Università degli Studi di Firenze
Dipartimento di Fisica e Astronomia
Dottorato di Ricerca in Fisica - XXIII Ciclo

Measurement of $\sigma(\mathbf{Z}/\gamma^{*}+\geq \mathbf{n}$ jets) in the electronic channel with the CMS detector

Tesi di Dottorato di Ricerca in Fisica
S.S.D. FIS/04

Simone Frosali

Relatore: **Prof. R. D'Alessandro**

Coordinatore: **Prof. A. Cuccoli**

Firenze, 30 Gennaio 2012

Contents

Introduction **1**

1 Electroweak physics and QCD at LHC **3**
- 1.1 The Standard Model . . . 3
- 1.2 Weak isospin and hypercharge . . . 5
- 1.3 Electroweak interaction . . . 7
  - 1.3.1 Electroweak lagrangian . . . 9
- 1.4 Higgs mechanism . . . 10
  - 1.4.1 Fermion mass terms . . . 13
- 1.5 Strong interaction . . . 14
  - 1.5.1 QCD lagrangian . . . 15
  - 1.5.2 Parton density functions . . . 15
  - 1.5.3 Infrared and collinear safe observables . . . 16
- 1.6 The $pp \to Z + jets$ process . . . 18
  - 1.6.1 Z boson production and decay . . . 19
  - 1.6.2 Associated jet production . . . 20

2 The CMS experiment at LHC **23**
- 2.1 The Large Hadron Collider . . . 23
  - 2.1.1 The LHC schedule . . . 26
- 2.2 The CMS detector . . . 28
  - 2.2.1 The Solenoid . . . 30
  - 2.2.2 The Tracker . . . 31
  - 2.2.3 The Electromagnetic Calorimeter (ECAL) . . . 34
  - 2.2.4 The Hadronic Calorimeter (HCAL) . . . 36
  - 2.2.5 The Muon System . . . 38
- 2.3 Trigger system . . . 40
  - 2.3.1 Level-1 Trigger . . . 41
  - 2.3.2 High Level Trigger . . . 42
- 2.4 CMS simulation and reconstruction software . . . 43
  - 2.4.1 Event simulation . . . 44
  - 2.4.2 Event reconstruction . . . 44
- 2.5 Reconstructed data sample . . . 46

3 MC production **49**
3.1 Event generator components 49
3.2 Matrix Element (ME) 52
3.3 Parton Shower (PS) 53
3.3.1 Merging Matrix Elements and Parton Shower 53
3.4 Hadronization 54
3.4.1 Lund string model 54
3.5 `PYTHIA` and `MADGRAPH` generators 55
3.6 MC samples produced 56

4 Electron reconstruction and Z(ee) events selection **59**
4.1 Standard electron reconstruction in CMSSW 59
4.1.1 Calorimetric energy reconstruction 59
4.1.2 ECAL driven seeding 62
4.1.3 Electron track reconstruction 63
4.1.4 Tracker driven seeding 68
4.2 Electron preselection 69
4.3 Electron charge determination 71
4.4 Electron momentum determination 73
4.4.1 The `GSFElectron` object 74
4.5 Electron Isolation 74
4.6 The `PATElectron` object 75
4.7 Electron HLT 76
4.8 Z(ee) events selection 77
4.9 Z candidate distributions after cuts 84

5 Jet reconstruction and selection in $Z \to ee + jets$ events **91**
5.1 Standard jet clustering algorithms in CMSSW 91
5.1.1 The SISCone algorithm 92
5.1.2 The k_t and anti-k_t algorithms 93
5.2 Jet reconstruction in CMS 95
5.2.1 Particle Flow jets (PF jets) 95
5.2.2 Jet position and momentum measurements 96
5.3 Jet energy corrections and p_T resolution 96
5.4 The `PATJet` object 102
5.5 Jet - lepton isolation 103
5.6 Z+jet events selection 103
5.6.1 Jet p_T and η distributions after cuts 104

6 Selection efficiency measurement **107**
6.1 The $Z \to ee + jets$ event selection efficiency 107
6.2 MC estimation of selection efficiency 108
6.3 Tag & Probe estimation of selection efficiency 112

6.3.1 Description of Tag & Probe tool 113
6.3.2 Global efficiency measurement with Tag & Probe 116
6.4 Final efficiency estimation 118

7 Cross section measurement **129**
7.1 Raw jet multiplicity 129
7.2 Signal and background extraction 131
7.3 Efficiency correction 135
7.4 Unfolding multiplicity distributions 135
7.5 Systematic uncertainties 138
7.5.1 Systematics of the efficiency 139
7.5.2 Systematics of the unfolding procedure 139
7.5.3 Systematics for the jet energy correction 140
7.5.4 Systematic from the luminosity measurement 142
7.6 Measurement of $\sigma(\mathrm{Z} + \geq n \text{ jets})$ 142

Conclusions **149**

Bibliography **153**

Introduction

The characterization of $Z/\gamma^* + jets$ production at the Large Hadron Collider (LHC), with the vector boson decaying leptonically, is one of the goals of early physics analyses of the Compact Muon Solenoid (CMS) experiment.

In the context of the Standard Model, the study of the production of electroweak bosons plus jets allows for tests of pQCD. The production cross section scales with the strong coupling constant for each additional jet. While current theoretical predictions at LO and at NLO are in good agreement with data, in order to establish guidance for higher order pQCD, comparison with data for a larger number of jets is needed. The study of the $Z/\gamma^* + jets$ process is also crucial for the understanding of the Higgs boson production background in the mass range 130 GeV $-$ 500 GeV, where the $H \to ZZ^* \to 4\ell$ is the most important channel for the Higgs boson detection. Also, $Z/W + jets$ events represent a background for many new physics searches, such as Super-Symmetry or the inclusive hadronic searches of Dark Matter, based on jets and missing energy. Finally these events are very useful for the calibration of the CMS calorimeters response, using the balancing of the jets and the recoiling Z boson.

The present work is focused on Z bosons decaying in the electronic channel. In particular, the pp collision data at $\sqrt{s} = 7$ TeV collected at CMS during all the 2010 are analyzed, for a total integrated luminosity of $\mathscr{L} = (36.2 \pm 1.4)\,\mathrm{pb}^{-1}$.

After a brief description of the Standard Model, with particular attention to the electroweak and QCD sectors, the CMS detector is presented together with the reconstruction algorithms that give us access to higher level objects built from the raw energy deposits recorded by the various subdetectors. The cuts adopted for selecting $Z \to e^+e^- + jets$ events are then described, and the kinematic variables of the reconstructed objects are compared with the Monte Carlo predictions for the selected events. Two different p_T cuts are taken into account for jets in the selected events: $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$.

Particular attention has been devoted to the efficiency measurement, performed by using a data driven method called the "Tag & Probe" method. A data driven measurement of the selection efficiency is indeed preferable, in order to avoid large systematic errors due to imperfections in the Monte Carlo simulations.

Finally the inclusive jet rate measurements obtained both for $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$ are shown and the cross section $\sigma\,(\mathrm{Z}+ \geq \mathrm{n\,jets})$ is evaluated for both cases. The results are compared with Monte Carlo predictions (considering MADGRAPH and PYTHIA generators with different tunes) and the systematic uncertainties that affect the measurement are discussed in detail.

Unless otherwise stated natural units $\hbar = c = 1$ are used throughout this work.

Chapter 1

Electroweak physics and QCD at LHC

In this chapter the Standard Model of electroweak and strong interaction is briefly described. The Standard Model [1] is a quantum field theory based on a $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$ local gauge symmetry. As shown in the following the above symmetry can be satisfied only if the fermion fields are massless: this fact contrasts with the experimental observation of massive fermions. A mechanism known as spontaneous symmetry breaking is used in the Standard Model to provide elementary particles with mass, without violating the local gauge symmetry. This mechanism requires the existence of a new field still unobserved, named Higgs field.

It will be shown how the request for a local gauge symmetry together with the spontaneous symmetry breaking of the $SU(2)_L \otimes U(1)_Y$ symmetry leads to the prediction of the existence of the weak gauge bosons Z and $W^\pm$ and how them and the fermion fields acquire mass through the interaction with the Higgs field. The unbroken $U(1)_{em}$ symmetry will be demonstrate to be responsible for the electromagnetic interaction mediated by the photon γ. Finally the strong interaction will be included in this picture, through the request of a local $SU(3)_c$ gauge symmetry, that is mediated by eight colored gluons.

Within this framework the phenomenology of the production of Z and $W^\pm$ bosons will be put into evidence, since their production rate at LHC is unprecedentedly high allowing very precise measurements. In particular the process $pp \to Z + jets$, which constitutes the subject of this thesis, will be described.

Unless otherwise stated natural units $\hbar = c = 1$ are used throughout this work.

1.1 The Standard Model

The Standard Model (SM) is the theory which provides the best description of the particle interaction phenomenology at the energies explored so far. The SM is built with six spin-$\frac{1}{2}$ particles called leptons and six spin-$\frac{1}{2}$ particles called quarks. They all constitute the matter fields and can interact through three fundamental forces, the electromagnetic force, the weak force and the strong force. The Gravitational force is not described by the SM, since it is not extensible to the general relativity. Leptons and quarks are classified in three generations:

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix} \quad \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix} \quad \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}, \tag{1.1}$$

$$\begin{pmatrix} u \\ d \end{pmatrix} \quad \begin{pmatrix} s \\ c \end{pmatrix} \quad \begin{pmatrix} t \\ b \end{pmatrix}. \tag{1.2}$$

There is no evidence for a fourth generation so far.

- Lepton doublets (Eq. (1.1)) are composed by a charged particle, electron e, muon μ and tauon τ, with electric charge $Q = -1$, and by a neutral particle, the electronic neutrino ν_e, the muonic neutrino ν_μ and the tauonic neutrino ν_τ. The first group of particles can interact via electromagnetic and weak forces, while the neutrinos only via the weak force.

- Quark doublets (Eq. (1.2)) are composed instead by a particle with electric charge $Q = +\frac{2}{3}$, up (u), charm (c) and top (t), and by a particle of electric charge $Q = -\frac{1}{3}$, down (d), strange (s) and bottom (b). The quarks can interact via electromagnetic, weak and strong forces.

For each particle a corresponding antiparticle exists, with the same mass but opposite quantum numbers. Ordinary matter consists of leptons and hadrons. The latters are particles composed by quarks and are classified in two categories: mesons, which are bound states of a quark and an antiquark, and baryons, which are bound states of three quarks. In the SM the three fundamental interactions are mediated by spin-1 bosons:

Strong : mediated by eight gluons g;

Weak : mediated by W^+, W^- and Z;

Electromagnetic : mediated by the photon γ.

Weak and electromagnetic forces are actually two manifestations of the same fundamental interaction, called electroweak interaction. The theory of electroweak interaction has been formulated by S.L. Glashow [2], A. Salam [3] and S. Weinberg [4] as an $SU(2) \otimes U(1)$ local gauge theory, and will be briefly described in the following section.

1.2 Weak isospin and hypercharge

In 1957 the Madame Wu [5] and Garwin-Lederman-Weinrich [6] experiments confirmed the parity violation for weak charged current interactions. Weak charged current interactions were proven experimentally to prefer final states with left handed particles or right handed antiparticles [7]. The absence of the "mirror image" states, i.e. right handed particles and left handed antiparticles, is a clear violation of parity invariance.

According to the Fermi's four fermion theory [8] of weak charged interactions, the invariant amplitude for weak interaction is given by the product of two conserved charged currents. For example, considering the $\nu_e e^- \to e^- \nu_e$ scattering process, the two conserved currents are:

$$J_\mu \equiv J^+_\mu = \bar{u}_\nu \gamma_\mu u_e \qquad \text{and} \qquad J^\dagger_\mu \equiv J^-_\mu = \bar{u}_e \gamma_\mu u_\nu \,, \tag{1.3}$$

where u_ν and u_e are respectively the neutrino and electron Dirac spinors and the $+$ and $-$ superscripts are to indicate the charge-raising and charge-lowering character of the currents, respectively. The weak interaction amplitude is therefore given by:

$$\mathfrak{M} = \frac{4G}{\sqrt{2}} (J^\mu)^+ (J_\mu)^- \,.$$

Charge conservation requires that $\mathfrak{M}$ is the product of a charge-raising and a charge-lowering current. The parity violation of the weak interaction shown in the experiments can be take into account by rewriting the charged currents in the following form:

$$J^+_\mu = \bar{u}_\nu \gamma_\mu \frac{1}{2}(1 - \gamma^5) u_e \,, \tag{1.4}$$

$$J^-_\mu = \bar{u}_e \gamma_\mu \frac{1}{2}(1 - \gamma^5) u_\nu \,. \tag{1.5}$$

The term $\frac{1}{2}(1-\gamma^5) = P_L$ is a projector that projects out just particles with left handed helicity and antiparticles with right handed helicity. If indeed the Dirac equation for massless fermions is decoupled introducing two-component spinors $\phi_1(\mathbf{p})$ and $\phi_2(\mathbf{p})$, the first describing left handed particles and right handed antiparticles, the second one describing instead right handed particles and left handed antiparticles [1]. The effect of the P_L projector on the neutrino spinor is shown in the following equation:

$$\frac{1}{2}(1-\gamma^5) u_\nu = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = \begin{pmatrix} \phi_1 \\ 0 \end{pmatrix} = (u_\nu)_L \,. \tag{1.6}$$

The P_L projector therefore projects out just ν_L and $\bar{\nu}_R$, where the subscripts L and R denote left handed and right handed spinors respectively. The $(u_\nu)_L$ term is the projected Dirac spinor, which represents both ν_L and $\bar{\nu}_R$. The currents shown in Eq. (1.4) and (1.5) are said to have a V-A form, where V-A is an acronym of *Vector* *minus Axial*. In fact these currents present a mixture of a vector term (γ^μ) and an axial-vector term ($\gamma^5\gamma^\mu$). This mixture automatically violates parity conservation, as shown above.

In the high energy limit it is possible to apply the same procedure to the massive fermions, obtaining from Eq. (1.4) and (1.5):

$$J^+_\mu = (\bar{u}_\nu)_L \gamma_\mu (u_e)_L \,, \qquad J^-_\mu = (\bar{u}_e)_L \gamma_\mu (u_\nu)_L \,.$$

Hereafter the Dirac spinors are denoted by using the particle names, e.g. $\bar{u}_\nu \equiv \bar{\nu}, u_e \equiv e$, etc. With this notation the weak charge currents become

$$J^+_\mu = \bar{\nu}_L \gamma_\mu e_L \,, \qquad J^-_\mu = \bar{e}_L \gamma_\mu \nu_L \,. \tag{1.7}$$

The two charged currents can be rewritten in an useful two-dimensional form. To this purpose the doublet

$$\chi_L = \begin{pmatrix} \nu \\ e^- \end{pmatrix}_L \tag{1.8}$$

is introduced. Also the "step-up" and "step-down" operators $\tau_\pm = \frac{1}{2}(\tau_1 \pm i\tau_2)$ are introduced, where τ_i with $i = 1, 2, 3$ are the Pauli spin matrices. The charged currents then become:

$$J^+_\mu = \bar{\chi}_L \gamma_\mu \tau_+ \chi_L \,, \tag{1.9}$$

$$J^-_\mu = \bar{\chi}_L \gamma_\mu \tau_- \chi_L \,. \tag{1.10}$$

In order to complete the $SU(2)$ invariance of the theory a third conserved current should exist with the form

$$J^3_\mu = \bar{\chi}_L \gamma_\mu \tau_3 \chi_L = \bar{\nu}_L \gamma_\mu \nu_L - \bar{e}_L \gamma_\mu e_L \,. \tag{1.11}$$

An "isospin" triplet of weak currents is therefore constructed, $J^\pm_\mu$ and J^3_μ. However, the J^3_μ current cannot be identified with the neutral current, since J^3_μ involves only left handed fermions, while the observed weak neutral current J^{nc}_μ has both left handed and right handed components [1]. Also the electromagnetic current involves both left and right handed components, and it does not couple with the chargeless neutrino, thus it is not easily connected to J^3_μ.

In order to save the $SU(2)_L$ symmetry (where the L subscript is to remind that the weak isospin current couples only left handed fermions) the electromagnetic current J^{em}_μ must be included in this picture. Since neither J^{nc}_μ or J^{em}_μ respects $SU(2)_L$ symmetry, the idea is to form two orthogonal combinations which do have definite transformation properties under $SU(2)_L$: one combination, J^3_μ, is to complete the weak isospin triplet, while the second, J^Y_μ, is unchanged by $SU(2)_L$ transformations, i.e. it is a weak isospin singlet. The weak hypercharge current is called J^Y_μ and is given by

$$J^Y_\mu = \bar{\psi} \gamma_\mu Y \psi \,. \tag{1.12}$$

Lepton	T	T^3	Q	Y
$\nu_{\ell,L}$	$\frac{1}{2}$	$\frac{1}{2}$	0	-1
e^-_L	$\frac{1}{2}$	$-\frac{1}{2}$	-1	-1
e^-_R	0	0	-1	-2

(a)

Quark	T	T^3	Q	Y
u_L	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{1}{3}$
d_L	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{1}{3}$	$\frac{1}{3}$
u_R	0	0	$\frac{2}{3}$	$\frac{4}{3}$
d_R	0	0	$-\frac{1}{3}$	$-\frac{2}{3}$

(b)

Table 1.1: Quantum numbers of lepton (a) (ℓ=e, μ, τ) and quark (b) ($[u, d]$=[u, d], [c, s], [t, b]) helicity states: electric charge Q_{em} in unit of e, weak isospin T with third axis projection T_3 and weak hypercharge Y are shown.

The weak hypercharge Y is defined by

$$Q = T^3 + \frac{Y}{2}\,, \tag{1.13}$$

where Q is the electric charge, while $T^3 = \int J^3_0(x)d^3x$ is the charge associated to J^3_μ and it is called the "weak isospin" quantum number. The electromagnetic current is expressed as a linear combination of J^3_μ and J^Y_μ:

$$J^{em}_\mu = J^3_\mu + \frac{1}{2}J^Y_\mu\,. \tag{1.14}$$

Just as Q generates the group $U(1)_{em}$, so the hypercharge operator Y generates a symmetry group $U(1)_Y$. The electromagnetic interaction is therefore incorporated by enlarging the symmetry group to $SU(2)_L \otimes U(1)_Y$. Eq. (1.14) represents the electroweak unification, that is the unification of weak and electromagnetic interaction. In Sec. 1.3 it will be shown how the weak neutral current can be expressed as an analogous linear combination. The charge, hypercharge and weak isospin quantum numbers of electrons and quarks are summarized in Table 1.1.

1.3 Electroweak interaction

To complete the electroweak unification the current-current interaction scheme used so far must be modified, assuming that the current-current structure is an effective interaction which results from the exchange of massive vector bosons with only a small momentum transfer. The QED basic interaction is given by

$$-\,ie(J^{em})^\mu A_\mu\,, \tag{1.15}$$

where the A_μ gauge fields represent the photon exchange. Just as the electromagnetic current is coupled to the photon, it can be assumed that the electroweak currents are coupled to vector bosons.

The SM consists of an isotriplet of vector fields W^i_μ coupled with strength g to the weak isospin currents $J^i_\mu = \bar{\chi}_L \gamma_\mu \frac{1}{2}\tau_i \chi_L$, together with a single vector field B_μ coupled to the weak hypercharge current J^Y_μ with strength conventionally taken to be $g'/2$. The basic electroweak interaction is therefore:

$$- ig(J^i)^\mu W^i_\mu - i\frac{g'}{2}(J^Y)^\mu B_\mu \,. \tag{1.16}$$

After the change of variables

$$W^\pm = \sqrt{\frac{1}{2}}(W^1_\mu \mp W^2_\mu)\,, \tag{1.17}$$

the representation with the charge-raising and charge-lowering currents is retrieved:

$$\begin{aligned} -i\frac{g}{\sqrt{2}}(J^+)^\mu W^+_\mu - i\frac{g}{\sqrt{2}}(J^-)^\mu W^-_\mu & \\ -ig(J^3)^\mu W^3_\mu - i\frac{g'}{2}(J^Y)^\mu B_\mu \,. & \end{aligned} \tag{1.18}$$

The fields $W^\pm_\mu$ describe massive charged bosons, whereas W^3_μ and B_μ are neutral fields. The first row of Eq. (1.18) describes therefore the charged current sector, while the second row describes the neutral current sector.

The electromagnetic interaction is embedded in Eq. (1.18). Indeed, when the boson masses are generated by symmetry breaking (see Sec. 1.4), the two neutral fields W^3_μ and B_μ must mix in such a way that the physical states, i.e. the mass eigenstates, are [1]:

$$A_\mu = B_\mu \cos\theta_W + W^3_\mu \sin\theta_W \quad (\text{massless})\,, \tag{1.19}$$

$$Z_\mu = -B_\mu \sin\theta_W + W^3_\mu \cos\theta_W \quad (\text{massive})\,, \tag{1.20}$$

where θ_W is called the Weinberg angle. After this change of variables, the neutral sector may be therefore rewritten as follows:

$$\begin{aligned} & - igJ^3_\mu (W^3)^\mu - i\frac{g'}{2}J^Y_\mu B^\mu \\ & \qquad = -i\left(g\sin\theta_W J^3_\mu + g'\cos\theta_W \frac{J^Y_\mu}{2}\right) A^\mu \\ & \qquad\qquad - i\left(g\cos\theta_W J^3_\mu - g'\sin\theta_W \frac{J^Y_\mu}{2}\right) Z^\mu \,. \end{aligned} \tag{1.21}$$

The first term is the electromagnetic interaction, and so, considering the Eq. (1.14), the expression in the brackets must be

$$eJ^{em}_\mu \equiv e(J^3_\mu + \frac{1}{2}J^Y_\mu)\,. \tag{1.22}$$

Therefore we have

$$g\sin\theta_W = g'\cos\theta_W = e\,, \tag{1.23}$$

that is the Weinberg angle is given by the ratio of the two independent group constants, $\tan\theta_W = g'/g$. Using Eq. (1.23) the weak neutral current interaction of Eq. (1.21) may be expressed in the form

$$-\,i\frac{g}{\cos\theta_W}(J^3_\mu - \sin^2\theta_W J^{em}_\mu)Z^\mu \equiv -i\frac{g}{\cos\theta_W}J^{nc}_\mu Z^\mu\,, \tag{1.24}$$

where the definition of the weak neutral conserved current J^{nc}_μ has been introduced as

$$J^{nc}_\mu = J^3_\mu - \sin^2\theta_W J^{em}_\mu\,. \tag{1.25}$$

The right handed component of J^{nc}_μ (the original problem) has been arranged to cancel with that in $\sin^2\theta_W J^{em}_\mu$ in order to leave a pure left handed J^3_μ of $SU(2)_L$, where $\sin^2\theta_W$ must be determined by experiment.

In conclusion, electromagnetic and weak neutral currents can be expressed as linear combination of J^3_μ and J^Y_μ. Electromagnetism and weak neutral current are therefore tightly bound and live in between the $SU(2)_L$ and the $U(1)_Y$ symmetries of the lagrangian.

1.3.1 Electroweak lagrangian

The lagrangian density which expresses the unified $SU(2)_L \otimes U(1)_Y$ electroweak theory may be expressed by a part containing the interaction terms shown in the previous sections and in another part containing kinematic terms. The first one can be also obtained by introducing the covariant derivatives in place of the ordinary derivatives in the massless fermion lagrangian[1] $\mathcal{L} = \sum_f \bar\psi i\gamma^\mu\partial_\mu\psi$ (where the sum is extended over all the fermions f). Covariant derivatives which preserve the local $SU(2)_L \otimes U(1)_Y$ gauge invariance have the following form, making use of the previously introduced gauge fields:

$$\mathcal{D}_\mu = \partial_\mu + ig\vec{W}_\mu \cdot \frac{\vec{\tau}}{2} - i\frac{g'}{2}YB_\mu\,. \tag{1.26}$$

After the replacement of the ordinary derivatives with the covariant ones, the lagrangian density becomes:

$$\mathcal{L} = \sum_f \bar\psi i\gamma^\mu\mathcal{D}_\mu\psi = \sum_f \bar\psi i\gamma^\mu\partial_\mu\psi + \mathcal{L}_{int}\,, \tag{1.27}$$

[1] The mass terms will be reintroduced in the following section in a gauge invariant way through the Higgs mechanism.

where the interaction term grouped the electromagnetic, charged weak current and neutral weak current terms described in the previous section

$$\mathcal{L}_{int} = \mathcal{L}_{em} + \mathcal{L}_{cc} + \mathcal{L}_{nc} \,. \tag{1.28}$$

Making use of Eq. (1.15), (1.18) and (1.24) the three interaction terms can be expressed as

$$\mathcal{L}_{em} = -ie(J^{em})^{\mu} A_{\mu} \,, \tag{1.29}$$

$$\mathcal{L}_{cc} = -i\frac{g}{\sqrt{2}}\big((J^{+})^{\mu} W^{+}_{\mu} + (J^{-})^{\mu} W^{-}_{\mu}\big) \,, \tag{1.30}$$

$$\mathcal{L}_{nc} = -i\frac{g}{\cos\theta_W}\big((J^{3})^{\mu} - \sin^2\theta_W (J^{em})^{\mu}\big) Z_{\mu} = -i\frac{g}{\cos\theta_W}(J^{nc})^{\mu} Z_{\mu} \,. \tag{1.31}$$

To complete the dynamics the second part of the lagrangian must be introduced, i.e. the kinematic terms of the gauge fields $\vec{W}_{\mu}$ and B_{μ}. The tensor associated to the gauge field $\vec{W}_{\mu}$ is

$$E^{\alpha}_{\mu\nu} = \partial_{\mu} W^{\alpha}_{\nu} - \partial_{\nu} W^{\alpha}_{\mu} - g\epsilon^{\alpha\beta\gamma} W^{\beta}_{\mu} W^{\gamma}_{\nu} \,, \tag{1.32}$$

and the corresponding lagrangian density is given by

$$\mathcal{L}_{W} = -\frac{1}{4}\vec{E}_{\mu\nu} \cdot \vec{E}^{\mu\nu} \,. \tag{1.33}$$

The kinematic terms for the gauge field B_{μ} is instead given by

$$\mathcal{L}_{B} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} \,, \qquad F_{\mu\nu} = \partial_{\mu}B_{\nu} - \partial_{\nu}B_{\mu} \,. \tag{1.34}$$

The electroweak lagrangian density, invariant under local gauge transformations of the group $SU(2)_L \otimes U(1)_Y$, can be therefore expressed as

$$\mathcal{L}_{EWK} = \sum_{f} \bar{\psi} i\gamma^{\mu}\mathcal{D}_{\mu}\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{4}\vec{E}_{\mu\nu} \cdot \vec{E}^{\mu\nu} \,, \tag{1.35}$$

where the sum is extended over all the fermions f.

