing b quarks is shown with yellow stripes while c fraction is represented by orange stripes. Results of the fit to the different templates are depicted in the plot as $f_{b,c,light}$ for the different flavour percentage inside the sample and with the corresponding scaling factor $c_{b,c,light}$.

5.6 Detector Level Event Distributions

In this Section the kinematic distribution of all the variables that characterize the $Z+b$ final state are presented. All the distributions are presented together with the MadGraph 5FS Monte Carlo full simulation comparison for the signal and all the background contributions, after the correction for the electrons, muons and $b-$jets scale factors and for the jet energy. Results are shown for both the $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$ final state with at least one jet of any flavour for the inclusive regime, $Z+$ jets topology, and with at least one $b-$tagged jet in the event in the $Z+b$ topology. At this stage these distributions are presented without any unfolding procedure, and so for this reason they are referred to as "detector level" distributions in order to emphasize the fact that all the detector effects are still present in data. The unfolding procedure that brings back the distributions presented in this Section to the "particle level" (i.e. bare of the detector effects) distribution will be discussed in the next Chapter. In the first part of this Section the inclusive regime variables (before the $b-$tagging) for leptons, Z bosons and jets are presented, showing the $Z+$ jets final state topology. An event display for the $Z(\to \mu^+\mu^-) + b\bar{b}$ recorded by CMS is presented in Fig. 5.6.1. The characterizing distributions for the $Z+$ jets final state are:

- the p_T and η of the most energetic (leading) and the second (subleading) leptons (Fig. 5.6.2-5.6.5);
- the dilepton invariant mass at the Z boson peak (Fig. 5.6.6):
- the dielectrons and dimuons p_T (Fig. 5.6.7).The p_T and η of the most energetic (leading) and the second (subleading) jet (Fig. 5.6.8-5.6.10);
- the jet multiplicity (Fig. 5.6.11);
- the scalar sum of all the jets momenta in the event, $H_T = \sum_j p_T^j$ (Fig. 5.6.12);

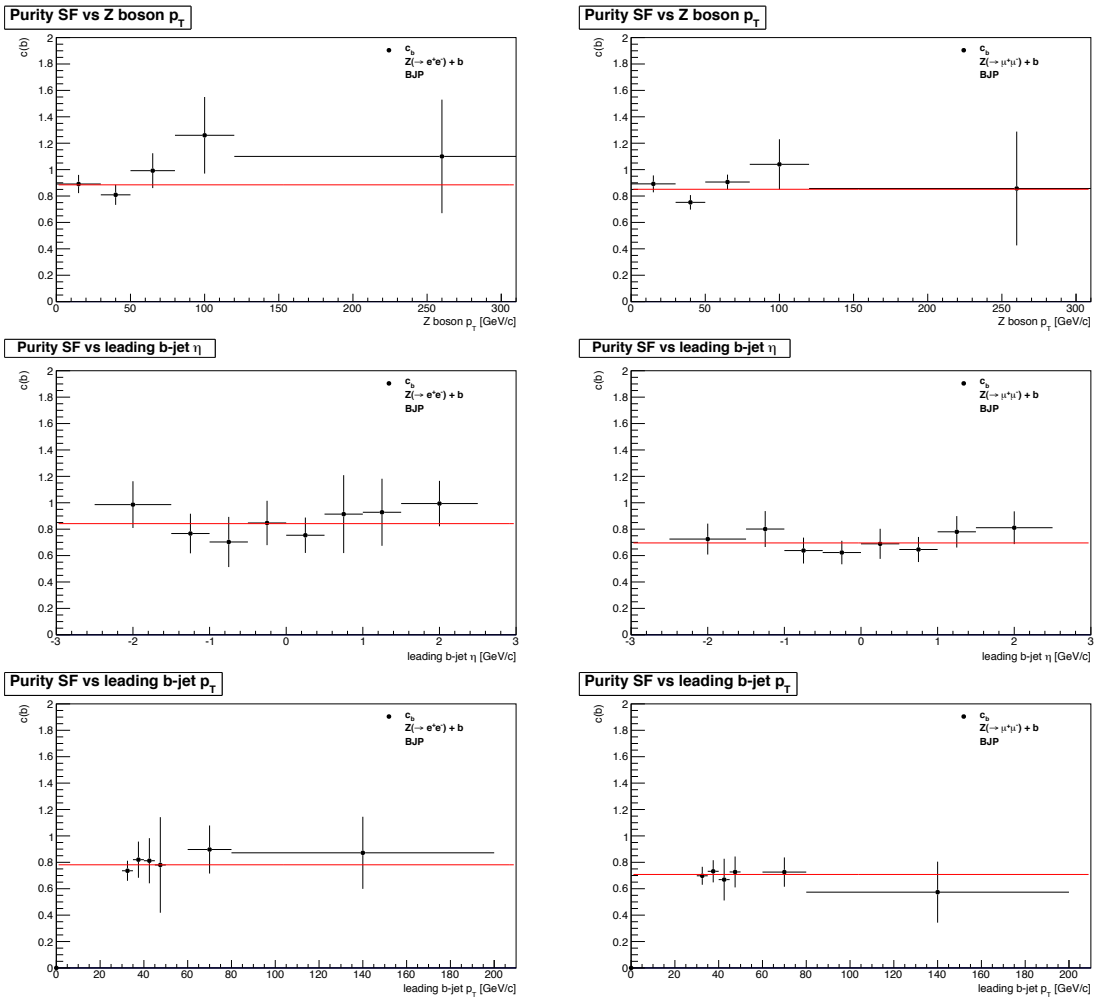

Figure 5.5.4: The b fraction scaling factor c_b extracted from the fit to the BJP discriminator after the $Z+b$ selection, with a Z decaying into electrons (left) and muons (right) final state as function of the Z boson transverse momentum (top), the leading $b-$jet pseudorapidity (center), leading $b-$jet transverse momentum (bottom).

- the azimuthal angle between the Z boson and the most energetic jet in the event, $\Delta\phi^{Zj} = |\phi_Z - \phi_{leading\ jet}|$ (Fig. 5.6.13).

In the second part of the Section the $b-$tag sample is presented, showing the Z plus (at least one) $b-$tagged jets in the event topology. To characterize the $Z + b$ final state the following distributions are presented:

- the invariant mass distribution of the dilepton pair with one $b-$tagged jet associated (Fig. 5.6.14);
- the p_T of the Z boson with one $b-$tagged jet associated (Fig. 5.6.15);
- the p_T and η of the most energetic (leading) $b-$jet in the event (Fig. 5.6.16);
- the number of $b-$tagged jets in the event (Fig. 5.6.17);
- the scalar sum of all the jet momenta in the event, $H_T = \sum_j p_T^j$ with one $b-$tagged jet associated (Fig. 5.6.18);
- the azimuthal angle between the Z boson and the most energetic $b-$jet in the event, $\Delta\phi^{Zb} = |\phi_Z - \phi_{leading\ b-jet}|$ (Fig. 5.6.19).

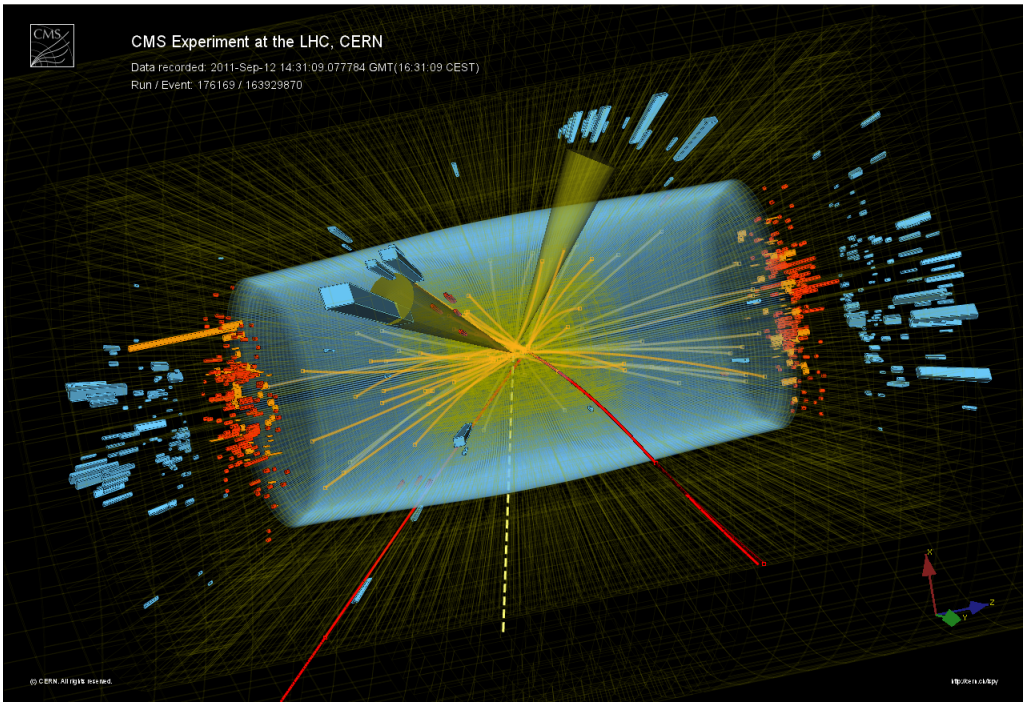

Figure 5.6.1: A CMS event showing two muons (red lines) coming from a $Z \to \mu^+\mu^-$ process in association with two b-tagged jets (yellow cones).

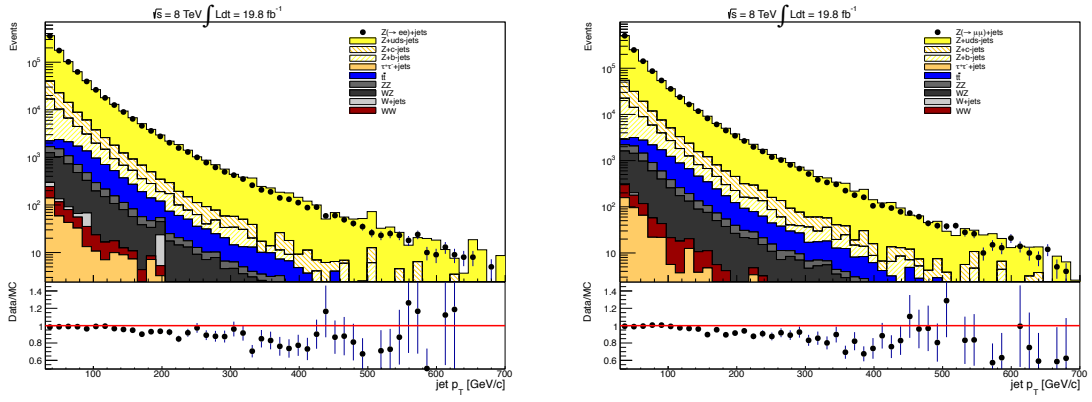

Figure 5.6.2: Transverse momentum distributions for the leading electron and muon in $Z+$jets events for $Z \to e^+e^-$ (left) and $Z \to \mu^+\mu^-$ (right).

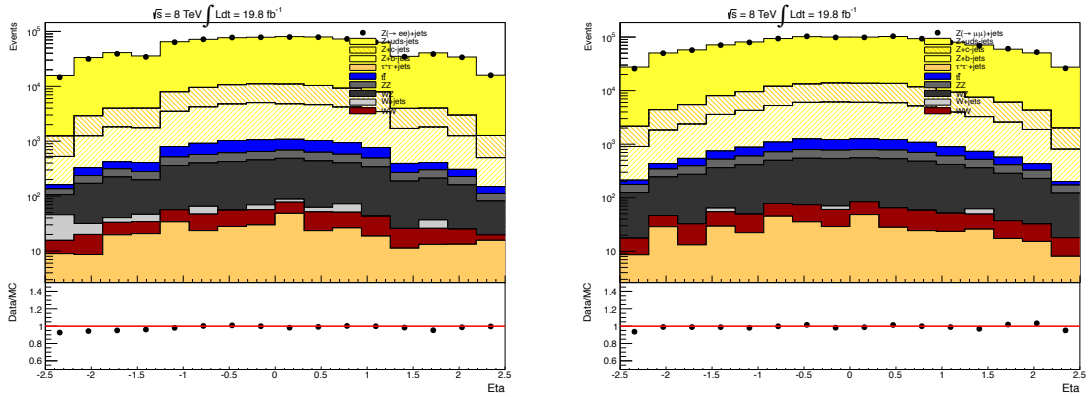

Figure 5.6.3: Pseudorapidity distributions for the leading electron and muon in $Z+$jets events for $Z \to e^+e^-$ (left) and $Z \to \mu^+\mu^-$ (right).

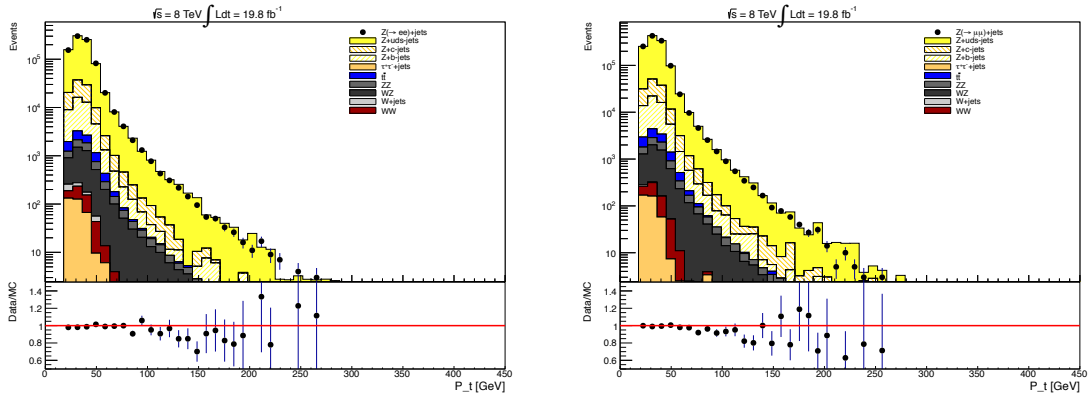

Figure 5.6.4: Transverse momentum distributions for the sub-leading electron and muon in $Z+$jets events for $Z \to e^+e^-$ (left) and $Z \to \mu^+\mu^-$ (right).

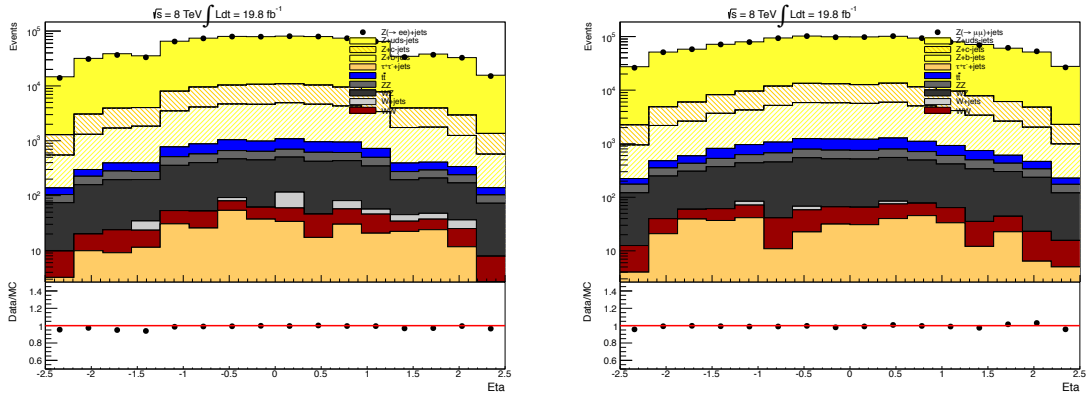

Figure 5.6.5: Pseudorapidity distributions for the sub-leading electron and muon in $Z+$jets events for $Z \to e^+e^-$ (left) and $Z \to \mu^+\mu^-$ (right) .

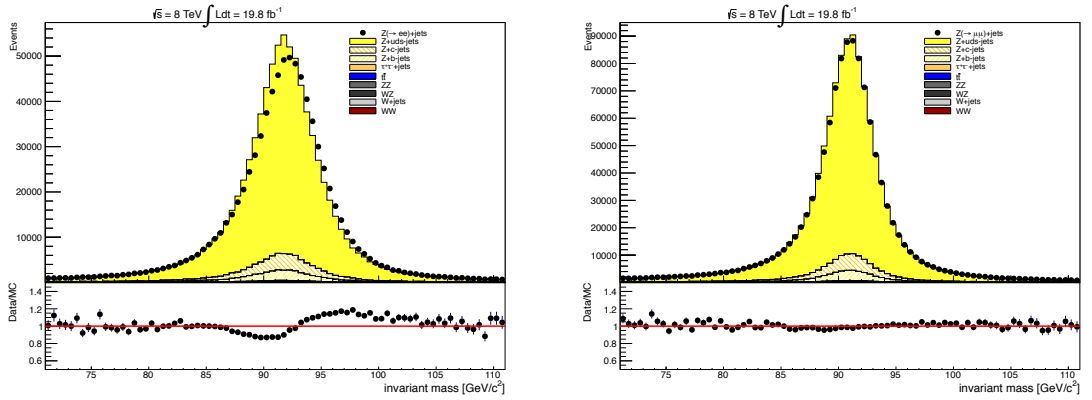

Figure 5.6.6: Invariant mass distributions for the e^+e^- pairs (left) and the $\mu^+\mu^-$ pairs (right) around the nominal Z boson mass.

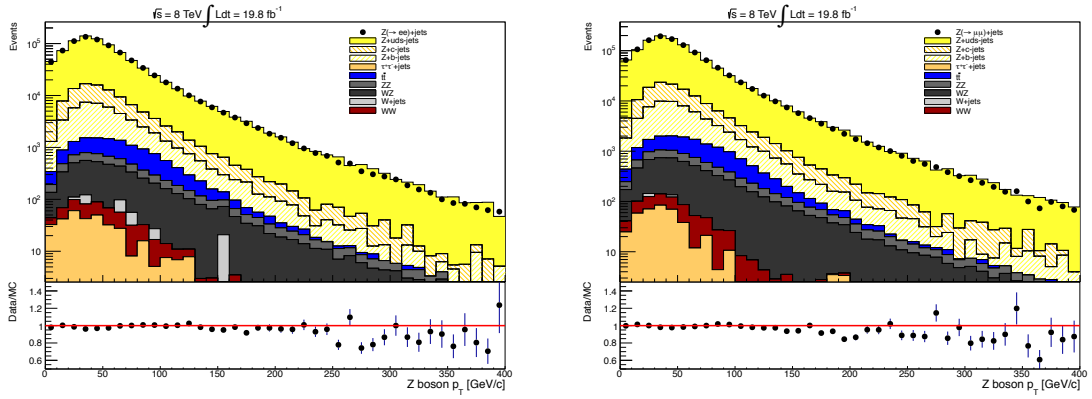

Figure 5.6.7: Z boson transverse momentum for the e^+e^- final state (left) and the $\mu^+\mu^-$ final state (right).

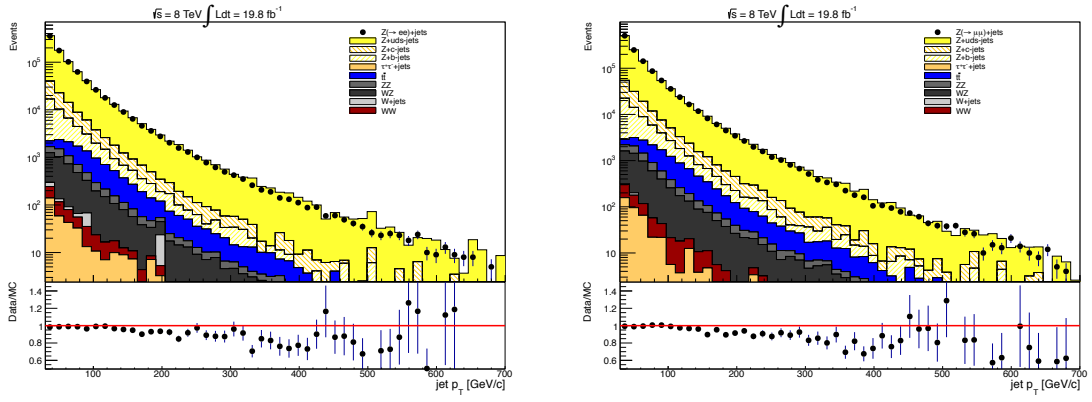

Figure 5.6.8: Transverse momentum distributions for the leading jet in $Z+$jets events for $Z \to e^+e^-$ (left) and $Z \to \mu^+\mu^-$ (right).

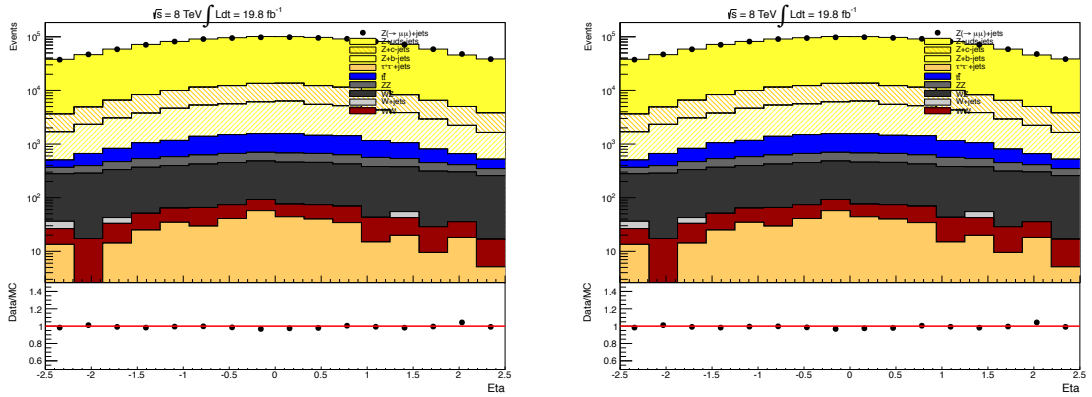

Figure 5.6.9: Pseudorapidity distribution for the leading jet in $Z+$jets events for $Z \to e^+e^-$ (left) and $Z \to \mu^+\mu^-$ (right).