1.4 Higgs mechanism

Lagrangian density described in Eq. (1.35) do not contain mass terms for the fermions. This is in striking contrast with what is observed experimentally. A mass term in such lagrangian densities indeed would break the $SU(2)_L \otimes U(1)_Y$ gauge symmetry, leading to an unrenormalizable theory which would lose all predictive power [1].

This problem can be overcome introducing new fields, known as Higgs fields, organized in doublets, whose potential is invariant under $SU(2)_L \otimes U(1)_Y$ local gauge transformations. The Higgs field lagrangian, in its global gauge invariant form, is written as

$$\mathcal{L}_{\text{Higgs}} = (\partial_\mu \phi^\dagger)(\partial^\mu \phi) - \mu^2 \phi^\dagger \phi - \lambda(\phi^\dagger \phi)^2, \tag{1.36}$$

where ϕ is a spin-$\frac{1}{2}$ spinor, μ is a complex parameter and λ is a real positive parameter. The corresponding hamiltonian density $\mathcal{H}$ is

$$\mathcal{H} = (\partial_\mu \phi^\dagger)(\partial^\mu \phi) + V\,, \tag{1.37}$$

where the potential V reads

$$V = \mu^2 \phi^\dagger \phi + \lambda(\phi^\dagger \phi)^2. \tag{1.38}$$

The fundamental state of the system is obtained through the minimization of the potential V. If $\mu^2 > 0$ the fundamental state is $\phi = 0$ and it preserves all the symmetries of the lagrangian. If $\mu^2 < 0$, the derivative of the potential with respect to $\phi^\dagger\phi$ leads to the following minimum condition:

$$\mu^2 + 2\lambda\phi^\dagger\phi = 0\,. \tag{1.39}$$

Eq. (1.39) can be satisfied in infinite different ways because a global $SU(2)_L \otimes U(1)_Y$ transformation leaves the $\phi^\dagger\phi$ product unaffected.

The lagrangian of Eq. (1.36) is not invariant under local $SU(2)_L \otimes U(1)_Y$ transformations because the derivatives do not linearly transform under a transformation depending on x^μ. In order to achieve the local $SU(2)_L \otimes U(1)_Y$ gauge invariance the ordinary derivatives must be replaced with the covariant ones of Eq. (1.26), which transform linearly under an $SU(2)_L \otimes U(1)_Y$ transformation.

The Higgs lagrangian, now $SU(2)_L \otimes U(1)_Y$ locally invariant by construction, is therefore:

$$\mathcal{L}_{\text{Higgs}} = (\mathcal{D}_\mu \phi)^\dagger (\mathcal{D}^\mu \phi) - \mu^2 \phi^\dagger \phi - \lambda(\phi^\dagger \phi)^2 - \frac{1}{4}\vec{E}_{\mu\nu} \cdot \vec{E}^{\mu\nu} - \frac{1}{4} F_{\mu\nu} F^{\mu\nu}, \tag{1.40}$$

where $\vec{E}_{\mu\nu}$ and $F_{\mu\nu}$ are defined in Eq. (1.32) and (1.34).

As in the globally invariant version of the lagrangian, if $\mu^2 < 0$ a degenerate fundamental state is found. By choosing a particular vacuum state the symmetry is broken. A fundamental state can be chosen with form

$$\phi_0 = \begin{pmatrix} 0 \\ \eta \end{pmatrix}, \qquad \eta = \sqrt{-\frac{\mu^2}{2\lambda}}. \tag{1.41}$$

Let's apply an x dependent perturbation to the vacuum state:

$$\phi = \begin{pmatrix} 0 \\ \eta + \frac{\sigma(x)}{\sqrt{2}} \end{pmatrix}. \tag{1.42}$$

Expanding the lagrangian density in Eq. (1.40) around the ground state and after the change of variables described in Eq. (1.17), (1.19) and (1.20), the following expression for the lagrangian density is obtained:

$$\begin{aligned}
\mathcal{L}_{\text{Higgs}} &= \frac{1}{2}\partial_\mu\sigma\partial^\mu\sigma + \mu^2\sigma^2 + \\
&-\frac{1}{4}A_{\mu\nu}A^{\mu\nu} + \\
&-\frac{1}{4}(W^{+\dagger}_{\mu\nu}W^{\mu\nu+} + W^{-\dagger}_{\mu\nu}W^{\mu\nu-}) + \frac{g^2\eta^2}{4}(W^{+\dagger}_{\mu}W^{\mu+} + W^{-\dagger}_{\mu}W^{\mu-}) + \\
&-\frac{1}{4}Z_{\mu\nu}Z^{\mu\nu} + \frac{g^2\eta^2}{4\cos^2\theta_W}Z_\mu Z^\mu + \\
&+ interaction\ terms,
\end{aligned} \tag{1.43}$$

The first line in Eq. (1.43) represents the Higgs boson scalar fields, with mass $\sqrt{-2\mu^2}$, the second line represents a massless field identified with the electromagnetic field, the third line represents $W^\pm$ fields, with mass $g\eta/\sqrt{2}$ and finally the fourth line represents the Z field with mass $g\eta/(\sqrt{2}\cos\theta_W)$. Thanks to the Higgs mechanism the mass terms for the vector bosons has been therefore obtained without breaking the local $SU(2)_L \otimes U(1)_Y$ gauge symmetry.

From Eq. (1.43) the mass of the vector bosons can be expressed as

$$m_W = \frac{g\eta}{\sqrt{2}}, \tag{1.44}$$

$$m_Z = \frac{g\eta}{\sqrt{2}\cos\theta_W} = \frac{m_W}{\cos\theta_W}. \tag{1.45}$$

Eq. (1.44) and Eq. (1.45) point out a dependence of the vector bosons masses from the θ_W angle and the parameter η, which is related to the G_F Fermi constant by the equation[2]:

$$\eta^2 = \frac{1}{2\sqrt{2}G_F} \simeq (174\,\text{GeV})^2. \tag{1.46}$$

It is therefore possible to estimate the W and Z masses starting from $\alpha = \frac{e^2}{4\pi}$, G_F and $\sin^2\theta_W$, the latter very precisely measured at LEP collider. Theoretical predictions show an excellent agreement with the experimental values obtained at CERN experiments SPS and LEP [9–11] and at Fermilab experiments CDF [12] and D0 [13], confirming therefore the model validity. In Table 1.2 the current measured mass values for vector bosons W e Z are shown.

Starting from Eq. (1.43) and (1.41) the mass of the Higgs boson may be expressed as

$$m_H = \sqrt{-2\mu^2} = 2\eta\sqrt{\lambda}. \tag{1.47}$$

[2]The relation between η and G_F is given by requiring the agreement of the Glashow, Weinberg and Salam model with the Fermi weak interaction theory.

	Mass (GeV)	**Full width (GeV)**
$\mathbf{m_W}$	80.399 ± 0.023	2.085 ± 0.042
$\mathbf{m_Z}$	91.1876 ± 0.0021	2.4952 ± 0.0023

Table 1.2: Current measured values of W and Z masses and full widths [14].

The Higgs boson mass depends therefore not only on the η parameter but also on the Higgs fields self coupling constant λ. The theoretical uncertainties on λ do not allow a precise prediction of the Higgs boson mass. The current limits are provided by LEP [15] and Tevatron [16] experiments:

$$\begin{aligned} m_H &> 114\,\text{GeV} \quad \text{(LEP)} \\ 158\,\text{GeV} < m_H &< 175\,\text{GeV} \quad \text{(Tevatron)} \end{aligned} \tag{1.48}$$

at the 95% of confidence level. The Higgs production at LHC follows the processes shown in Fig. 1.1, where the gluon-gluon fusion constitutes the more relevant contribution. There are also many decay channels, whose width depends on the Higgs boson mass as shown in Fig. 1.2.

1.4.1 Fermion mass terms

Now it can be demonstrated how mass terms for the fermions can arise through the interaction with the Higgs field, without breaking the $SU(2)_L \otimes U(1)_Y$ symmetry. Introducing an interaction term of the fermions with the Higgs field, in the context of spontaneous symmetry breaking, makes it possible to assign masses to the fermions.

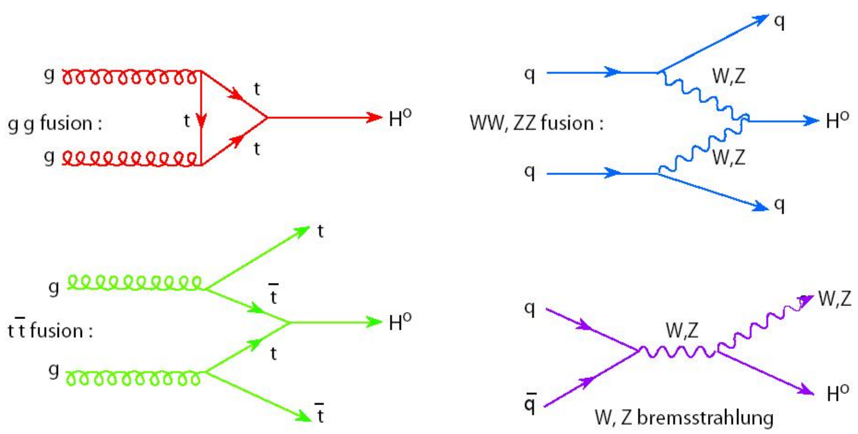


Figure 1.1: Main Higgs boson production processes.

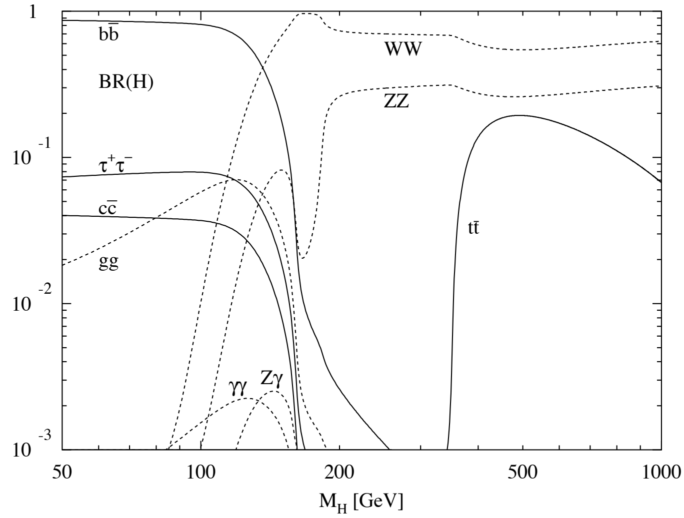


Figure 1.2: Higgs boson decay branching ratios versus boson mass.

This can be done via a Yukawa coupling with coupling constant g_f, with form:

$$\mathcal{L}_{\text{mass}}^{\text{fermion}} = -\sum_f g_f(\bar{\psi}_L^f \phi \psi_R^f + \bar{\psi}_R^f \phi^\dagger \psi_L^f)\,, \tag{1.49}$$

where ψ_L and ψ_R are fermion's helicity eigenstates.

For example Eq. (1.49) reads for the electrons:

$$\mathcal{L}_{\text{mass}}^{\text{electron}} = -g_\ell(\bar{\chi_L}\phi\chi_R + \bar{\chi_R}\phi^\dagger\chi_L)\,, \tag{1.50}$$

which can be divided in two terms by using Eq. (1.42) for the Higgs field and Eq. (1.8) for the lepton doublet:

$$\mathcal{L}_{\text{mass}}^{\text{electron}} = -g_\ell\eta(\bar{e}_L e_R + \bar{e}_R e_L) - \frac{g_\ell}{\sqrt{2}}(\bar{e}_L e_R + \bar{e_R} e_L)\sigma(x)\,, \tag{1.51}$$

g_ℓ being the Yukawa coupling constant for the lepton family. The mass term for the electron is therefore:

$$m_e = g_\ell \eta\,. \tag{1.52}$$

1.5 Strong interaction

Out of the quark model proposed by Gell-Mann [17] in 1964, the idea of the "colour" quantum number was proposed by Han and Nambu [18] in 1965 to avoid the apparent paradox that the quark model seemed to require a violation of the Pauli exclusion principle to describe hadron spectroscopy. Quantum Chromo Dynamics (QCD) was then quantized as a gauge theory with $SU(3)_c$ symmetry in

1973 by Fritzsch [19], Gross and Wilczec [20], Weinberg [21]. The colour symmetry produces new massless gauge bosons called *gluons* (in number of eight, as the $SU(3)$ generators), which mediate the strong interaction with coupling constant α_S.

QCD coupling constant α_S ranges over several orders of magnitude when moving from hard, i.e. large momentum transfer processes, to soft processes [22]. Its value grows as the momentum transfer decreases. This effect is known as "asymptotic freedom", and it justifies the use of perturbation theory (perturbative QCD or pQCD) when describing hard processes. At small energies (large distances), where the value of the coupling constant becomes large, the theory behaves in a non-perturbative way; in such a regime the isolated quark or gluon cross sections vanish and are replaced by bound state dynamics. This effect is known as "confinement" and it justifies the non-observation of free quarks and gluons.

1.5.1 QCD lagrangian

QCD is an $SU(3)_c$ gauge theory whose lagrangian is written in the following form:

$$\mathcal{L}_{QCD} = \mathcal{L}_{\text{invar}} + \mathcal{L}_{\text{gauge fix}} + \mathcal{L}_{\text{ghost}} \,. \tag{1.53}$$

$\mathcal{L}_{\text{invar}}$ is invariant under local $SU(3)_c$ transformations and reads:

$$\mathcal{L}_{\text{invar}} = \sum_f \bar{\psi}_f \big(i\gamma_\mu \mathcal{D}^\mu - m_f\big)\psi_f - \frac{1}{4} F^a_{\mu\nu} F^{\mu\nu}_a \,, \tag{1.54}$$

where f runs over the six quark fields, $\mathcal{D}^\mu$ is the covariant derivative:

$$\mathcal{D}_\mu = \partial_\mu + igA^a_\mu T_a \,, \tag{1.55}$$

and

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu - gC_{abc} A^b_\mu A^c_\nu \,, \tag{1.56}$$

where A^a_μ are the fields of the eight colored gluons, T_a are the eight generators of $SU(3)$, C_{abc} are the structure constants that define the commutation rules of the $SU(3)$ generators.

$\mathcal{L}_{\text{gauge fix}}$ and $\mathcal{L}_{\text{ghost}}$ in Eq. (1.53) are terms needed for technical reasons connected to how the quantization of the QCD lagrangian is performed [23].

1.5.2 Parton density functions

The cross section for a $pp \to N$ process at a hadronic collider is conveniently expressed as

$$d\sigma_{pp\to N} = \int_0^1 dx_1 \int_0^1 dx_2 f_1(x_1, \mu_F^2) f_2(x_2, \mu_F^2) d\hat{\sigma}_{pp\to N}(\mu^2) \tag{1.57}$$

In this expression $\hat{\sigma}$ is the parton level cross section, x_1 and x_2 are the momentum fraction of the proton momentum carried by the two colliding partons, $f_{1,2}$ are the

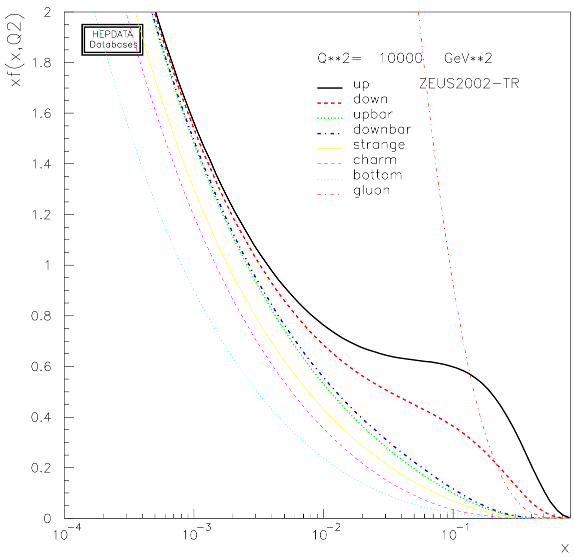


Figure 1.3: Distribution of $xf(x)$ as a function of the momentum fraction x at the electroweak scale ($\mu^2 = 10,000\,\mathrm{GeV}^2$) [24].

parton density functions (PDFs) describing the probability that a parton carries momentum fraction $x_{1,2}$ and μ is the factorization scale, as to say the scale at which the separation between the hard perturbative interaction and the long distance, non-perturbative, evolution of the produced partons takes place. The PDFs for quarks and gluons at the electroweak scale $\mu^2 = 10,000\,\mathrm{GeV}^2$ as obtained from fits based on HERA and Tevatron data are shown in Fig. 1.3.

PDFs evolution with scale is governed by the DGLAP equation [25], as long as $\alpha_S(\mu)$ remains in the perturbative validity region. DGLAP equation allows global fits of a variety of data taken from different experiments, at different scales. Two collaborations are the main provider of global PDFs fits, CTEQ [26] and MRST [27].

1.5.3 Infrared and collinear safe observables

One of the most remarkable successes of pQCD is the prediction of jet inclusive cross section in e^+e^- collisions. Jets are collimated sprays of hadrons. Even if hadrons are the result of a non perturbative process involving the quarks and gluons produced in the hard process, pQCD is very good at predicting jet cross section. The reason is that quarks and gluons are produced in the "bulk" of the process, and involve a high momentum transfer. Quarks and gluons originated in the hard process then undergo the non-perturbative hadronization process, but this happens at much lower energies (and much higher distances). Hadronization happens too late to modify substantially the topology of the event. This is an example of the "factorization" properties of QCD cross section calculation that will be shown in the next section.

Let's consider the e^+e^- collider case. The lowest order process contributing to the inclusive jet cross section is $e^+e^- \to q\bar{q}$, with a γ or a Z boson exchanged in the s-channel. The leading order cross section, in case of γ exchange is [28]

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\cos\theta} = 3\sum_q \frac{\pi\alpha^2 Q_q^2}{2s}(1+\cos^2\theta)\,, \tag{1.58}$$

where θ is the emission angle of the q line with respect to the direction along the beams, Q_q is the charge of the quark and s is the center of mass energy.

When moving to the next to leading order we have to account for possible additional gluon emission out of the quark or antiquark lines, leading to a $q\bar{q}g$ final state as shown in Fig. 1.4:

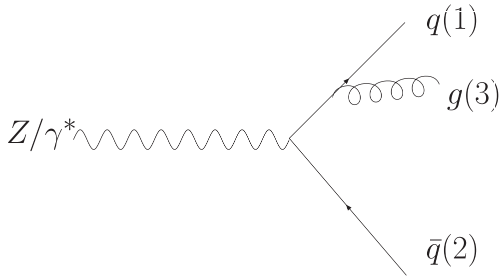


Figure 1.4: A $q\bar{q}g$ final state.

The corresponding cross section, considering quarks and gluons on the mass shell, is conveniently expressed if the energy fractions x_i are introduced as follows

$$x_i = \frac{2E_i}{\sqrt{s}}, \qquad i = 1,2,3\,, \tag{1.59}$$

where $1,2$ denote q and $\bar{q}$ respectively, and 3 denotes the gluon, as shown in Fig. 1.4. With this notation the cross section for emission of an additional gluon is

$$\frac{\mathrm{d}\sigma}{\mathrm{d}x_1\mathrm{d}x_2} \propto \alpha_S \frac{x_1^2 + x_2^2}{(1-x_1)(1-x_2)}\,. \tag{1.60}$$

Since $1-x_1 = x_2E_3(1-\cos\theta_{2,3})/\sqrt{s}$ and $1-x_2 = x_1E_3(1-\cos\theta_{1,3})/\sqrt{s}$, Eq. (1.60) is divergent when either the gluon is soft or when it is collinear with the line from which it originates.

These divergences are of course not physical, but only due to the fact that only next to leading order real emission contributions have been considered. The recipe to cure these divergences is to use a regularization procedure, introducing a cutoff: in this way divergences are replaced with large logarithms, function of the cutoff. Then, in this particular case, taking into account also the virtual diagrams completely cancels divergences at NLO.

Divergences like the ones encountered here always arise whenever a real emission line is added. If a quantity is free of such divergences it is called an *infrared-collinear safe* quantity. To make it possible to compare observables measured from experiments with theoretical prediction it is important to define observables in an *infrared-collinear safe* fashion. An observable $O(p_1, p_N)$, function of N measured momenta p_i, is infrared safe if it is unchanged by adding a soft particle, i.e.

$$O(p_1, \dots, p_N) = O(p_1, \dots, p_N, \epsilon), \quad \text{where } \epsilon^2 \text{ is small}, \tag{1.61}$$

while it is collinear safe if it is unchanged by the splitting of a four-momentum, i.e.

$$O(p_1, \dots, p_i, \dots, p_N) = O(p_1, \dots, p_{i1}, p_{i2}, \dots, p_N), \quad \text{where } p_{i1} + p_{i2} = p_i\,. \tag{1.62}$$

For what concerns jets, an infrared-collinear safe jet measurement must keep therefore unchanged the number of jets in an event if a soft particle is added or if the four-momentum of a particle is split into two. A good jet reconstruction algorithm must therefore produce infrared-collinear safe objects. Such an algorithm is called IRC-safe algorithm. In Sec. 5.1 of Chapter 5 the two main IRC-safe jet reconstruction algorithms adopted by CMS experiment will be described.

1.6 The $pp \rightarrow Z + jets$ process

The present work concerns the study of the $pp \rightarrow Z + jets$ process, and it is focused on Z bosons decaying to electrons only. The importance of studying such a process is related to the following reasons:

- In the context of the SM, the study of the production of electroweak bosons with N jets allows for tests of pQCD. The production cross section scales with the strong coupling constant for each additional jet. While current theoretical predictions at Leading Order (LO) and at Next to Leading Order (NLO) are in good agreement with data, in order to establish guidance for higher order pQCD, comparison with data for a larger number of jets is needed.

- In particular the study of the $pp \rightarrow Z + jets$ process is crucial for the understanding of the Higgs boson production background since $H \rightarrow ZZ^* \rightarrow 4\ell$ is one of the most promising channels for the Higgs boson detection in the mass range $130\,\text{GeV} - 500\,\text{GeV}$, with the exception of a small interval near $160\,\text{GeV}$ where the $H \rightarrow ZZ^*$ branching ratio has a big drop due to the opening of the WW on-shell production [24].

- The leptonic decays of W and Z accompanied by jets offer also the opportunity to search for new physics beyond the SM. Many extensions of the SM predict new particles with electroweak couplings that decay into the SM gauge bosons W, Z, and γ, accompanied by jets. Any production of new heavy particles with quantum numbers conserved by the strong interaction and electroweak couplings is likely to contribute to signatures with one or more electroweak gauge bosons. Additional jets will always be present at some level from initial-state radiation, and may also be products of cascade decays of new heavy particles.

- The $Z + jets$ channel is also a possible background for new physics processes. The $Z \to \nu\bar{\nu} + jets$ process is indeed an irreducible background to inclusive hadronic searches of Dark Matter, based on jets and missing energy [29].

- Finally, from a detector point of view, the electronic channel of Z decay allows to calibrate the electromagnetic calorimeter and to refine the track reconstruction algorithms, thanks to its clear signature.

1.6.1 Z boson production and decay

The dominant production mechanism for vector bosons in pp collisions is the weak Drell-Yan production process [30], where a quark and an antiquark annihilate to form a vector boson: the reaction $pp \to W+X$ is dominated by the annihilation of the $u\bar{d} \to W^+$ and $d\bar{u} \to W^-$ while the $pp \to Z+X$ is dominated by the annihilation of the $u\bar{u}$, $d\bar{d} \to Z$. Calculations of the total production cross sections for W and Z bosons incorporate parton cross sections, parton density functions, higher-order QCD effects, and coupling factors of the different quarks and antiquarks to the W and Z bosons.

Current cross section calculations are limited by uncertainties in PDFs, as well as higher-order QCD and electroweak radiative corrections. Due to the perturbative QCD limitations, PDFs are determined experimentally either from deep inelastic lepton-nucleon scattering experiments (like ZEUS and H1 at HERA) or from other hadron scattering experiments (e.g. Tevatron). Combining Fig. 1.3, which shows the PDF distributions at the electroweak scale, with the LHC kinematic plot shown in Fig. 1.5, one can see that over the measurable rapidity range ($|\eta| < 2.4$) gluons are the dominating partons. The scattering mostly happen therefore between *sea* quarks generated by the $g \to q\bar{q}$ splitting process.