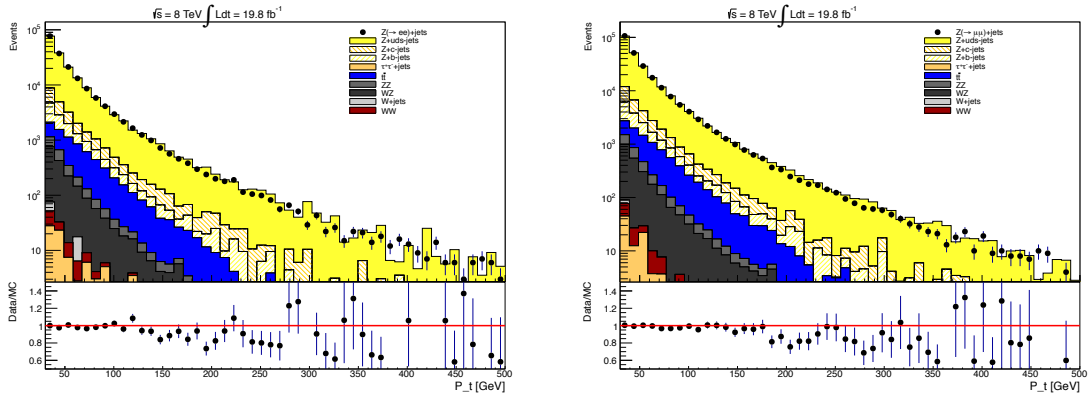

Figure 5.6.10: Transverse momentum distribution for the sub-leading jet in $Z+$jets events for $Z \to e^+e^-$ (left) and $Z \to \mu^+\mu^-$ (right).

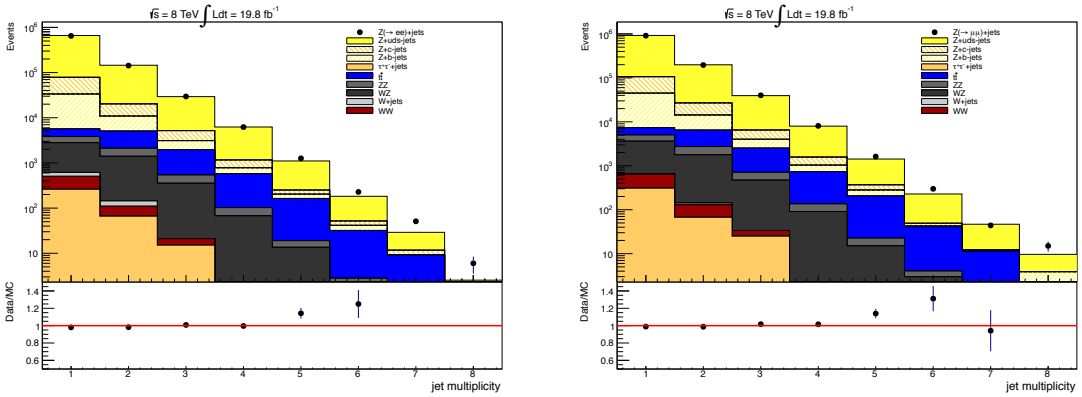

Figure 5.6.11: Jet multiplicity distribution in $Z+$jets events for $Z \to e^+e^-$ (left) and $Z \to \mu^+\mu^-$ (right).

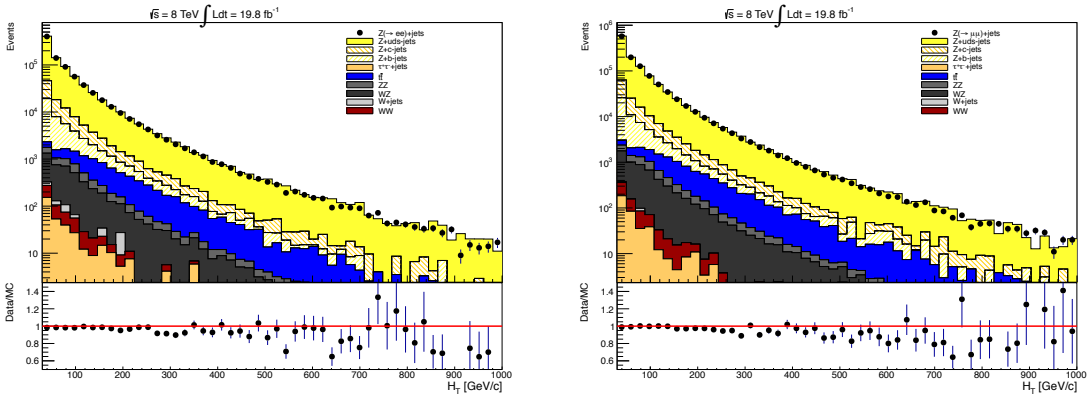

Figure 5.6.12: The total hadronic energy of the event H_T in $Z+$jets events for $Z \to e^+e^-$ (left) and $Z \to \mu^+\mu^-$ (right).

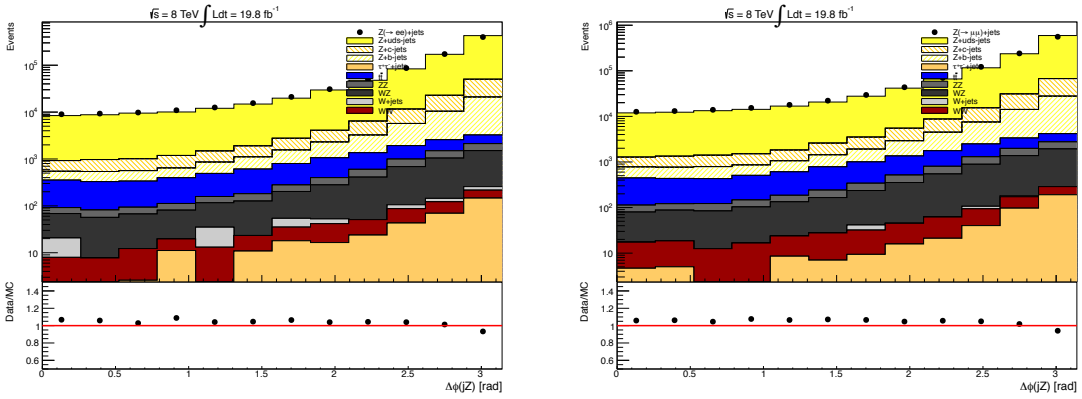

Figure 5.6.13: The azimuthal angle of the Zj system in $Z+$jets events for $Z \to e^+e^-$ (left) and $Z \to \mu^+\mu^-$ (right).

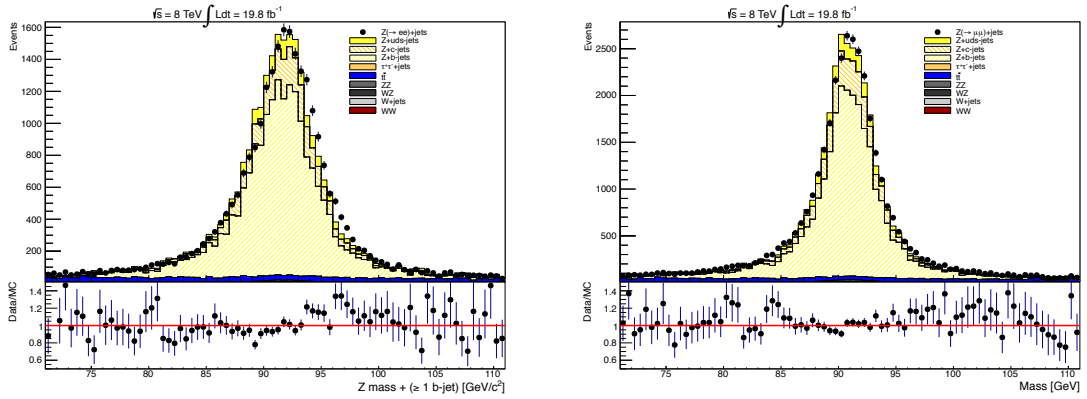

Figure 5.6.14: Invariant mass distributions for the e^+e^- pairs (left) and the $\mu^+\mu^-$ pairs (right) around the nominal Z boson mass with at least one associated $b-$tagged jet.

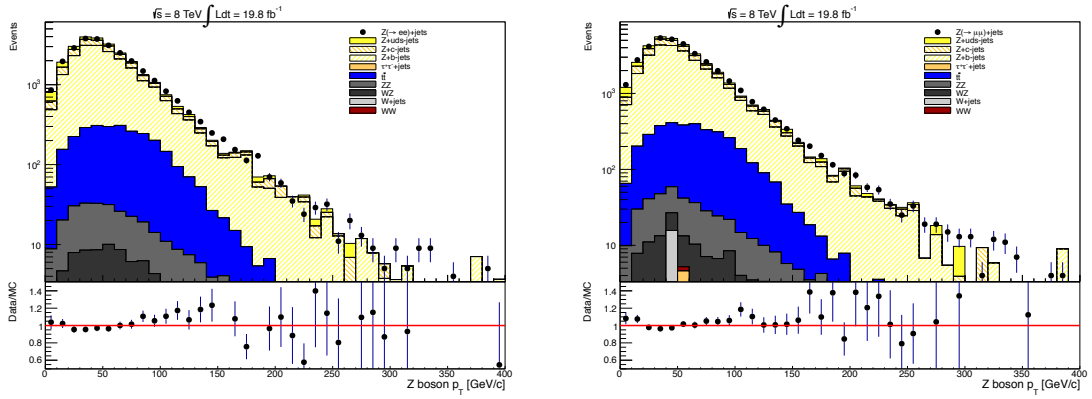

Figure 5.6.15: Transverse momentum distributions for the e^+e^- pairs (left) and the $\mu^+\mu^-$ pairs (right) with at least one $b-$tagged jet associated.

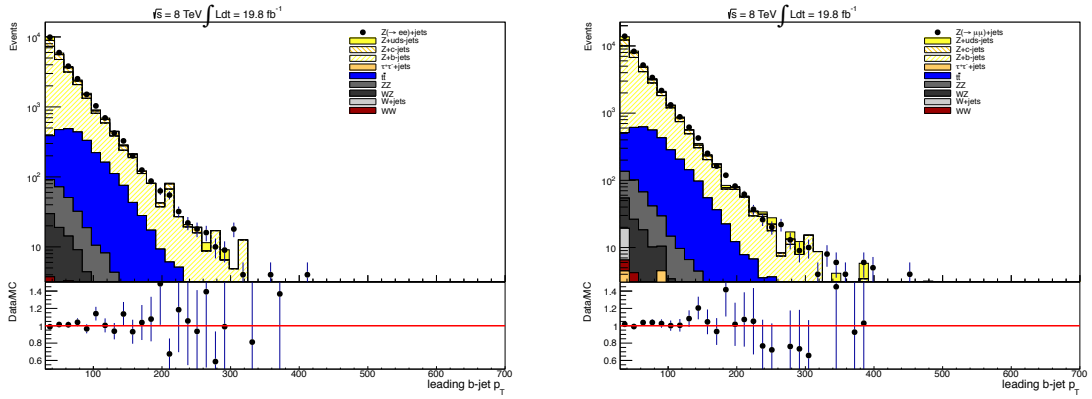

Figure 5.6.16: Transverse momentum distributions for the leading $b-$jet in $Z + b$ events for $Z \to e^+e^-$ (left) and $Z \to \mu^+\mu^-$ (right).

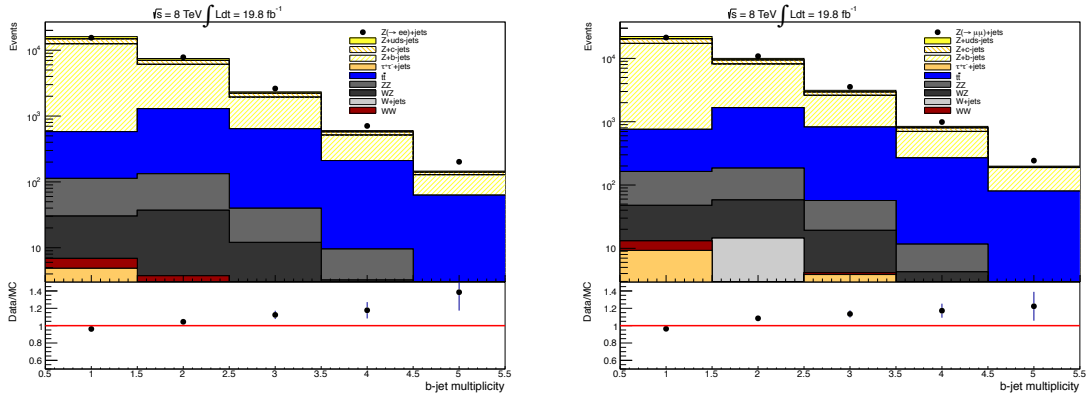

Figure 5.6.17: $b-$jet multiplicity in $Z+b$ events for $Z \to e^+e^-$ (left) and $Z \to \mu^+\mu^-$ (right).

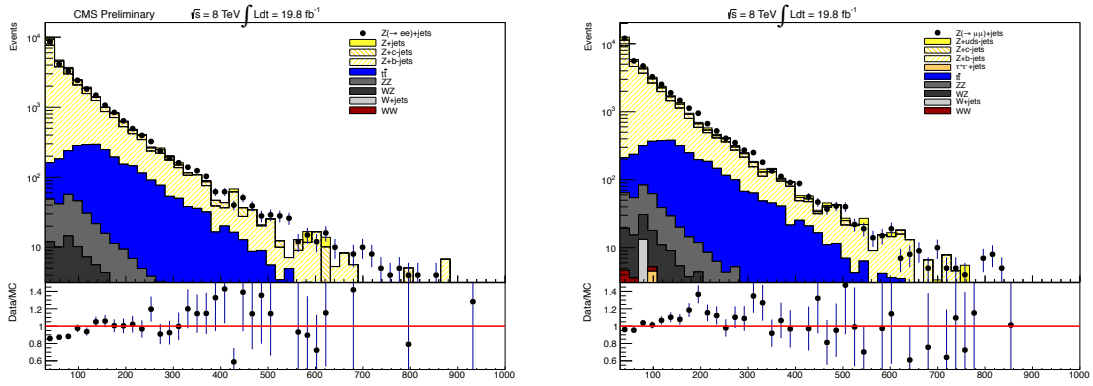

Figure 5.6.18: H_T distributions in $Z+b$ events for $Z \to e^+e^-$ (left) and $Z \to \mu^+\mu^-$ (right).

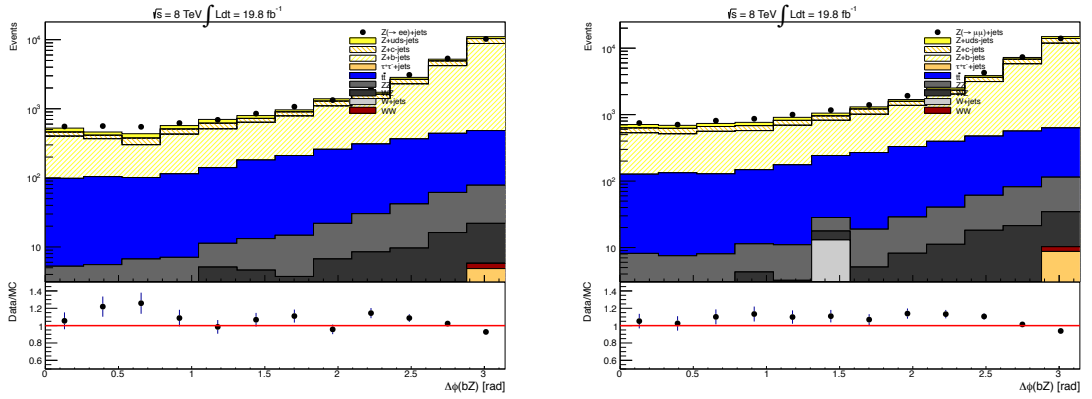

Figure 5.6.19: The azimuthal angle of the Zb system in $Z+b$ events for $Z \to e^+e^-$ (left) and $Z \to \mu^+\mu^-$ (right).

Chapter 6

Differential $Z + b$ Cross Section Measurement

In order to compare the Z+jets and $Z+b$ observables presented in Chapter 5 (detector level distributions) to the theoretical predictions of the corresponding differential cross sections in perturbative QCD, the measured quantities must be corrected for the detector effects that belong to experimental measurements. The unfolding procedure is a statistical method that aims to deconvolve measured distributions from these effects giving back the particle level corrected cross section. In this Chapter I will present the general features of the problem, and I will dedicate particular attention to the two methods employed in this analysis to obtain the differential $Z + b$ cross sections: the Singular Value Decomposition and the Bayesian method. I will focus on the strategy adopted and the physics choices taken to build up the unfolding matrix and how to choose the regularization parameter of the algorithm. Finally, I will show the unfolded distributions comparing the unfolded data with the Monte Carlo truth information. The comparison of the final results with several theoretical predictions for the $Z + b$ final state, exploiting also the different b quark descriptions inside the proton, will be presented after the discussion and the evaluation on the different sources of systematic uncertainties. In the end, the measured differential cross section ratios between the inclusive and the $b-$tagged samples as a function of the kinematic observables of the event will be shown and compared to the theoretical predictions.

6.1 Data Unfolding

Given a measured distribution $f(x)$ for the observable x, the true event-by-event measurement for the observable under consideration is only *ideally* the histogram of x. The detectors experimentally used to perform a given measurement induce sizable effects in the determination of $f(x)$, for instance:

- smearing due to the finite resolution and accuracy of the variable x,
- non linear response of detector components,
- acceptance, limited and depending on the different variables to be measured,
- radiative effects at parton level (initial and final state radiation).

The most common example is the measurement of a distribution $g(y) = s(y) + b(y)$ made by the sum of the signal s and the background b of the kinematic variable y. In order to obtain the *true* (theoretical) distribution $f(y)$, considering all the effects pointed out before, the transformation between the measured distribution and the theoretical one is[1]

$$g(y) = \int K(y,x) f(x) \mathrm{d}x,$$

where the kernel function $K(y,x)$ represents the generic sum of all the possible detector effects, describing the *response* of the detector for every y value at a given x. The inversion of this integral equation and the determination of the true distribution $f(x)$ is called *unfolding.* In common experimental situations, the numerical interpretation of the problems leads to a discrete version of the inversion problem. While using histograms, the response kernel function becomes a *response matrix* and the functions are represented by vectors:

$$\mathbf{y}{=}\mathbf{A}(\mathbf{y},\mathbf{x}) \cdot \mathbf{x},$$

where $\mathbf{y}$ is the measured data histogram and $\mathbf{x}$ to be determined. The transition $\mathbf{y}{\rightarrow}\mathbf{x}$ is described by the a_{ij} elements of the $\mathbf{A}$ matrix that represents the probability to observe an entry in the bin i of the histogram $\mathbf{y}$, if the true value x comes from the bin j of the histogram $\mathbf{x}$. The simplest way to invert the unfolding matrix is the so called *bin-by-bin* correction, that consists in the correction of the measured data by the efficiency evaluated bin by bin as a function of the distribution under study. This naive approach is not reliable when the difference between the true and the measured distributions becomes large. More sophisticated approaches are the *Singular Value Decomposition* and the *Bayesian* unfolding.

6.1.1 Unfolding Techniques

Two different approaches can be followed to calculate the response matrix using different mathematical methods. The *Singular Value Decomposition* [103] (SVD) is based on the regularization of the response matrix through the analysis of the minima of its singular values as a function of the regularization parameter k_{SVD}. This method is found to be the most robust unfolding procedure from a mathematical point of view, and it will be taken as the standard algorithm for the analysis. The Bayesian approach of the unfolding is an iterative algorithm based on the application of the Bayes theorem [104] to calculate the response matrix. The algorithm is regularized by the parameter that accounts for the number of iteration, k_{Bayes}. The implementation of the two methods in the analysis relies on the `RooUnfold` package [105].

6.1.2 Building the Response Matrix

Unfolding data through one of the algorithms presented requires a response matrix that describes the bins migration of the distribution of a chosen observable between the detector level reconstruction and the relative particle level (Monte Carlo truth) information. This matrix is built up using the Monte Carlo distribution of the observable under consideration after correcting for all the proper scale factors, mapped

[1] Fredholm integral equation of the first kind [102].

to the corresponding generator level distribution. The reconstructed observables are always simulated using MadGraph as the default. This matrix is then applied to the data distribution by means of one of the chosen unfolding algorithm, using the `RooUnfold` package. The result of this procedure gives as output the data distribution *corrected* for the migrations induced by any of the detector level effects described in Sec. 6.9. In order to fill the response matrix, the very same selection criteria at generator level must be applied according to the selection after reconstruction:

- generated $p_T^{jet} > 30$ GeV
- generated $|\eta^{jet}| < 2.5$
- at least one jet in the event
- opposite sign dielectron and dimuon invariant mass in the range [71;111] GeV.

The b quarks are selected by looking into the event history for the presence of a B hadron. In addition, to account the QED final state radiation emission in the Z boson decay, the generated $e^{\pm}, \mu^{\pm}$ objects are composed of a *bare* lepton (without any radiative emission) and all the photons inside a cone of $\Delta R = 0.1$ [106]. The final object is then obtained by summing on the quadrimomentum p_μ of its constituents creating a *dressed* lepton. The association between the reconstructed jet and the generated one is done by selecting only reconstructed jets that have a geometrical distance in the (η, ϕ) plan of $\Delta R < 0.5$ from the complete list of the generated jets. The response matrices calculated for the most relevant observables in the Z + jets and $Z + b$ final state are reported in Fig. 6.1.1-6.1.5.

6.1.3 Unfolding Validation

In order to validate the unfolding procedure several checks have been made. The first tests (*identity check*s) consist in the application of the unfolding matrix to the same reconstructed Monte Carlo used in the matrix building. By doing this, it is expected to have, by definition, exactly the same Monte Carlo truth distribution of the matrix itself. The ratio between the unfolding output and the generator level Monte Carlo gives exactly one for each distribution, demonstrating that no biases are introduced by the unfolding procedure in the analysis. Another way to validate the unfolding (*closure test*) is to run the algorithm on the distribution obtained from a signal Monte Carlo that is *not* the same one used to derive the response matrix: for this analysis the test is performed employing MadGraph for the matrix calculation and sherpa or powheg for the matrix application. These tests produce a very good agreement between unfolded and truth distributions, showing in general no dependence on the particular Monte Carlo used for the unfolding. Examples of the closure test in the leading $b-$jet momentum distribution are presented in Fig. 6.1.6.

6.1.4 The Choice of the Regularization Parameter

For both the SVD and the Bayesian unfolding, a choice for the regularization parameter must be done. The best value of the k_{SVD} parameter is selected by finding the largest value up to which the errors associated to the unfolding remain small

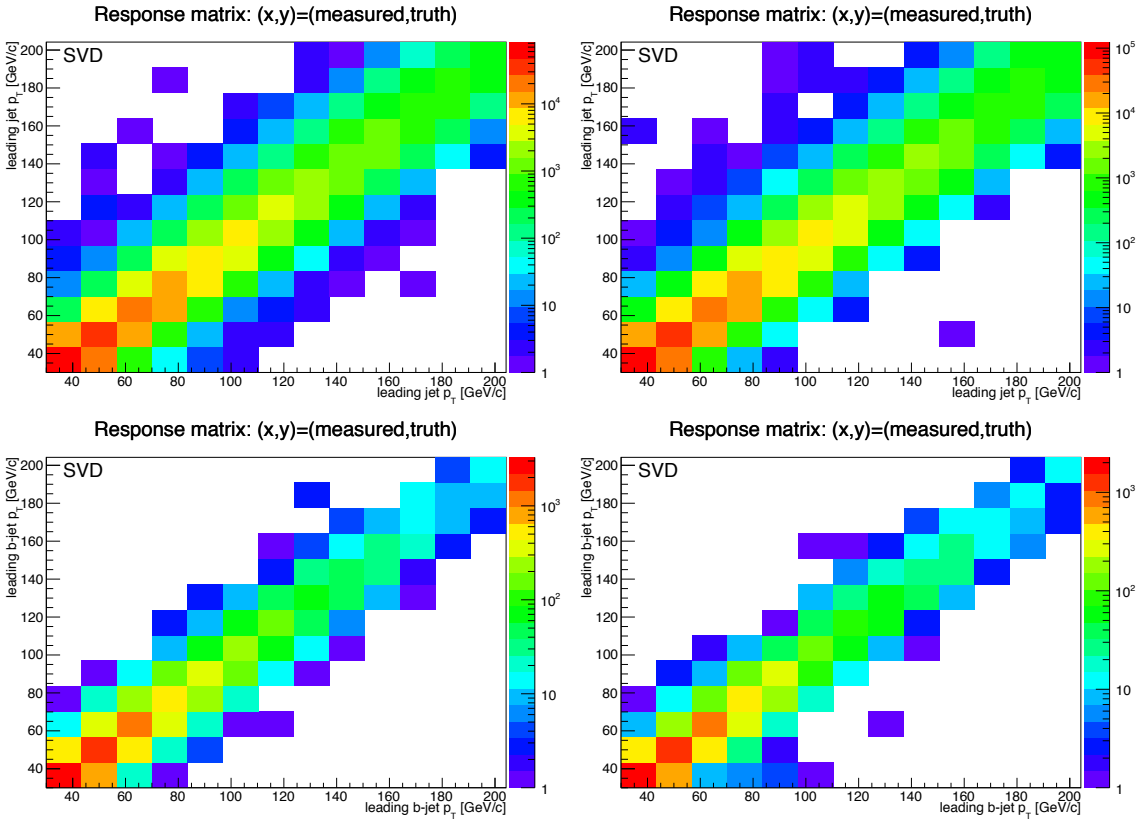

Figure 6.1.1: Response matrix for the leading jet momentum in electrons (top left) and muons (top right) final state and for the leading $b-$jet momentum in electrons (bottom left) and muons (bottom right) final state.

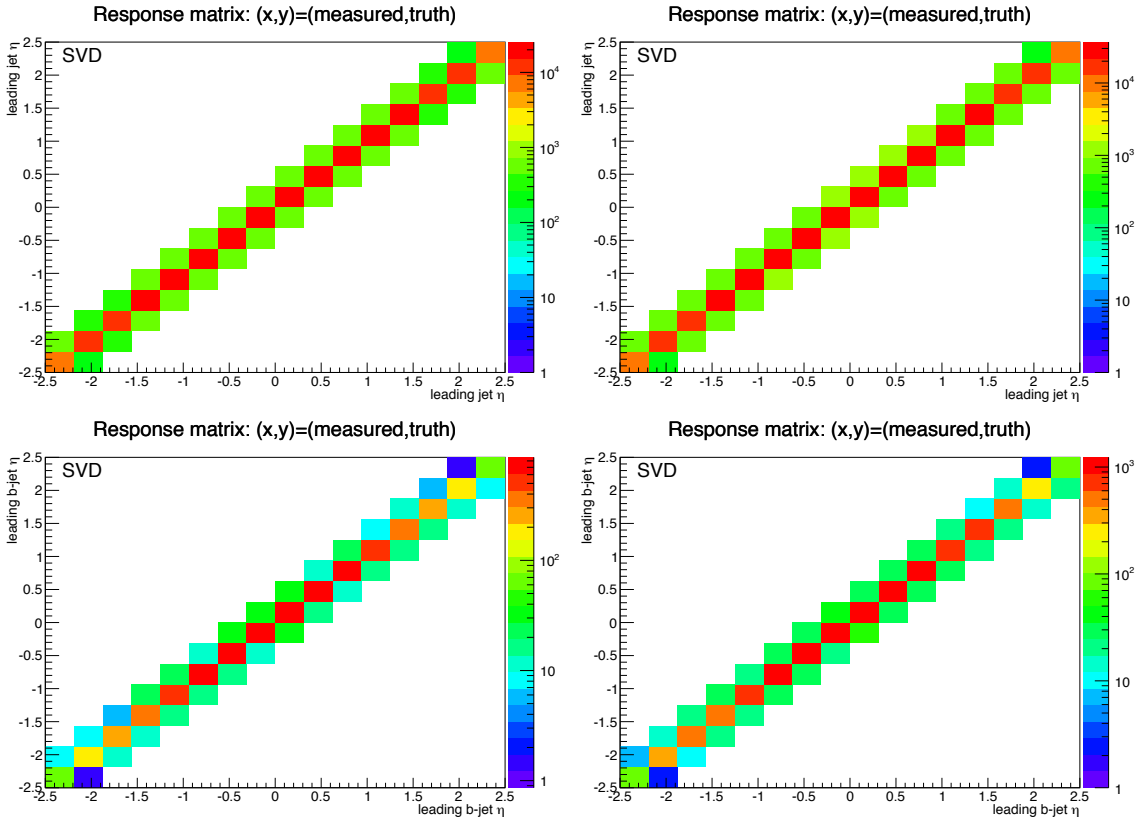

Figure 6.1.2: Response matrix for the leading jet pseudorapidity in electrons (top left) and muons (top right) final state and for the leading $b-$jet pseudorapidity in electrons (bottom left) and muons (bottom right) final state.

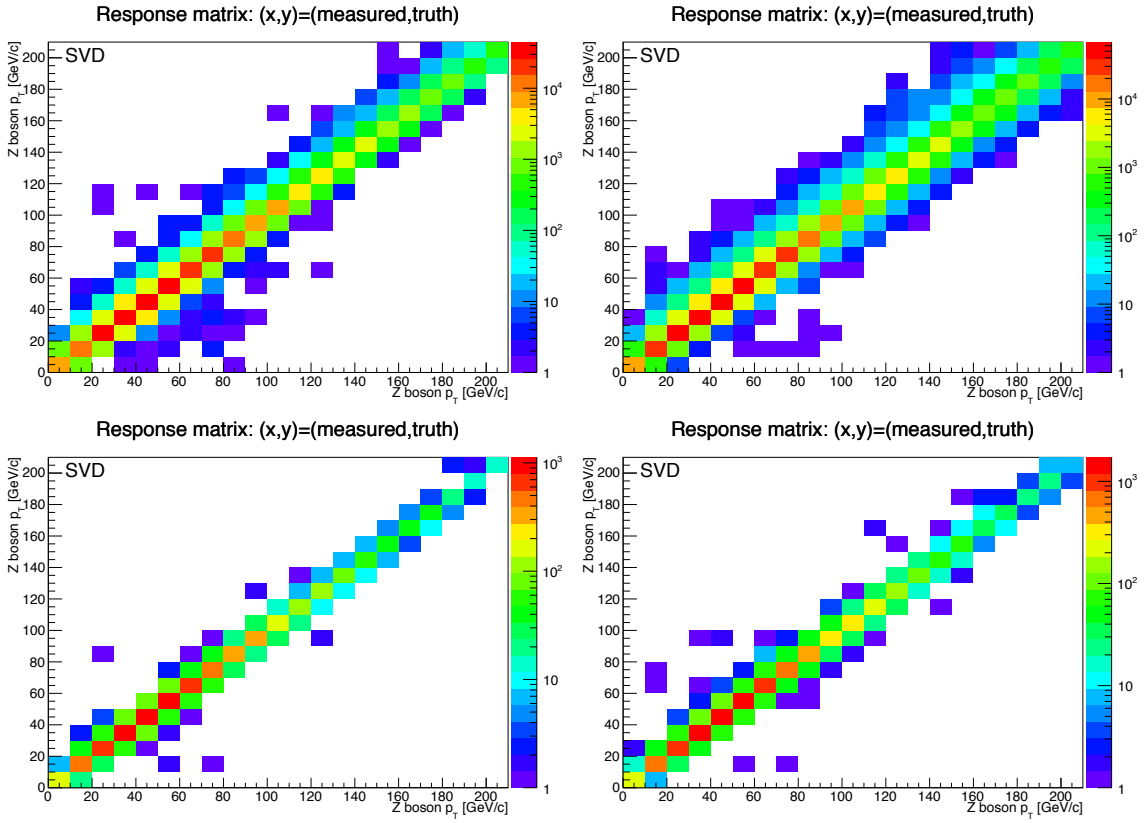

Figure 6.1.3: Response matrix for the leading Z momentum in electrons (top left) and muons (top right) final state and for the Z momentum in electrons (bottom left) and muons (bottom right) plus $b-$jets final state.

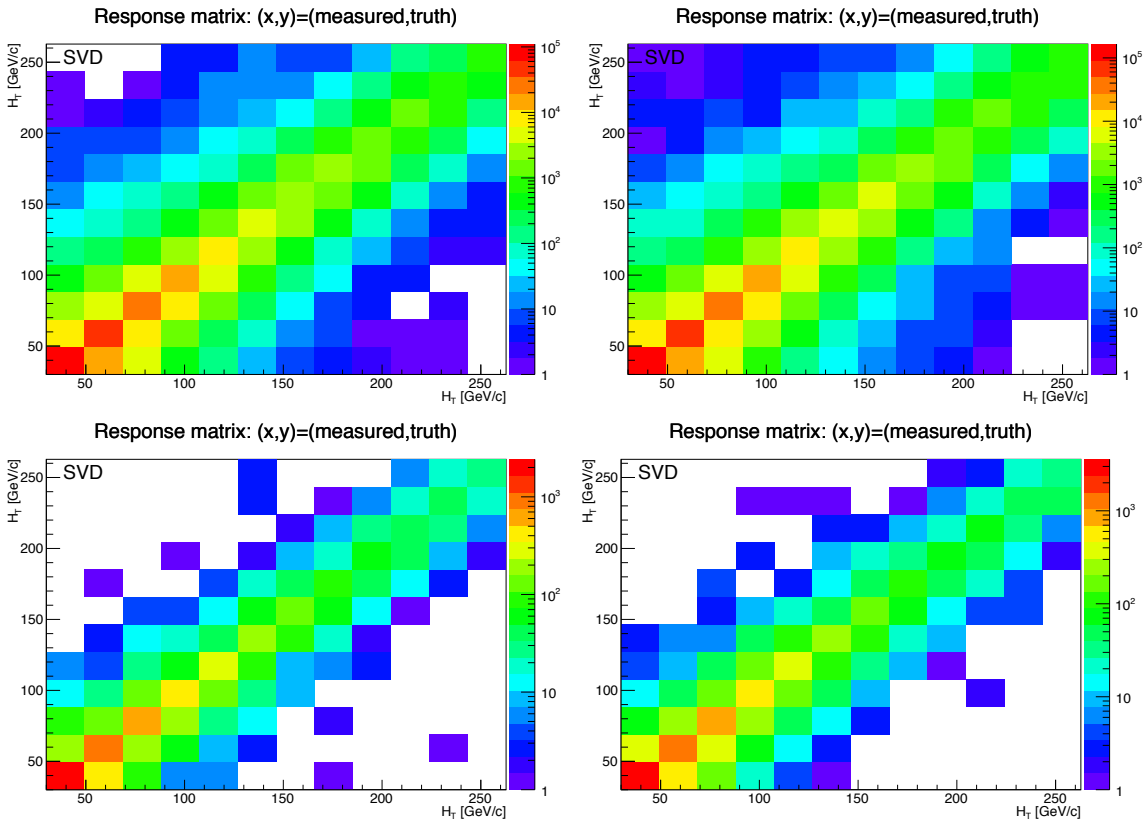

Figure 6.1.4: Response matrix for H_T in electrons (top left) and muons (top right) final state and for H_T + $b-$jets in electrons (bottom left) and muons (bottom right) final state.

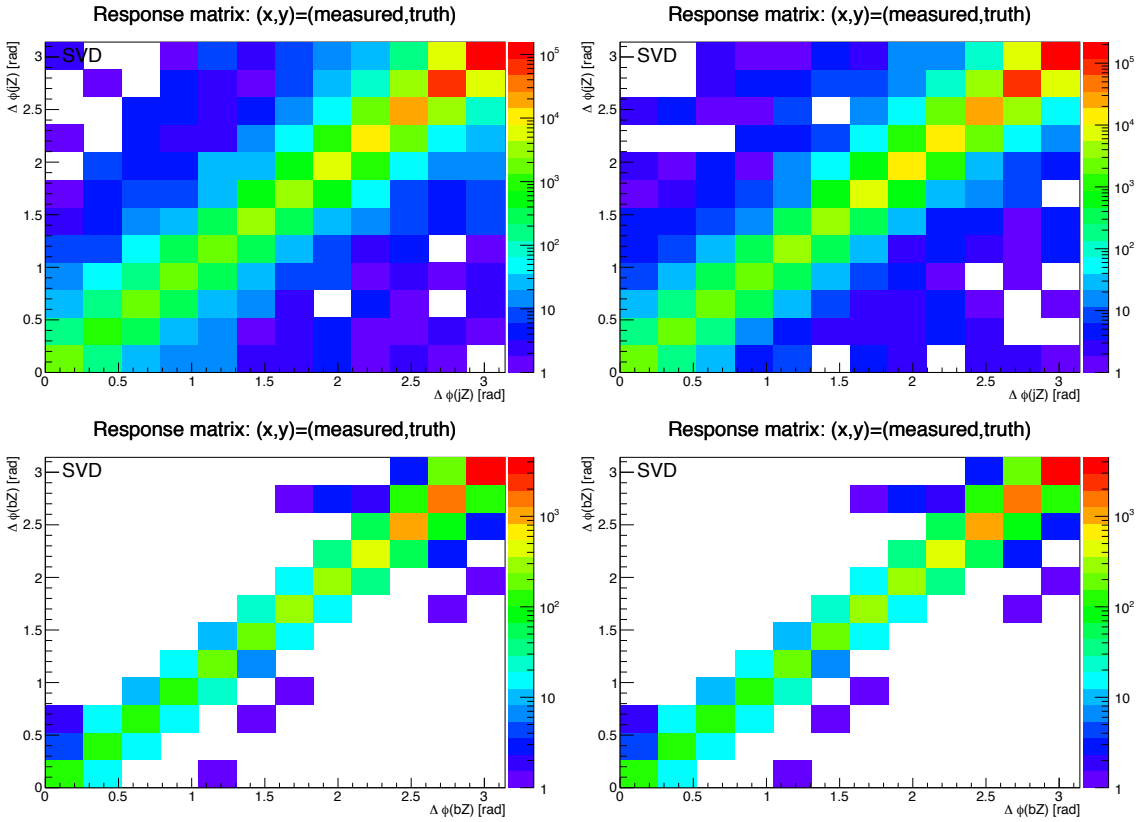

Figure 6.1.5: Response matrix for $\Delta\phi(Zj)$ in electrons (top left) and muons (top right) final state and for $\Delta\phi(Zb)$ in electrons (bottom left) and muons (bottom right) final state.

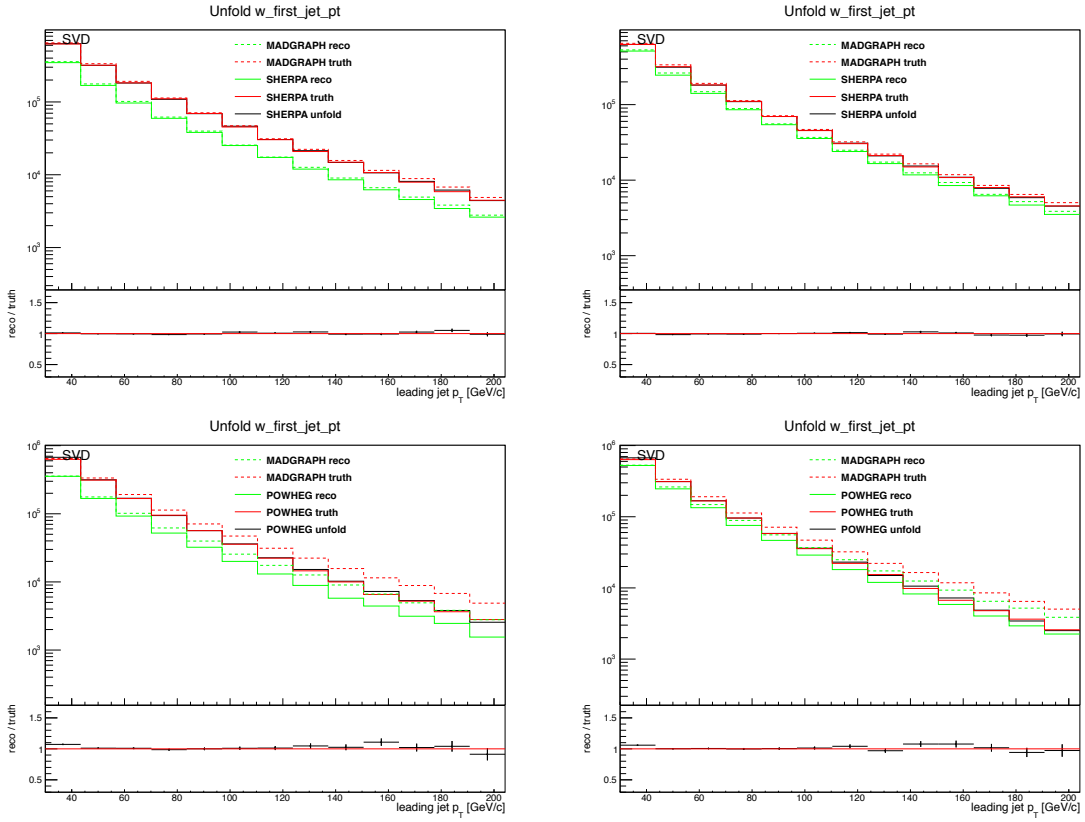

Figure 6.1.6: Leading jet momentum unfolding closure test: the unfolding is applied to SHERPA (top left for electrons final state and top right for muons final state) and POWHEG (bottom left for electrons final state and bottom right for muons final state). The ratio between unfolding results and truth Monte Carlo shows very good agreement, validating the procedure.

SVD Regularization	$k_{SVD}^{Z(\to e^+e^-)+\textbf{jets}}$	$k_{SVD}^{Z(\to \mu+\mu^-)+\text{jets}}$	$k_{SVD}^{Z(\to e^+e^-)+b}$	$k_{SVD}^{Z(\to \mu^+\mu^-)+b}$
$p_T^{b(\text{jet})}$	5	5	6	6
$\eta^{b(\text{jet})}$	3	3	2	2
p_T^Z	9	9	10	10
H_T	10	10	6	6
$\Delta\phi^{Zb(\text{jet})}$	3	3	8	8

Table 6.1: Choice of the regularization parameters in the SVD unfolding for each kinematic observable studied in the analysis.

compared to the statistical ones. From the SVD unfolding theory, it is known that the range of validity of the regularization parameter is

$$k_{SVD} \in [2, N_{bins}] \, .$$

As it can be seen in dedicated studies [105], small values would bias the unfolding result towards the Monte Carlo truth distribution, while too large values would give a result that is dominated by unphysical statistical fluctuations. It has been found that the choice of the regularization parameter depends on the observable under consideration. For this reason, in order to select the final values of the regularization parameter in the SVD unfolding, the relative closure tests have been studied for each distribution. The best k_{SVD} chosen is the one that gives the most reliable closure test in each distribution, i.e. the ratio between the unfolded results (made using a different Monte Carlo between the one used for the response matrix calculation) and the generator level distributions is the closest to the unity. The final regularization parameters chosen for the observables under study are presented in Table 6.1 for the inclusive Z + jets and the $Z + b$ final state in the dielectron and dimuon case. For the Bayes unfolding, the default value $k_{Bayes} = 4$ suggested by RooUnfold has been used. Since the SVD approach appeared more stable, this algorithm has been chosen for the presentation of the results.

6.2 Unfolded Distributions

The results of the unfolding procedure, performed with the SVD method, are presented for the following observables, defining the particle level Z+jets and Z+$b-$jets final state (Fig. 6.2.1-6.2.5) :

- the leading jet and $b-$jet transverse momentum
- the leading jet and $b-$jets pseudorapidity
- the Z boson momentum in association with at least one jet or $b-$jet
- the H_T, evaluated in Z + jets and $Z + b-$jets
- the $\Delta\phi(Zj)$ and $\Delta\phi(Zb)$ between the Z boson direction and the leading jet and $b-$jet direction.