As previously shown in Eq. (1.57), for a generic process $p_i p_j \to Z + X$, where p_i and p_j are the interacting partons of the protons, the cross section can be calculated as follows:

$$\sigma_Z = \sum_{i,j} PDF(\chi_i, \chi_j, Q^2) \otimes \sigma_{p_i p_j \to Z+X} \,, \tag{1.63}$$

where χ_i and χ_j are the proton momentum fraction that p_i and p_j are carrying respectively. The Z/γ^* inclusive production cross section as computed by FEWZ [32] package at NNLO is:

$$\sigma_Z^{incl} = (3048 \pm 132)\,\mathrm{pb}\,, \tag{1.64}$$

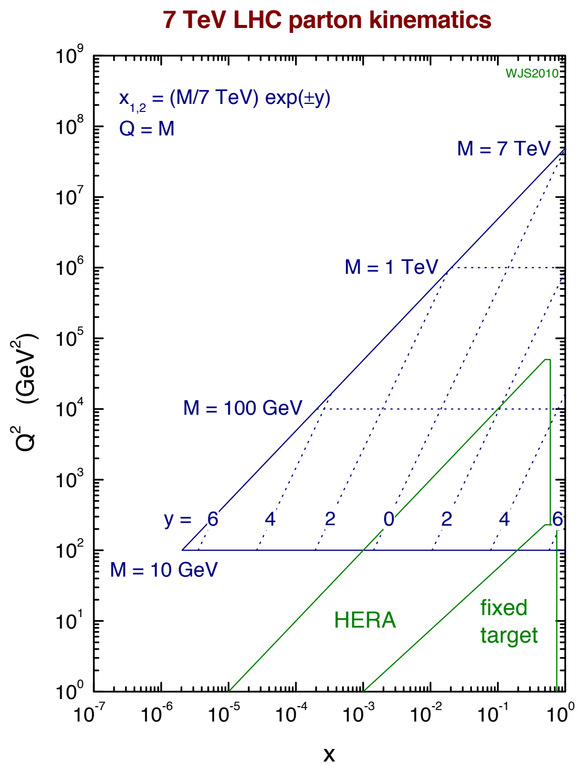


Figure 1.5: The LHC kinematic plane showing the relation between the parton variables (x, Q^2) and kinematic variables corresponding to a final state of mass M produced with rapidity y at LHC for 7 TeV collision energy [31].

if the mass of the boson is restricted to be above 50 GeV and only leptonic decay channels are considered. The W and Z production cross sections times the electronic branching ratio as a function os the collider energy are shown in Fig. 1.6, along with experimental observations from previous experiments.

The Z boson decays hadronically with a branching ratio of almost 70%. Invisible decays to neutrinos account for 20% of the decays while the remaining are leptonic decays to electrons, muons and taus, in almost equal amounts. In Table 1.3 a detail list of the Z decay branching ratios is shown. As stated before this study is focused on the electronic channel, which account for approximately 3.4% of the total branching ratio.

1.6.2 Associated jet production

The majority of Z bosons are produced at rest in the transverse plane or with very little transverse momentum. Production of hard outgoing partons (jets) gives a transverse momentum to the Z boson, leading to more complex and interesting events as stated at the beginning of this Section.

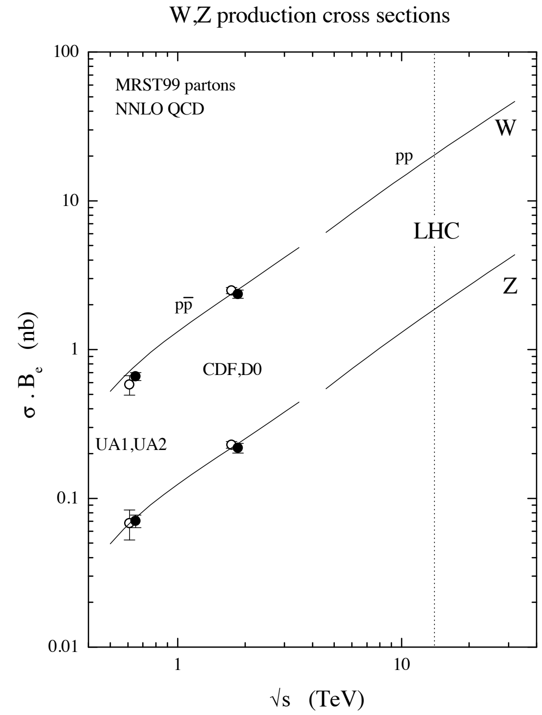


Figure 1.6: Predictions fro the total W, Z production cross section times electronic branching ratio in $p\bar{p}$ and pp collisions as a function of the collider energy $\sqrt{s}$. Experimental measurements from UA1, UA2, CDF and D0 experiments are also shown [33].

Decay Mode	Branching Ratio
e^+e^-	$(3.363 \pm 0.004)\%$
$\mu^+\mu^-$	$(3.366 \pm 0.004)\%$
$\tau^+\tau^-$	$(3.370 \pm 0.008)\%$
invisible	$(20.00 \pm 0.06)\%$
hadrons	$(69.91 \pm 0.06)\%$

Table 1.3: Summary of Z decay modes [14].

The number of processes contributing to these final states increases as the number of jets goes up. For example, while there are only 9 processes contributing to $Z + 1\,jets$ events, for $Z + 4\,jets$ events there are 485 tree-level processes [34]. For high Q^2 the value of the strong coupling constant is $\alpha_s \ll 1$, therefore multiple exchanges involving terms in $\alpha_s^2, \alpha_s^3, ...$ have decreasing cross sections. Feynman diagrams of Z production in association with 1 and 2 outgoing jets in the final state are shown in Fig. 1.7, the first originating from a quark-gluon interaction and the second from a gluon-gluon interaction. Due to the difficulty in calculating the ex-

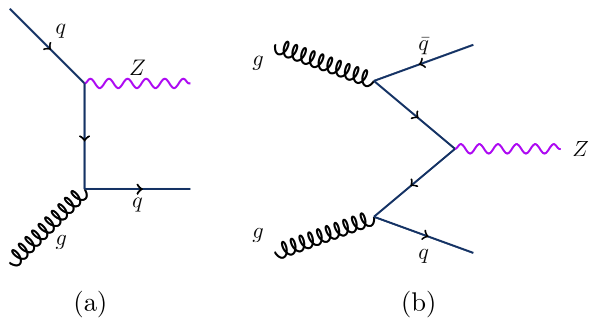


Figure 1.7: Quark-gluon (a) and gluon-gluon (b) Z production Feynman diagrams, with 1 and 2 jets in the final state respectively.

tra loops involved in higher order calculations, it is very difficult to calculate cross sections even at NLO. Current state-of-the-art calculations go up to $Z + 3\,jets$ at NLO [35]. Precise measurements allow to validate pQCD predictions and to help to properly model and constrain background for other searches.

Chapter 2

The CMS experiment at LHC

The Large Hadron Collider (LHC) [36–39] is the most powerful collider ever built. It started its activity on October 21st 2008 and it will run for the next two decades. LHC investigates processes with really tiny cross sections, down to a few femtobarns.

The two main reasons that drove the choice of a hadron collider instead of an electron collider like LEP [9–11] are the possibility to reach center of mass energies much higher than LEP and the wider energy spectrum that can be explored by a hadron machine. The first goal can be achieved thanks to the lower amount of synchrotron radiation emitted by circulating hadrons. Compared to electron, the energy loss is reduced by a factor of $\mathcal{O}(m_e/m_p)^4$. The second issue is desirable for a machine involved in the discovery of new physics, and can be achieved thanks to the composite nature of protons, despite the production of many low energies particles in a complex environment.

In this chapter the main characteristics of LHC are briefly reviewed, and in particular the LHC experiment this work is involved in, i.e. the Compact Muon Solenoid (CMS) [40], will be described.

2.1 The Large Hadron Collider

The LHC accelerator has been installed in the underground tunnel which housed the LEP electron-positron collider until 2000. Two counter circulating proton beams flow in the 27 km LHC ring, placed near Geneva at a depth varying from 50 m to 175 m. LHC is projected to produce 7 TeV proton beams, resulting in an energy in the center of mass frame equal to 14 TeV. This energy is 70 times the energy reached by LEP and 7 times the energy reached by Tevatron [41], the hadron collider placed at Fermilab of Chicago (USA) which was the most powerful accelerator before the LHC. Also Lead nuclei may be accelerated by LHC, at the energy of 2.76 TeV/nucleon in the center of mass frame.

The acceleration is performed in several stages. The proton injection starts at the duo-plasmatron, which is the proton source. A linear accelerator (LINAC)

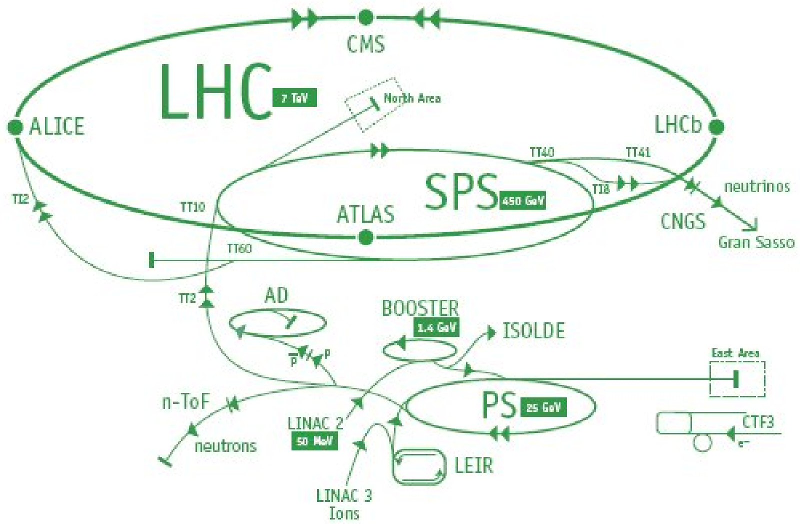


Figure 2.1: A schematic view of the LHC accelerator complex.

boosts the protons to energy of 750 KeV using Radio Frequency Quadruples. A $30\,\mu$s pulse is then ejected into the Proton Synchrotron Booster (PSB) which increases the energy to 1.4 GeV . The LHC bunch train with 25 ns spacing starts in the Proton Synchrotron (PS), where the energy increases to 25 GeV . Then the protons are accelerated up to 450 GeV by the Super Proton Synchrotron (SPS), and they are finally injected into the LHC. In the LHC ring, the acceleration continues until the protons reach the energy of 3.5 TeV (7 TeV in the future). A schematic description of the LHC accelerator complex and its services is shown in Fig. 2.1.

The two beams collide in four interaction points, where the four main experiments are built. Two general purpose experiments, called ATLAS [42] and CMS [40], perform general Standard Model measurements and seek for new physics; one experiment called LHCb [43] is dedicated to B meson physics and it will carry out precise measurements of CP violation; one experiment called ALICE [44] investigates heavy ion physics. This collider is capable of investigating mass scales from the order of a few GeV, as in the case of B meson physics, up to a few TeV, for the discovery of new vector bosons or quark compositeness. In order to extend the LHC capability to explore new physics rare processes an enormous effort has been made to raise the proton momentum as much as possible. In particular, a very sophisticated magnet system is needed to keep such high momentum protons within the machine orbit. The magnetic field needed to keep the protons on a circular orbit is given by Eq. (2.1):

$$B[T] = \frac{p[\,\mathrm{GeV}\,]}{0.3\rho[m]}\,, \tag{2.1}$$

where B is the magnetic field (expressed in Tesla), p is the momentum (expressed in GeV) and ρ is the orbit radius (expressed in meters). For a circumference of about 27 km, the magnetic field needed for 7 TeV protons is about 5.4 T. Actually, since LHC is made of curved and rectilinear sections, the bending magnetic superconductor dipoles have to produce an 8.3 T magnetic field. This value is close to the technological edge for superconducting magnets nowadays. To reach this huge magnetic field intensity the dipoles are kept below 1.9 K by a superfluid helium cooling system.

Since the beam energy is limited by the bending power of the magnetic system and by the circumference of the machine, another handle to raise the rate of interesting and rare events is the luminosity $\mathcal{L}$. The event rate n for a process with cross section σ is

$$n = \mathcal{L}\sigma \, . \tag{2.2}$$

The luminosity is connected to the beam properties with the following approximated formula [14]:

$$\mathcal{L} = f \frac{n_1 n_2}{4\pi\sigma_x\sigma_y} \, , \tag{2.3}$$

where n_1 and n_2 are the number of particles in beam 1 and 2 respectively, f is the collision frequency, σ_x and σ_y are transverse dimensions of the beams. At the LHC nominal luminosity the proton bunches will collide at a frequency of about 40 MHz, corresponding to a spatial separation between bunches of about 7.5 m. The transverse dimensions of the beam can be squeezed down to 15 μm.

The need for such a high luminosity has driven the choice of a proton-proton collider, instead of a proton-antiproton. In fact, even if a proton-antiproton machine has the advantage that both beams can be kept in the same beampipe, to produce the number of antiprotons needed to reach the desired luminosity is an unfeasible task.

In the hard proton proton collision, with high transferred momentum, the center of mass energy $\sqrt{\hat{s}}$ is connected to the total center of mass energy $\sqrt{s}$ as:

$$\sqrt{\hat{s}} = \sqrt{x_1 x_2 s} \, , \tag{2.4}$$

where x_1 and x_2 are the energy fractions of the two partons participating in the hard scattering.

The center of mass of the two hardly interacting partons is not motionless in the experiment frame, but rather it is on average boosted along the direction defined by the colliding beams. For this reason boost invariant observables ave very important to characterize the event. One of such observables is the transverse momentum p_T, defined as the projection of the momentum vector on a plane perpendicular to the beam axis. Lorentz boost indeed does not transform coordinates in the plane orthogonal to the boost direction.

Another useful observable is the rapidity y, defined as

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z} = \tanh^{-1} \left(\frac{p_z}{E} \right) , \tag{2.5}$$

where E is the particle energy and p_z the projection of particle momentum along the beam direction. Under a boost along z with speed β, y undergoes the following transformation: $y \to y - \tanh^{-1}\beta$, hence rapidity differences are invariant and therefore also the shape of the rapidity distribution $\mathrm{d}N/\mathrm{d}y$ is invariant. In the ultrarelativistic approximation the rapidity y is the same as the pseudorapidity η defined as

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right). \tag{2.6}$$

It is often useful to refer to pseudorapidity as it depends only on the direction of the three-vector.

2.1.1 The LHC schedule

On September 10th 2008 the LHC circulated for the first time a single proton beam, constituted by a single proton bunch. During a test performed on September 19th 2008 an accident occurred in sector 3 and 4: a quench in about 100 bending magnets caused a loss of approximately six tonnes of liquid helium, which was vented into the tunnel, and a temperature rise of about 100 K in some of the affected magnets. The accident forced a shutdown of about one year, in order to repair the damaged sectors and to check all the system. Therefore a revision of the plans for the first years of running was also necessary. In particular the beam energy and luminosity have been lowered for the first year of running. On November 20th 2009 LHC restarted its activity with one beam circulating at 450 GeV of energy, and on November 23rd for the first time two counter circulating beams flew around the LHC ring and the first collisions at 900 GeV in the center of mass frame were observed. On November 30th 2009, with proton beams at 1.18 TeV, LHC has become the most powerful accelerator ever built, and on December 16th the first collisions at $\sqrt{s} = 2.36\,\mathrm{TeV}$ were observed. After a few months of stop, a new plan for the first period of running was approved at the Chamonix conference on February 2010. The main decision taken in that context was to run for about the first 24 months at 3.5 TeV per beam, i.e. with collisions at 7 TeV in center of mass frame.

On March 30th 2010 LHC started its physics program with the first collisions at 3.5 TeV per beam. The luminosity has been increased during all 2010, starting from a value of $5\cdot10^{27}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ up to about $2\cdot10^{32}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ at the end of the running period planned for the 2010, i.e. November 3rd. In Fig. 2.2 the instant luminosity produced during all the 2010 is shown. The integrated luminosity collected by the CMS experiment for the first year of activity is shown in Fig. 2.3, and it corresponds to a total amount of about $47\,\mathrm{pb}^{-1}$ delivered to CMS and $43\,\mathrm{pb}^{-1}$ actually recorded by the experiment.

In the fall of 2010 LHC produced also its first heavy ion collisions in center of ALICE, ATLAS and CMS experiments. From November 8th up to December 6th, Pb ion collisions took place indeed in the LHC ring at an energy of 574 TeV in

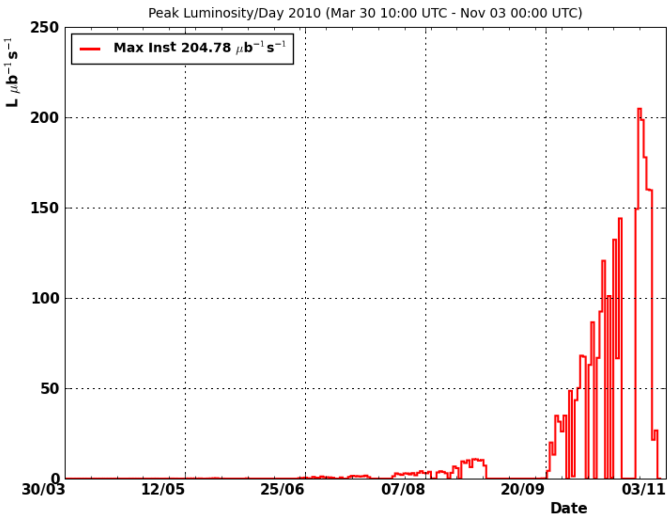


Figure 2.2: Maximum instantaneous luminosity per day delivered to CMS during stable beams at 7 TeV centre-of-mass energy, expressed in $\mu b^{-1} \cdot s^{-1}$. The plot is related to the 2010 data taken period.

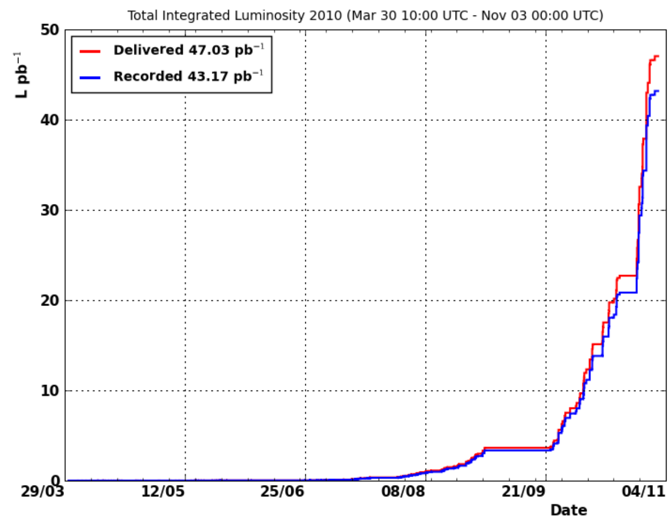


Figure 2.3: Integrated luminosity versus time delivered to (red), and recorded by CMS (blue) during stable beams at 7 TeV centre-of-mass energy. The plot is related to the 2010 data taken period and the integrated luminosity is expressed in pb^{-1}.

Circumference	26.659 km
Maximum dipole field	8.33 T
Magnet temperature	1.9 K
Beam energy at injection	450 GeV
Beam energy at collision (nominal)	7 TeV
Beam energy at collision in 2010	3.5 TeV
Maximum luminosity (nominal)	$10^{34}\text{cm}^{-2}\text{s}^{-1}$
Maximum luminosity reached in 2010	$2.07 \cdot 10^{32}\text{cm}^{-2}\text{s}^{-1}$
Number of proton bunches (nominal)	2808
Maximum number of proton bunches in 2010	368
Maximum number of colliding proton bunches in 2010	348
Number of proton per bunch	10^{11}
Bunch separation in time (nominal)	24.95 ns
Bunch separation in time in 2010	150 ns
Collision frequency (nominal)	40.08 MHz
Crossing angle	300 μrad
Bunch length (r.m.s.)	7.5 cm
Transverse beam size at impact point	15 μm
Energy loss per turn (at 14 TeV)	7 keV
Total radiated power per beam (at 14 TeV)	3.8 kW
Stored energy per beam (at 14 TeV)	350 MJ

Table 2.1: LHC technical parameters for proton-proton collisions. For some parameters both the nominal and the startup values are shown.

center of mass frame, i.e. 2.76 TeV per nucleon, for a total integrated luminosity recorded by CMS of about $8.7\mu\text{b}^{-1}$.

In Table 2.1 some LHC technical parameters are shown.

2.2 The CMS detector

The Compact Muon Solenoid experiment (CMS) [40] is a general purpose LHC experiment. Its main feature is the 4T superconducting solenoidal magnet; such a strong magnetic field permits a compact design of the apparatus. The main design priorities of CMS were a redundant muon tracking system, a good electromagnetic calorimeter and a high quality inner tracking system.

The structure of CMS is typical of general purpose collider detectors. It consists of several cylindrical detecting layers, coaxial with the beam direction (*barrel* region), closed at both ends with disks (*endcap* region). Figs. 2.4 and 2.5 show two schematic views of the CMS detector, that has a full lenght of 21.6 m, a diameter of 15 m, and a total weight of 12500 tons.

The coordinate frame used in CMS is a right-handed triad, with the x axis pointing towards the LHC centre, y axis directed upward along the vertical and z axis along the beam direction with the direction required to complete the right-handed

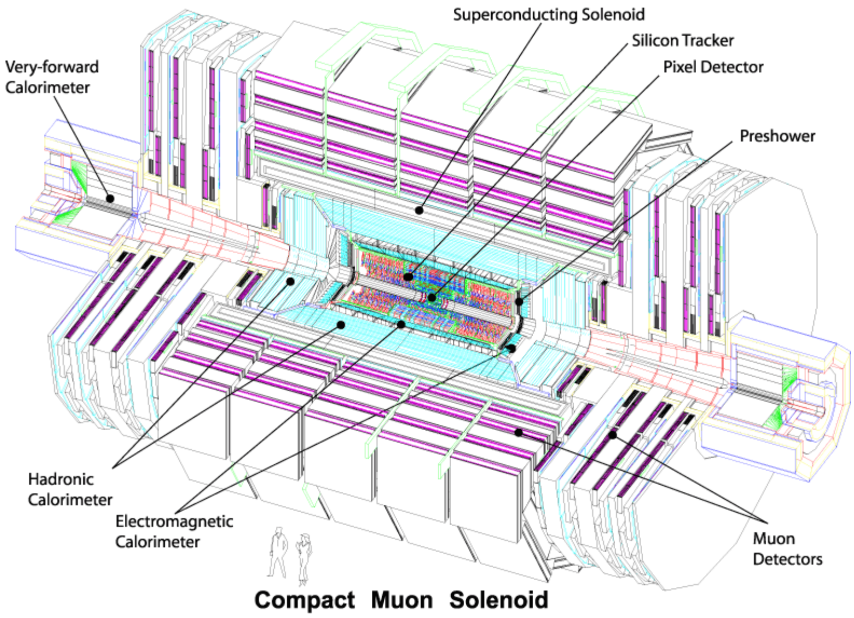


Figure 2.4: A view of the CMS detector with its subdetectors labeled.

triad. The cylindrical symmetry of CMS design and the invariant description of proton-proton physics suggest the use of a pseudo-angular reference frame, given by the triplet (r, ϕ, η), where r is the distance from the z axis, ϕ is the azimuthal angle, measured starting from the x axis positive direction, η is defined in Eq. (2.6).

CMS is made up of four main subdetectors:

- **Silicon Tracker: $r < 1.2$ m, $|\eta| < 2.5$**. It is made of a Silicon Pixel vertex detector and a surrounding Silicon Microstrip detector, with a total active area of about 215 m^2. It is used to reconstruct charged particle tracks and vertices.

- **ECAL: 1.2 m $< r <$ 1.8 m, $|\eta| < 3$**. It is an Electromagnetic Calorimeter to precisely measure electrons and photons.

- **HCAL: 1.8 m $< r <$ 2.9 m, $|\eta| < 5$**. It is a Hadronic Calorimeter for jet direction and energy measurement.

- **Muon System: 4 m $< r <$ 7.4 m, $|\eta| < 2.4$**. It is a composite tracking system for muons. It consists of Drift Tubes (DT) in the barrel region and Cathode Strip Chambers (CSC) in the endcaps. A complementary system of Resistive Plate Chambers (RPC) is used both in the barrel and in the endcaps.

The Silicon Tracker, ECAL and HCAL are located inside the magnetic coil. Muon Chambers are located in the magnet return yoke. In the following sections a brief description of each component is given.

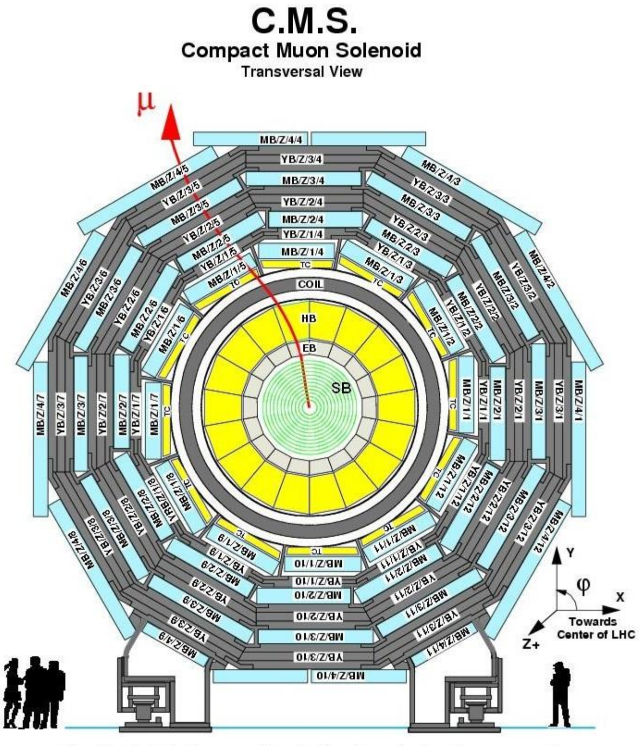


Figure 2.5: A transverse view of the CMS detector.

2.2.1 The Solenoid

The CMS magnet [45], which houses the tracker, the electromagnetic and the hadronic calorimeters, is the biggest superconducting solenoid ever built in the world. The solenoid achieves a magnetic field of 3.8 T in the free bore of 6 m in diameter and 12.5 m in length. The energy stored in the magnet is about 2.6 GJ at full current. The superconductor is made of four Niobium-Titanium layers. In case of a quench, when the magnet looses its superconducting property, the energy is dumped to resistors within 200 ms. The magnet return yoke of the barrel has 12-fold rotational symmetry and is assembled of three sections along the z-axis; each is split into 4 layers (holding the muon chambers in the gaps). Most of the iron volume is saturated or nearly saturated, and the field in the yoke is around the half (1.8 T) of the field in the central volume.