The effect of the application of the response matrix built using the reconstructed Monte Carlo simulation of Z+jets and Z+$b-$jets events using MadGraph+pythia,

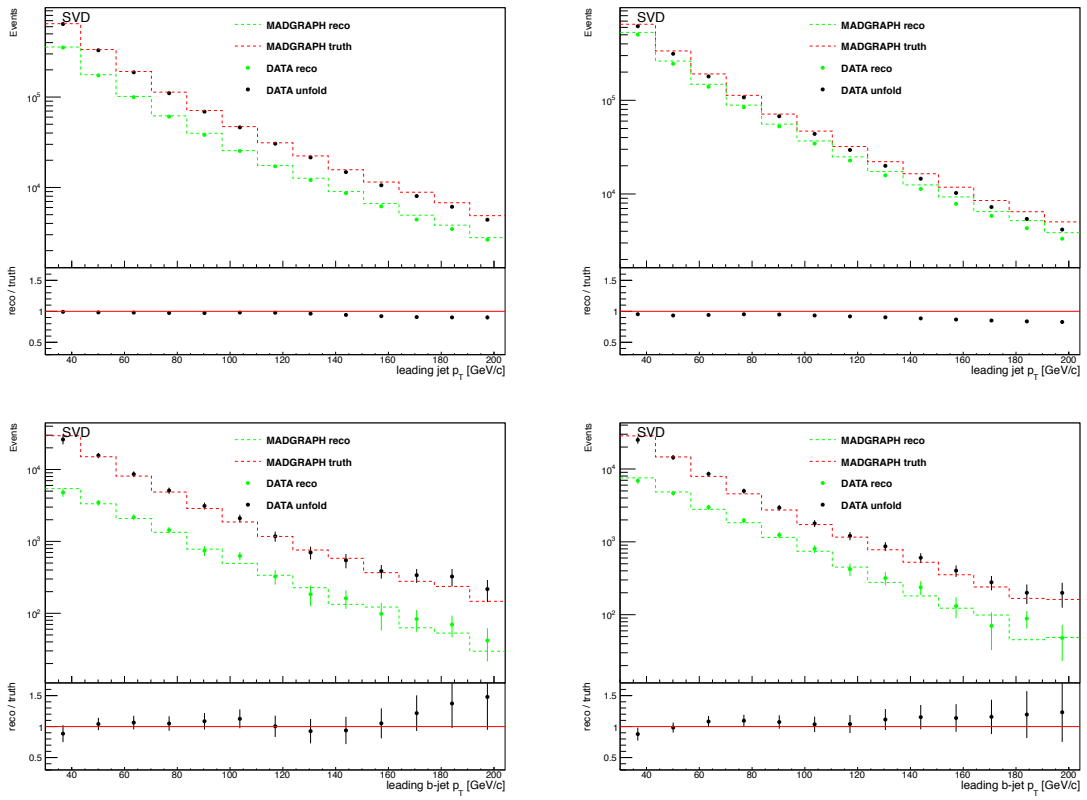

Figure 6.2.1: Unfolding results for the leading jet momentum in electrons (top left) and muons (top right) final state and for the leading $b-$jet momentum in electrons (bottom left) and muons (bottom right) final state.

and the matrix element NLO calculation of MadGraph for the generator part, is presented by superimposing the three distributions (data, reconstructed Monte Carlo and truth Monte Carlo) for each of the above mentioned distribution. Results are always shown for the $Z \to ee$ and $Z \to \mu\mu$ final state separately.

6.3 Systematic Uncertainties

Many sources of systematic uncertainty arise in the flow of the $Z+$jets and $Z+b$ analysis. The different sources can be summarized in the following list:

- Jet Energy Correction and Resolution
- pile-up reweighting procedure
- unfolding procedure
- VV background subtraction
- top background subtraction
- b purity estimation
- scale factors calculations

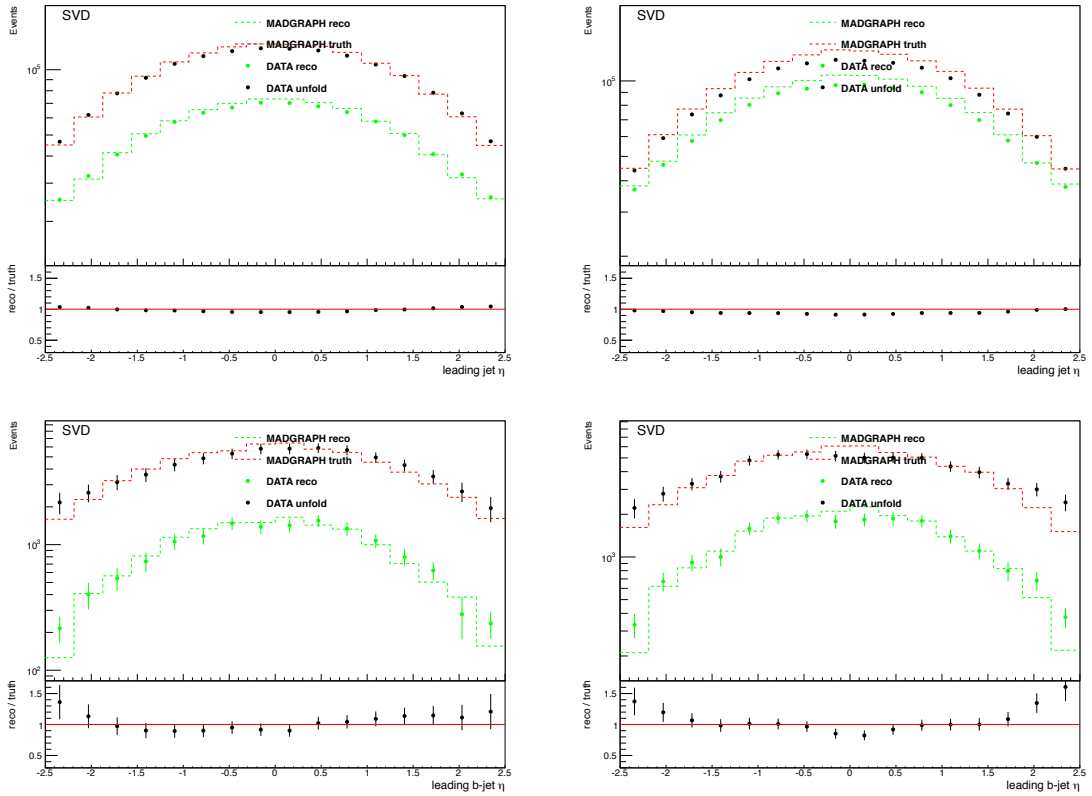

Figure 6.2.2: Unfolding results for the leading jet η in electrons (top left) and muons (top right) final state and for the leading $b-$jet η in electrons (bottom left) and muons (bottom right) final state.

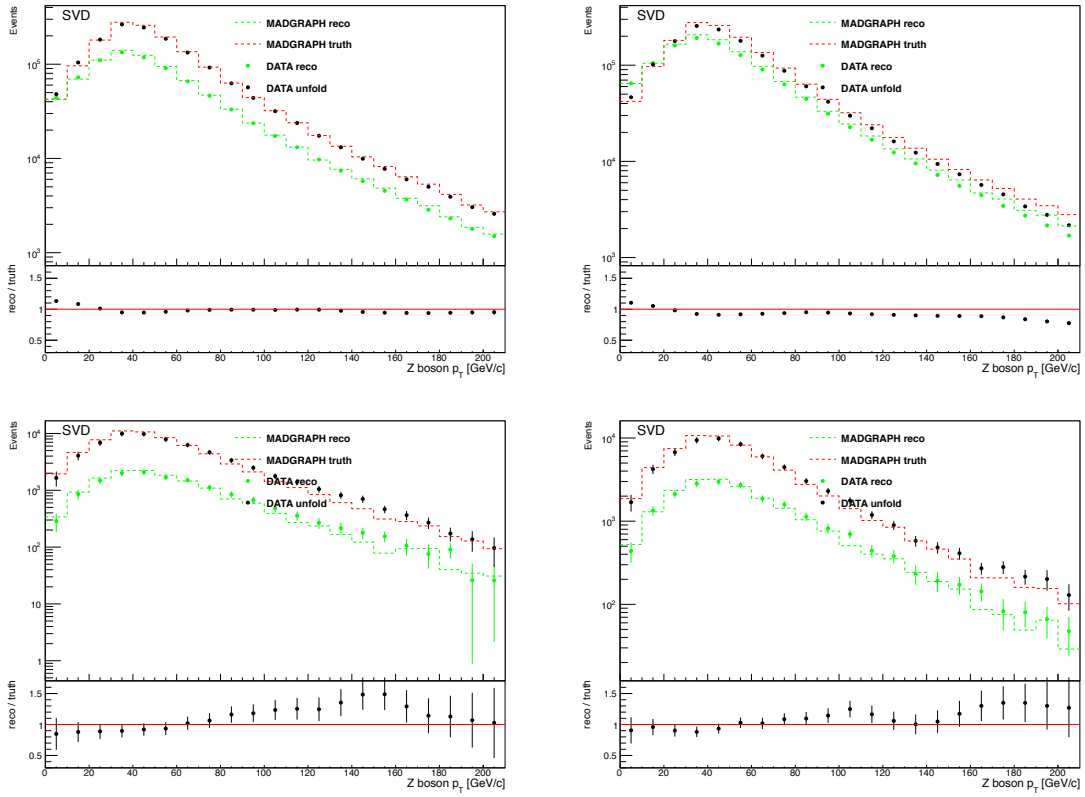

Figure 6.2.3: Unfolding results for the leading Z momentum in electrons (top left) and muons (top right) final state and for the Z momentum in electrons (bottom left) and muons (bottom right) plus $b-$jets final state.

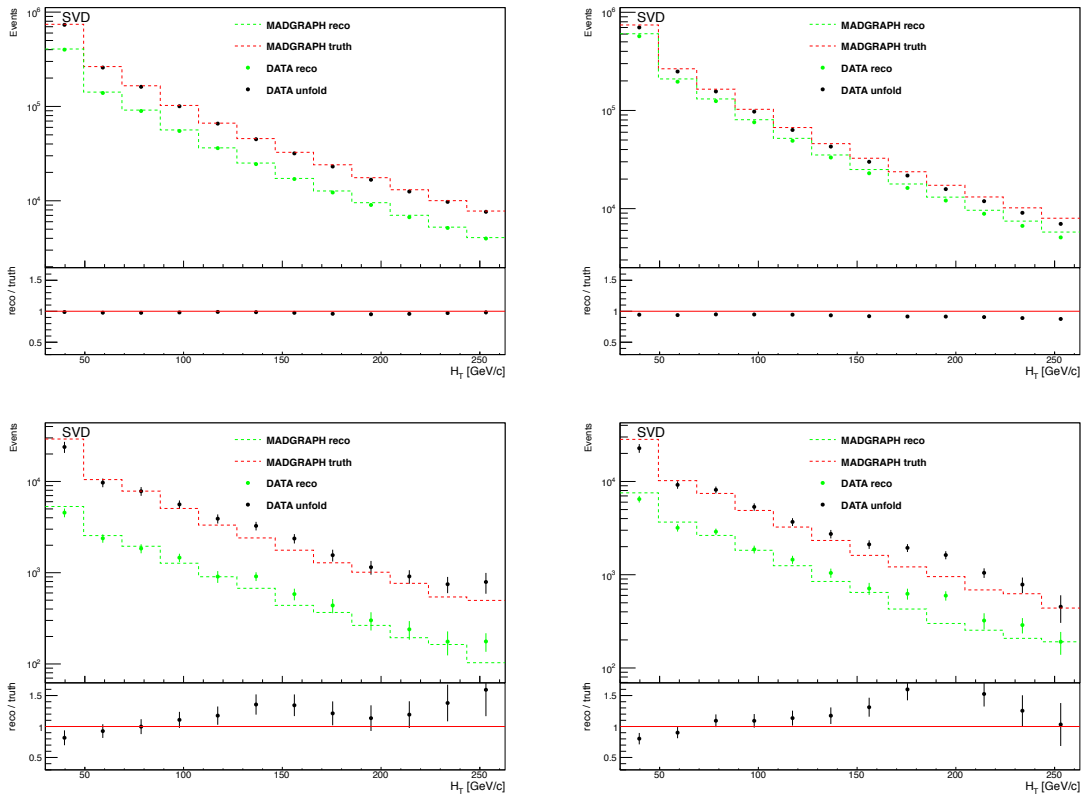

Figure 6.2.4: Unfolding results for H_T in electrons final state (top left) and muons (top right) and for H_T + $b-$jets in electrons final state (bottom left) and muons (bottom right).

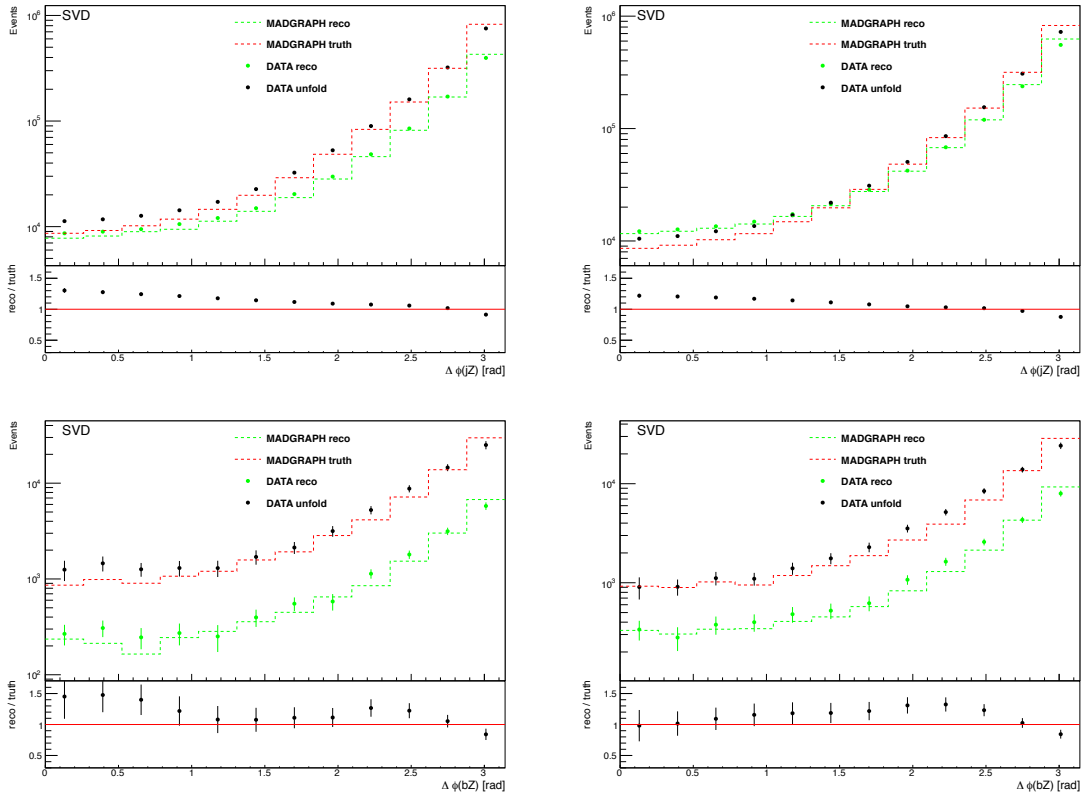

Figure 6.2.5: Unfolding results for $\Delta\phi(Zj)$ in electrons (top left) and muons (top right) final state and for $\Delta\phi(Zb)$ in electrons (bottom left) and muons (bottom right) final state.

- luminosity

Each contribution will be described in this Section, and the total sum of the single uncertainties will be added in quadrature together with the statistical error in the final cross section measurement, to be presented in Sec. 6.4.

Jet Energy Correction and Resolution

Jet energy correction uncertainties affect the reconstruction of the jet energy, and are related to the different types of corrections described in Sec. 4.2, such as the uniformity of the detector response in pseudorapidity, the momentum uniformity in the energy reconstruction, the flavour dependence in the jet composition, the pile up subtraction. Dedicated measurements and results for the single Jet Energy Correction sources of systematic uncertainties are documented in Ref. [107]. The resulting uncertainty depends on the transverse momentum and pseudorapidity of the jet under consideration. The systematic effect due to the application of each type of JEC is estimated by comparing the distributions obtained using the standard JEC and its variation by a quantity that depends on the p_T, η of the jet under consideration. The evaluation strategy for this source of uncertainty is to shift all reconstructed jet transverse momenta by ± 1 standard deviation of the measured jet energy scale uncertainty. The same procedure is applied to evaluate the systematic uncertainty of the Jet Energy Resolution. The JEC source affects the transverse momentum of the leading b-jet with a maximum contribution of 4.6%, while the JER ranges from 0.5% to 3.2%.

Pile-Up Reweighting

The pile-up reweighting procedure has a related systematic uncertainty coming essentially from the error on the measurement of the total inelastic cross section in the Monte Carlo and from the luminosity measurement [108]. The combination of these uncertainties propagates itself on the expected number of interactions, estimated to be around 4%. This corresponds to an overall scale error on the pile-up distribution, and so it is needed to measure the effect of these errors through the results. The uncertainty in the $Z + b$ analysis due to these effects is estimated by shifting the overall mean of the distribution of the number of interactions up or down by 5%, then redoing the reweighting and evaluating the corresponding variation. The average systematic error in the observables of the analysis is smaller than 3%.

Unfolding

The systematic error due to the unfolding procedure is evaluated by building up a new response matrix whose elements are weighted for the ratio between the pre-unfolded data and the reconstructed Monte Carlo. The difference between the unfolded results obtained by the application of this weighted response matrix and the standard one is considered as a systematic error. This uncertainty ranges from 5% to 25%. Another source of systematic uncertainty in the unfolding is the effect of the statistic errors on the elements of the response matrix. This contribution is estimated by calculating the response matrix errors. The statistical error associated

to the response matrix calculations ranges from 2.2% to 17%. Finally, the uncertainty on the variation of the total cross section between unfolded distribution is calculated. This contribution is found to be negligible in this analysis ($< 0.1\%$).

VV Background

The systematic uncertainty due to the diboson background ($VV = ZZ, WW, WZ$) cross section has been evaluated by varying the theoretical cross section of these processes by the $\pm 15\%$ of their central value. This contribution is found to be about 0.2%.

Top Background Extraction

The contribution to systematic uncertainty of the top background evaluation is given by the errors on the normalization, extracted from the fit to the $e^+e^-, \mu^+\mu^-$ invariant mass distribution with the $e\mu$ method described in Subsec. 5.4.2. The systematic uncertainty related to the $t\bar{t}$ fit results has a maximum value of about 11%.

b Fraction Estimation

Systematic error on the purity scaling factors estimation is evaluated from the errors given by the fit to the b fraction using the Jet Probability Discriminator distribution (Sec. 5.5). The systematic uncertainty related to the b purity extraction procedure in bins of the leading $b-$jet momentum has an average value of 3%.

Efficiency Calculation

Systematic uncertainties in the efficiency calculations come from the $Z \to \mu\mu$ and $Z \to ee$ Tag & Probe measurements of data/MC efficiency scale factors for lepton reconstruction, identification, isolation and trigger in each channel. Uncertainty in the $b-$tagging scale factors correction comes from the pile-up contamination (estimated by varying of 5% the uncertainty on estimated pile-up in data), cut on the lepton transverse momentum, (moving the selection cut from 5 to 9 GeV), gluon splitting rate in Monte Carlo ($g \to b\bar{b}$, varied by $\pm 50\%$) and energy fraction carried by the B mesons in the b hadronization (varied by $\pm 5\%$) [109]. The global effect of the systematic uncertainty related to the scale factor application is 2% in the muons final state and 1.5% in the dielectron final state, while the b-tagging amounts to 3%.

Luminosity

The systematic error on the LHC luminosity delivered to CMS in the 2012 proton-proton physics run, evaluated using the cluster counting technique in the silicon pixel detector (as described in details in Ref. [110]), amounts to 2.6%.

A breakdown of all the relative statistics (evaluated separatly in data and Monte Carlo) and systematics error sources above mentioned, together with the relative total error expressed in percentage, is given as an illustrative example for the leading b-jet differential cross section measurement in Tables 6.2 - 6.3, for the combined electrons and muons final state. All the uncertainties on the other observables measured in the analysis are listed in Appendix A.

p_T^b **bin**	**data**	**MC**	$t\bar{t}$ **fit**	b **fit**	**unfolding**	**total stat.**
1	1.7	4.5	0.0	6.8	0.2	7.2
2	1.5	3.7	0.2	3.5	0.5	4.1
3	1.6	3.9	0.3	2.8	1.0	3.7
4	1.8	4.3	0.4	2.6	1.2	3.8
5	2.2	5.1	0.5	2.4	1.0	4.1
6	2.6	6.0	0.6	2.3	0.6	4.5
7	3.1	7.2	0.7	2.8	0.6	5.4
8	3.7	8.4	0.7	3.1	2.0	6.2
9	4.1	9.4	0.7	2.9	4.1	6.7
10	4.4	10.0	0.6	2.2	6.1	7.3
11	5.0	11.2	0.5	1.4	7.9	8.6
12	6.4	14.2	0.4	0.9	9.3	10.9
13	7.5	16.7	0.4	0.5	10.2	12.6

Table 6.2: Statistical errors contributions (in percent) for the leading b jet differential cross section measurement.

p_T^b bin	efficiency	JEC	JER	pile-up	b-tagging	luminosity	unfolding	total syst.
1	1.5	1.4	0.6	0.3	3.0	2.6	0.2	15.4
2	1.5	1.3	0.5	0.3	3.0	2.6	0.5	6.3
3	1.5	2.0	0.8	0.5	3.0	2.6	1.0	6.9
4	1.5	4.0	0.9	0.8	3.0	2.6	1.2	9.2
5	1.5	2.6	1.1	0.6	3.0	2.6	1.0	10.2
6	1.5	1.9	1.2	0.4	3.0	2.6	0.6	12.0
7	1.5	2.4	0-8	1.0	3.0	2.6	0.6	10.7
8	1.5	3.4	1.5	1.4	3.0	2.6	2.0	13.6
9	1.5	3.7	2.2	1.3	3.0	2.6	4.1	15.4
10	1.5	3.9	1.5	0.9	3.0	2.6	6.1	17.4
11	1.5	4.1	2.1	1.7	3.0	2.6	7.9	24.1
12	1.5	4.6	5.6	2.7	3.0	2.6	9.3	34.8
13	1.5	4.6	7.7	2.9	3.0	2.6	10.2	37.9

Table 6.3: Systematic errors contributions (in percent) for the leading b jet differential cross section measurement.

6.4 Differential Cross Sections Measurement

The unfolded results presented in Sec. 6.6 represent the differential cross sections (expressed in picobarn/GeV) for the production of Z+jets and Z+b events evaluated as a function of the observables x = leading jet p_T, leading jet η, Z boson p_T, $\Delta\phi_{Zj}$, H_T :

$$\sigma(Z(\rightarrow \ell\ell) + \text{jets}) = \frac{N^{Z+jets} - (N_{bkg}^{VV} + N_{bkg}^{tt})}{\epsilon_\ell \cdot \int \mathcal{L} \cdot \mathrm{d}t}$$

$$\sigma(Z(\rightarrow \ell\ell) + (\geq 1)b\text{-jet}) = \frac{N^{Z+(\geq 1b-jet)} - (N_{bkg}^{VV} + N_{bkg}^{tt} + N_{charm} + N_{light})}{\epsilon_\ell \cdot \epsilon_b \cdot \int \mathcal{L} \cdot \mathrm{d}t}.$$

The expected number of diboson (VV) and top pairs ($t\bar{t}$) background events, N_{bkg}, and the estimated number of events containing charm and light (u, d, s) jets in the case of the b-tagged observables, have been subtracted from the Z + jets and $Z + b -$ jets samples. The number of signal events is then divided by the integrated luminosity and the lepton and b-tagging efficiencies calculated as described in Sec. 4.2. The unfolded differential cross section measurements are compared to several perturbative QCD theoretical calculations of the associated production of a Z boson and b quarks. Theoretical predictions at leading order are computed with the MadGraph and powheg plus pythia 6 generator setups. The theoretical cross sections have been rescaled to the NNLO calculations when available. The parton density functions adopted for MadGraph+pythia 6 is CTEQ6L1, while the CT10 is used inside powheg+pythia 6. The systematic uncertainties coming from the choice of the PDF set inside the generators it is not included in the final results, and it is part of the future improvement of the anlysis. All these predictions have the 5-flavour scheme (5FS) approach in the description of the b quark production, assuming $m_b = 0$ and including the gluon splitting contribution $g \rightarrow b\bar{b}$ inside the b quark PDF. A comparison between the 5FS interpretation of $Z+b$ and the 4FS one, with $m_b \neq 0$ and b quark PDF set to zero inside the MadGraph matrix element calculation has been made, comparing all the differential cross section results in the two different approaches.

Combination of the electrons and muons channels

The final values of the unfolded differential cross sections as a function of the kinematic observables describing the Z + jets and $Z + b$ final states, together with the relative ratio between MadGraph 4FS, MadGraph 5FS and powheg theoretical predictions are presented in Figs. 6.6.1-6.6.5 combining the $Z \rightarrow e^+e^-$ and the $Z \rightarrow \mu^+\mu^-$ final state, giving the $Z \rightarrow l^+l^-$ results. A very good agreement between the results obtained in the electrons final state and muons final state is found, as it is shown in Fig. 6.4.1. The combination is made by mean of the appropriate weighted sum of the cross sections measured in the single channel. The complete set of the single channel results is presented in Appendix B. Systematics uncertainties in the combined cross section are calculated by considering the eventual correlation between the electrons final state and the muons final state for each contribution. The uncertainties related to the Jet Energy Correction and Resolution, as well as the ones coming from the pile-up reweighting, the unfolding systematics, the b- tagging, the background subtraction and the luminosity are fully correlated. Uncertainties com-

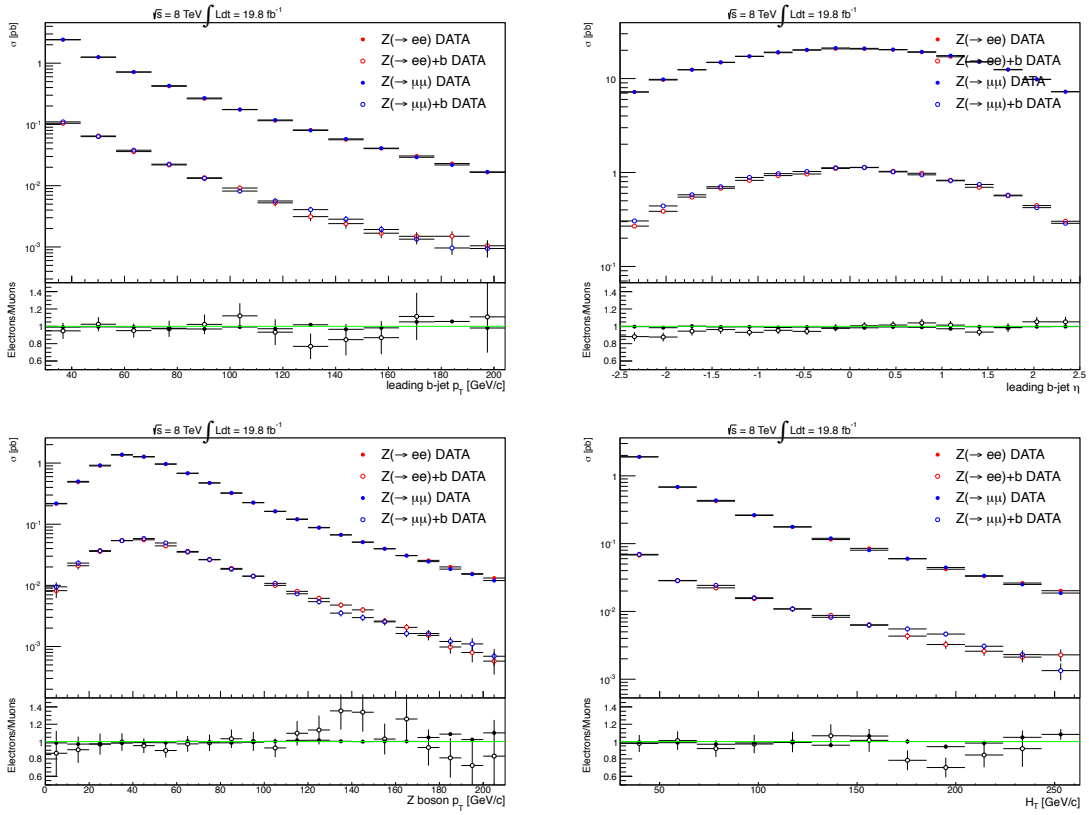

Figure 6.4.1: Comparison between electrons and muons results in the $Z+b$ selection for the variables used in the final cross section results.

ing from the efficiency evaluation, b-jets fraction and $t\bar{t}$ fits, as well as the statistic uncertainty of data and Monte Carlo samples are treated as uncorrelated.

6.5 Differential Cross Section Ratios

The differential cross section ratio

$$R = \frac{\sigma(Z + b\text{-jets})}{\sigma(Z + \text{jets})}$$

between events containing a Z boson decaying into dielectron and dimuon final state in association with at least one $b-$jet or one jet, as a function of the kinematic observables of the event x ($\frac{\mathrm{d}R}{\mathrm{d}x}$) are presented in Fig. 6.6.6-6.6.10. Results on the differential cross section ratios are compared with the theoretical predictions of MadGraph 4FS and MadGraph 5FS plus pythia6 and powheg+pythia6. This is the first measurement of the differential cross section ratios of $Z+b$ and $Z+$jets in proton-proton collisions up to now. Previous measurements for this ratio have been presented by the D0 collaboration in $p\bar{p}$ collision at the Tevatron [111].

6.6 Results Interpretation and Perspectives

The 8 TeV measured cross sections for the associated production of a Z boson and at least one jet (upper part of Fig. 6.6.1-6.6.5), after the unfolding, are in very good agreement with the LO calculation made by MadGraph 5FS event generator interfaced with pythia6 for the parton shower in each kinematic observable presented. The Monte Carlo theoretical predictions are normalized to the integrated luminosity in data using a cross section value that includes the next-to-next-to-leading order (NNLO) corrections. The only observed difference (even if very tiny) between data and MadGraph 5FS can be found in the angular difference $\Delta\phi_{Zj}$ between the Z boson direction and the most energetic jet in the event direction, where a difference of $\sim 10\%$ in the data/theory ratio is measured, and in the low momentum region of the Z boson. This effect would probably disappear with a more precise calculation, i.e. a next-to-leading-order (NLO) version of MadGraph. The comparison between data and powheg+pythia6 shows a huge disagreement almost everywhere. This is probably due to the fact that the NLO matrix element calculation of powheg is limited to the emission of one jet. The case where more than one jet is present is evaluated by taking the additional jets from the parton shower (pythia6), giving a poor description of the process. Difference of the order of tens of percent can be seen especially in the observables that require explicitly the presence of additional jets, such as H_T (Fig. 6.6.4), where the energy of all the jets in the event is measured, and in the $\Delta\phi_{Zj}$ distributions. Reasonable agreement can be seen in the $b-$jet momentum and pseudorapidity and in the Z boson momentum. In order to have a NLO precision for the additional jets a possible solution is to use the *Multi-Scale Improved NLO* (MINLO) [112] approach implemented together with powheg, giving the NLO precision up to the third jet. This is one of the future studies provided for the development of the analysis. In the $Z + (\geq 1)b-$jet final state, a very good agreement between the LO MadGraph 4FS and 5FS calculation, using respectively the b quark mass explicitly in the calculation of the matrix element and inside the b PDF, is observed in all the kinematic variables of the event. In particular, the 4FS calculation appears to describe more precisely the low momentum region of the leading $b-$jet and Z boson, as well as the H_T. In order to better understand the relevance of the quark mass effects in the $Z + b$ final state, new observables with more sensitivity to the structure of the b production will be investigated. An example of possible study in this direction is the angular correlation between jets when more than one jet is tagged as a b in the final state. In this framework an essential part of the analysis would be focused on the study of the separation between subsequent $b-$jets produced in association with the Z boson. The comparison of the measured distribution with new generators will be also essential to understand the topology of $Z + b$ in more details. The relative differential cross section ratios (Fig. 6.6.6-6.6.10) between the inclusive and $b-$tagged samples, give an indication of the precision in the current knowledge between the two different final state topologies. The best description of the ratios achieved in this work is found when using MadGraph 4FS as theoretical prediction, showing a good agreement in all the distributions, with a Data/Theory difference of the order of (0.2 - 0.4) %. The most important part of the spectrum from the physics point of view, for Beyond the Standard Model searches and Higgs studies, in the ratios measurements is the high p_T of the $b-$jets and their central η regions. In this regions the agreement with the theoretical predictions has

to be futher investigated and it is an ongoing development of the anlysis. A general less precise description of the cross section ratios is found while using the massless approach (MadGraph 5FS) and powheg, for which the same issue discussed for the inclusive case naturally propagates to the relative ratios measured. Finally, the theoretical uncertainties coming from the renormalization scale chosen to describe the process and the usage of different PDF sets are an essential part of the final results that is currently under investigation in this analysis.

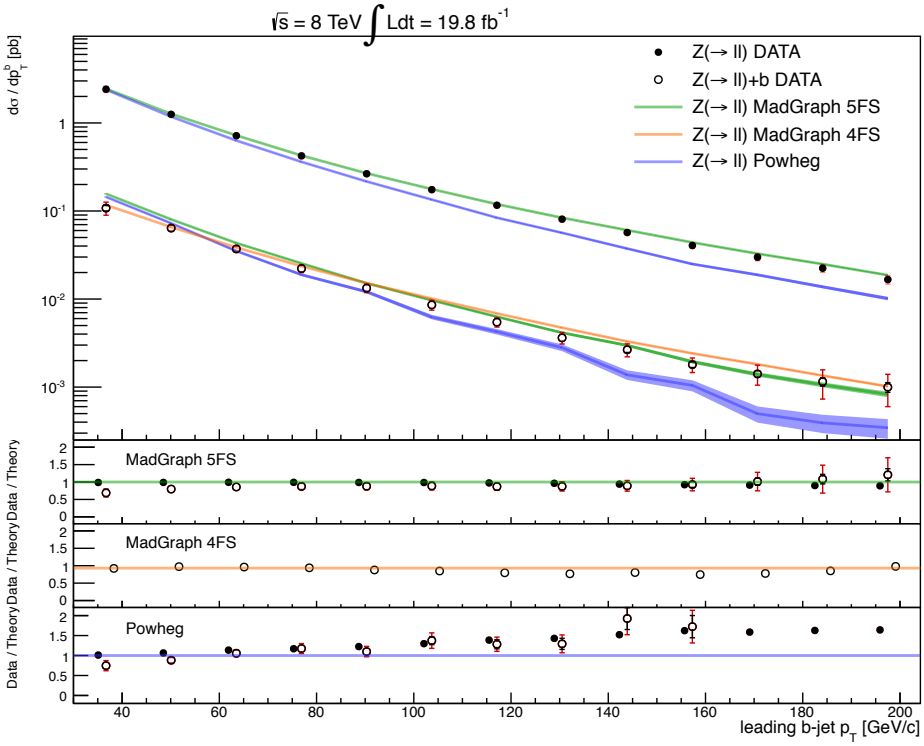

Figure 6.6.1: Unfolded differential $Z+$jets (upper part of the plot) and $Z+b-$jets (lower part of the plot) cross section as a function of the leading jet p_T, compared with the different theoretical predictions (coloured lines).

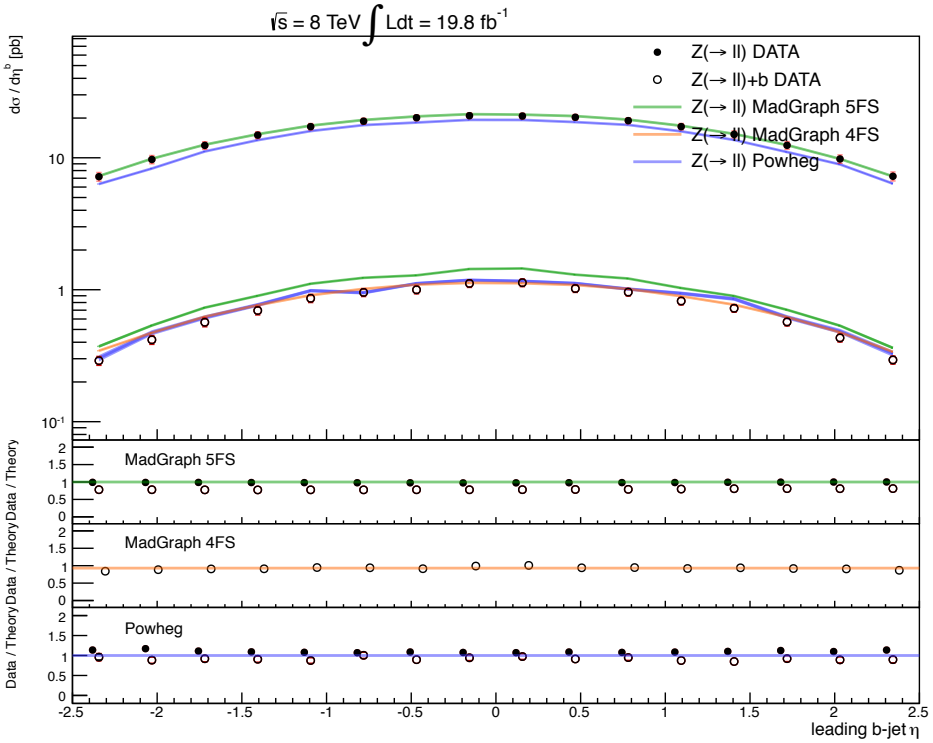

Figure 6.6.2: Unfolded differential $Z+$jets (upper part of the plot) and $Z+b-$jets (lower part of the plot) cross section as a function of the leading jet η, compared with the different theoretical predictions (coloured lines).

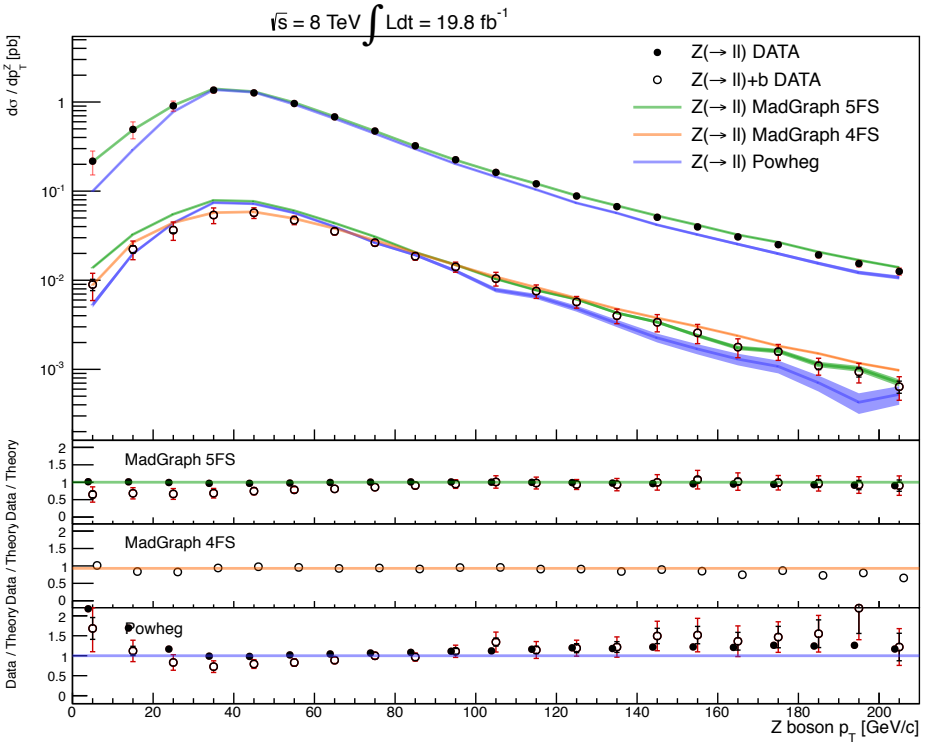

Figure 6.6.3: Unfolded differential $Z+$jets (upper part of the plot) and $Z+b-$jets (lower part of the plot) cross section as a function of the Z boson p_T, compared with the different theoretical predictions (coloured lines).

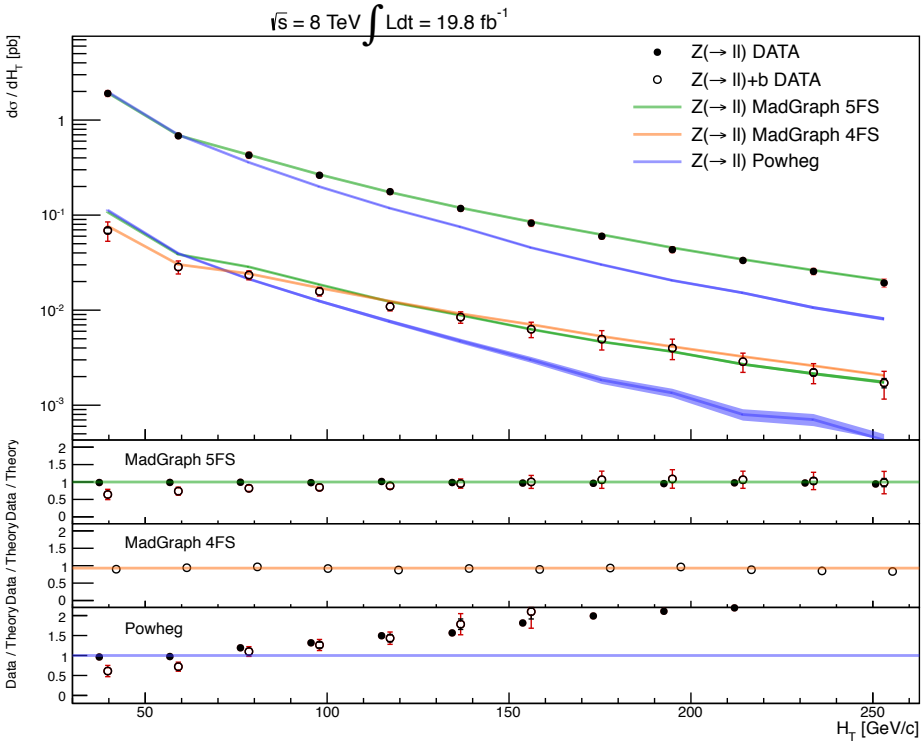

Figure 6.6.4: Unfolded differential $Z+$jets (upper part of the plot) and $Z+b-$jets (lower part of the plot) cross section as a function of H_T, compared with the different theoretical predictions (coloured lines).

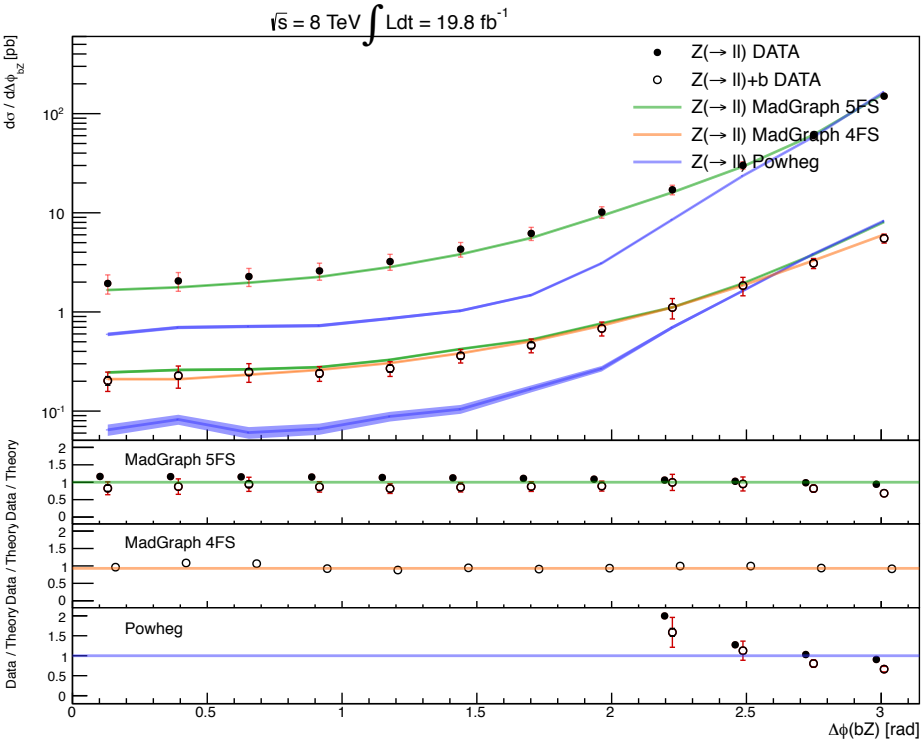

Figure 6.6.5: Unfolded differential $Z+$jets (upper part of the plot) and $Z+b-$jets (lower part of the plot) cross section as a function of $\Delta\phi(Zb)$, compared with the different theoretical predictions (coloured lines).