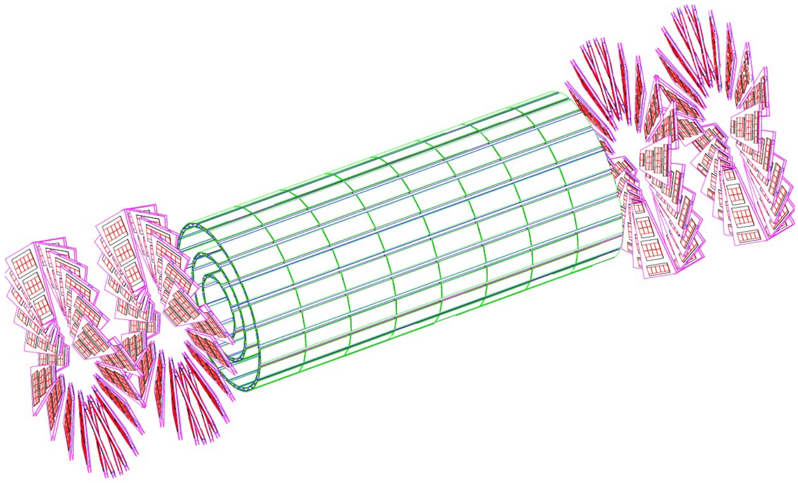


Figure 2.6: A schematic view of the pixel vertex detector.

2.2.2 The Tracker

The Silicon Tracker [46,47] is the CMS innermost detector and covers the region $|\eta| < 2.5$, $r < 120$ cm. It consists of a Silicon Pixel detector and a surrounding Silicon Microstrip detector.

Its goal is to provide a precise momentum estimate for charged particles, and to allow a precise determination of the position of secondary vertices. LHC events are very complex, and track reconstruction comes as a complex pattern recognition problem. In order to ease pattern recognition two requirements are fundamental:

- low detector occupancy,
- large hit redundancy.

The low hit occupancy is achieved with a highly granular detector, while the redundancy is achieved with a large number of detecting layers.

In the following the two Tracker subdetectors are explained.

The Pixel Vertex detector

The pixel detector [48] is a fundamental device for impact parameter measurements. It is also extremely important as a starting point in reconstructing charged particle tracks. It covers the region $|\eta| < 2.5$ and it is organized into three 53 cm long barrel layers, positioned at $r = 4.4, 7.3$ and 10.2 cm, and two disks per each side, placed at $z = \pm 34.5$ cm and ± 46.5 cm covering radii between 6 and 15 cm to guarantee at least two crossed layers per track coming from the centre of the detector within the fiducial angle $|\eta| < 2.5$. The overall number of readout channels is about 60 millions. A schematic view of the pixel detector is shown in Fig. 2.6.

Each layer is composed with modular detector units, containing a 250μm thin segmented sensor plate with highly integrated readout chips. Since both $r\phi$ and

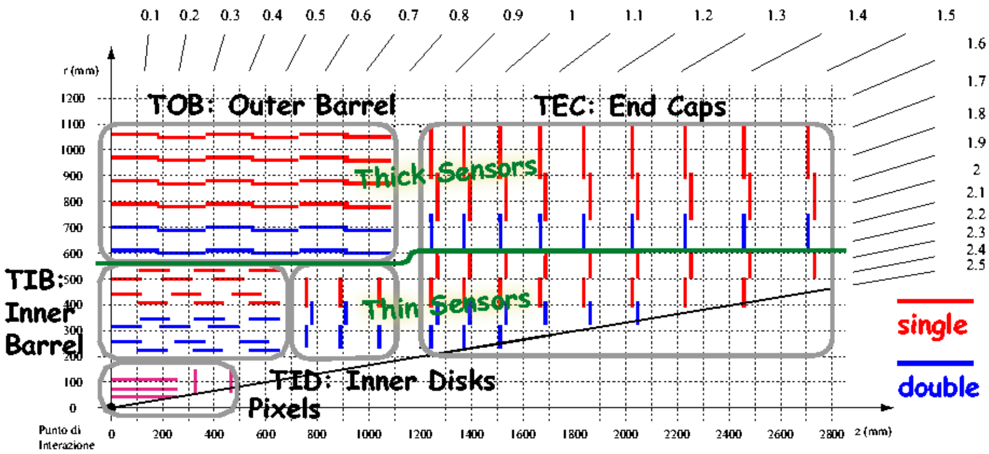


Figure 2.7: An $r - z$ schematic view of a sector of the Silicon Strip Tracker. The location of single sided and double sided detectors is put into evidence.

z coordinates are important for vertex finding and impact parameter resolution, a rectangular pixel shape has been chosen to optimize both measurements. The pixels consist of n^+ implant over a n-type substrate sensor, with a size of $100 \times 150\,\mu\text{m}^2$ and are combined with analog signal readout to profit of charge sharing effects among pixels and improve position resolution by interpolation. The charge sharing between pixels is enhanced by the Lorentz drift of charge carriers, improving in this way the intrinsic hit resolution down to $10 - 15\,\mu$m in the transverse plane, far below the $150\,\mu$m width of each n^+ implant. The detectors placed on the disks are rotated with an angle of $20°$ around the central radial axis to benefit of charge sharing improved both in r and $r\phi$ directions by induced Lorentz effects.

The Silicon Microstrip detector

The Silicon Strip Tracker (SST) is made of ten barrel layers and twelve endcap disks on each side. It has about 10 millions readout channels. The SST covers a tracking volume up to $r = 1.1$ m with a length of 5.4 m and is divided in four parts, as it is shown in Fig. 2.7: TIB (*Tracker Inner Barrel*), TID (*Tracker Inner Disks*), TOB (*Tracker Outer Barrel*) and TEC (*Tracker EndCap*). As indicated in Fig. 2.7 some of the layers are equipped with *single sided* detectors, some with *double sided* detectors. Single sided detectors can provide the particle's impact point position in the direction perpendicular to the strips. Double sided detectors can provide both coordinates on the detector surface, as they are made with two single sided detectors glued back-to-back with an angle of 100 mrad between the strips directions. Inner layers are equipped with 300 μm thick sensors, while outer layers are equipped with 500 μm thick sensors, for a total of 15.148 sensors.

The high flux of radiation through the tracker sensors causes damages. Pixel and microstrip detectors and readout electronics are radiation hard. Nevertheless, the pixel detector, which is exposed to the highest flux per unit area, will need to be replaced at least once during LHC lifetime. In order to limit the effect of radiation damage on sensor performances the tracker is meant to be run at low temperature (-10°C).

The material budget in the tracker is limited as possible, as the electron energy loss due to bremsstrahlung and nuclear interactions of hadrons need to be kept as low as possible. This is needed so as not to spoil tracking performances and to keep the number of photons that get converted into an e^+e^- pair through interaction with the material as low as possible. The tracker depth in terms of radiation length X/X_0[1] and in terms of interaction length λ/λ_0[2] as obtained from the full simulation of the tracker is shown in Fig. 2.8 as a function of η. The material budget is higher in the region $1 < |\eta| < 2$ due to the presence of cables and services in this region.

The alignment of the tracker modules is very important to obtain a high spatial resolution. Deviations are caused by assembly inaccuracy, deformations due to cooling and stress from the magnetic field. Therefore, three methods are used for the tracker alignment. The geometry was determined during assembly to an accuracy of 80 to 150 μm. An infrared laser system is used for continuous monitoring of the position of selected tracker modules. The final alignment is done with tracks from well known physics processes, e.g. cosmic muons, or di-muons from J/Ψ, Υ and Z^0.

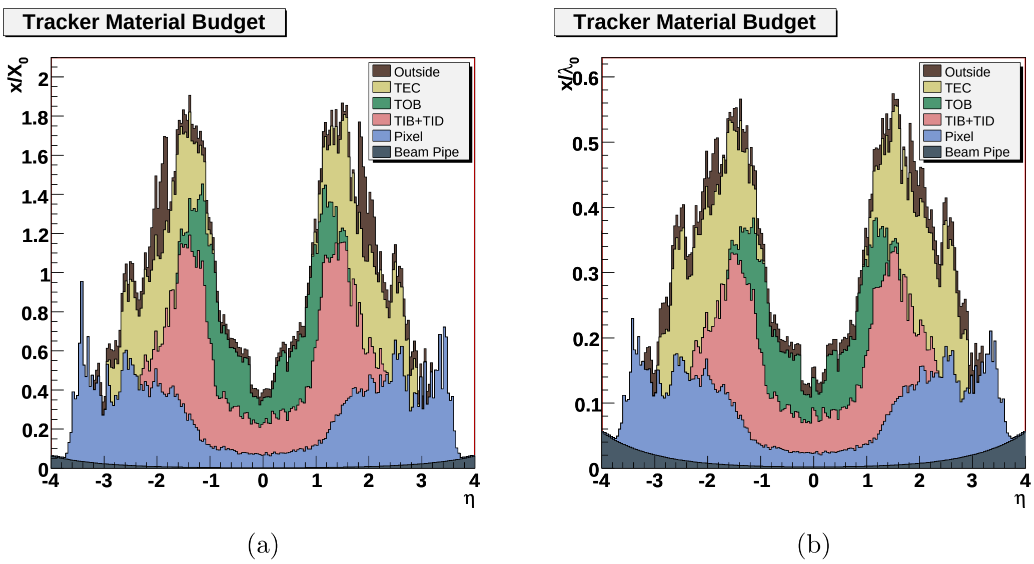


Figure 2.8: (a) Radiation length and (b) interaction length of the tracker as a function of η. Contributions from different components are put into evidence.

[1] X_0 is the distance over which a high energy electron reduces its energy to a fraction $(1-1/e)$ of the initial energy.

[2] λ_0 is the mean free path of a hadron before having an interaction when traversing a material.

2.2.3 The Electromagnetic Calorimeter (ECAL)

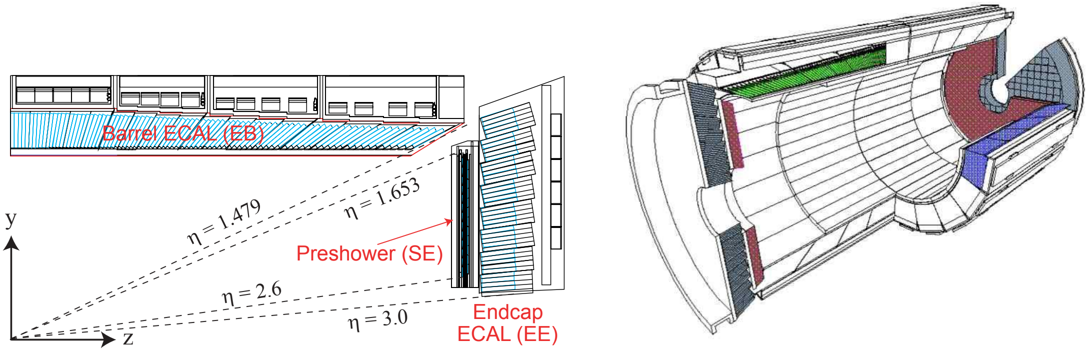
The main goal of an electromagnetic calorimeter is to identify electrons and photons and to precisely measure their energy. The CMS *Electromagnetic CALorimeter* (ECAL) [49, 50] is a homogeneous calorimeter with cylindrical geometry and consists of 75848 Lead Tungstate ($PbWO_4$) scintillating crystals, divided into an *Ecal Barrel* (EB) with 61200 crystals and two *Ecal Endcaps* (EE), each containing 7324 crystals. In Fig. 2.9 a schematic representation of ECAL is shown.

The *barrel* inner radius is 129 cm, while the length is 630 cm and the η extent is $|\eta| < 1.479$. It consists of 36 *supermodules*, each one with a length equal to the half of the barrel length. Each supermodule consists of a matrix of 20×85 crystals in the (ϕ, η) plane, covering an azimuthal angle of 20°. Supermodules are divided into 4 modules along the η direction, and each module is in turn divided in *submodules*. Submodules are the basic units of ECAL and consist of a matrix of 5×2 crystals. Crystals in the barrel region are tapered shaped, with a $2.2\,\mathrm{cm}\times 2.2\,\mathrm{cm}$ front face and 23 cm length, and they are positioned at a radius of 1.24 m. The $\Delta\eta \times \Delta\phi$ granularity in the barrel is 0.0175×0.0175. The depth in radiation lengths in the barrel region is about 26 X_0. The crystals are grouped in 5×5 matrices called *trigger towers*, providing information to the trigger system. To avoid that cracks might align with particle trajectories, the crystal axes are tilted by 3° with respect to the direction from the interaction point (IP), both in ϕ and in η.

Each *endcap* covers the region $1.479 < |\eta| < 3$ and consists of two halves called *Dees*. All the crystals have the same shape ($22 \times 2.47 \times 2.40$ cm^3) and they are grouped in structures of 5×5 crystals called *supercrystals*. The $\Delta\eta \times \Delta\phi$ granularity in the endcaps varies from 0.0175×0.0175 to 0.05×0.05. Unlike in the barrel, where the crystals are arranged in a $\eta - \phi$ geometry, the endcap crystals are arranged in a $x - y$ geometry. To ensure good hermeticity, the outer perimeter of the ECAL endcaps has been studied in order to give an overlap of half crystal between the barrel and the endcaps. Moreover, in order to avoid the presence of gaps pointing to the interaction point, the crystal axes are oriented to point 1300 mm beyond the IP. On the inner side of the endcaps two *preshower* detectors are placed, in order to improve the π^0/γ separation and the vertex identification. The preshower, which covers the region $1.653 < |\eta| < 2.6$, is a sampling calorimeter consisting of two lead converters (2 X_0 and 1 X_0 thick respectively) followed by silicon strips with a pitch of less than 2 mm. The strips following the two absorbers are disposed in orthogonal way. The presence of a preshower (a total of 3 X_0 of lead) in the endcap region allows the use of slightly shorter crystals (22 cm), keeping the total radiation length more than 26 X_0.

Different reasons brought to the choice of the $PbWO_4$ as active medium for ECAL. Its short radiation length ($X_0 = 0.89$ cm) and Moliere Radius[3] ($R_M = 2.19$

[3]The Moliere Radius (R_M) gives an estimate of the transverse development of an electromagnetic shower: on average 90% of the energy released by the shower lies in a cylinder with radius



(a) (b)

Figure 2.9: Schematic representation of ECAL: (a) $r - z$ plane projection, (b) 3D view.

cm) allow to build a compact and high granularity calorimeter. An important aspect is the fast response ($\sim$ 80% of the light is collected within 25 ns), which is compatible with the high LHC rate. Finally, the $PbWO_4$ has a good intrinsic radiation hardness, which makes it suitable to work in the hard LHC environment. The main drawback of the $PbWO_4$ crystals is the low light yield ($\sim$ 10 photo-electrons/MeV), which makes an internal amplification for the photodetectors necessary. The photodetectors for ECAL have to be radiation hard, fast and able to operate in the strong CMS magnetic field. The devices which match these characteristics and that have been chosen for the electromagnetic calorimeter are the *Avalanche PhotoDiodes* (APDs) [49, 51] for the barrel and the *Vacuum PhotoTriodes* (VPTs) [49, 52] for the endcaps.

The energy resolution of a homogeneous calorimeter is usually written as

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c\,, \tag{2.7}$$

where a, b and c represent the *stochastic*, *noise* and *constant* terms of the energy resolution respectively. The values of the a, b and c parameters obtained from test beam measurements are [53]

$$\frac{\sigma_E}{E} = \frac{2.8\%\,[\mathrm{GeV}^{1/2}]}{\sqrt{E}} \oplus \frac{0.12\,\mathrm{GeV}}{E} \oplus 0.3\%\,, \tag{2.8}$$

where the energy E is expressed in GeV. The *stochastic term* a includes the contribution of the fluctuations in the number of electrons which are produced and collected. The fluctuations are poissonian and this term takes into account the

R_M .

crystal light emission, the light collection efficiency and the quantic efficiency of the photo-detector.
The *noise term b* includes contributions from the electronic noise, both due to the photodetector and to the preamplifier, and from pile-up events. The contributions change at the different pseudorapidities and with the luminosity of the machine.
The *constant term c* is the dominating term at high energies and it includes many different contributions. Among them, the most important are:

- the stability of the operating conditions, such as the temperature and the high voltage. Both the scintillation mechanism and the APD gain are affected by the temperature and the response for a given energy deposit varies with the temperature of the calorimeter with a slope which is around $-4\%/°C$ for the barrel. The stability of the temperature within 0.05°C is required to keep the contribution to the constant term below 0.1%,

- the presence of dead materials in front of the crystals and the rear leakage of the electromagnetic shower. Anyway, the 25 radiation length featured by the crystals keeps this effect negligible up to energies in the TeV range,

- the longitudinal non uniformity of the crystal light yield. A strong focusing effect of the light takes place due to the tronco-pyramidal shape of the crystals and to the high refractive index, so the light collection is not uniform. The light collection was corrected by mechanically abrading one lateral face of each crystal to an extent able to keep non-uniformity at the level of $0.35\%/X^0$, keeping the contribution to the constant term c below 0.3%,

- the intercalibration errors, which is the dominant contribution,

- the radiation damage of the crystals, which changes their response to a certain amount of deposited energy when exposed to high radiation dose rates.

2.2.4 The Hadronic Calorimeter (HCAL)

The CMS *Hadronic CALorimeter* (HCAL) [54], together with the electromagnetic calorimeter, makes a complete calorimetric system for the jet energy and direction measurement. Also, thanks to its hermeticity, it can provide a precise measurement of the missing transverse energy E_T^{miss}.
HCAL is a sampling calorimeter and covers the region $|\eta| < 5$. It is divided in four subdetectors: the *Barrel Hadronic Calorimeter* (HB) and the *Endcap Hadronic Calorimeter* (HE), both placed inside the magnetic coil, the *Outer Hadronic Calorimeter* (HO, or *Tail Catcher*), placed in the barrel region outside the magnetic coil, and finally the *Forward Hadronic Calorimeter* (HF), consisting of two units placed in the very forward region outside the magnetic coil. In Fig. 2.10 a schematic view of HCAL is shown.

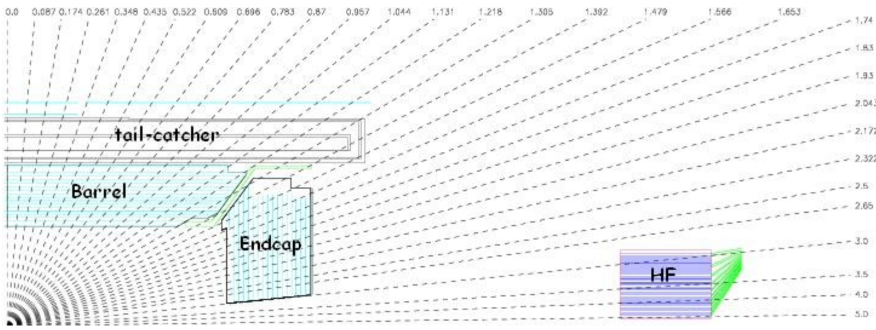


Figure 2.10: A schematic $r - z$ view of a quadrant of the CMS hadronic calorimeter HCAL.

In order to maximize particle containment for a precise missing transverse energy measurement, the amount of absorber material inside the magnetic coil was maximized, reducing therefore the amount of the active material. Since HCAL is mostly placed inside the magnetic coil, a non-magnetic material like brass was chosen as absorber. HB and HE are therefore made with brass absorber layers interleaved with plastic scintillators (*Wavelength shifters*, WLS) coupled to transparent optical fibers, which transmit the light to the HPD (*Hybrid PhotoDiodes*) photodetectors.

Here is a detailed description of the four HCAL subdetectors.

- **HB**: with a length of 9 m, it extends in the region of radius 178 cm $< r <$ 288 cm and pseudorapidity $|\eta| < 1.4$, surrounding ECAL. It consists of 2304 calorimetric towers, with a granularity of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. Each tower consists of 15 brass layers, with a thickness of 50 mm each, arranged in the direction parallel to the beam. At the edges there are 2 steel layers in order to guarantee structural stability. Interleaved with the absorber layers there are 17 plastic scintillator layers of 3.7 mm thickness, except for the inner one which is 9 mm thick. The optical fibers of each tower are all sent to the same photodetector.

- **HE**: it extents in the pseudorapidity region $1.3 < |\eta| < 3.0$, partially overlapping with HB. The empty region between the two detectors is used for the services and it does not point toward the interaction point, in order to preserve the calorimeter hermeticity. The η segmentation varies from 0.87 for the towers at lower η to 0.35 for the towers close to the beam line. For each endcap there are 2304 calorimetric towers, which consists of 19 plastic scintillator layers of 3.7 mm thickness, interleaved with 78 mm thick brass absorber layers.

- **HO**: since the HCAL Barrel dimensions are limited by the presence of the solenoid, the HO detector is added in the barrel region, outside the magnetic coil, in order to improve the energy measurement of the most energetic hadronic showers. It consists of several plastic scintillator layers, which increase the actual calorimetric thickness to more than 10 interaction lenghts λ_0, thus reducing the tails of the energy resolution distribution. The scintillators in the HO have a thickness of 10 mm. The granularity is $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$, i.e. the same as for the HB, in order to have a 1-1 correspondence between the HB calorimetric towers and the HO segments. The light coming from the scintillators is collected by WLS fibers and sent toward the photodetectors placed on the return yoke.

- **HF**: in order to guarantee the coverage of the system up to $|\eta| = 5$, the HF detector was placed in the very forward region, at a distance of 11.2 m from the interaction point. This detector is optimized for the identification of those processes which produce very forward jets, in particular processes which deal with the Higgs boson and SUSY particles production. The two HF cylindrical units have a length of 1.65 m and an active radius of 1.4 m. HF is a sampling calorimeter consisting of quartz fibers sandwiched between iron absorbers. The choice of these materials was due to the high radiation dose of the forward region, which does not allow the use of plastic scintillators. With particles crossing the quartz fibers emit Čerenkov light and this light is detected by the radiation resistant photomultipliers. The whole HF consists of 900 towers and 1800 readout channels.

The energy resolution in the different geometrical regions of HCAL may be parametrized as follows, by using one stochastic term and one constant term:

$$barrel/endcap: \frac{\sigma_E}{E} = \frac{90\% \, [\mathrm{GeV}^{1/2}]}{\sqrt{E}} \oplus 4.5\% \,, \tag{2.9}$$

$$forward: \frac{\sigma_E}{E} = \frac{172\% \, [\mathrm{GeV}^{1/2}]}{\sqrt{E}} \oplus 9.0\% \,, \tag{2.10}$$

where the energy is expressed in GeV.

2.2.5 The Muon System

The CMS muon system [55] is dedicated to the identification and measurement of high p_T muons, in combination with the tracker. The system is placed outside the magnetic coil, embedded in the return yoke, to fully exploit the 1.8 T return flux.

The system consists of three independent subsystems (Fig. 2.11):

- Drift Tubes (DT) are placed in the barrel region, where the occupancy is relatively low (< 10 Hz/m^2).
- Cathode Strip Chambers (CSC) are in the endcaps, where the occupancy is higher (> 100 Hz/m^2).

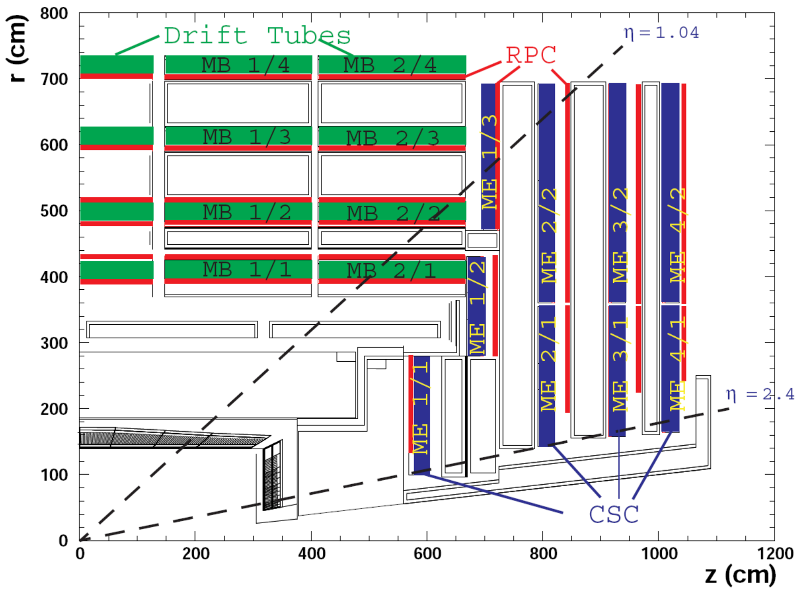


Figure 2.11: A schematic view of a quadrant of the CMS muon system.

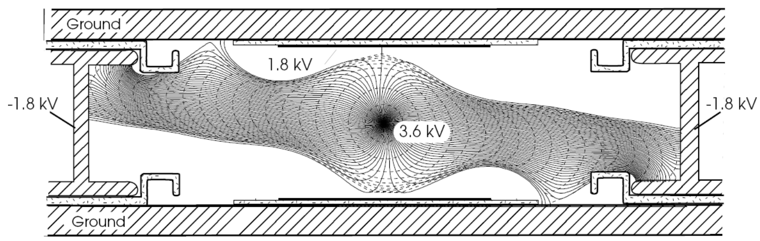


Figure 2.12: A schematic representation of a drift tube chamber. Drift lines in presence of magnetic field are also shown.

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

The Drift Tube system is made of chambers consisting of twelve layers of drift tubes each, packed in three independent substructures called *superlayers*. In each superlayer two chambers have anode wires parallel to the beam axis, two have perpendicular wires. Thus, each superlayer can provide two measurements of the $r-\phi$ coordinate and two measurements of the z coordinate of the track hit positions. Each chamber (Fig. 2.12) is made of two parallel aluminum plates with "I" shaped spacer cathodes, isolated from the aluminum plates with polycarbonate plastic. Chambers are filled with a gas mixture of Ar(85%) and CO_2(15%). The position resolution is about 100 μm in both $r\phi$ and rz.