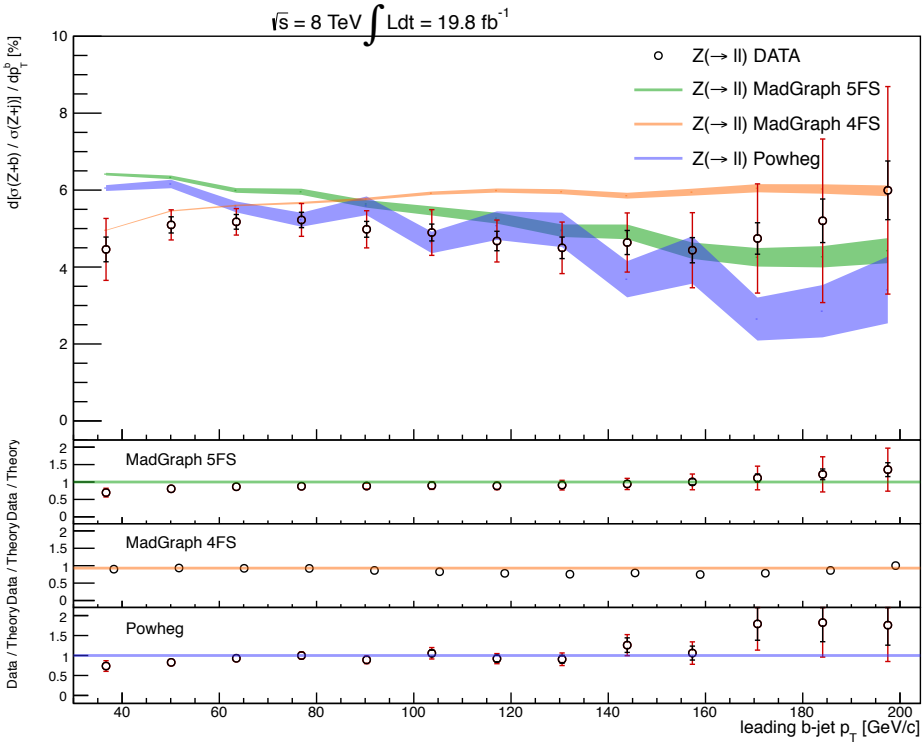

Figure 6.6.6: Differential cross section ratio between the $Z + b-$jets and the $Z +$ jets cross section as a function of the leading jet p_T, compared with the different theoretical predictions (coloured lines).

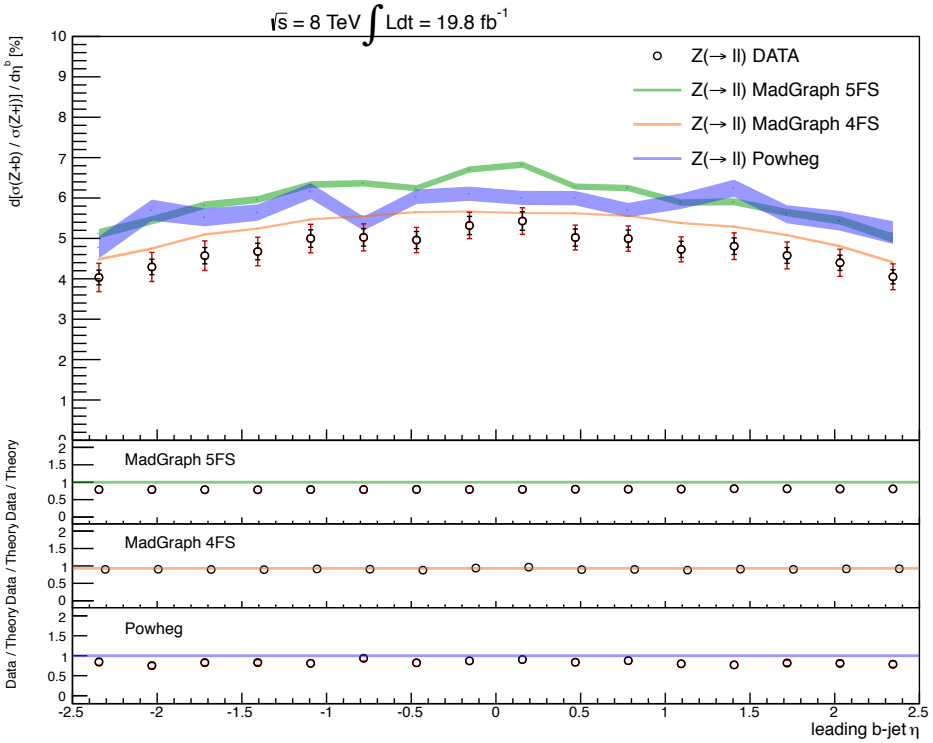

Figure 6.6.7: Differential cross section ratio between the $Z + b-$jets and the $Z +$jets cross section as a function of the leading jet η, compared with the different theoretical predictions (coloured lines).

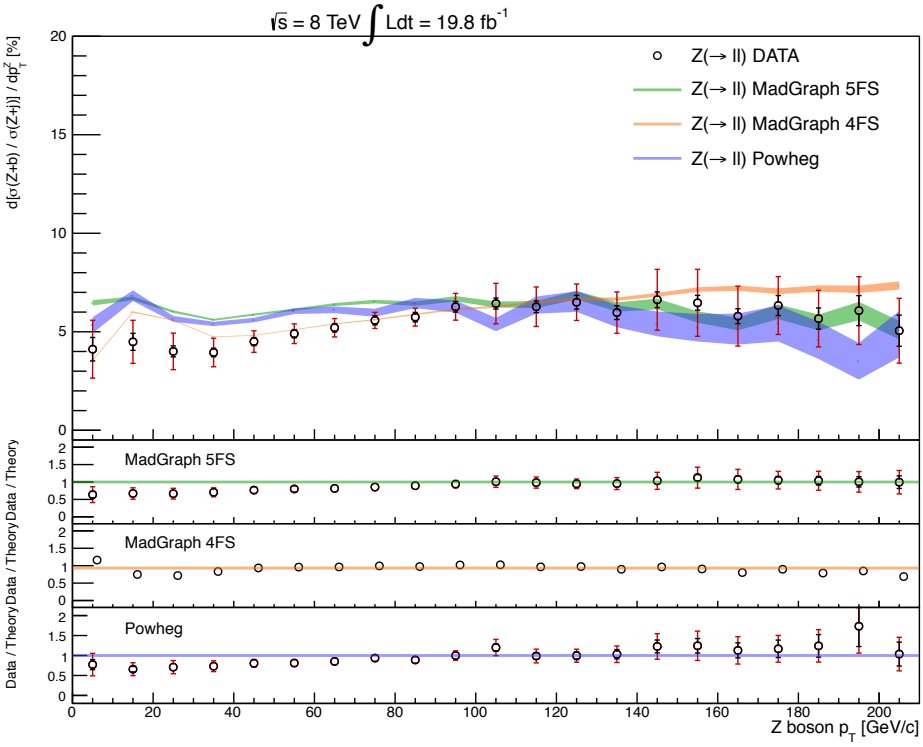

Figure 6.6.8: Differential cross section ratio between the $Z+b-$jets and the $Z+$jets cross section as a function of the Z boson p_T, compared with the different theoretical predictions (coloured lines).

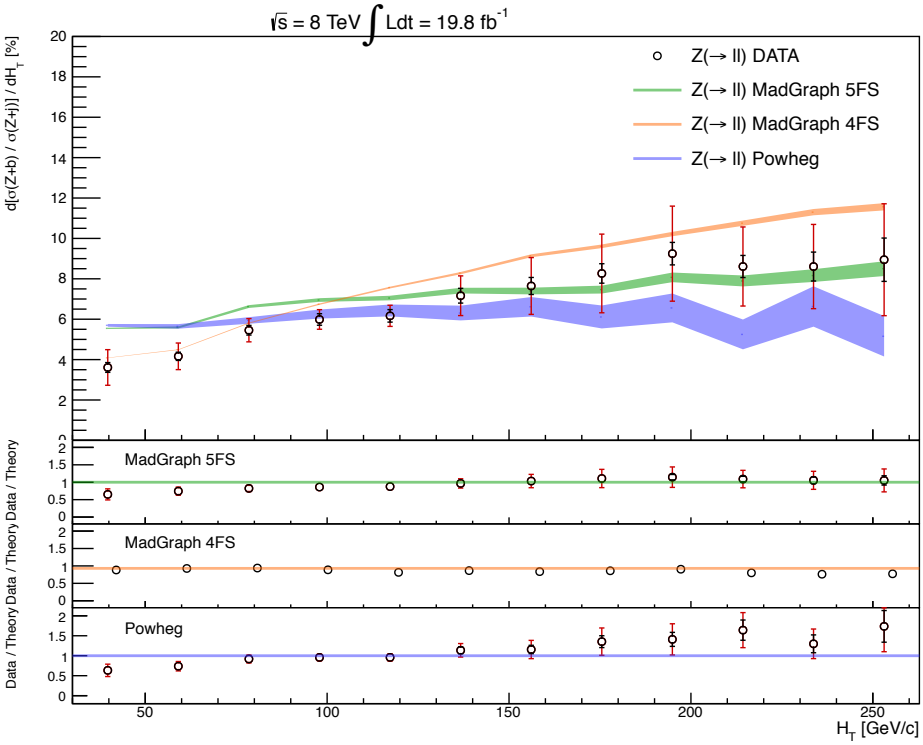

Figure 6.6.9: Differential cross section ratio between the $Z+b-$jets and the $Z+$jets cross section as a function of H_T, compared with the different theoretical predictions (coloured lines).

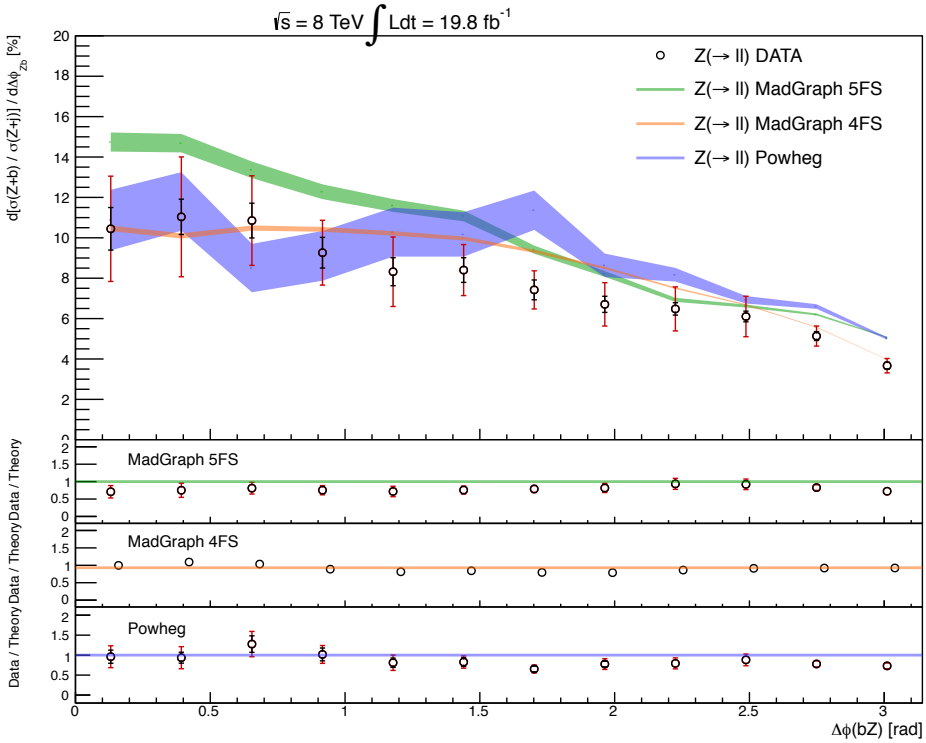

Figure 6.6.10: Differential cross section ratio between the $Z + b-$jets and the $Z +$ jets cross section as a function of $\Delta\phi(Zb)$, compared with the different theoretical predictions (coloured lines).

Conclusions

The Standard Model process of the associated production of jets and $b-$jets with a Z boson decaying into lepton pairs ($\ell = e, \mu$) has been measured in LHC pp collisions at $\sqrt{s} = 8$ TeV with the CMS experiment, using a total integrated luminosity of about 20 fb^{-1}. A special attention is dedicated to the top-antitop background, estimated using data-driven techniques. The b quarks content of events has been evaluated by fitting the different flavours contributions from the Monte Carlo simulation. The measured cross sections in data are compared to several theoretical predictions after the unfolding procedure. The cross sections have been measured as a function of many kinematic observables describing the full $Z + b$ event topology, such as the momentum and pseudorapidity of the most energetic $b-$jet, the Z boson momentum, the transverse energy sum of all the jets in the event and the azimuthal angular difference between the Z boson direction and the leading jet direction. In particular, the comparison between unfolded data and the matrix element plus parton shower leading order prediction in perturbative QCD using the b quark mass set to zero (5FS) or included inside the b PDF (4FS) with MadGraph event generator is tested. Good agreement between data and theoretical prediction has been observed for all the measured distributions. The comparison made with the $Z+$ 1 jet calculation at the next-to-leading order precision in powheg shows some discrepancies due to the kinematic limits inside the generator, as described in details in Sec 6.6. Finally, the differential cross section ratios between the inclusive $Z+$jets and $Z + b-$jets final state has been measured and compared with the theoretical expectations. The most important systematic uncertainties for this analysis are found to be the ones related to the unfolding procedure and to the jet energy corrections. This is the first 8 TeV measurement for the $Z + b$ differential cross sections up to date. Many important applications of this work can be found in the framework of the understanding of b quarks in perturbative QCD, the Higgs boson study, and New Physics searches.

Appendix A

Systematic Uncertainties

p_T^b **bin**	**data**	**MC**	$t\bar{t}$ **fit**	b **fit**	**unfolding**	**total stat.**
1	1.7	4.5	0.0	6.8	0.2	7.2
2	1.5	3.7	0.2	3.5	0.5	4.1
3	1.6	3.9	0.3	2.8	1.0	3.7
4	1.8	4.3	0.4	2.6	1.2	3.8
5	2.2	5.1	0.5	2.4	1.0	4.1
6	2.6	6.0	0.6	2.3	0.6	4.5
7	3.1	7.2	0.7	2.8	0.6	5.4
8	3.7	8.4	0.7	3.1	2.0	6.2
9	4.1	9.4	0.7	2.9	4.1	6.7
10	4.4	10.0	0.6	2.2	6.1	7.3
11	5.0	11.2	0.5	1.4	7.9	8.6
12	6.4	14.2	0.4	0.9	9.3	10.9
13	7.5	16.7	0.4	0.5	10.2	12.6

Table A.1: Statistical error contributions (in percent) for the leading b-jet differential cross section measurement in $Z(\rightarrow \ell\ell) + b$.

p_T^b **bin**	**JEC**	**JER**	**pile-up**	**unfolding**	**total syst.**
1	1.4	0.6	0.3	0.2	15.4
2	1.3	0.5	0.3	0.5	6.3
3	2.0	0.8	0.5	1.0	6.9
4	4.0	0.9	0.8	1.2	9.2
5	2.6	1.1	0.6	1.0	10.2
6	1.9	1.2	0.4	0.6	12.0
7	2.4	0-8	1.0	0.6	10.7
8	3.4	1.5	1.4	2.0	13.6
9	3.7	2.2	1.3	4.1	15.4
10	3.9	1.5	0.9	6.1	17.4
11	4.1	2.1	1.7	7.9	24.1
12	4.6	5.6	2.7	9.3	34.8
13	4.6	7.7	2.9	10.2	37.9

Table A.2: Systematic errors contributions (in percent) for the leading b jet differential cross section measurement in $Z(\rightarrow \ell\ell) + b$.

η^b **bin**	**data**	**MC**	$t\bar{t}$ **fit**	b **fit**	**unfolding**	**total stat.**
1	1.0	2.7	0.2	4.3	1.0	4.5
2	1.0	2.6	0.2	4.3	0.9	4.5
3	1.0	2.6	0.2	4.2	0.9	4.4
4	0.9	2.5	0.2	4.2	0.9	4.4
5	0.9	2.4	0.2	4.2	0.8	4.3
6	0.8	2.2	0.3	4.1	0.8	4.3
7	0.8	2.1	0.3	4.1	0.7	4.3
8	0.7	2.0	0.3	4.1	0.7	4.2
9	0.7	1.9	0.3	4.1	0.7	4.2
10	0.8	1.9	0.3	4.1	0.7	4.2
11	0.8	2.0	0.3	4.1	0.7	4.2
12	0.9	2.1	0.3	4.1	0.7	4.2
13	0.9	2.2	0.3	4.1	0.8	4.3
14	0.9	2.3	0.2	4.1	0.8	4.3
15	1.0	2.4	0.2	4.1	0.8	4.3
16	1.0	2.5	0.2	4.0	0.9	4.3

Table A.3: Statistical error contributions (in percent) for the leading $b-$jet pseudorapidity differential cross section measurement in $Z(\rightarrow \ell\ell) + b$.

η^b **bin**	**JEC**	**JER**	**pile-up**	**unfolding**	**total syst.**
1	1.9	0.2	0.2	1.0	6.4
2	1.9	0.2	0.2	0.9	6.3
3	1.8	0.2	0.2	0.9	6.2
4	1.8	0.2	0.2	0.9	6.0
5	1.7	0.2	0.2	0.8	5.7
6	1.7	0.2	0.2	0.8	5.5
7	1.7	0.2	0.2	0.7	5.3
8	1.7	0.2	0.2	0.7	5.3
9	1.7	0.2	0.2	0.7	5.2
10	1.7	0.2	0.2	0.7	5.1
11	1.7	0.3	0.2	0.7	5.0
12	1.8	0.3	0.2	0.7	5.1
13	1.8	0.4	0.2	0.8	5.1
14	1.9	0.4	0.2	0.8	5.3
15	2.0	0.4	0.2	0.8	5.5
16	2.0	0.4	0.2	0.9	5.6

Table A.4: Systematic errors contributions (in percent) for the leading b jet pseudorapidity differential cross section measurement in $Z(\rightarrow \ell\ell) + b$.

p_T^Z **bin**	**data**	**MC**	$t\bar{t}$ **fit**	b **fit**	**unfolding**	**total stat.**
1	5.2	13.4	0.2	12.9	3.8	14.5
2	2.9	7.6	0.2	8.9	2.2	9.6
3	2.2	5.8	0.2	6.4	1.6	6.9
4	1.8	4.3	0.2	5.0	1.6	5.2
5	1.6	4.2	0.2	4.0	1.6	4.3
6	1.6	4.3	0.2	3.6	1.6	4.4
7	1.7	4.5	0.3	3.3	1.7	4.5
8	1.8	4.5	0.3	3.0	2.2	4.7
9	1.9	4.7	0.4	2.8	2.3	4.6
10	2.1	4.8	0.4	2.5	2.3	4.9
11	2.2	5.2	0.4	2.4	2.4	4.3
12	2.5	5.4	0.4	2.6	3.5	4.7
13	2.8	6.3	0.3	2.7	3.7	5.9
14	3.1	7.2	0.3	2.7	4.5	5.6
15	3.3	7.5	0.2	1.9	4.5	6.5
16	3.3	7.6	0.2	2.0	4.6	5.4
17	3.7	8.1	0.2	2.8	4.2	6.3
18	4.1	10.2	0.1	2.7	6.3	7.2
19	5.0	12.3	0.1	3.5	7.4	9.1
20	7.0	17.4	0.1	3.8	9.1	12.8
21	9.0	22.3	0.1	4.0	12.2	15.6
22	10.0	25.5	0.1	4.1	13.2	17.5

Table A.5: Statistical error contributions (in percent) for Z boson momentum differential cross section measurement in $Z(\rightarrow \ell\ell) + b$.

p_T^Z **bin**	**JEC**	**JER**	**pile-up**	**unfolding**	**total syst.**
1	14.5	7.8	4.8	21.0	30.6
2	11.3	4.9	2.2	15.3	21.6
3	8.8	2.1	0.6	19.0	22.2
4	4.3	0.5	0.6	17.63	19.2
5	1.4	1.1	0.7	11.2	12.9
6	0.6	0.5	0.7	8.0	10.5
7	1.1	0.4	0.6	5.1	7.6
8	1.1	0.3	0.4	2.2	6.3
9	0.9	0.4	0.2	5.3	8.7
10	0.9	0.6	0.3	2.4	12.2
11	1.2	0.3	0.3	5.5	17.8
12	1.3	0.4	0.3	10.6	16.2
13	1.4	1.5	0.4	15.8	13.8
14	1.4	1.5	1.6	14.3	17.6
15	1.4	0.6	2.7	11.8	21.8
16	1.5	0.6	2.0	15.3	23.4
17	1.6	0.7	1.7	19.7	23.3
18	1.7	0.8	0.5	21.9	18.1
19	1.8	0.9	2.1	20.2	19.4
20	1.9	1.3	4.3	10.0	21.5
21	2.5	1.6	6.2	7.1	25.1
22	2.7	1.8	7.3	5.8	27.7

Table A.6: Systematic errors contributions (in percent) for Z boson momentum differential cross section measurement in $Z(\rightarrow \ell\ell) + b$.

H_T **bin**	**data**	**MC**	$t\bar{t}$ **fit**	b **fit**	**unfolding**	**total stat.**
1	1.7	4.7	0.0	6.3	1.4	6.7
2	1.9	4.9	0.1	4.1	1.8	4.8
3	1.9	4.8	0.1	3.3	2.6	4.3
4	2.1	5.0	0.3	3.3	2.7	4.6
5	2.1	5.4	0.4	3.4	2.6	4.7
6	2.3	5.7	0.6	2.1	2.7	4.9
7	2.4	5.9	0.6	2.1	3.2	5.4
8	2.6	6.3	0.7	2.3	3.4	5.5
9	3.3	6.6	0.8	2.3	3.5	5.9
10	3.4	7.3	0.9	2.4	3.9	5.9
11	4.3	10.3	0.9	2.6	5.8	8.0
12	7.3	15.3	1.0	2.6	8.5	11.2
13	8.1	17.9	1.0	2.7	9.8	12.9

Table A.7: Statisticals error contribution (in percent) for the H_T differential cross section measurement in $Z(\rightarrow \ell\ell) + b$.