Cathode Strip Chambers are multiwire proportional chambers with segmented cathodes (Fig. 2.13). Each chamber can provide both hit position coordinates. Chambers are filled with a gas mixture of Ar(40%), CO_2(50%), CF_4(10%). The chamber spatial resolution is about 80-85 μm.

Resistive Plate Chambers are made of parallel bakelite planes, with a bulk re-

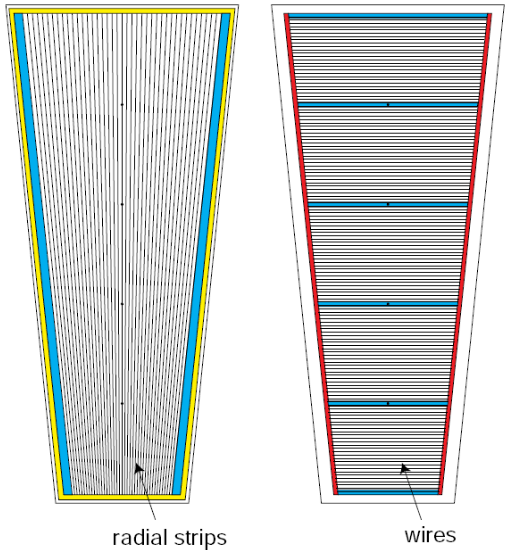


Figure 2.13: A schematic representation of CSC cathode panel (left) and anode panel (right).

sistivity of $10^{10} \div 10^{11}$ Ωcm. The gap between the plates if filled with a mixture of $C_2H_2F_4$ (94.5%) and i-C_4H_{10}. They operate in avalanche mode. Those chambers have limited spatial resolution, but they have excellent timing performances. They are mainly used for bunch crossing identification.

2.3 Trigger system

LHC will produce interactions at 40 MHz frequency, but only a small fraction of these events can be written on disk. On the one hand the speed at which data can be written to mass storage is limited, on the other hand the vast majority of events produced is not interesting, because it involves low transferred momentum interactions (minimum bias events). Thus, a trigger system is needed to save interesting events at the highest possible rate. The expected rate of events written to disk is foreseen to be 100 Hz.

CMS has chosen a two-level trigger system, consisting of a Level-1 Trigger (L1) [56] and a High Level Trigger (HLT) [57]. Level-1 Trigger runs on dedicated processors, and accesses coarse level granularity information from calorimetry and muon system. A Level-1 Trigger decision has to be taken for each bunch crossing within 3.2 μs. Level-1 Trigger task is to reduce the data flow from 40 MHz to 100 kHz.

The High Level Trigger is responsible for reducing the L1 output rate down to the target rate of 100 Hz. HLT code runs on a farm of commercial processors and can access to the full granularity information of all the subdetectors.

The main characteristics of the CMS trigger system will be described in the

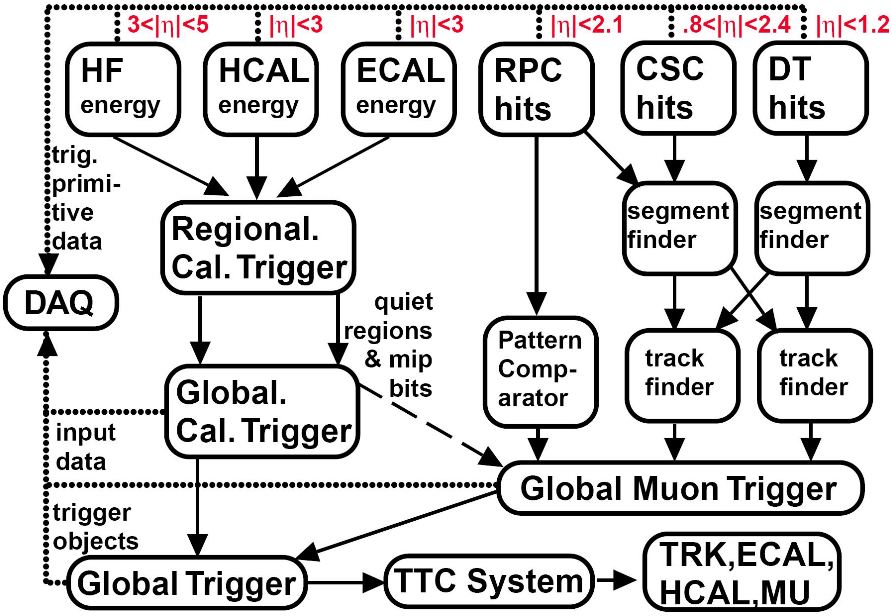


Figure 2.14: Level-1 trigger components.

following.

2.3.1 Level-1 Trigger

The Level-1 trigger is responsible for the identification of electrons, muons, photons, jets and missing transverse energy. It has to have a high and carefully understood efficiency. Its output rate and speed are limited by the readout electronics and by the performances of the *Data AcQuisition* (DAQ) [58] system.

It consists of three main subsystems:

- L1 Calorimeter Trigger;
- L1 Muon Trigger;
- L1 Global Trigger.

The L1 Global Trigger is responsible for combining the output of L1 Calorimeter Trigger and L1 Muon Trigger and for making the decision. L1 Muon Trigger is actually a composed system itself: information from RPC, CSC and DT specific triggers are combined in the so called L1 Global Muon Trigger. The organization of CMS Level-1 Trigger is schematically summarized in Fig. 2.14.

L1 Calorimeter Trigger

The input for L1 Calorimeter Trigger is calorimeter towers, as to say clusters of signals collected both from ECAL and HCAL. Towers are calculated by calorimeter high level readout circuits, called Trigger Primitive Generators. The Regional

Calorimeter Trigger finds out electron, photon, τ and jet candidates along with their transverse energy and sends them to the Global Calorimeter Trigger.

The Global Calorimeter Trigger sorts the candidates according to their transverse energy and sends the first four to the L1 Global Trigger.

L1 Muon Trigger

The RPC trigger electronics builds Track Segments, gives an estimate of the p_T and sends these segments to the Global Muon Trigger. It also provides the CSC logic unit with information to solve hit position ambiguities in case two or more muon tracks cross the same CSC chamber.

The CSC trigger builds Local Charged Tracks (LCT), as to say track segments made out of the cathode strips only. A p_T value and a quality flag are assigned to the LCTs. The best three LCTs in each sector of nine CSC chambers are passed to the CSC Track Finder, that uses the full CSC information to build tracks, assign them a p_T and a quality flag and sends them to the Global Muon Trigger.

DTs are equipped with Track Identifier electronics, which is able to find groups of aligned hits in the four chambers of a super-layer. Those Track Segments are sent to the DT Track Correlator that tries to combine segments from two superlayers, measuring the ϕ coordinate. The best two segments are sent to the DT Track Finder that builds tracks and sends them to the Global Muon Trigger.

The Global Muon Trigger sorts the RPC, CSC and DT muon tracks and tries to combine them. The final set of muons is sorted according to the quality, and the best four tracks are passed to the L1 Global Trigger.

L1 Global Trigger

The L1 Global Trigger is responsible for collecting objects created from the Calorimeter and Muon Triggers and for making a decision whether to retain the event or not. If the event is accepted the decision in sent to the Timing Trigger and Control System, that commands the readout of the remaining subsystems.

In order to take the decision, the L1 Global Trigger sorts the ranked objects produced by calorimetry and muon system and checks if at least one of the thresholds in the Level-1 Trigger table is passed. The Level-1 trigger rate for the low luminosity period is about 50 kHz.

2.3.2 High Level Trigger

The High Level Trigger is designed to reduce the Level-1 output rate to the rate of 100 events/s that is going to be written to mass storage. HLT code runs on commercial processors and performs reconstruction using the information from all subdetectors. Data read from subdetectors are assembled by a builder unit and then assigned to a switching network that dispatches events to the processor farm. The CMS switching network has a bandwidth of 1Tbit/s.

This simple design ensures maximum flexibility to the system, the only limitation being the total bandwidth and the number of processors. The system can be easily upgraded adding new processors or replacing the existing ones with faster ones as they become available. Since the algorithm implementation is fully software, improvements to the algorithms can be easily implemented and do not require any hardware intervention.

Event by event, the HLT code is run on a single processor, and the time available to make a decision is about 300 ms. The real time nature of this selection imposes several constraints on the resources an algorithm can use. The reliability of HLT algorithms is of capital importance, because events not selected by the HLT are lost.

In order to efficiently process events the HLT code has to be able to reject not interesting events as soon as possible; computationally expensive algorithms must be run only on good candidates for interesting events. In order to meet this requirement the HLT code is organized in a virtually layered structure:

- Level 2: uses only complete muon and calorimetry information;
- Level 2.5: uses also the pixel information;
- Level 3: makes use of the full information from all the tracking detectors.

Each step reduces the number of events to be processed in the next step. The most computationally expensive tasks are executed in the Level 3; time consuming algorithms such as track reconstruction are only executed in the region of interest. Besides, since the ultimate precision is not required at HLT, track reconstruction is performed on a limited set of hits, and is stopped once the required resolution is achieved.

2.4 CMS simulation and reconstruction software

The CMS simulation and reconstruction software, CMSSW [59], is a C++ [60] framework that can be configured via Python [61] scripts.

CMS Event Data Model (EDM) is based on the concept of `Event`. An `Event` is a C++ class that contains the information about a physics event, both raw level data and reconstructed quantities. Reconstruction algorithms can access information from the `Event` and put reconstructed quantities in it. `Event`*s* can be read from and written to ROOT [62] files.

CMSSW can be run feeding the desired Python configuration script into the executable `cmsRun`. The configuration file contains the *modules*, as to say the algorithms that the user wants to run and it specifies the order in which they need to be run. The executable reads in the configuration file and, using a plugin manager, finds out in which libraries the modules to be run are defined and loads them.

2.4.1 Event simulation

Event generation in CMSSW can be done with many event generator programs. Those programs can be run from within the framework, using dedicated interface libraries. The configuration of the event generators is performed feeding `cmsRun` with the appropriate configuration file containing the flags to be set in the event generator. The event generator is responsible for filling the HepMC [63] record with all the information about the currently generated event. The HepMC record is then captured by the CMSSW framework and stored in the `Event`. At this level the produced objects are said to be in GEN format.

After the event has been generated the simulation of the detector follows. The first step in the simulation of instrumental effects is the smearing of the vertex position. The event primary vertex, that is placed by the event generator at the origin of CMS coordinate system, is smeared according to the expected *pp* impact point distribution per bunch crossings. The next step is the simulation of the interaction of particles with the detector. The description of these interactions is achieved using GEANT4 [64]. Once energy deposits and multiple scattering effects in the CMS subdetectors are simulated, the simulation of the signals produced by the subdetectors follows. This step is known as "*Digitization*", and the objects produced at this level are said to be in SIM format.

2.4.2 Event reconstruction

At the beginning of reconstruction chain, real data consists of signals in each of the subdetectors. Data at this step are said to be in RAW format. Starting from RAW data the reconstruction of events follows. At this step of the event reconstruction also the simulated events (GEN-SIM) are in RAW format. With this approach exactly the same algorithms that will be used on real data can be run on simulated samples.

The reconstruction chain is defined by the user via a Python script, which selects the input files (usually ROOT files) and the modules to be executed and their execution order, sets the parameters used by the modules and finally selects which files have to be produced (usually one or more other ROOT files) and which objects have to be stored into them. The objects produced at this level are in RECO format or, after a further filter and reorganization, in AOD (*Analysis Object Data*) format, which is a higher level object with respect to the RECO.

CMS is also equipped with a set of tools whose goal is to make CMS physics analysis easier and more efficient. These tools are collected into a *Physics Analysis Toolkit*, or PAT [65]. The Physics Analysis Toolkit is a layer built on and within the CMSSW framework, with the aim of simplifying analysis operations by providing easier access to high-level information, as well as tools to perform common analysis tasks.

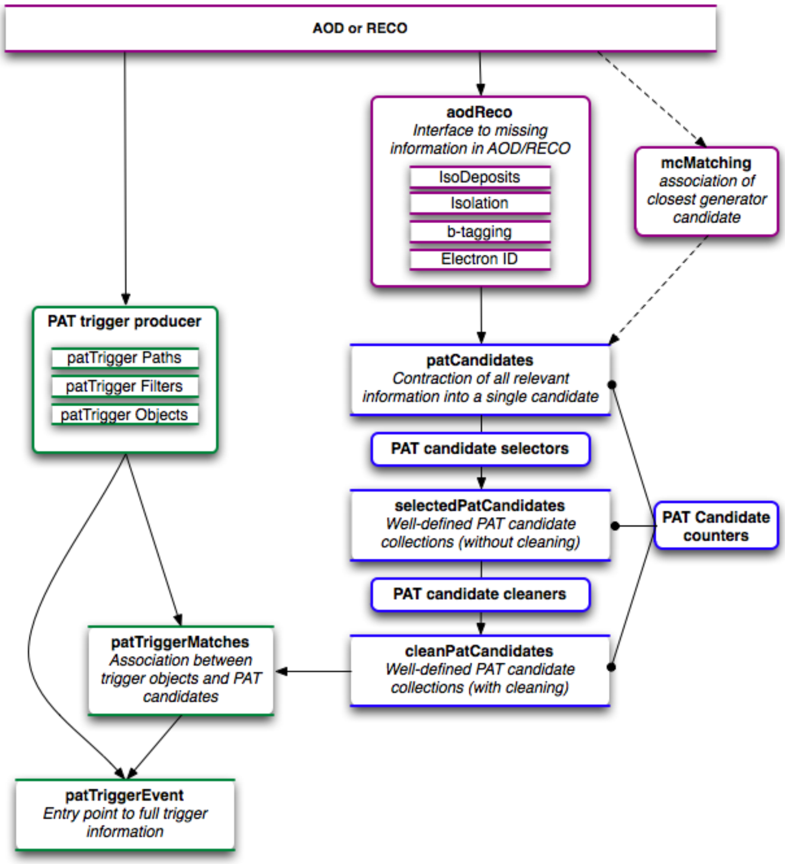


Figure 2.15: Scheme of the Physics Analysis Toolkit (PAT) workflow.

Starting from data in RECO (or AOD) format, the PAT algorithm creates a *candidate* for each particle present in the event, e.g. electron, jets etc., and produces ROOT files called *PAT-tuples* ready for the analysis. The algorithm proceeds through the following steps, as shown in Fig. 2.15:

- *Pre-creation*: some preliminary information are added to the RECO (or AOD) files, like correspondences with generator information in case of simulations.

- *Candidate creation*: significant available information are combined in a candidate object.

- *Candidate selection*: several extra selections can be applied by the user on the created candidates.

- *Candidate disambiguation*: some of the physics measurements, like energy deposits in the calorimeter might be reconstructed several times as different physics objects. For example, a cluster in the electromagnetic calorimeter can be interpreted as a photon, electron or jet. It might therefore be present in several different candidate collections at the same time. In this step further information is added that allows to distinguish different physics objects.

- *PAT Trigger Event*: besides the main PAT production sequence, trigger information is re-keyed into a human readable form. Also the PAT trigger matching provides the opportunity to connect PAT objects with trigger objects for further studies. Thus the user can easily figure out exactly which object(s) fired a given trigger.

The produced PAT-tuples maintain full event provenance. Even if very flexible in their content the objects are always well defined by the configuration file they have been created.

2.5 Reconstructed data sample

As described in Sec. 2.1.1, the total amount of data delivered to CMS in 2010 is about 47 pb^{-1}, while the amount of data actually recorded by CMS is about 43 pb^{-1}. The difference between delivered to and recorded by CMS data is due to the DAQ efficiency and to the dead time. The progressive rise of LHC instant luminosity from the beginning of 7 TeV collisions to the rest of 2010 (shown in Fig. 2.2), made necessary the gradual increase of the L1 trigger thresholds and the used HLTs with tighter selection criteria. In Sec. 4.7 the triggers that are required for the $Z \to ee$ events selection will be described.

Besides the first data filtering operated by the trigger system, a further decrease of the available data sample is due to the run selection, i.e. the selection of runs and luminosity sections (i.e. part of a single run) where CMS was fully operative without any particular problem to some subdetector. These runs and luminosity sections are considered *good* and should be processed. Within the CMS electron group, where the study of $pp \to Z(ee) + jets$ shown in this thesis took place, the official selection for the 2010 data taken used for the analysis leads to a total integrated luminosity available [66]

$$\mathscr{L} = (36.2 \pm 1.4)\,\mathrm{pb}^{-1}\,. \tag{2.11}$$

The data sample was reconstructed making use of the PAT software package described in Sec. 2.4.2, producing therefore PAT-tuple files which are the starting point for the analysis described in the following chapters. The pre-selection criteria applied on electrons and jets in the PAT phase, together with the trigger requirements, will be described in Chapters 4 and 5.

In order to reduce the size of the produced PAT-tuples and the CPU time required to process them in the further steps of the analysis, a first data skimming is applied on data sample in the PAT phase. Not all the events are therefore stored in the PAT-tuples, but only the events with the following characteristics:

- presence of at least one couple of e^+e^- (null charge of the couple required),
- $50\,\mathrm{GeV} < m_{e^+e^-} < 130\,\mathrm{GeV}$, where $m_{e^+e^-}$ is the invariant mass of the electron-positron couple.

The initial number of events collected in the selected runs and luminosity sections is equal to 74885719, while the skimmed events stored in the PAT-tuples are 730522.

Chapter 3

MC production

For the $pp \to Z + jets$ cross section measurement, the use of Monte Carlo simulations is needed in order to study the effects of the CMS acceptance on the process, the reconstruction efficiency (to be compared with the data driven efficiency that will be described in Chapter 6) and the effects of the reconstruction on the jet multiplicity.

In the following sections the main aspects of the computations that lead to the generation of realistic events will be briefly described. In particular the generators used in this thesis for signal and background simulations will be described, i.e. `PYTHIA` [67] and `MADGRAPH` [68, 69]. Finally the characteristics of the Monte Carlo datasets used will be shown.

3.1 Event generator components

The structure of events produced at high energy colliders is extremely complex, and numeric simulations are necessary to effectively simulate realistic events. Monte Carlo event generators are complex computer programs that subdivide the problem of producing realistic events into a sequence of tasks that can be handled separately with the help of both analytic and numeric computation.

Different event generators exist that implement computations at different levels of precision and with different techniques. Typically, the highest precision calculations, that take into account several orders in perturbation theory, are only available for a limited number of processes, thus making it hard to derive predictions on inclusive quantities. On the other hand these quantities can often be described with reasonable precision with programs that implement lower order calculations.

A schematic representation of the different components (and calculation steps) that are implemented in event generators is shown in Fig. 3.1. The production of hadron-hadron collision events is the result of the following chain of calculations:

- The first step is the calculation of cross sections for the selected processes. Cross sections are calculated for a pair of incoming partons (quarks and glu-

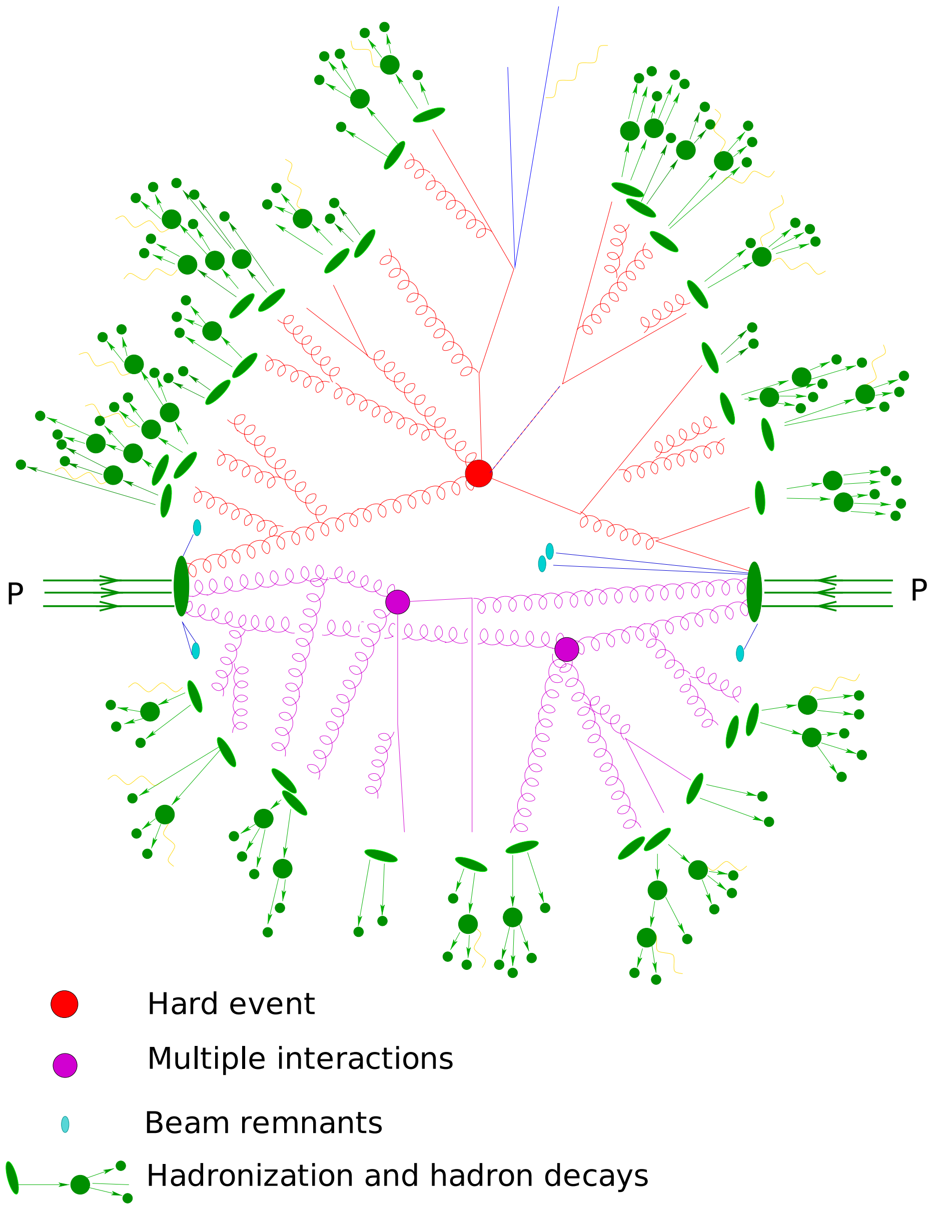


Figure 3.1: A schematic representation of the generation of an event in a typical event generator [70]. Partons from the two incoming hadrons participate in the hard scattering and in softer multiple interactions. Hadron remnants are treated. Quarks and gluons are turned into hadrons by hadronization and then hadrons decay.

ons) extracted from the colliding hadrons. This step is performed by using the *Matrix Element* (ME) method, which calculates the matrix element associated to the Feynman diagrams of the process. Matrix element will be described in Sec. 3.2.

- The event production starts with two colliding hadrons with given momenta. One parton out of each hadron is selected to enter the scattering process we are interested in. This step is often referred to as hard scattering generation. Final state partons and leptons are produced according to the calculated differential cross sections.

- Resonances produced in the hard event are allowed to decay.

- When two partons take part in the hard event, accelerated colour charges are present, thus bremsstrahlung can occur. This effect is called *Initial State Radiation* (ISR). To simulate ISR, knowledge of the parton density function is needed. In case of Z boson production for example, the ISR emission is responsible for the non null Z transverse momentum, due to the four-momentum conservation. An example of ISR emission in Z production process is shown in Fig. 3.2. The ISR is simulated with the *Parton Shower* (PS) algorithm, which will be described in Sec. 3.3.

- Also the final state partons can produce further radiation, called *Final State Radiation* (FSR). Such radiation is simulated by the Parton Shower algorithm.

- In addition to the partons taking part in the hard interaction, several other parton pairs can interact during a hadron-hadron collision, giving rise to interactions with smaller transferred momentum. These *Multiple Parton Interactions* (MPI) contribute to the so called underlying structure of the event (*underlying event*). Such interactions need to be simulated too if we want to produce realistic events, and ISR and FSR need to be simulated for these collisions too.

- Leftovers of the interacting hadrons need to be simulated to balance the colour charge and four momentum conservation. The beam remnant handling is thus another step in the event generation.

- The calculations described so far are carried out in the perturbative regime, but, as the produced partons move apart from each other, the coupling constant gets stronger and stronger and confinement effects take place. When the coupling constant is strong enough quark-antiquark pairs are produced from the vacuum and the partons turn into hadrons. This generation step is referred to as *hadronization*, and is calculated by using empirical models (see Sec. 3.4).

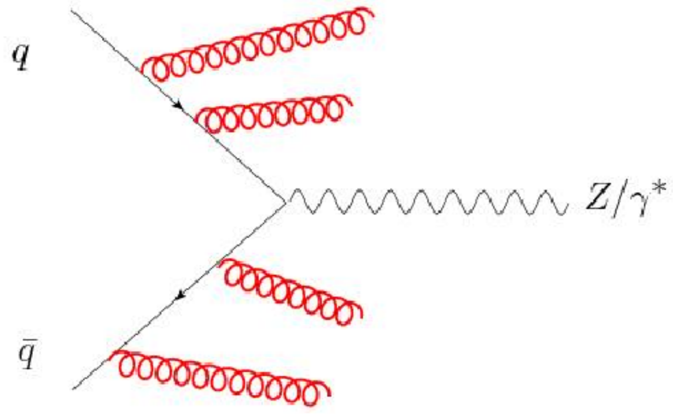


Figure 3.2: Example of the Initial State Radiation in a process of Z boson production.

- Finally, the event generator takes care of decaying τ leptons and B-hadrons; in general particles with very short lifetime are allowed to decay by the generator itself. Those that live enough to reach the detector are left undecayed.

The main steps of the event generation, i.e. Matrix Element, Parton Shower and Hadronization, will be briefly described in the following sections.