H_T **bin**	**JEC**	**JER**	**pile-up**	**unfolding**	**total syst.**
1	3.2	0.9	0.2	20.8	22.0
2	0.9	0.9	0.6	13.3	14.8
3	4.7	0.4	0.3	5.0	9.9
4	6.1	1.6	0.6	2.3	9.8
5	3.6	2.7	0.9	3.2	8.4
6	2.6	0.9	0.8	10.0	12.6
7	5.6	2.6	1.7	14.1	17.9
8	6.8	1.3	1.8	19.2	22.4
9	6.1	1.5	1.1	21.3	23.7
10	6.3	2.6	1.1	19.4	22.1
11	6.3	3.9	0.4	18.5	22.4
12	7.2	6.4	1.0	24.6	30.4
13	7.1	7.1	1.5	31.8	37.6

Table A.8: Systematic errors contributions (in percent) for the H_T differential cross section measurement in $Z(\rightarrow \ell\ell) + b$.

$\Delta\phi^{Zb}$ **bin**	**data**	**MC**	$t\bar{t}$ **fit**	b **fit**	**unfolding**	**total stat.**
1	5.8	13.8	0.6	5	5.7	10.1
2	4.1	9.5	0.6	5	3.6	7.9
3	4.0	9.2	0.6	5	3.5	8.9
4	4.0	9.4	0.6	6	3.4	8.2
5	4.1	9.4	0.7	6	3.6	8.4
6	3.6	8.4	0.7	5	3.5	6.2
7	3.6	7.6	0.7	4	3.4	6.6
8	2.8	6.4	0.6	4	3.4	4.0
9	2.2	5.4	0.5	3	2.6	4.7
10	1.9	4.1	0.4	3	2.2	4.3
11	1.6	4.0	0.3	3	1.6	4.3
12	1.2	3.2	0.2	4	1.2	4.7

Table A.9: Statistical error contributions (in percent) for the $\Delta\phi^{Zb}$ differential cross section measurement in $Z(\rightarrow \ell\ell) + b$.

$\Delta\phi^{Zb}$ **bin**	**JEC**	**JER**	**pile-up**	**unfolding**	**total syst.**
1	5.1	1.7	1.4	12.1	19.7
2	5.3	3.2	1.0	20.5	23.9
3	5.5	2.3	2.5	15.8	19.8
4	4.5	1.6	1.6	9.4	14.9
5	4.7	2.3	1.1	8.9	14.7
6	5.7	2.3	1.8	8.0	13.7
7	4.3	1.5	1.6	10.5	14.4
8	4.5	1.4	0.4	11.9	14.8
9	3.1	0.1	0.5	11.8	22.7
10	3.1	0.6	0.3	21.5	20.7
11	2.9	0.5	0.4	19.8	10.8
12	2.3	0.5	0.0	8.5	9.2

Table A.10: Systematic errors contributions (in percent) for the $\Delta\phi^{Zb}$ differential cross section measurement in $Z(\rightarrow \ell\ell) + b$.

Appendix B

Single Channel Cross Section Results

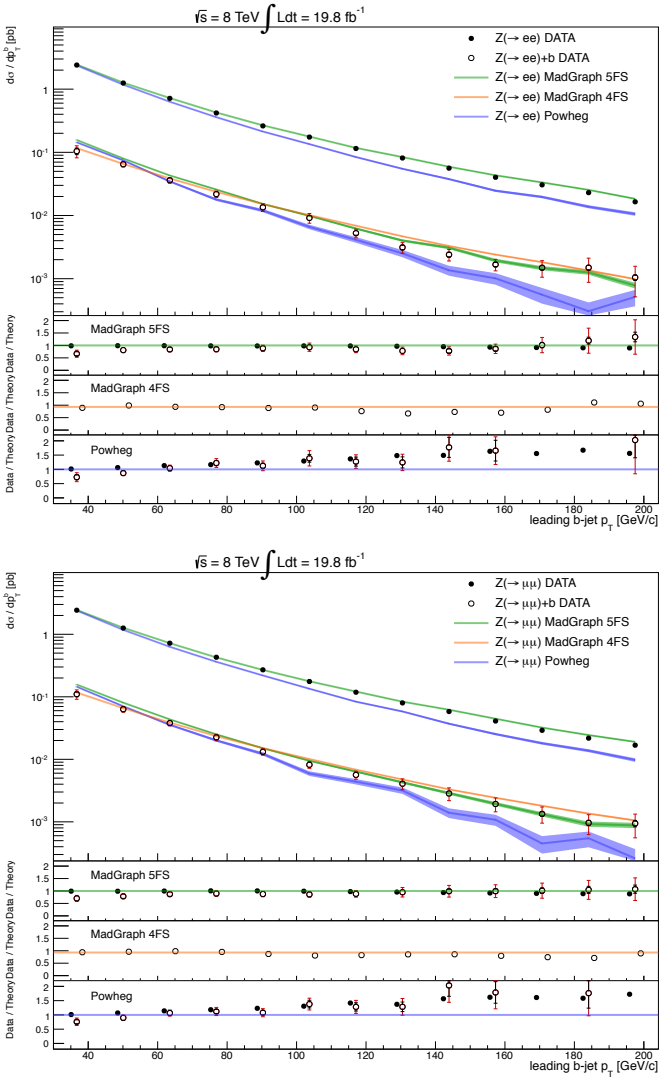

Figure B.1: Unfolded differential Z + jets (upper part of the plot) and $Z + b-$jets (lower part of the plot) cross section as a function of the leading jet p_T for electrons (top) and muons (bottom) final state, compared with the different theoretical predictions (coloured lines).

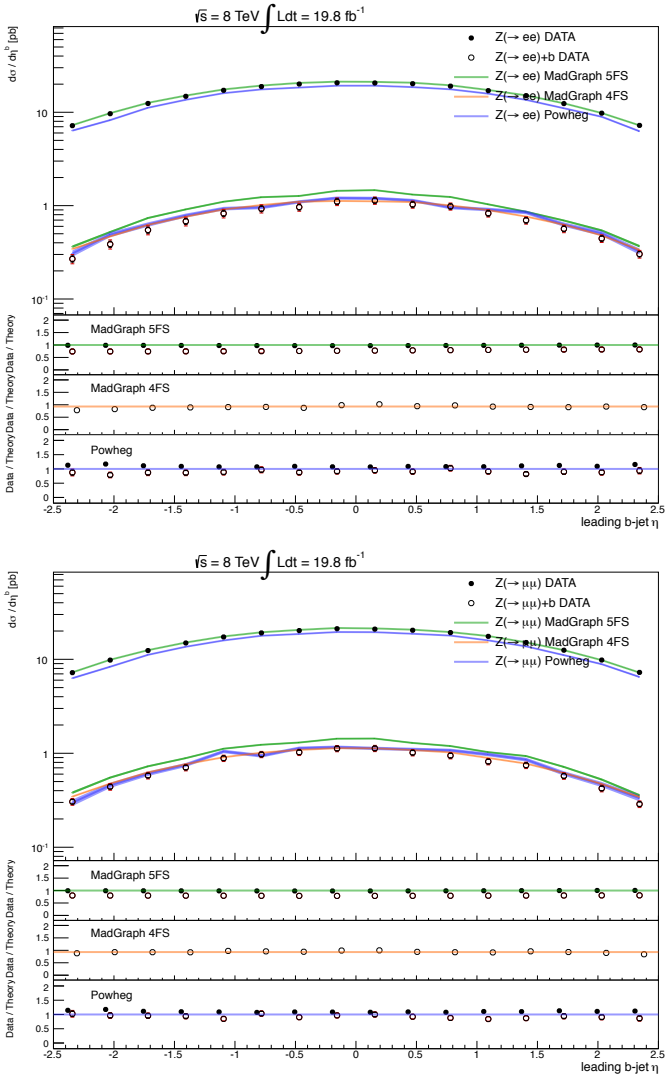

Figure B.2: Unfolded differential Z + jets (upper part of the plot) and $Z + b-$jets (lower part of the plot) cross section as a function of the leading jet η for electrons (top) and muons (bottom) final state, compared with the different theoretical predictions (coloured lines).

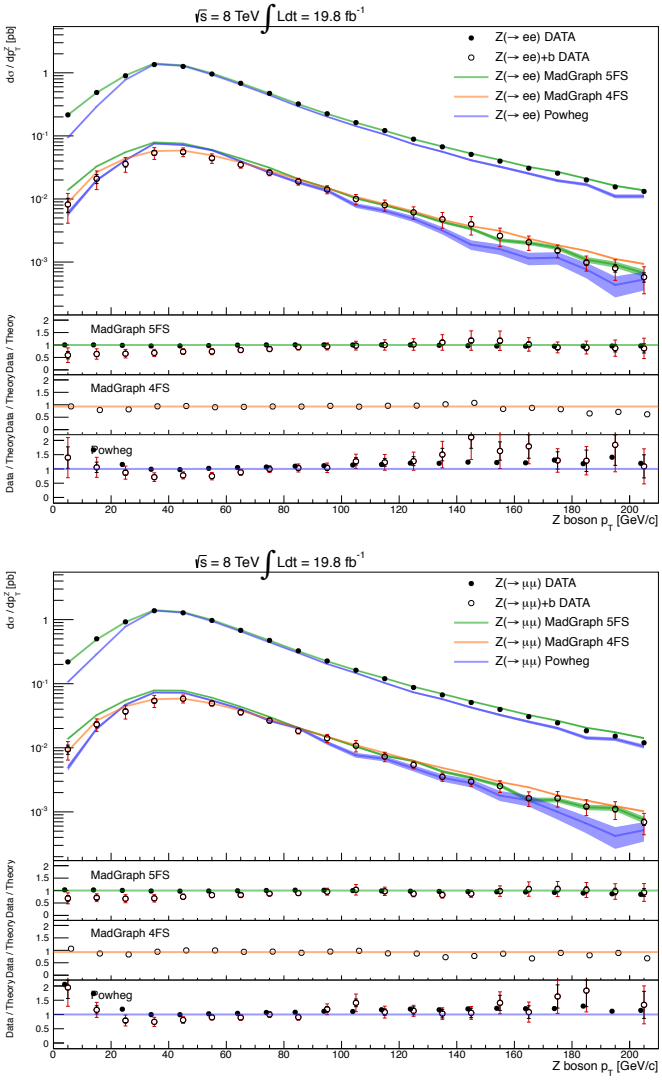

Figure B.3: Unfolded differential Z + jets (upper part of the plot) and $Z + b-$jets (lower part of the plot) cross section as a function of the Z boson p_T for electrons (top) and muons (bottom) final state, compared with the different theoretical predictions (coloured lines).

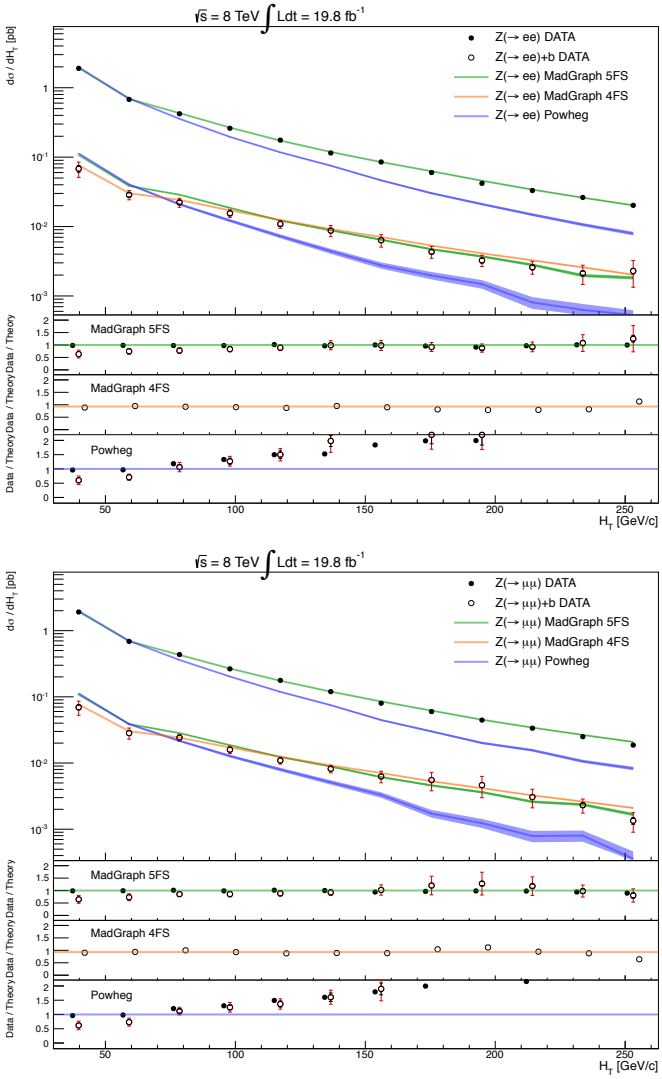

Figure B.4: Unfolded differential Z + jets (upper part of the plot) and $Z + b-$jets (lower part of the plot) cross section as a function of H_T for electrons (top) and muons (bottom) final state, compared with the different theoretical predictions (coloured lines).

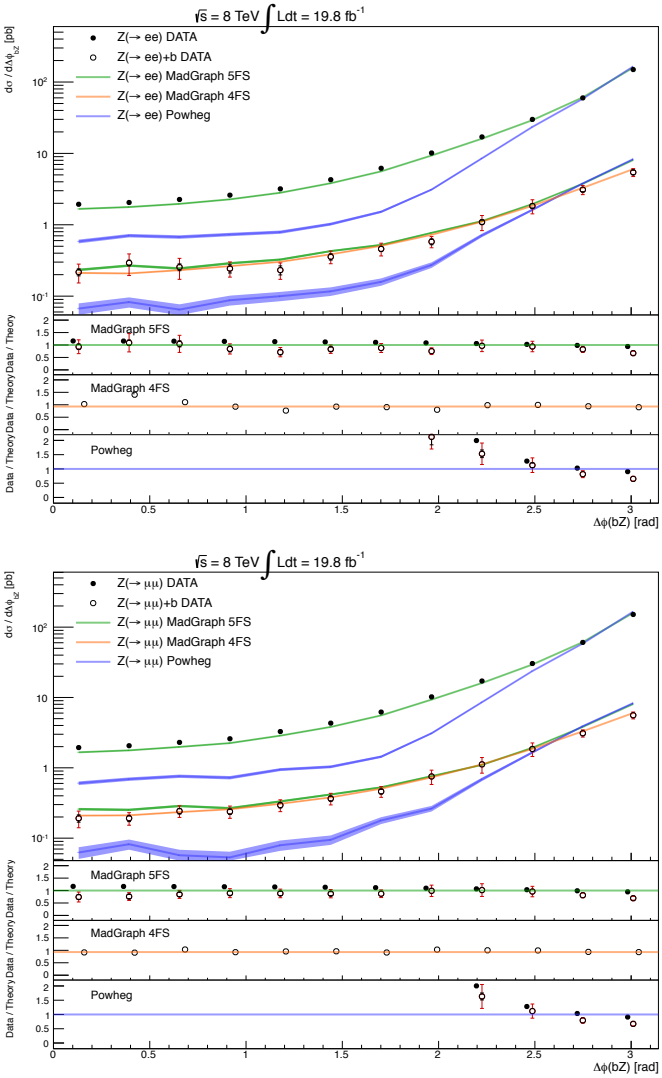

Figure B.5: Unfolded differential Z + jets (upper part of the plot) and $Z + b-$jets (lower part of the plot) cross section as a function of $\Delta\phi(Zb)$ for electrons (top) and muons (bottom) final state, compared with the different theoretical predictions (coloured lines).

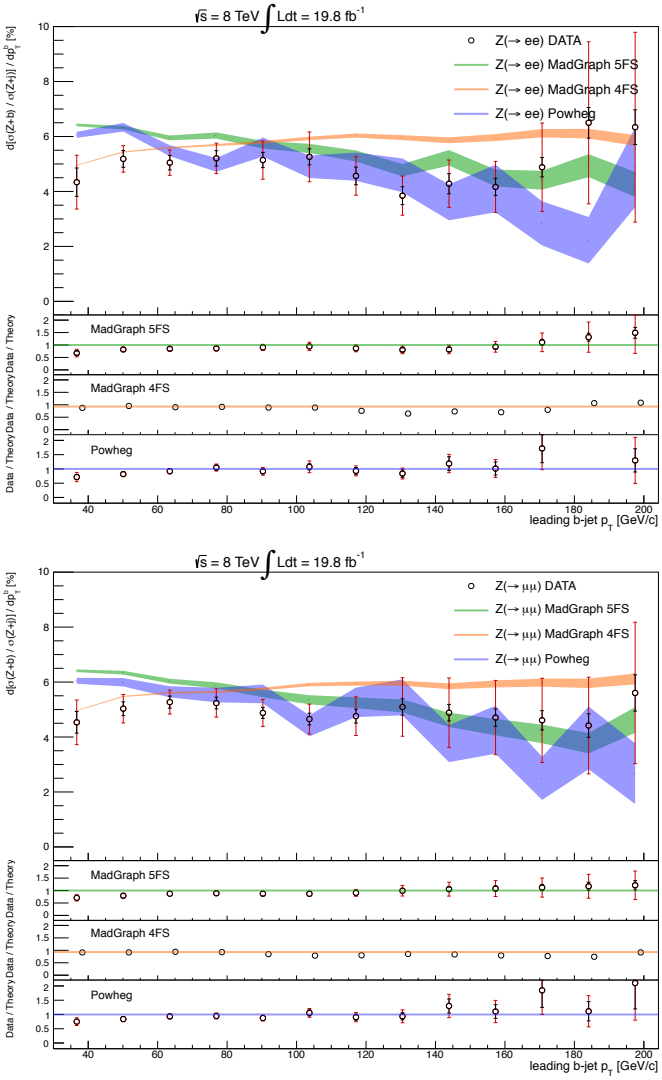

Figure B.6: Differential cross section ratio between the $Z+$jets (upper part of the plot) and $Z+b-$jets (lower part of the plot) cross section as a function of the leading jet p_T for electrons (top) and muons (bottom) final state, compared with the different theoretical predictions (coloured lines).

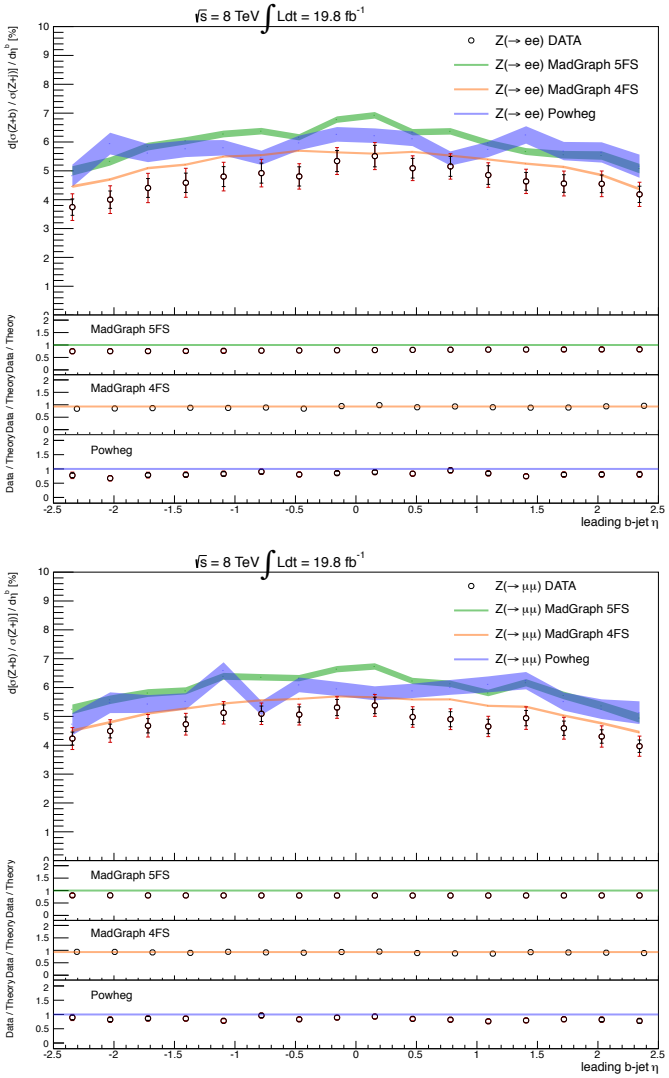

Figure B.7: Differential cross section ratio between the $Z+$jets (upper part of the plot) and $Z+b-$jets (lower part of the plot) cross section as a function of the leading jet η for electrons (top) and muons (bottom) final state, compared with the different theoretical predictions (coloured lines).

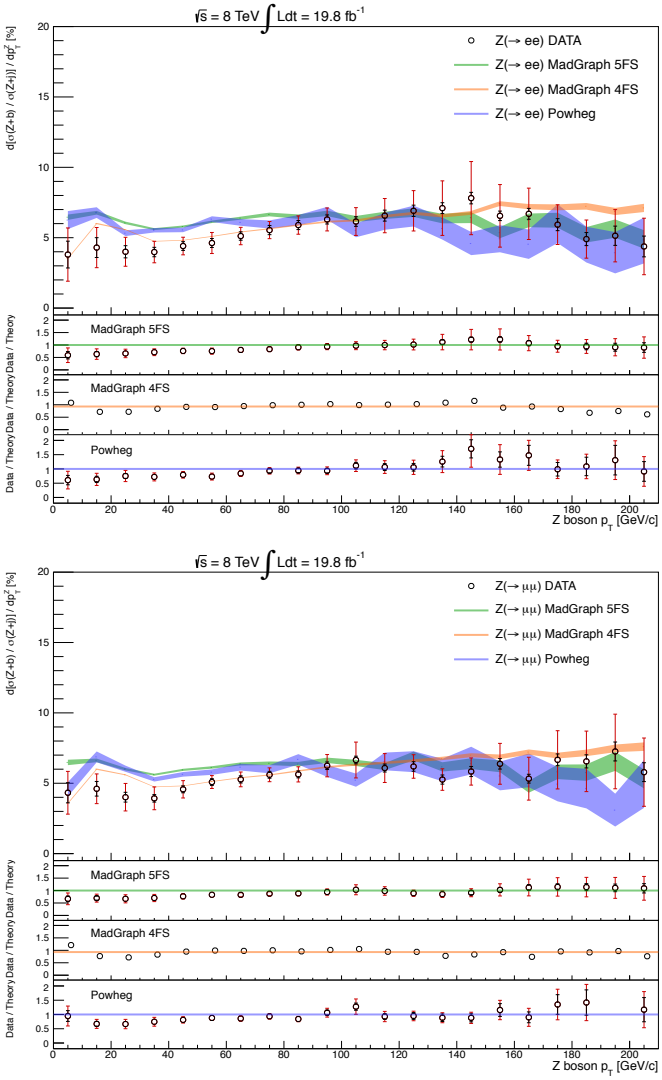

Figure B.8: Differential cross section ratio between the$Z$ + jets (upper part of the plot) and $Z + b-$jets (lower part of the plot) cross section as a function of the Z boson p_T for electrons (top) and muons (bottom) final state, compared with the different theoretical predictions (coloured lines).