3.2 Matrix Element (ME)

The first step in the generation of an event is the calculation of the hard processes cross sections. General purpose event generators can perform such calculations for a vast variety of processes. Nevertheless it is often useful to interface such generators with dedicated hard process libraries in order to produce particular events, such as Supersimmetry (SUSY) processes for example.

The state-of-the-art in the field of matrix element (ME) calculation is NLO, with all the virtual loop corrections included. Loop calculations are complex and they are available for a limited number of processes. For this reason tree-level matrix element calculations still play an important role in the simulation of events produced at hadron colliders. Tree-level cross section calculations can be performed up to several (in order of eight) partons in the final state [68].

The main problems with tree-level matrix elements are the soft and collinear divergences (Sec. 1.5.3). Since at tree-level the loop corrections that would cancel these divergences are omitted, the phase space has to be carefully tailored to avoid the problematic regions. This means that the matrix element cross section calculations are performed away from soft and collinear divergences.

In order to produce realistic events, phase-space regions omitted from the matrix element calculations need therefore to be recovered, with care to avoid divergences.

This is done in a quite effective way by using Parton Shower calculations.

3.3 Parton Shower (PS)

When treating $2 \to n$ processes, tree-level Matrix Elements suffer from divergences in the soft and collinear regions. The splittings that suffer from these divergences are $q \to qg$, $\bar{q} \to \bar{q}g$, $g \to gg$: the two first processes have a QED counterpart, while the third comes from the non-abelian nature of QCD. The splitting $g \to q\bar{q}$ does not suffer from the soft divergence. The tree-level divergences would be removed including also virtual corrections. Such calculations, however, are extremely complex and are available only for a limited set of processes.

Parton Shower algorithm offers an alternative way both to handle the complexity of several successive branchings and to remove soft and collinear divergences. The parton showers are described by the algorithm as a succession of elementary events $a \to bc$, where each event can happen with a certain probability. The Parton Shower machinery thus handles the divergences of the original Matrix Element by imposing the conservation of total probability. The parton cascade is evolved down to a certain virtuality, of the order of 1 GeV^2. After that, non perturbative effects take place and hadronization is applied.

It should be noticed that the parton shower machinery relies on a collinear approximation of the matrix element, thus it should perform well in the description of the evolution of jets, but one cannot expect it to give a precise answer for the description of well separated parton configurations.

3.3.1 Merging Matrix Elements and Parton Shower

Matrix element and parton shower calculations have different virtues and different applicability limits. We can summarize some of the main facts about the ME calculations as follows:

- as long as tree-level is concerned, these calculations can be performed up to several (order of eight) partons in the final state;
- ME are good at describing well separated parton configurations;
- ME calculations are exact to a given order in perturbation theory.

However:

- ME cross sections have divergences in the soft and collinear regions, thus they can not describe the internal structure of a jet;
- since hadrons is what it is observed in experiments, fragmentation models need to be applied to the partons. To use bare ME partons would imply the need to tune these models for each center of mass energy; this fact limits the applicability of bare ME calculations.

On the other hand, Parton Showers:

- are universal; given the basic hard process, the parton shower technique will produce realistic parton configurations;
- are derived in the collinear limit, and handle divergences by requiring conservation of total probability. This makes them particularly suited to describe the evolution of jets;
- can be used to evolve partons down to a common scale; this removes the need of tuning fragmentation models at different scales.

However, since they are derived in the collinear approximation, they may fail in efficiently filling the phase space for well separated parton configurations.

From the above description it is clear that a combined use of ME and PS would make it possible to take advantages of the qualities of the two approaches in the phase space regions where each performs better [71–73].

Several prescriptions exist to perform ME-PS matching avoiding double-counting or holes in the phase space. Care must be taken indeed to avoid that a configuration with n partons emerging from the ME is produced also by an $(n-1)$-partons ME plus an additional hard emission coming from the PS.

3.4 Hadronization

After the Parton Shower step of the event generation what remains is a set of partons with virtualities of the order of the cutoff scale at which the shower was stopped. Hadronization is the step in which partons are turned into hadrons. The process is non-perturbative and at the present moment it is described by several models. The one used by `PYTHIA` and `MADGRAPH` is the Lund string model [74].

3.4.1 Lund string model

Quark and antiquarks produced in the shower move apart from each other transferring part of their energy to the colour field that connects them. As they move apart the color field lines tighten and acquire a string shaped configuration. The energy stored per unit length in the colour field tends therefore to be uniform, as shown in Fig. 3.3. When enough energy is stored in the string it can break up into a quark antiquark pair.

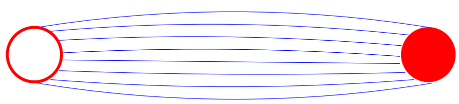


Figure 3.3: Schematic representation of the color field as the string forms.

With this simple mechanism the formation of mesons is described. The flavour of the $q\bar{q}$ pair that results from the string break up is assigned with probabilities tuned to data. The formation of baryons is more complex and it requires considering a three quark final state in which two of the quarks are close and form one of the two end points of the string [74].

3.5 PYTHIA and MADGRAPH generators

MC samples used in this thesis have been produced by using PYTHIA 6.4 and MADGRAPH 4.4.12 generators.

PYTHIA [67] is a general purpose generator, which has been used extensively at LEP, HERA and at Tevatron for e^+e^-, ep and $p\bar{p}$ physics. It contains a large subprocess library covering Standard Model physics but also SUSY, Technicolor and other Exotics processes. PYTHIA uses ME method for the cross section calculation at LO, while for higher order diagrams, like ISR and FSR gluon emission, it uses the PS method. This recipe is inefficient in describing jets with a high emission angle, due to the PS inefficiency for well separated partons (see Sec. 3.3.1). To reduce this effect PYTHIA implements a corrected PS approach, in which the first emission from the shower is ME-corrected.

MADGRAPH [68, 69] uses instead ME also for events with more than one jet, even if considering only real diagrams. The jet energy distribution is therefore described better than with PYTHIA. Nevertheless it does not consider the contributions carried to cross section calculation by the virtual diagrams, e.g. gluon exchange between incoming quarks. The cross calculation is thus performed at LO, as in PYTHIA. The event information (particle IDs, momenta, spin etc.) is then interfaced with PYTHIA which handles the rest of the generation steps (involving parton showering, hadronization etc.).

The last step of the event generation, i.e. the hadronization, is performed both for PYTHIA and MADGRAPH by using the Lund string model, as described in Sec. 3.4.1.

The generation of the underlying event is a complicated process. For its description several phenomenological models exist, with various degrees of sophistication. The increased activity in the underlying event of a hard scattering collision over that observed in soft collisions cannot be explained solely by ISR. Multiple parton interactions provide a natural way of explaining this increased activity. A hard scattering is more likely to occur when the hard cores of the beam hadrons overlap and this is also when the probability of a MPI is greatest. The transverse region is sensitive to the underlying event and the MPI parameters have to be tuned to fit the data. Besides the underlying event, also the hadronization step have to be tuned to data, as stated in Sec. 3.4.1. Since the tuning belongs to generation steps simulated with the PS, only PYTHIA needs to be tuned. MADGRAPH samples will be tuned by using the same PYTHIA tunes.

At CMS, two PYTHIA tunes have been applied to the generated MC samples and were studied: the Z2 tune and the D6T tune. Z2 tune [75], which seems to agree better with collected data as will be shown in the following Chapters, uses the parton shower ordered in transverse momentum and was tuned to the first collision data at the LHC. Therefore the description of the underlying event does not rely solely on extrapolations to LHC energies but already is able to properly describe the first minimum bias data obtained at the LHC. D6T tune [76] employs the virtuality ordered shower and it is a relatively old tune, based on measurements at the Tevatron collider and older machines. It uses extrapolations in order to make predictions for LHC energies.

For this thesis the Z2 tune was used both for signal and background simulated samples. For the signal sample also D6T tune was used as complementary for systematics studies.

3.6 MC samples produced

To model the data, a set of MADGRAPH and PYTHIA Monte Carlo samples provided by the CMS generator group was used. All the samples used are simulated at $\sqrt{s} = 7$ TeV, i.e. the energy in the center of mass frame of the 2010 LHC collisions, and include pileup corresponding to the expected pileup in 2010 collision data.

Signal

In order to have a more efficient jet simulation, the signal $pp \to Z + jets$ was modeled with a MADGRAPH Drell-Yan dataset. Z decays in the three leptonic families are included in the dataset and the kinematic cut $M_{l^+l^-} > 50$ GeV was applied to the lepton pair. For the normalization to the integrated luminosity of data, the NNLO cross section (determined with FEWZ code [32]) was used. As stated in the previous section, two different datasets were used for the signal simulation, one using the Z2 tune and the other using the D6T tune. Z2 was the main tune used for the analysis, while D6T tuned dataset was used only for systematics studies. Finally, in order to compare the obtained results also with PYTHIA generated $Z + jets$ events, also a PYTHIA sample $Z \to e^+e^- + jets$ is used for the signal, as will be shown in Chapter 7, with $M_{e^+e^-} > 20$ and Z2 tune.

Background

For what concerns the background, seven contributes were considered.

- Four contributions belong to the EWK background:
    - $W + jets \to l\nu_l$,
    - $WW \to anything$,

- $WZ \rightarrow anything$,
- $ZZ \rightarrow anything$.

The first of them was simulated with MADGRAPH and includes W decays in all the three leptonic families. The other EWK samples are simulated with PYTHIA . To scale them to the data luminosity, NNLO cross section was used for $W + jets$ while NLO cross section was used for the other EWK samples.

- Another contribution is given by

  - $pp \rightarrow t\bar{t}$.

  This sample was modeled with PYTHIA and the NLO cross section was used for the normalization.

- The last two contributions belong to the QCD background and are called:

  - QCD ElectroMagnetic (EM) enriched,
  - QCD $b, c \rightarrow e$.

  The QCD EM enriched background includes jets and it is enriched with electromagnetic contributions (i.e. electrons and photons), while the QCD $b, c, \rightarrow e$ background represents the background of electrons coming from heavy quark decays. These two background samples, which are mutually exclusive, are simulated with PYTHIA in three distinct regions of transverse momentum of the partons produced in the hard scattering: $20\ \mathrm{GeV} < \hat{p}_T < 30\ \mathrm{GeV}$, $30\ \mathrm{GeV} < \hat{p}_T < 80\ \mathrm{GeV}$ and $80\ \mathrm{GeV} < \hat{p}_T < 170\ \mathrm{GeV}$. Each sample is normalized to the data luminosity by using the LO cross section. Due to the very high cross sections of the QCD processes, the simulated statistics for 4 out of 6 QCD samples is less than the available data statistics, and therefore they have been multiplied by a normalization factor greater than 1, as shown in Table 3.2.

As done for data samples (see Sec. 2.5), a first data skimming is applied on MC sample in the PAT phase, in order to reduce the size of the produced PAT-tuples and the CPU time required to process them in the further steps of the analysis. The skimming requirements are the same applied on data, i.e.

- presence of at least one couple of e^+e^- (null charge of the couple required),
- $50\,\mathrm{GeV} < m_{e^+e^-} < 130\,\mathrm{GeV}$, where $m_{e^+e^-}$ is the invariant mass of the electron-positron couple.

The full list of Monte Carlo samples and their cross sections is summarized on Table 3.1 for signal and Table 3.2 for background. The total number of processed events, the number of skimmed events and the scale factor for the normalization to the 36.2 pb^{-1} of data luminosity are also shown in the tables.

Signal							
generator	**tune**	**process**	**kinematic cuts**	**events**	**skim. events**	**σ(pb)**	**norm. scale**
MADGRAPH	Z2	$Z \to l^+l^- + jets$	$M_{ll} > 50\,\text{GeV}$	4857046	4857046	3048 (NNLO)	0.023
MADGRAPH	D6T	$Z \to l^+l^- + jets$	$M_{ll} > 50\,\text{GeV}$	1500000	1500000	3048 (NNLO)	0.073
PYTHIA	Z2	$Z \to e^+e^- + jets$	$M_{ee} > 20\,\text{GeV}$	2127607	2127607	1666 (NNLO)	0.028

Table 3.1: List of MADGRAPH signal samples. "Norm. scale" is the normalization factor to the 36.2 pb^{-1} of data luminosity.

Background								
	generator	**tune**	**process**	**kinematic cuts**	**events**	**skim. events**	**σ(pb)**	**norm. scale**
EWK	MADGRAPH	Z2	$W \to l\nu_l + jets$	None	15161497	33190	31314 (NNLO)	0.074
	PYTHIA	Z2	$WW \to anything$	None	2050240	36104	43 (NLO)	$7.5 \cdot 10^{-4}$
	PYTHIA	Z2	$WZ \to anything$	None	2185664	68199	18.2 (NLO)	$3.0 \cdot 10^{-4}$
	PYTHIA	Z2	$ZZ \to anything$	None	2113368	103960	5.9 (NLO)	$1.0 \cdot 10^{-4}$
$t\bar{t}$	PYTHIA	Z2	$t\bar{t}$	None	1099550	73955	157.5 (NLO)	0.005
QCD	PYTHIA	Z2	QCD EM Enriched	20 GeV $< \hat{p}_T <$ 30 GeV	36375274	31735	2454400 (LO)	2.429
	PYTHIA	Z2	QCD EM Enriched	30 GeV $< \hat{p}_T <$ 80 GeV	71834019	261589	3867500 (LO)	1.938
	PYTHIA	Z2	QCD EM Enriched	80 GeV $< \hat{p}_T <$ 170 GeV	8073559	132060	139500 (LO)	0.622
	PYTHIA	Z2	QCD $b, c \to e$	20 GeV $< \hat{p}_T <$ 30 GeV	2243439	3582	132160 (LO)	2.121
	PYTHIA	Z2	QCD $b, c \to e$	30 GeV $< \hat{p}_T <$ 80 GeV	1995502	18944	136850 (LO)	2.469
	PYTHIA	Z2	QCD $b, c \to e$	80 GeV $< \hat{p}_T <$ 170 GeV	1043390	48514	9360 (LO)	0.323

Table 3.2: List of MADGRAPH and PYTHIA background samples. "Norm. scale" is the normalization factor to the 36.2 pb^{-1} of data luminosity.

Chapter 4

Electron reconstruction and Z(ee) events selection

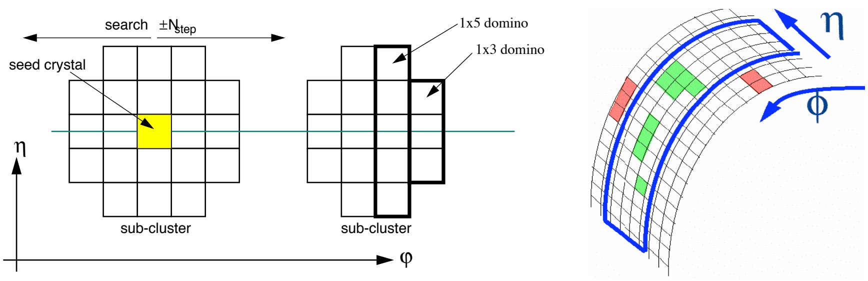
Electrons are selected in CMS through their charged hits in the Tracker and their energy deposit in ECAL. The reconstruction of electrons therefore uses information from the Pixel detector, the Silicon Strip Tracker and the Electromagnetic Calorimetry. A brief description of these detectors has been given in Chapter 2. Here the standard offline electron reconstruction algorithms are shown, together with the applied criteria for selecting $Z \rightarrow ee$ events. Unless specified, the word "electron" refers in the following to both electron and positron, as well as the symbol "ee" is used to denote an electron-positron couple.

4.1 Standard electron reconstruction in CMSSW

The electron reconstruction relies on the combination of both Tracker and Calorimetry information. Three main steps can be identified: the energy "clustering" in ECAL, the track-seeding and finally the inward-outward track reconstruction. The algorithms used for these three steps are implemented in the CMSSW Electron Photon package and they are employed both for HLT and offline analysis.

4.1.1 Calorimetric energy reconstruction

Electron reconstruction begins by clustering the ECAL energy of the electron. A single electron generates a shower which develops on several ECAL crystals. For a supermodule of the ECAL barrel in the test beam, electrons with an energy of 120 GeV impinging at the center of a crystal deposit about 97% of their incident energy in a 5×5 crystal window [77]. Due to the presence of the tracker material in front of the Electromagnetic Calorimeter, there is also a significant bremsstrahlung emission which is responsible for a further energy spread in the bending plane, i.e. along the ϕ direction. It is therefore necessary to collect the energy of all the crystals involved in these processes. The starting point is the search for crystals with



(a) (b)

Figure 4.1: Domino construction (a) and supercluster construction (b) steps of Hybrid algorithm [80].

energy above a certain threshold, which are called *seeds*. The seeds are sorted in descending order of E_T and only the most energetic seed is kept among the adjacent ones. Starting from each seed the energy deposits are then grouped by using two different reconstruction algorithms [78,79]: the *Hybrid* algorithm, used in the barrel, and the *Multi*5×5 algorithm, employed in the endcaps. Two different algorithms are necessary to take into account the different geometry of ECAL crystals and the different magnetic field map in barrel and endcap regions.

The Hybrid algorithm

In the barrel region clustering is obtained by using the Hybrid algorithm, which takes advantage of the $\eta - \phi$ projective geometry of the crystals to perform the collection of both the energy in individual showers and of the set of showers compatible with a bremsstrahlung emission. This is done by collecting energy within a rectangular window extended in the ϕ direction.

The Hybrid algorithm operates as follows. A list of seed crystals with $E_T > 1$ GeV is first constructed. Starting from a seed crystal, a *cluster* is defined as an ensemble of ϕ-contiguous "dominos", which have collected an energy larger than 0.1 GeV. Each domino consists of 5 crystals with the same ϕ value (which corresponds to a domino width of 0.087 in η). Valleys, where less than 0.1 GeV are collected in a domino, separate different clusters.

The dominos are then clustered in ϕ in order to form *superclusters*. Each distinct cluster of dominos grouped in the supercluster is requested to have a seed domino with energy greater than 0.35 GeV. The ϕ roads are allowed to extend up to ± 17 crystals around the seed, which corresponds to ± 0.174 rad. The hybrid supercluster is made up of a series of showers at constant η but spread in the ϕ-direction. In Fig. 4.1 the domino and the supercluster construction steps are shown. Each energy deposit can be well contained in a 5×5 crystal window.

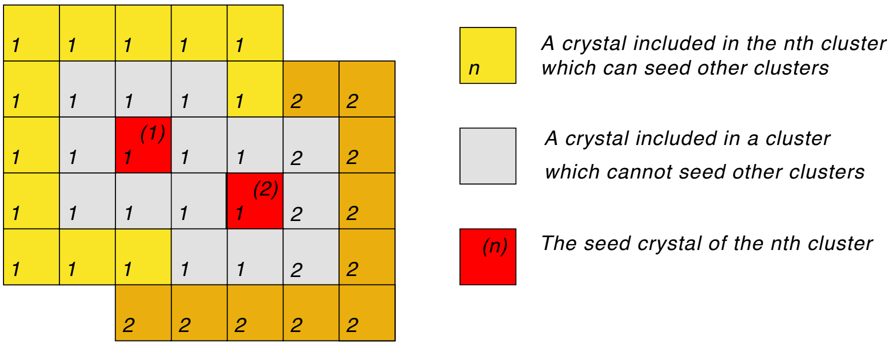


Figure 4.2: An illustration of two overlapping Multi5$\times$5 clusters. Crystals indicated in yellow are eligible to seed further Multi5 $\times$ 5 clusters, provided they are local maxima in energy [78].

The Multi5 $\times$ 5 algorithm

Since the crystals in the endcaps are not arranged in an $\eta - \phi$ projective geometry as in the barrel, the hybrid algorithm cannot be applied there. The same idea of collecting energy deposits within a window in η and ϕ must be therefore implemented differently. This is achieved by using the Multi5 $\times$ 5 algorithm, which operates as follows on the set of crystals sorted in descending order of E_T.

The Multi5 $\times$ 5 algorithm starts from a seed crystal with $E_T > 180\,\mathrm{MeV}$, and checks if it is a local maximum in energy by comparing its energy to the energy of its 4 neighbours by side in a Swiss Cross pattern. If the crystal is not a local maximum, the algorithm continues by searching for other seeds. Around each seed it constructs a 5 $\times$ 5 matrix of crystals, including only crystals that do not already belong to a cluster.

To allow closely overlapping showers due to bremsstrahlung to be collected, the outer 16 crystals of the 5$\times$5 matrix may seed a new matrix. In case of overlaps, the overlapping crystals are associated to the cluster with largest seed E_T. An example of the result of this process is shown in Fig. 4.2.

To produce the final supercluster by recovering bremsstrahlung, a rectangular window in η and ϕ extended in the ϕ-direction is created around energy deposits with transverse energy above 1.0 GeV. Other energy deposits falling within the window are added to form the supercluster. This procedure is performed in descending order of E_T of the energy deposits and an energy deposit may only belong to one cluster.

In the endcaps, the region $1.6 < |\eta| < 2.6$ is covered by the preshower detector. Electrons and photons reconstructed in this region will deposit some fraction of their energy in the preshower, so this energy must be measured and added to each cluster. This is done by summing the energy measured from the preshower strips intersected at the position extrapolated between each energy deposit in the calorimeter and the primary vertex. This energy is added to each endcap supercluster before any energy scale corrections are applied.

Energy scale corrections

To perform a precise measurement of electron or photon energy at the primary vertex, a number of calibration and energy corrections have to be applied [78, 79]. The electron/photon energy is computed as the sum of the signal amplitude measured in the channels included in the supercluster, corrected as follows:

$$E_{e/\gamma} = F \times \sum_{cluster} G \cdot c_i \cdot A_i \,,$$

where A_i are the amplitudes measured in each crystal in ADC counts, c_i are the crystal intercalibration constants, G is global scale calibration term and F represents the supercluster energy correction.

The intercalibration procedures aim at determining the c_i coefficients. The largest sources of variations are due to the light yield difference between the crystals, and to the APD and VPT gain variation (see Chapter 2 Sec. 2.2.3). These variations are reduced by calibration with radioactive sources, test beam and cosmic rays (only for the barrel). The final step of the intercalibration is performed with collision data, by using the symmetry of energy deposits in ϕ and the $\pi^0, \eta \rightarrow \gamma\gamma$ decays.

The factor F is composed by three different factors:

$$F = C_{EB}(\eta) \times f(brem) \times f(\eta, E_T)\,,$$

- $C_{EB}(\eta)$ function corrects for the leakage of energy from the exposed faces of the barrel crystals, which increase with η. This correction therefore is only applied to barrel crystals.
- $f(brem)$ aims at correcting for the response of the clustering algorithm to the shower, where *brem* describes the spread of the electromagnetic cascade, i.e. the dimensions of the cluster.
- $f(\eta, E_T)$ is a residual correction applied to all reconstructed supercluster, due to the nonlinear distribution of matter in the detector and of the energy dependence.

The energy resolution obtained with barrel supermodules in a test beam for 120 GeV electrons is shown in Fig. 4.3, before and after all relevant corrections have been applied.

4.1.2 ECAL driven seeding

After superclusters have been reconstructed, the electron track reconstruction is seeded by matching ECAL superclusters with hit pairs or triplets from the innermost

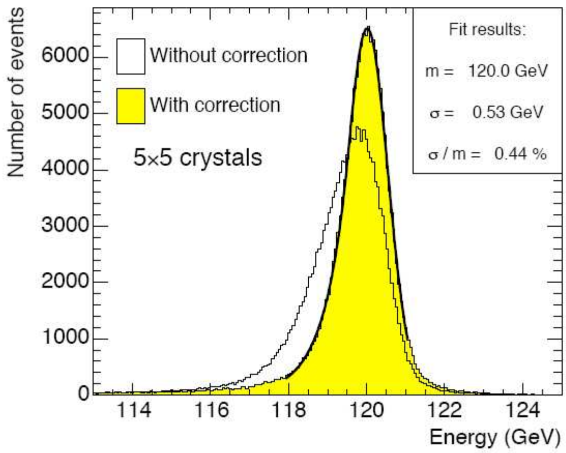


Figure 4.3: Energy resolution measured with barrel supermodules in a test beam with 120 GeV electrons [79].

tracker layers, i.e. using an ECAL driven seeding [81]. In the barrel region the hits may belong to the TIB layers, but at least 1 hit must be a pixel hit (to reduce to number of reconstructed electrons coming from conversions). In the endcaps also pixelless seeds are allowed, composed by TID or TEC hits.

The adopted procedure takes advantage of the fact that the supercluster position is on the helix of the initial electron trajectory and therefore it is possible to predict the position of the hits backpropagating the helix parameters through the magnetic field towards the innermost part of the measured trajectory. This strategy allows for an efficient filtering of background from jets faking electrons.

Windows in ϕ and z (or transverse radius r_T in the forward region) are defined in order to match the hits of each trajectory seed, taking into account both sign charge hypotheses. In case of triplets, at least two out of the three hits are required to be matched. Once a hit is matched on the first layer, this information is used to refine the helix parameters and a second hit is looked for in the second layer using smaller windows. In order to further reduce the contamination of fake electrons from jets, the first ϕ window is made E_T dependent. In Fig. 4.4 the seeding efficiency as a function of the generated electron p_T is shown.

The output of the ECAL driven seeding is the starting point of the electron track reconstruction.

4.1.3 Electron track reconstruction

Once the track seed has been identified, the full track reconstruction is performed. The standard algorithm used in CMS for the track reconstruction is the *Kalman Filter* (KF) [82]. The starting point is the trajectory model for a relativistic charged particle in a magnetic field, i.e. a helix with the axis parallel to the

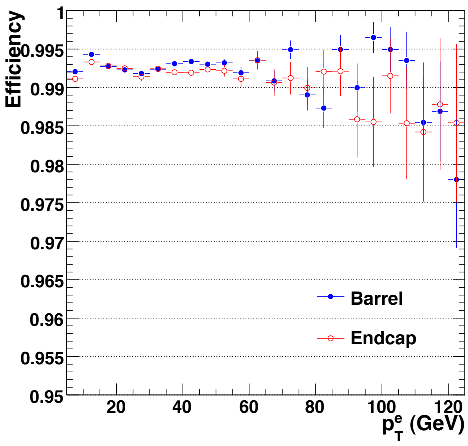


Figure 4.4: Electron seeding efficiency as a function of the generated electron p_T. Electrons are from a simulated sample of $Z \to ee$ decays for the signal and from a sample of QCD di-jet events with $p_T^{hat} = 15 - 170\,\text{GeV}$ for the background. The preselection cuts described in Sec. 4.2 have been applied [81].

direction of the magnetic field. The number of parameters needed to describe such a trajectory is five. Therefore tracks are described by a five dimensional state vector $\vec{y}$ containing the information about the momentum, the direction and the position at each point of the trajectory. The state vector can be written as a function of a coordinate, $\vec{y} = \vec{y}(z)$, and its evolution as a function of z can be described by a set of differential equations. It is anyway sufficient to consider the state vector in a discrete set of points, e.g. the intersection of the track with the silicon sensors. In this way the problem reduces to a discrete set of equations

$$\vec{y}(z_k) \equiv \vec{y}_k = \vec{f}_{k-1}(\vec{y}_{k-1}) + \vec{w}_{k-1} \tag{4.1}$$

where $\vec{f}_{k-1}(\vec{y}_{k-1})$ is the track propagator from the detector $k-1$ to the sensor k and the random variable $\vec{w}_{k-1}$ is the noise which incorporates a random disturbance of the track between z_{k-1} and z_k. The state vector $\vec{y}$ is not directly observed and the quantities $\vec{m}$ measured by the detector are functions of the state vector with a distortion due to the measurement noise $\vec{\epsilon}_k$

$$\vec{m}_k = \vec{h}_k(\vec{y}_k) + \vec{\epsilon}_k \tag{4.2}$$

and should be inserted in Eq. (4.1) in place of $\vec{y}$.

The track fitting with the Kalman Filter proceeds throughout three steps: the *filtering*, i.e. the estimate of the state vector with a local measurement, the *prediction* of the state vector in the future, and the *smoothing*, i.e. the estimate of the state vector in the past by using all the measurements collected up to the present time. The track vector is extrapolated from the $k-1$ detector to the k detector by means of the track model, then the extrapolated state vector is updated with the measurement on the detector k. The covariance matrix of the extrapolated state vector is computed by error propagation and the covariance matrix of the noise between the detectors $k-1$ and k is added to the propagated one. The smoothing of the estimate state vector can be done running two filters both inward-outward and outward-inward and then combining both the predictions with the measured value.

The Kalman Filter is a least square estimator, where the theoretical model of the trajectory is not a priori known and has to be obtained and updated every iteration of the filter. It is the optimal filter when the system is linear and both $\vec{w}_k$ and $\vec{\epsilon}_k$ are Gaussian random variables, otherwise non linear filters may do a better job as described in the following.

The bremsstrahlung problem

The main problem related to the electron track reconstruction is the bremsstrahlung emission in the tracker, which strongly affects both the momentum and the energy measurements. The material budget before ECAL varies as a function of the pseudorapidity and it has its maximum around $|\eta| = 1.5$, as shown in Fig. 2.8 of Chapter 2. Such a large amount of material causes a significative bremsstrahlung emission. As expected such bremsstrahlung emission is directly related to the tracker material, as shown in Fig. 4.5 where the number of emitted photons is plotted as a function of electron pseudorapidity for 30 GeV p_T electrons.

For what concerns the tracking, when a photon is emitted the track gets more curved than predicted from the most probable value, hence biasing the estimation towards lower p_T values. This effect depends on the hardness of the photon which is radiated, so the tail is more evident for higher p_T tracks. Again, a late radiation has only a small effect on the reconstructed track parameters. Another important

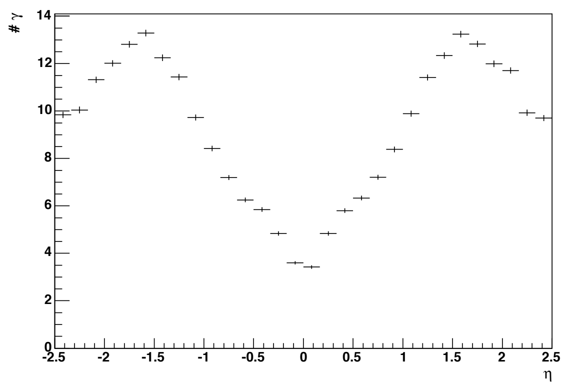


Figure 4.5: Average number of bremsstrahlung photons emitted as a function of the electron pseudorapidity for 30 GeV p_T simulated electrons. Only photons with energy higher than 10 MeV have been generated [83].

consequence of the bremsstrahlung emission on tracker measurements is the low number of hits in the tracks, because the reconstruction is often stopped by the hard χ^2 cut in the reconstructed track updating when hard photon are emitted.

The Gaussian Sum Filter algorithm

The noise $\vec{w}$, present in Eq. (4.1), takes into account both the multiple scattering and the energy loss. In case of high energy electrons, the ionization loss can be neglected with respect to the energy loss due to bremsstrahlung. The latter can be described by using the Bethe-Heitler model [84], in which the probability density function $f(z)$ of the electron energy loss is expressed by

$$f(z) = \frac{(-\ln z)^{c-1}}{\Gamma(c)}\,, \tag{4.3}$$

where $c = t/\ln 2$, t being the thickness in units of radiation length of the material that the electron crosses, z is the fraction of energy remaining after the material layer is traversed and finally $\Gamma(c) = \int_0^{+\infty} c^{-1}e^{-c}\mathrm{dt}$ is the Legendre notation of the Euler Γ-function. While the multiple scattering can be well described with a single Gaussian, such approximation doesn't hold for the Bethe-Heitler distribution, so the use of the Kalman Filter described above is not correct in the case of electron reconstruction. An adapted algorithm is therefore necessary for electron tracks: it is the *Gaussian Sum Filter* (GSF) algorithm [85] and it is essentially a non linear generalization of the Kalman Filter, which approximates the distribution of the process noise by a mixture of N Gaussians. The resulting filter is a weighted sum of Kalman Filters running in parallel, with weights depending on the observations. Each Kalman Filter corresponds to one of the components of the mixture:

$$\sum_{i=0}^{N} w_i \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(z-\mu_i)}{2\sigma_i^2}} \equiv f(z)\,. \tag{4.4}$$

In this way it is possible to fit the weights, means and standard deviations of each gaussian applying Kalman Filter on single gaussian terms. Since the Gaussian Sum Filter corresponds to N Kalman Filter running in parallel it brings to an increment of the combinatorial tries and consequently higher usage of CPU time.

For the electron tracking with Gaussian Sum Filter, in order to preserve efficiency and to follow the electron trajectory in case of bremsstrahlung emission, a very loose χ^2 compatibility is required in the building steps of the electron tracking, with a cut value of 2000. The combinatories are limited by requiring at most 5 candidate trajectories at each layer and at most one layer with missing hit. Finally, in order to reduce the probability of connecting a primary electron to a leg from a photon conversion, a high χ^2 penalty (90) is used in the cases of missing hit [81]. In Fig. 4.6 the number of collected hits for electron tracks reconstructed by using both the Gaussian Sum Filter and standard Kalman Filter used for MIPs is shown.

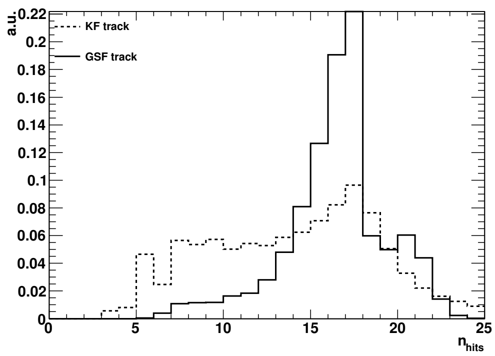
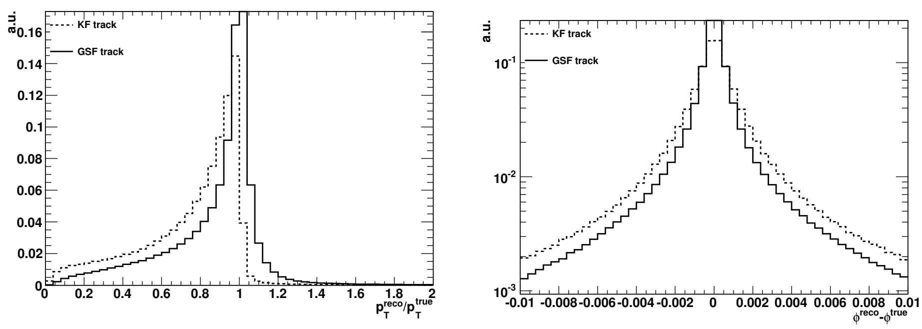


Figure 4.6: Number of reconstructed hits per track for electrons from simulated $Z \to ee$ decays: distribution as obtained with the dedicated tracking procedure used for electrons (solid line) and with the standard Kalman Filter used for MIPs (dashed line) [81].



(a) (b)

Figure 4.7: Electron track momentum parameters residual distributions at the innermost track position for both the dedicated GSF electron tracking (solid line) and the standard Kalman Filter used for MIPs (dashed line): transverse momentum magnitude (a) and momentum ϕ direction (b) . Electrons are from a simulated sample of $Z \to ee$ decays. [81].

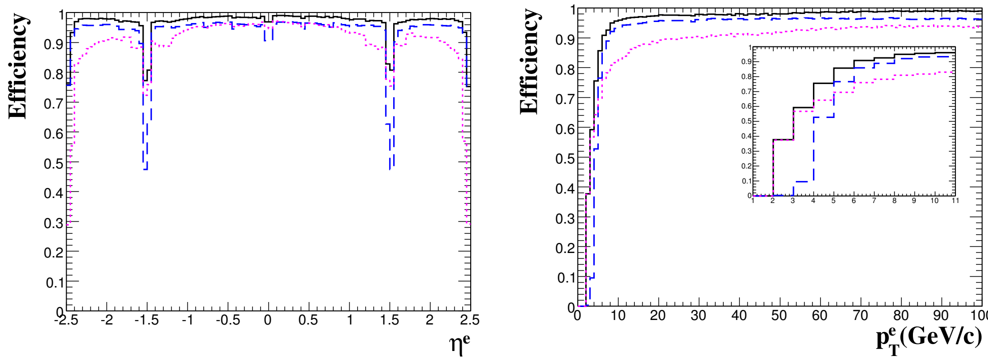
Fig. 4.7 instead presents the comparison between the track momentum parameters as obtained from the GSF fit and from the standard Kalman Filter procedure used for MIPs. The results are shown for electrons from a simulated sample of $Z \to ee$ decays. One can see that the GSF estimate is more precise, in particular for the ϕ direction. The transverse momentum reconstruction shows a less biased measurement for tracks having been subject to bremsstrahlung emission, while a similar resolution is observed from the right hand side of the distribution.

4.1.4 Tracker driven seeding

The ECAL driven electron seeding is very efficient for isolated electrons with $p_T \gtrsim 10\,\text{GeV}$, as shown in Fig. 4.4. At lower p_T the ϕ window used for the superclusters can be too small for electrons which radiate, since at this energy electron and photon clusters are more separated than for high momentum electrons. Moreover, in case of electrons in jets, the energy collected in the superclusters may include some neutral contributions from the jets, biasing the energy measurement used to seed electron tracks. For these reasons, the ECAL driven algorithm is complemented by a tracker driven seeding algorithm [81, 86], in order to increase the efficiency for low p_T and non isolated electrons. This algorithm starts from the high purity tracks, and makes use of the Particle Flow (PF) event reconstruction, which exploits the fine ECAL granularity. The detailed description of the PF algorithm and its commissioning can be found in References [87–89]. The PF event reconstruction combines the information from all subdetectors to identify and individually reconstruct all particles produced in the collisions.

The tracker driven seeding algorithm can be illustrated with two extreme cases. When an electron does not radiate energy by bremsstrahlung it gives rise to a single cluster in ECAL and its track is well reconstructed by the standard Kalman Filter. The track can then be matched with a cluster and its momentum compared to the cluster energy forming an E/p ratio. If this ratio is close to the unity, the seed of the track is promoted to electron seed. Alternatively, when an electron radiates a significant amount of energy due to the bremsstrahlung, the standard Kalman Filter is not able to follow the change of curvature, and the track has a small number of hits and a large χ^2. Thus, using the tracker as a preshower and exploiting the differences between a pion track and an electron track reconstructed with the standard Kalman Filter algorithm, the electron track can be selected. The variety of situations between the two extreme cases requires a more sophisticated treatment. In practice the algorithm applies a first selection based on the number of hits associated to the tracks and on the value of χ^2_{KF} (chi-square of the Kalman Filter track). Then a light GSF refit is carried out. The GSF refit χ^2_{GSF}, the χ^2_{KF}/χ^2_{GSF} ratio, the number of hits, the energy loss as measured by the track, as well as the quality of the ECAL cluster track matching are fed into a *MultiVariate Analysis* producing a global identification variable called MVA [86] which is used as a discriminator, as described in the next section.

Both the ECAL and tracker driven seeding algorithms are used for all the p_T and η ranges. In case both of them are successful, the ECAL seeding is taken. The use of tracker driven seeding brings additional efficiency both for isolated and non isolated electrons, in particular in the ECAL crack regions ($\eta \simeq 0$ and $|\eta| \simeq 1.5$) and at low p_T, as shown in Fig. 4.8.



(a) (b)

Figure 4.8: Electron seeding efficiency (solid line) as a function of (a) generated electron η and (b) generated electron p_T for a sample of electrons with uniform distribution in η and p_T and for $p_T > 2$ GeV. The individual contributions from the ECAL driven (dashed line) and from the tracker driven (dotted line) seeding algorithms are also shown, as well as a zoom of the region $p_T < 11$ GeV. [81].

4.2 Electron preselection

Electron candidates are built from the reconstruction of GSF tracks and their associated superclusters. In the case of electrons with ECAL driven seeding, the associated supercluster is simply the supercluster that initiated the seed reconstruction. For the electrons with seeds found only by the tracker seeding algorithm, a tracker driven bremsstrahlung recovery algorithm and identification of the "electron cluster" developed in the context of the Particle Flow reconstruction is used. The tracker driven algorithm runs on all GSF tracks to produce superclusters by grouping together Particle Flow clusters which are matched with presumed "photon" lines, tangent to the electron trajectory at any of the tracker measurements layers. The electron cluster, defined as the cluster matched with the outermost track state, is finally added to the supercluster. This procedure leads to a new collection of superclusters that are used to build the electron candidates.

Electron candidates, formed by the association of a GSF track and its associated supercluster, are then preselected using available track-cluster matching observables in order to reduce the rate of jets faking electrons. The preselection is made very loose so as to efficiently reconstruct electrons and satisfy a large number of possible analyses [81]. For electrons that have an ECAL driven seed, the following cuts are already applied at the seeding level:

- $E_T > 4\,\mathrm{GeV}$, where E_T is the supercluster transverse energy,
- $H/E < 0.15$, where H is the energy deposited in the HCAL towers in a cone of radius $\Delta R = 0.15$ centered on the electromagnetic supercluster position and E is the energy of the electromagnetic supercluster.

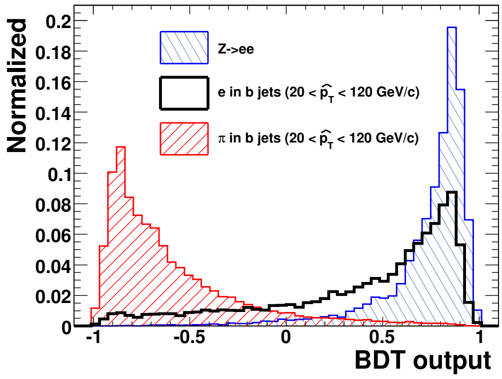


Figure 4.9: Output of the multivariate analysis for isolated electrons in a simulated $Z \to ee$ sample (blue), for non-isolated electrons in b jets (black) and for pions (red). The histograms are normalized to unity [86].

In addition to this selection, the following requirements are also applied only on the electron candidates with ECAL driven seeds:

- $|\Delta\eta_{in}| = |\eta_{SC} - \eta_{in}^{extrap.}| < 0.02$, where η_{SC} is the energy weighted position in η of the supercluster and $\eta_{in}^{extrap.}$ is the η coordinate of the position of closest approach to the supercluster position, extrapolated from the innermost track position and direction,

- $|\Delta\phi_{in}| = |\phi_{SC} - \phi_{in}^{extrap.}| < 0.15$, where ϕ_{SC} is the energy weighted position in ϕ of the supercluster and $\phi_{in}^{extrap.}$ is the ϕ coordinate of the position of closest approach to the supercluster position, extrapolated from the innermost track position and direction.

In case of tracker driven only electrons, the global identification variable MVA obtained from the multivariate analysis is used. For the electron candidates preselection with multivariate analysis many more variables are taken into account than for the tracker driven seeding, like the shape of the ECAL clusters, the fraction of radiated energy, the resolution σ_{p_T}/p_T and the χ^2 of the GSF tracks etc. [86]. Electron candidates in these cases are required to satisfy:

- MVA > -0.4, where MVA is the output of multivariate analysis.

The distribution of the MVA variable used for the preselection of electrons with tracker driven only seed is shown on Fig. 4.9 for simulated electrons in b-jets and from $Z \to ee$ decays, as well as for pions in b-jets. A very good separation between electrons and pions is achieved when the electrons are isolated. The electron-pion separation remains good also for electrons in jets [81].

Fig. 4.10 shows the electron reconstruction efficiency after preselection as a function of generated electron η and p_T for isolated electrons with uniform η and p_T distributions with $p_T > 2\,\text{GeV}$.

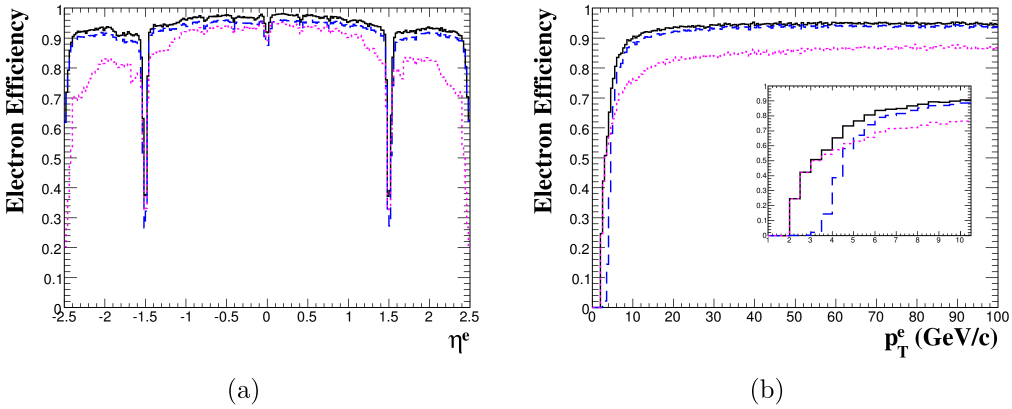


Figure 4.10: Electron efficiency after preselection (solid line) as a function of (a) generated electron η and (b) generated electron p_T for a sample of di-electrons events with uniform distribution in η and p_T and with $p_T > 2\,\text{GeV}$. The individual contributions from ECAL seeded electrons (dashed line) and from tracker seeded electrons (dotted line) are also shown, as well as a zoom of the region $p_T < 10.5\,\text{GeV}$ [81].

4.3 Electron charge determination

Electron charge identification suffers from the conversion of radiated photons and more generally from the showering of primary electrons, in particular when this happens early in the detector. Bremsstrahlung in matter can indeed lead to deflections and kinks in the track of the electron and this results in a wrong measurement of the sign of the track curvature. For these reasons, the charge mis-identification (or charge mis-ID) almost linearly increases in the region $1.1 < |\eta| < 2.5$, following the distribution of the material budget of the pixel detectors which reaches $\simeq 0.6 X_0$ at $|\eta| = 2.5$. Moreover, since the electron charge is determined looking at the track bending in the magnetic field, the charge mis-ID from the GSF track charge also increases as a function of p_T (the higher the p_T, the lower the bending radius) [81].

The charge determination can be improved by combining several charge estimates in a majority method that takes the value from the two out of three estimates that are in agreement. The three charge estimates used are:

- GSF track charge,
- general track charge,
- supercluster charge.

For the firsts two estimators, the electron charge is evaluated looking at the track bending. The electron general track is obtained by matching the GSF track

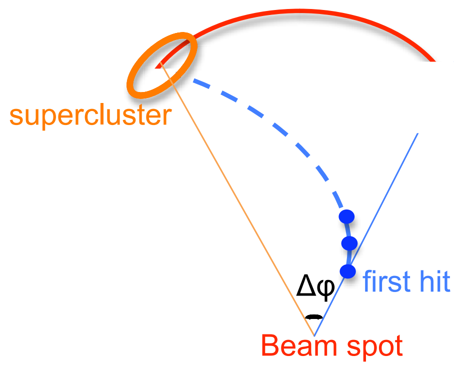
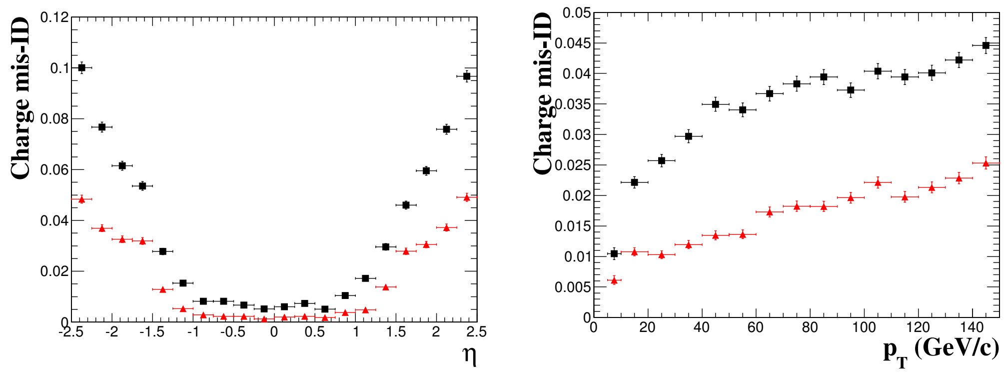


Figure 4.11: The supercluster charge is determined by looking at the ϕ separation between the vector joining the beam spot and the supercluster position (orange line) and the one joining the beam spot and the first hit of the electron track (blue line).



(a) (b)

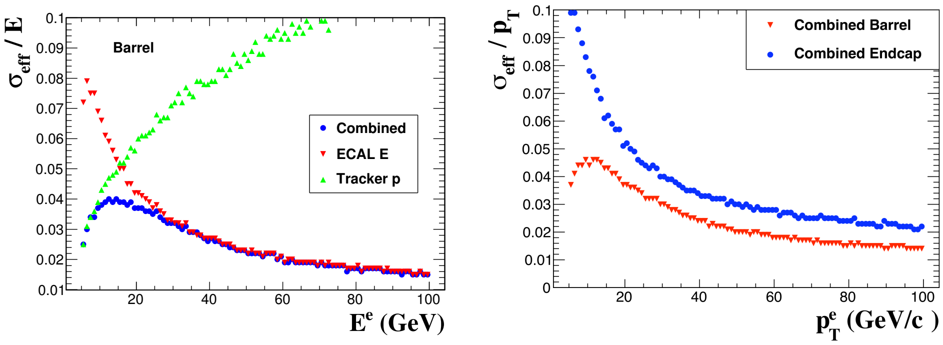
Figure 4.12: Electron charge mis-ID as obtained from the GSF track charge (squares, black) and from the majority method (triangles, red) as a function of (a) generated electron η and (b) generated electron p_T. The sample used is made of simulated back-to-back electrons with uniform p_T and η distributions [81].

with general tracks as reconstructed for pions and muons, asking for at least one hit shared in the innermost part of the tracks. The supercluster charge is obtained by computing the sign of the ϕ difference between the vector joining the beam spot and the supercluster position and the one joining the beam spot and the first hit of the electron track, as shown in Fig. 4.11. The charge identification performance as obtained both from the curvature of the GSF electron track and from the majority method is shown on Fig. 4.12 as a function of η and p_T, from a sample of simulated back-to-back electrons with uniform p_T and η distributions. One can see a significant improvement in the charge determination obtained by using the majority method, by a factor ~ 2 or more over the entire p_T range.

4.4 Electron momentum determination

The electron momentum magnitude is obtained from the combination of the ECAL and the tracker measurements, so as to take advantage of the track momentum estimate particularly in the low energy region and/or in the ECAL crack regions. Following [77, 81], the measurements from ECAL and tracker are either combined or only one measurement is used according to their sensitivity to bremsstrahlung induced effects. A weighted mean is used that involves the error determination on the supercluster energy and the error on the track momentum from the GSF fit. The electron momentum magnitude is therefore defined as this weighted mean of E and p when $|E/p-1| < 2.5 \cdot \sigma_{E/p}$, with weights computed as the normalized inverse of the variance of each measurement. In all the other cases the ECAL measurement is used, except for some particular cases for which the tracker measurement is taken (e.g. for electrons in the ECAL cracks with $E < 60\,\mathrm{GeV}$ and $E/p < 1 - 2.5 \cdot \sigma_{E/p}$) [81].