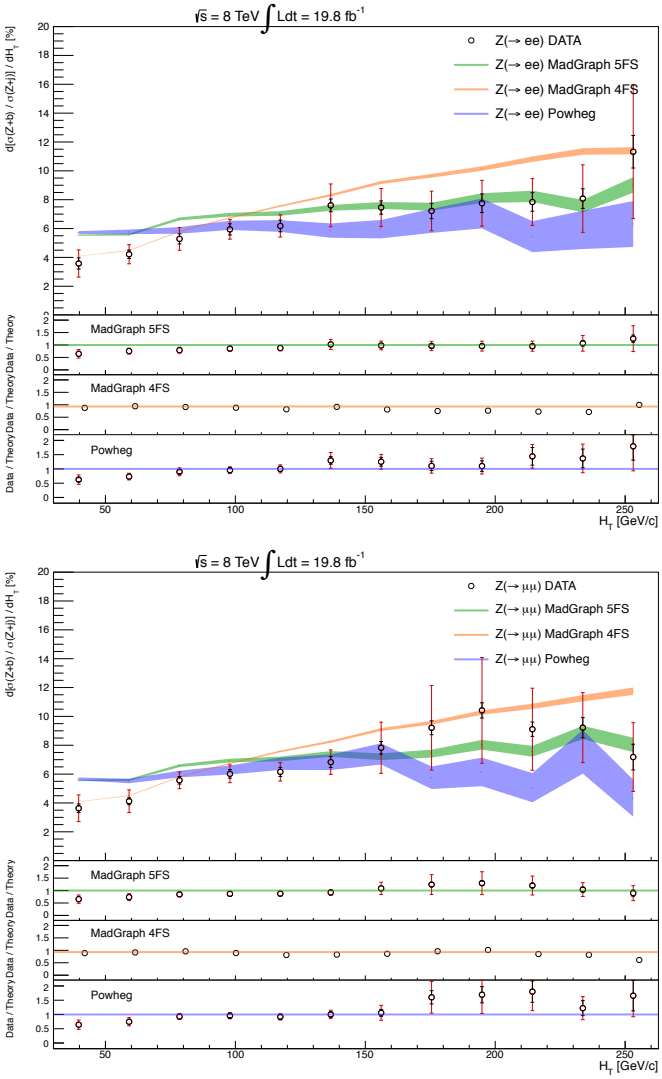

Figure B.9: Differential cross section ratio between the $Z+$jets (upper part of the plot) and $Z+b-$jets (lower part of the plot) cross section as a function of H_T for electrons (top) and muons (bottom) final state, compared with the different theoretical predictions (coloured lines).

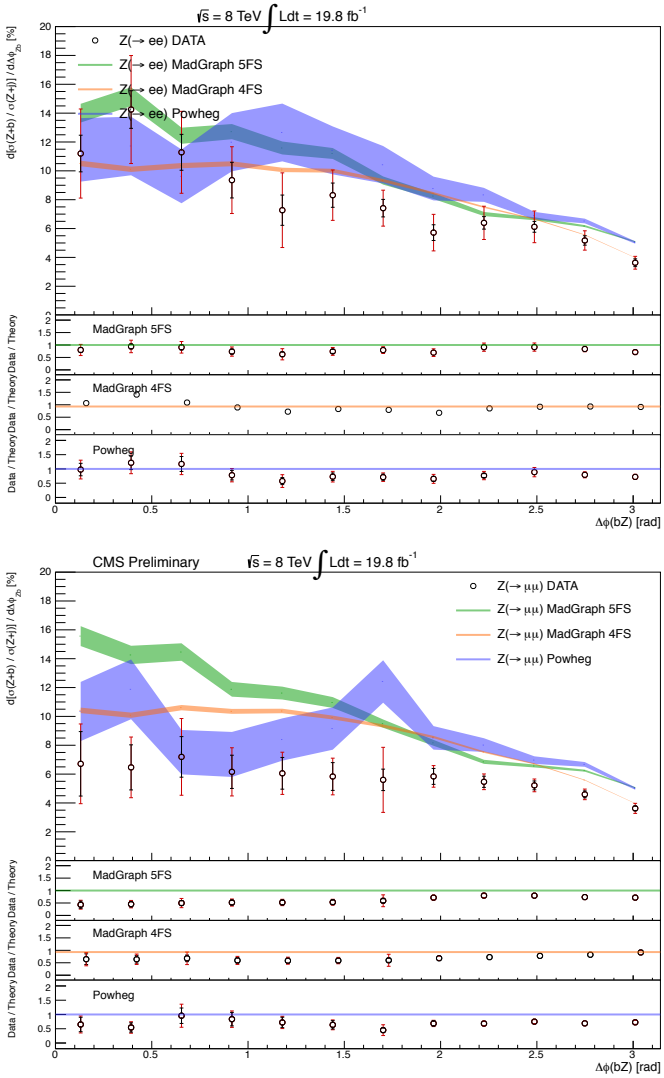

Figure B.10: Differential cross section ratio between the $Z+$jets (upper part of the plot) and $Z+b-$jets (lower part of the plot) cross section as a function of $\Delta\phi(Zb)$ for electrons (top) and muons (bottom) final state, compared with the different theoretical predictions (coloured lines).

List of Figures

List of Tables

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: 367–377, 1968.

[3] Steven Weinberg. A Model of Leptons. *Phys.Rev.Lett.*, 19:1264–1266, 1967.

[4] Nicola Cabibbo. Unitary Symmetry and Leptonic Decays. *Phys.Rev.Lett.*, 10: 531–533, 1963.

[5] Makoto Kobayashi and Toshihide Maskawa. CP Violation in the Renormalizable Theory of Weak Interaction. *Prog.Theor.Phys.*, 49:652–657, 1973.

[6] J. Beringer et al. Review of Particle Physics (RPP). *Phys.Rev.*, D86:010001, 2012.

[7] Peter W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys.Rev.Lett.*, 13:508–509, 1964.

[8] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, 1964.

[9] 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.

[10] Walter Van Doninck. Neutral Current Induced Reactions in the GARGAMELLE Experiment. 1976.

[11] N. Cabibbo, L. Maiani, G. Parisi, and R. Petronzio. Bounds on the Fermions and Higgs Boson Masses in Grand Unified Theories. *Nucl.Phys.*, B158:295–305, 1979.

[12] R. Barate et al. Search for the standard model Higgs boson at LEP. *Phys.Lett.*, B565:61–75, 2003.

[13] TEVNPH (Tevatron New Phenomena and Higgs Working Group). Combined CDF and D0 Upper Limits on Standard Model Higgs Boson Production with up to 8.6 fb^{-1} of Data. *FERMILAB-CONF-11-354-E*, 2011.

[14] Serguei Chatrchyan et al. Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV. *JHEP*, 06:081, 2013.

[15] 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.

[16] Sidney R. Coleman and S.L. Glashow. Departures from the eightfold way: Theory of strong interaction symmetry breakdown. *Phys.Rev.*, 134:B671–B681, 1964.

[17] S.L. Glashow, J. Iliopoulos, and L. Maiani. Weak Interactions with Lepton-Hadron Symmetry. *Phys.Rev.*, D2:1285–1292, 1970.

[18] D.J. Gross and Frank Wilczek. Ultraviolet Behavior of Nonabelian Gauge Theories. *Phys.Rev.Lett.*, 30:1343–1346, 1973.

[19] Yuri 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.

[20] V.N. Gribov and L.N. Lipatov. Deep inelastic $e-p$ scattering in perturbation theory. *Sov.J.Nucl.Phys.*, 15:438–450, 1972.

[21] Guido Altarelli and G. Parisi. Asymptotic Freedom in Parton Language. *Nucl.Phys.*, B126:298, 1977.

[22] Albert De Roeck. Physics results from the first electron - proton collider HERA. *DESY Conference Note*, 1995.

[23] R.B. Drucker. Electroweak results from NuTeV. *eConf*, C9808031:11, 1998.

[24] A. van Hameren, C.G. Papadopoulos, and R. Pittau. Automated one-loop calculations: A Proof of concept. *JHEP*, 0909:106, 2009.

[25] C.F. Berger, Z. Bern, L.J. Dixon, F. Febres Cordero, D. Forde, et al. An Automated Implementation of On-Shell Methods for One-Loop Amplitudes. *Phys.Rev.*, D78:036003, 2008.

[26] A. Bredenstein, A. Denner, S. Dittmaier, and S. Pozzorini. NLO QCD corrections to pp to $t\bar{t}\ b\bar{b}$ + X at the LHC. *Phys.Rev.Lett.*, 103:012002, 2009.

[27] Stefano Catani, Leandro Cieri, Giancarlo Ferrera, Daniel de Florian, and Massimiliano Grazzini. Vector boson production at hadron colliders: A fully exclusive QCD calculation at NNLO. *Phys.Rev.Lett.*, 103:082001, 2009.

[28] Radja Boughezal, Fabrizio Caola, Kirill Melnikov, Frank Petriello, and Markus Schulze. Higgs boson production in association with a jet at next-to-next-to-leading order in perturbative QCD. *JHEP*, 1306:072, 2013.

[29] Leandro Cieri. NNLO QCD results for diphoton production at the LHC and the Tevatron. *arXiv:1305.5101*, 2013.

[30] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-$k(t)$ jet clustering algorithm. *JHEP*, 0804:063, 2008.

[31] S. Dimopoulos, S. Raby, and Frank Wilczek. Supersymmetry and the Scale of Unification. *Phys.Rev.*, D24:1681–1683, 1981.

[32] J.A. Casas, J.R. Espinosa, and M. Quiros. Improved Higgs mass stability bound in the standard model and implications for supersymmetry. *Phys.Lett.*, B342:171–179, 1995.

[33] V.C. Rubin, N. Thonnard, and Jr. Ford, W.K. Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 /R = 4kpc/ to UGC 2885 /R = 122 kpc/. *Astrophys.J.*, 238:471, 1980.

[34] Linda M. Carpenter, Andre Nelson, Chase Shimmin, Tim M.P. Tait, and Daniel Whiteson. Collider searches for dark matter in events with a Z boson and missing energy. *arXiv:1212.3352*, 2012.

[35] Laurent Canetti, Marco Drewes, and Mikhail Shaposhnikov. Matter and Antimatter in the Universe. *New J.Phys.*, 14:095012, 2012.

[36] B. Pontecorvo. Mesonium and anti-mesonium. *Sov.Phys.JETP*, 6:429, 1957.

[37] Y. Fukuda et al. Measurements of the solar neutrino flux from Super-Kamiokande's first 300 days. *Phys.Rev.Lett.*, 81:1158–1162, 1998.

[38] J. Schechter and J.W.F. Valle. Neutrino Masses in SU(2) x U(1) Theories. *Phys.Rev.*, D22:2227, 1980.

[39] Max Baak and Roman Kogler. The global electroweak Standard Model fit after the Higgs discovery. *arXiv:1306.0571*, 2013.

[40] Fernando 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.

[41] John M. Campbell, R. Keith Ellis, F. Maltoni, and S. Willenbrock. Associated production of a Z Boson and a single heavy quark jet. *Phys.Rev.*, D69:074021, 2004.

[42] Fabio Maltoni, Giovanni Ridolfi, and Maria Ubiali. b-initiated processes at the LHC: a reappraisal. *JHEP*, 1207:022, 2012.

[43] 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.

[44] Fabio Maltoni, Thomas McElmurry, and Scott Willenbrock. Inclusive production of a Higgs or Z boson in association with heavy quarks. *Phys.Rev.*, D72: 074024, 2005.

[45] Torbjorn Sjostrand. Monte Carlo Generators. *High-energy physics. Proceedings, European School, Aronsborg, Sweden, June 18-July 1, 2006*, pages 51–74, 2006.

[46] T. Stelzer and W.F. Long. Automatic generation of tree level helicity amplitudes. *Comput.Phys.Commun.*, 81:357–371, 1994.

[47] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. *JHEP*, 0605:026, 2006.

[48] Matt Dobbs and Jorgen Beck Hansen. The HepMC C++ Monte Carlo event record for High Energy Physics. *Comput.Phys.Commun.*, 134:41–46, 2001.

[49] Stefan Hoeche, Frank Krauss, Nils Lavesson, Leif Lonnblad, Michelangelo Mangano, et al. Matching parton showers and matrix elements. *arXiv:hep-ph/0602031*, 2006.

[50] Valentin Hirschi, Rikkert Frederix, Stefano Frixione, Maria Vittoria Garzelli, Fabio Maltoni, et al. Automation of one-loop QCD corrections. *JHEP*, 1105: 044, 2011.

[51] Stefano Frixione, Fabian Stoeckli, Paolo Torrielli, Bryan R. Webber, and Chris D. White. The MCaNLO 4.0 Event Generator. *arXiv:1010.0819*, 2010.

[52] T. Gleisberg, Stefan. Hoeche, F. Krauss, M. Schonherr, S. Schumann, et al. Event generation with SHERPA 1.1. *JHEP*, 0902:007, 2009.

[53] 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.

[54] Carlo Oleari. The POWHEG-BOX. *Nucl.Phys.Proc.Suppl.*, 205-206:36–41, 2010.

[55] Rafael C. Lopes de Sa. Precise measurements of the W mass at the Tevatron and indirect constraints on the Higgs mass. *arXiv:1204.3260*, 2012.

[56] The ALEPH, DELPHI, L3, OPAL Collaborations, the LEP Electroweak Working Group. Electroweak Measurements in Electron-Positron Collisions at W-Boson-Pair Energies at LEP. *Phys. Rept.*, 532:119, 2013.

[57] Lawrence J. Hall, David Pinner, and Joshua T. Ruderman. A Natural SUSY Higgs Near 126 GeV. *JHEP*, 1204:131, 2012.

[58] Serguei Chatrchyan et al. Search for a Higgs boson decaying into a b-quark pair and produced in association with b quarks in proton-proton collisions at 7 TeV. *Phys.Lett.*, B722:207–232, 2013.

[59] B. Holdom, W.S. Hou, T. Hurth, M.L. Mangano, S. Sultansoy, et al. Four Statements about the Fourth Generation. *PMC Phys.*, A3:4, 2009.

[60] Serguei Chatrchyan et al. Search for a Vector-like Quark with Charge 2/3 in $t + Z$ Events from pp Collisions at $\sqrt{s} = 7$ TeV. *Phys.Rev.Lett.*, 107:271802, 2011.

[61] 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.

[62] J.A. Aguilar-Saavedra, R. Benbrik, S. Heinemeyer, and M. Perez-Victoria. A handbook of vector-like quarks: mixing and single production. *arXiv:1010.0819*, 2013.

[63] Guido Altarelli and Martin W. Grunewald. Precision electroweak tests of the standard model. *Phys.Rept.*, 403-404:189–201, 2004.

[64] ALEPH Collaboration, CDF Collaboration, D0 Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration, SLD Collaboration, LEP Electroweak Working Group, Tevatron Electroweak Working Group, SLD Electroweak and Heavy Flavour Groups. Precision Electroweak Measurements and Constraints on the Standard Model. *LEPEWWG-2010-01, TEVEWWG-2010-01*, 2010.

[65] M. Beccaria, N. Orlando, G. Panizzo, F.M. Renard, and C. Verzegnassi. The Relevance of polarized bZ production at LHC. *Phys.Lett.*, B713:457–461, 2012.

[66] D. Choudhury, Timothy M.P. Tait, and C.E.M. Wagner. Beautiful mirrors and precision electroweak data. *Phys.Rev.*, D65:053002, 2002.

[67] Kunal Kumar, William Shepherd, Tim M.P. Tait, and Roberto Vega-Morales. Beautiful Mirrors at the LHC. *JHEP*, 1008:052, 2010.

[68] Lyndon Evans and Philip Bryant. LHC Machine. *JINST*, 3:S08001, 2008.

[69] G. Aad et al. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.

[70] S. Chatrchyan et al. The CMS experiment at the CERN LHC. *JINST*, 3: S08004, 2008.

[71] K. Aamodt et al. The ALICE experiment at the CERN LHC. *JINST*, 3: S08002, 2008.

[72] Jr. Alves, A. Augusto et al. The LHCb Detector at the LHC. *JINST*, 3: S08005, 2008.

[73] CMS Collaboration. *CMS tracker design report: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997.

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

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

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

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

[78] CMS Collaboration. *CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems*. Technical Design Report CMS.

[79] Sergio Cittolin, Attila Rácz, and Paris Sphicas. *CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project*. Technical Design Report CMS. CERN, Geneva, 2002.

[80] CMS Collaboration. *Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET*. Number CMS-PAS-PFT-09-001. 2009. Geneva, Apr 2009.

[81] CMS Collaboration. *Particle-flow commissioning with muons and electrons from J/ψ and W events at 7 TeV*. Number CMS-PAS-PFT-10-003. 2010. Geneva, 2010.

[82] CMS Collaboration. *Tracking and Primary Vertex Results in First 7 TeV Collisions*. Number CMS-PAS-TRK-10-005. 2010. Geneva, 2010.

[83] CMS Collaboration. Determination of jet energy calibration and transverse momentum resolution in CMS. Technical report, Jul 2011. P11002. 67 pp.

[84] Matteo Cacciari and Gavin P. Salam. Pileup subtraction using jet areas. *Phys.Lett.*, B659:119–126, 2008.

[85] CMS Collaboration. Jet energy resolution in CMS at $\sqrt{s}$=7 TeV. (CMS-PAS-JME-10-014), 2011.

[86] Didar Dobur. Jets and missing transverse energy reconstruction with cms. *arxiv:0904.0391*, 2009.

[87] https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookMetAnalysis.

[88] Vardan Khachatryan et al. Measurements of Inclusive W and Z Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV. *JHEP*, 1101:080, 2011.

[89] Serguei Chatrchyan et al. Identification of b quark jets with the CMS experiment. *JINST*, 8:P04013, 2013.

[90] Christian Weiser. A combined secondary vertex based b-tagging algorithm in CMS. Technical Report CMS-NOTE-2006-014, CERN, Geneva, Jan 2006.

[91] Thomas Speer, Kirill Prokofiev, R Frühwirth, Wolfgang Waltenberger, and Pascal Vanlaer. Vertex fitting in the CMS tracker. Technical Report CMS-NOTE-2006-032, CERN, Geneva, Feb 2006.

[92] ALEPH Collaboration. A precise measurement of $Z \to b\bar{b}/Z \to hadrons$. *Physics Letters B*, 313:535 – 548, 1993.

[93] G Borisov and C Mariotti. Fine tuning of track impact parameter resolution of the DELPHI detector. *Nucl. Instrum. Methods Phys. Res., A*, 372(CERN-OPEN-96-011), 1996.

[94] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. MadGraph 5 : Going Beyond. *JHEP*, 1106:128, 2011.

[95] S. Agostinelli et.al. Geant 4 simulation toolkit. *Nucl.Inst.Meth.A*, 506(3):250 – 303, 2003.

[96] Ryan Gavin, Ye Li, Frank Petriello, and Seth Quackenbush. FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order. *Comput.Phys.Commun.*, 182:2388–2403, 2011.

[97] John M. Campbell. $W/Z + B, \bar{B}$ jets at NLO using the Monte Carlo MCFM. 2001.

[98] John M. Campbell and R.K. Ellis. MCFM for the Tevatron and the LHC. *Nucl.Phys.Proc.Suppl.*, 205-206:10–15, 2010.

[99] John M. Campbell, R. Keith Ellis, and Ciaran Williams. Vector boson pair production at the LHC. *JHEP*, 1107:018, 2011.

[100] Serguei Chatrchyan et al. Measurement of the $Z/\gamma*+b$-jet cross section in pp collisions at 7 TeV. *JHEP*, 1206:126, 2012.

[101] Serguei Chatrchyan et al. Measurement of the cross section and angular correlations for associated production of a Z boson with b hadrons in pp collisions at $\sqrt{s} = 7$ TeV. *arXiv:0904.0391*, 2013.

[102] http://mathworld.wolfram.com/FredholmIntegralEquationoftheSecondKind.html.

[103] Andreas Hocker and Vakhtang Kartvelishvili. SVD approach to data unfolding. *Nucl.Instrum.Meth.*, A372, 1996.

[104] G. D'Agostini. A multidimensional unfolding method based on Bayes theorem. *Nucl.Instrum.Meth.*, A362, 1995.

[105] T. Adye. Unfolding algorithms and tests using RooUnfold. *arxiv:1105.1160*, May 2011.

[106] V.Candelise et al. Characterization of the final state radiation in Z boson decay to electrons. *CMS-AN-2012-204.*, 2012.

[107] https://twiki.cern.ch/twiki/bin/view/CMS/JECUncertaintySources.

[108] https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupSystematicErrors.

[109] https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagPOG.

[110] CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update. Technical Report CMS-PAS-LUM-13-001, CERN, Geneva, 2013.

[111] Victor Mukhamedovich Abazov et al. Measurement of the ratio of differential cross sections $\sigma(p\bar{p} \to Z + bjet)/\sigma(p\bar{p} \to Z + jet)$ in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. *Phys.Rev.*, D87:092010, 2013.

[112] Keith Hamilton, Paolo Nason, and Giulia Zanderighi. MINLO: Multi-Scale Improved NLO. *JHEP*, 1210:155, 2012.

[113] Serguei Chatrchyan et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys.Lett.*, B716:30–61, 2012.