As expected, the tracker measurement prevails at low energies as well as in the regions where the precision of the ECAL measurement is poor. The performances of the combined electron momentum are illustrated in Fig. 4.13 which presents the normalized momentum effective RMS of the combined estimate as well as of the ECAL and tracker measurements alone for electrons in the ECAL barrel. Electrons



(a) (b)

Figure 4.13: Performances of the combined momentum estimate: (a) effective momentum resolution for the ECAL, the tracker and the combined momentum estimates as a function of the electron generated energy for electrons in the ECAL barrel and (b) effective transverse momentum resolution for electrons in the ECAL barrel and electrons in the ECAL endcaps. Electrons are from a simulated sample of di-electron events with uniformly distributed transverse momentum between 2 and 150 GeV [81].

are from a simulated sample of di-electron events with uniformly distributed transverse momentum between 2 and 150 GeV. The precision is clearly improved by using the combined estimate with respect to the ECAL only measurement for energies below $\simeq 25 - 30\,\mathrm{GeV}$, and reaches an effective RMS of $\sim 1.5\%$ for $E = 100\,\mathrm{GeV}$. This value of the effective RMS is larger than the one predicted by the Eq. (2.8) of Chapter 2 for $E = 100\,\mathrm{GeV}$ ($\sim 0.5\%$). The energy resolution of the reconstructed electrons shown in Fig. 4.13 indeed takes into account also the bremsstrahlung emission that occurs between the impact point and ECAL and that must be recovered by the electron reconstruction algorithms. On the contrary the resolution shown in Eq. (2.8) is an intrinsic property of the ECAL crystals and has been measured with electron beams incident directly on the crystals [53].

The normalized effective transverse momentum resolution for electrons in the ECAL barrel and endcaps is also shown in Fig. 4.13.

4.4.1 The `GSFElectron` object

The CMSSW object which contains a reconstructed electron candidate which pass the preselection cuts is called `GSFElectron`, because it is associated to an instance of GSF track. In this object are stored all the information about the electron candidate, like the charge (obtained from the majority method), the momentum, the direction evaluated as the $\eta - \phi$ coordinates extrapolated to the vertex, and also the reference to the associated GSF track and ECAL supercluster. All the variables used for the preselection are also stored in the object. The `GSFElectrons` are qualified as "ECAL-driven" or "tracker-driven" depending if they have been reconstructed starting from seeds found by using the ECAL driven or the tracker driven algorithm.

4.5 Electron Isolation

A very useful tool available in the pp environment to distinguish between leptons which are produced in high p_T process and leptons produced copiously in the background QCD or in other jet rich processes is the lepton isolation. Electrons coming from a Z boson decay are expected to be fairly isolated.

The isolation variable is a check of how much activity accompanies a lepton as it traverses through the detector. This activity can be measured in three different subdetectors of CMS: the tracker, ECAL and HCAL. A well defined isolation variable will accurately sum the energy surrounding a particle while making sure at the same time that none of the energy associated with the actual particle leaks into this sum.

For the present analysis the *Combined Relative Isolation* [90] was used, which combines the activity in all the three subdetectors in a unique isolation variable. The isolation deposits *TrkIso*, *ECALIso* and *HCALIso* in the three subdetectors

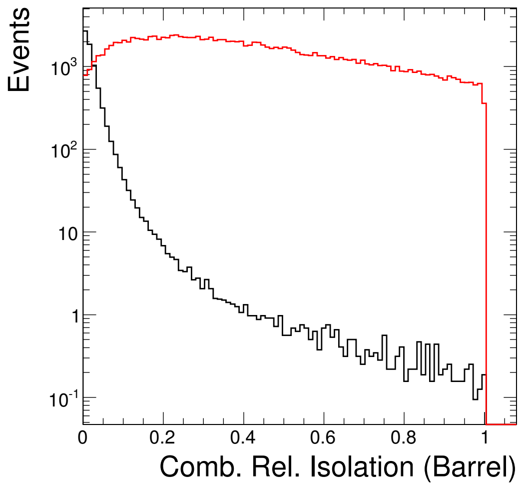


Figure 4.14: Combined Relative Isolation evaluated for $p_T > 20\,\mathrm{GeV}$ barrel electrons, taken from all the QCD `PYTHIA` samples (red line) and from the $Z + jets$ `MADGRAPH` sample (black line). For this plot the MC samples are normalized to a reference luminosity of 50 pb^{-1} (see Sec. 3.6 for a MC datasets description).

are defined as the scalar sum of the transverse momentum of all the reconstructed objects around the electrons (tracks or calorimetric towers), within a cone of radius $\Delta R = 0.3$, where $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$, and considering a veto cone of $\Delta R = 0.015$ around the electron to remove the energy associated with the actual particle [90]. The Combined Relative Isolation variable is therefore defined as:

$$CombRelIso = \frac{TrkIso + ECALIso + HCALIso}{p_T}, \tag{4.5}$$

where p_T is the transverse momentum of the electron expressed in GeV. Such an isolation variable takes into account that a high lepton p_T is itself a signature of interesting events, so we may relax the isolation cut according to p_T. A comparison between the CombRelIso variable as measured in the Barrel for the MC signal sample (`MADGRAPH` *Z+jets* sample) and for the MC QCD background samples (all the `PYTHIA` *QCD* samples) is shown in Fig. 4.14. This plot shows that an appropriate choice of cut on the CombRelIso variable can have a great impact on the background rejection.

4.6 The `PATElectron` object

As described in Sec. 2.4.2, data and MC are reconstructed by using the Physics Analysis Toolkit. The reconstructed events are therefore organized in PAT-tuples, which contain several `PATObjects`, one for each reconstructed physic object, e.g. electrons, jets, muons etc.

For what concerns the electrons, a `PATElectron` object is defined, which contains all the information stored in the `GSFElectron` object together with several extra data, for example:

- Isolation deposits, as defined in the previous section.
- Match on generator level (for MC samples). The matching of a RECO particle with the correspondent GEN particle is performed by looking for a GEN particle within a cone of ΔR around the RECO particle.
- Match on trigger level. The matching of a RECO particle with the trigger primitives (i.e. particles which turned on one or more trigger bits) is performed by looking for a primitive within a cone of ΔR around the RECO particle.
- Shower shape variables, like $\sigma_{i\eta i\eta}$ [91]. This variable represents the width of the ECAL cluster along the η direction and it is defined as:

$$\sigma_{i\eta i\eta} = \frac{\sum_i^{5\times5} w_i(\eta_i - \bar{\eta}_{5\times5})^2}{\sum_i^{5\times5} w_i}\,, \tag{4.6}$$

where the index i runs over all the crystals in a 5×5 block of crystals centered on the seed crystal, η_i is the η position of the i^{th} crystal, measured in terms of crystal indexes within the cluster, $\bar{\eta}_{5\times5}$ is the energy weighted mean η of the 5×5 block of crystals and w_i is the weight of the i^{th} crystal and is defined as

$$w_i = 4.2 + \ln(E_i/E_{5\times5})\,, \tag{4.7}$$

where E_i and $E_{5\times5}$ are the energy of the i^{th} and the 5×5 block of the crystal respectively.

Cluster shape variables as $\sigma_{i\eta i\eta}$, together with the ratio between hadronic and electromagnetic energy H/E and track matching variables $\Delta\phi_{in}$ and $\Delta\eta_{in}$, are all called *electron identification* variables. All these information and many others are stored in the `PATElectron` object, which is the starting object of our analysis.

In order to reduce the stored information, the following preselection requirements are also applied on the electrons to be stored as `PATElectrons`:

- electron $p_T > 8\,\mathrm{GeV}$,
- electron $\eta < 3$.

4.7 Electron HLT

The progressive rise of LHC instant luminosity from the beginning of 7 TeV collisions to the rest of 2010 made necessary the gradual change of the L1 trigger thresholds and High Level Triggers (HLT) used (see Sec. 2.3), from more relaxed trigger criteria to tighter ones. To extract $Z \to ee$ events, a set of triggers corresponding to the lowest threshold unprescaled single electron trigger available was used.

For the first 8.3 pb^{-1} of data the delivered instantaneous luminosity allowed for a lower L1 trigger threshold. For this first set of data the `L1_SingleEG5` was therefore used for HLT seeding, requiring one electromagnetic object above a threshold of $E_T > 5\,\mathrm{GeV}$. An increase in the collision rate made it necessary to move to the higher threshold `L1_SingleEG8` path as L1 seed for High Level electron triggers, which required an electromagnetic object with $E_T > 8\,\mathrm{GeV}$. In Table 4.1 all the High Level Triggers used for data analysis of this work are shown, along with their respective L1 trigger seeds. Two different matching windows may be used by the HLT algorithm while searching for pixel hits consistent with a track: a large window with very relaxed cuts ("LW") and a small window with tighter cuts ("SW"). In addition, as shown in Table 4.1, the single electron energy threshold required by the HLTs used increased from $E_T > 10\,\mathrm{GeV}$ up to $E_T > 17\,\mathrm{GeV}$ during 2010. The HLT selections were also tightened as data rates were increasing in order to keep the recorded data rates manageable. In Table 4.2 and Table 4.3 the additional cuts required by the High Level Triggers starting from run number 141956 are shown.

For the analysis of MC samples, only one of the above High Level Triggers is required, i.e. the `HLT_Ele17_SW_TighterEleIdIso1_L1R_v3`. This different approach however does not affect the efficiency estimation, since for the present analysis the efficiency obtained from MC simulations is rescaled to an efficiency measurement obtained by using a data driven method, as will be described in Chapter 6. Besides, for what concerns the comparison between data and MC, the possible discrepancy between the two different approaches is expected to be within a few percents.

4.8 Z(ee) events selection

In this section the selection cuts used to select $Z \to ee$ events are described. For early LHC data taking a cut based approach was chosen due to its simplicity

Run range	High Level Trigger	Level-1 seed
136033 - 139980	HLT_Ele10_LW_L1R	L1_SingleEG5
140058 - 141882	HLT_Ele15_SW_L1R	L1_SingleEG5
141956 - 144114	HLT_Ele15_SW_CaloEleId_L1R	L1_SingleEG5
146428 - 147116	HLT_Ele17_SW_CaloEleId_L1R	L1_SingleEG8
147196 - 148058	HLT_Ele17_SW_TightEleId_L1R	L1_SingleEG8
148819 - 149064	HLT_Ele17_SW_TighterEleIdIso1_L1R_v2	L1_SingleEG8
149181 - 149442	HLT_Ele17_SW_TighterEleIdIso1_L1R_v3	L1_SingleEG8

Table 4.1: High Level electron triggers used to extract $Z \to ee$ events from data.

HLT type	H/E	$\Delta\eta_{in}$	$\Delta\phi_{in}$	$\sigma_{i\eta i\eta}$
CaloEleId	0.15	-	-	0.014 (0.035)
TightEleId	0.15	0.01	0.08	0.012 (0.032)
TighterEleIdIso1	0.05	0.008 (0.007)	0.1	0.011 (0.031)

Table 4.2: High Level Trigger identification requirements by path. Thresholds for barrel (endcap, if different) are shown. The definitions of these variables are discussed in Sec. 4.6.

HLT type	$ECALIso/p_T$	$HCALIso/p_T$	$TrkIso/p_T$
CaloEleId	-	-	-
TightEleId	-	-	-
TighterEleIdIso1	0.125 (0.075)	0.05	0.15 (0.1)

Table 4.3: High Level Trigger isolation requirements by path. Thresholds for barrel (endcap, if different) are shown. The definitions of these variables are discussed in Sec. 4.5.

and good efficiency. Ultimately, multivariate techniques may provide higher performance, but a cut based selection can be a useful tool to understand the data and compare them directly with the Monte Carlo. For what concerns the isolation and electron identification variables the cuts have been optimized in order to retain the signal on inclusive $W \to e\nu$ events in Monte Carlo samples with various degrees of efficiency, called Working Points [92]. A Working Point is therefore identified by the correspondent single electron efficiency in simulated $W \to e\nu$ events. For this analysis an asymmetric selection is applied on the two electrons originating from the Z boson, in terms of p_T, isolation and electron identification variables.

For the *leading electron* (i.e. the electron of the couple with higher p_T) a p_T tighter cut is required and Working Point 80 (WP80, 80% of single electron efficiency) is used. For the *second electron*, with lower p_T, a looser cut is applied on p_T and the Working Point 95 (WP95, 95% of single electron efficiency) is required. These selection criteria are chosen in order to obtain at the same time good efficiency and good background rejection.

In the following the selection cuts are described in detail, in the same order in which they are applied. In the plots shown hereafter, the MC samples described in Sec. 3.6 are reported as follows:

- QCD background, which groups all the QCD EM Enriched and QCD b, $c \to e$ contributions,
- $t\bar{t} + jets$ background, which corresponds to the $t\bar{t}$ `PYTHIA` sample,
- EWK background, which groups $Wl\nu$, WW, WZ and ZZ contributions,

- $Z \to ee$ signal, which corresponds to the $Z \to l^{+}l^{-} + jets$ `MADGRAPH` sample with Z2 tune.

In Chapter 6 the measurement of the selection efficiency, obtained by using both MC and data driven methods, will be shown.

1. Acceptance (Acc)

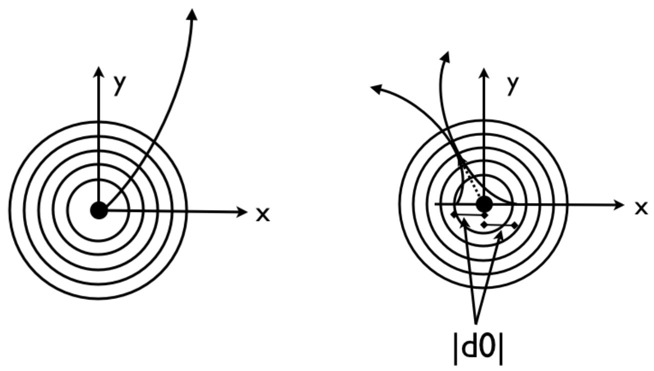
With the term "acceptance" a series of cuts that are mainly related to the electron kinematic variables is indicated, and that can be implemented directly on the output of an event generator. In other words, acceptance cuts are supposed to have little dependence on the detector. In this first step the requests are:

- two electrons with opposite charge;
- electron η must be in the fiducial region of ECAL and Tracker: $|\eta| < 1.44 \cup 1.57 < |\eta| < 2.50$. In this way the transition region between ECAL barrel and endcaps is excluded, which presents worse reconstruction efficiency due to the presence of a high quantity of material budget which increases the bremsstrahlung emission (see Sec. 4.1.3). The limit of $|\eta| < 2.5$ is due to the Tracker acceptance;
- the leading electron p_T must satisfy the requirement $(p_T)_{lead} > 20\,\text{GeV}$;
- the second electron p_T must satisfy the requirement $(p_T)_{sec} > 10\,\text{GeV}$;
- the invariant mass of the electron couple must be within the mass range $60\,\text{GeV} < m_{ee} < 120\,\text{GeV}$.

2. Trigger (Trg)

For each run range of Table 4.1 the associated High Level Trigger as shown in the list is required to be fired. As stated in Sec. 4.7 several triggers are needed since HLT changed during data taking, according to the luminosity raising. In addition to the trigger bit ON, the matching of the leading electron of the couple with the online trigger object is also required. The matching is performed following geometric criteria, i.e. requiring that the leading electron and the object that fired the trigger are separated in the $\phi-\eta$ plane by a distance $\Delta R < 0.5$. The information on trigger matching is stored in the PAT-tuples.

For what concern the MC samples only the tighter HLT of Table 4.1 is required, both for trigger bit and trigger matching as done for data. However, as stated in Sec. 4.7, the effect of this simplification does not affect the final efficiency estimation, which will be obtained by using a data driven method (described in Chapter 6), and for the data-MC comparison the discrepancy between the two different approaches is expected to be within a few percents.



(a) (b)

Figure 4.15: Scheme of the impact parameter d_0 for an electron coming from the Interaction Point (a) and for an electron coming from a photon conversion (b).

Conversion Rejection	WP80	WP95
Missing Hits $\leq$	0	1
$\sqrt{\Delta\cot^2(\theta) + \mathrm{Dist}^2}$	0.02	-

Table 4.4: Conversion Rejection cuts applied on leading (WP80) and second (WP95) electrons.

3. Impact Parameter (Imp)

The impact parameter d_0 is defined as the minimum distance of a track from the primary vertex, measured in the transverse plane $x - y$. The electrons of the selected couple are required to have $|d_0| < 0.02\,\mathrm{cm}$.

The electrons coming from photon conversions have on average higher values of d_0 than the electrons coming from the Z. In fact the photon conversions happen in the beam pipe or in the tracker. This effect is shown in Fig. 4.15, where the electron candidate track is extrapolated toward the nominal Interaction Point. Such a cut limits therefore the collection of electrons coming from photon conversions as well as of electrons coming from the decays of long lived particles. In Fig. 4.16 the impact parameter distribution for the electrons coming from Z candidates is shown, together with the cut applied on it. For this plot all the selection cuts have been applied but the one on the impact parameter.

4. Conversion Rejection (Conv)

In order to further reject the background due to electrons from photon conversions in tracker material a conversion rejection tool is applied [93]. Electrons with missing hits in front of the innermost valid track are rejected as originated from a

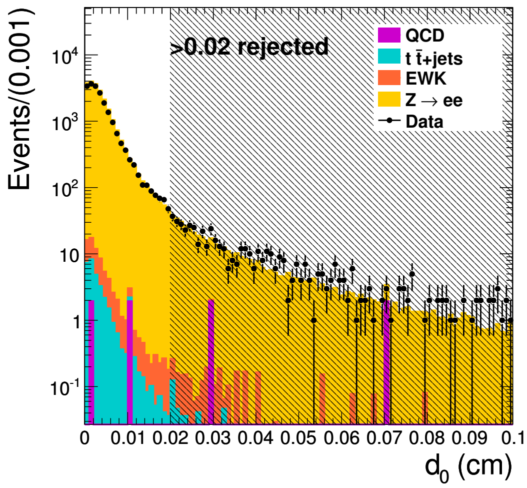


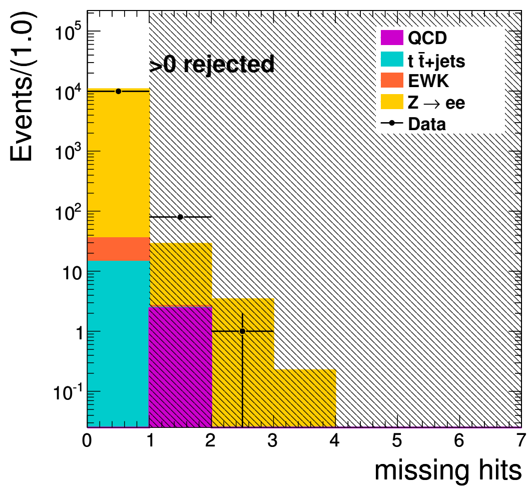
Figure 4.16: Impact Parameter: distribution of the d_0 variable (expressed in cm) for the electrons of the selected couples. All selections applied but the cut on d_0.

conversion that occurred in the tracker material. The tracks are also inspected to locate possible conversion partner tracks. To be identified as a conversion partner, the track must have:

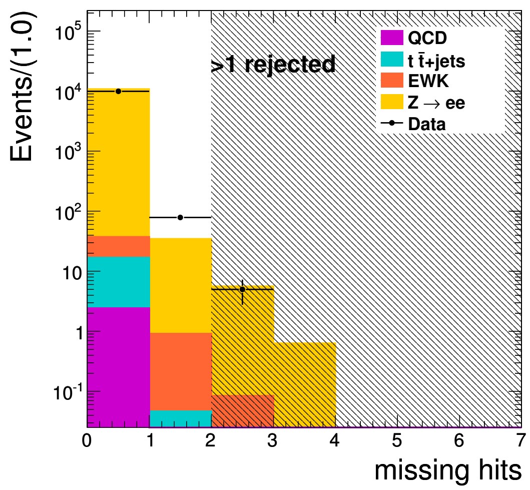
- opposite sign as the electron track,
- approximately the same $\cot(\theta)$ as the electron track,
- small distance ("Dist") in the transverse plane, where "Dist" is defined as the distance in the $x - y$ plane between the two tracks when the track in question and the electron GSF track would be parallel when extrapolated. All neighboring tracks in a $\Delta R = 0.3$ cone are considered.

For GSF electrons, as the ones considered in the present analysis, the cut on $\cot(\theta)$ and Dist is applied looking at the quadratic sum between $\Delta\cot(\theta) = |\cot(\theta_{el}) - \cot(\theta_{trk})|$ and Dist. A track is therefore considered a conversion partner track, and then is rejected, if $\sqrt{\Delta\cot^2(\theta) + \mathrm{Dist}^2}$ is lower than a certain value.

Different selections are chosen for WP80 (leading) and WP95 (second) electrons in terms of missing hits, while the cut on $\sqrt{\Delta\cot^2(\theta) + \mathrm{Dist}^2}$ is applied only on the WP80 electrons. The cut values are shown in Table 4.4. In Fig. 4.17 the distributions of the missing hits for the leading and the second electrons are shown, together with the cuts applied. For this plot all the analysis selections have been applied but the conversion rejection itself. In Fig. 4.18 the distribution of $\sqrt{\Delta\cot^2(\theta) + \mathrm{Dist}^2}$ for the leading electrons is shown, with all the cuts applied but the isolation and the conversion rejection itself. Also the isolation cut is not required in Fig. 4.18 in order to highlight the effect of the conversion rejection on the QCD background.



(a)



(b)

Figure 4.17: Conversion Rejection: missing hits for leading (a) and second (b) electrons. All selections applied but the conversion rejection.

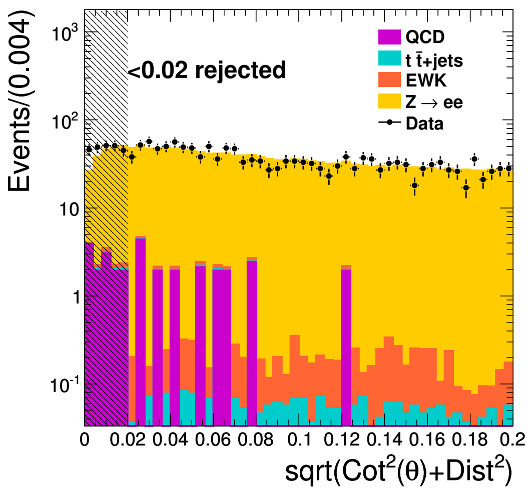
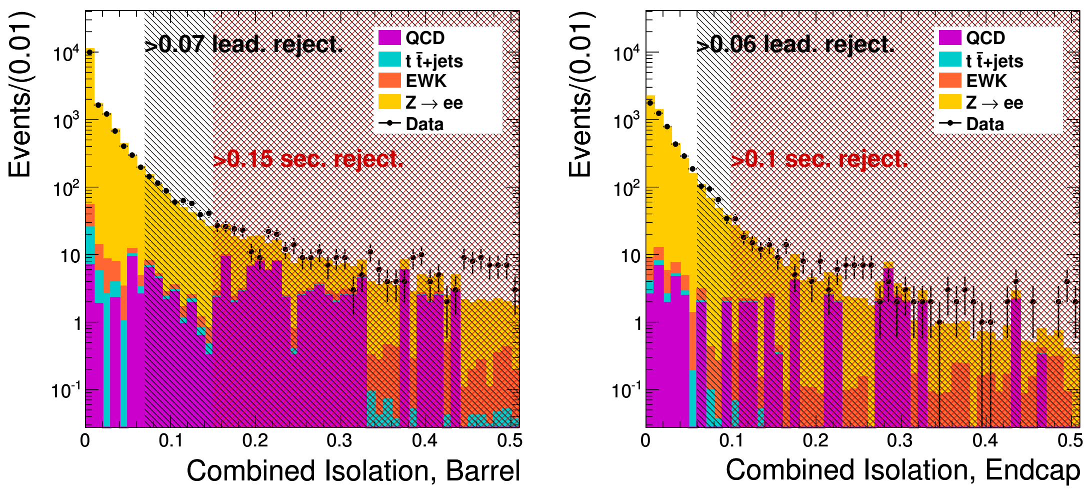


Figure 4.18: Conversion Rejection: distribution of the quadratic sum between $\cot(\theta)$ and Dist for leading electrons. No cuts applied on second (WP95) electrons. All selections applied but the isolation and the conversion rejection itself. Also the isolation cut is not required in order to highlight the effect of the conversion rejection on the QCD background.

5. Isolation (Iso)

As described in Sec. 4.5 the electron isolation provides the most important contribution on the background rejection. For this work the Combined Relative Isolation (CombRelIso) variable as defined in Eq. (4.5) is used, which combines the



(a) (b)

Figure 4.19: Isolation: distribution of the Combined Relative Isolation variable for Barrel (a) and Endcap (b) regions. Cuts on leading (black) and second (red) electrons are also shown. All selections applied but isolation one.

Comb. Rel. Isolation	WP80	WP95
Barrel	0.07	0.15
Endcaps	0.06	0.1

Table 4.5: Combined Relative Isolation cuts applied on leading (WP80) and second (WP95) electrons. Barrel and Endcap cuts are shown.

activity measured in all the three subdetectors in a unique isolation variable. Electrons with a Combined Relative Isolation higher than an appropriate value are considered as not isolated electrons and then rejected.

Different sets of isolation cuts are chosen for WP80 and WP95 electrons, depending also on detector regions (barrel and endcaps). In Table 4.5 the cuts applied on the CombRelIso variable are shown, both for the leading (WP80) and the second (WP95) electrons of the couples. In Fig. 4.19 the distribution of the CombRelIso variable is shown both for barrel and endcaps together with the applied cuts, when all the analysis selections have been required but the electron isolation itself.

6. Electron Identification (EiD)

The electron identification [94] makes use of several quality cuts in order to distinguish true electrons from fakes, and to distinguish between electrons coming from a Z decay from jets faking electrons. The variables used for the electron identification are H/E, $\Delta\eta_{in}$, $\Delta\phi_{in}$ and $\sigma_{i\eta i\eta}$. They concern respectively the electron energy fraction deposited in hadronic calorimeter (which is expected to be small), the geometrical matching between the electron track and the supercluster, and the calorimeter shower shape. A more detailed description of these variables is already given in Sec. 4.2 and Sec. 4.6.

Selection cuts on these variables are conveniently tuned for WP80 and WP95 electrons and for detector regions (barrel or endcaps). Table 4.6 shows the cuts applied on each electron identification variable. In Fig. 4.20, Fig. 4.21, Fig. 4.22 and Fig. 4.23 the distributions of each electron identification variable are shown, when all the analysis selections have been required but the electron identification itself.

4.9 Z candidate distributions after cuts

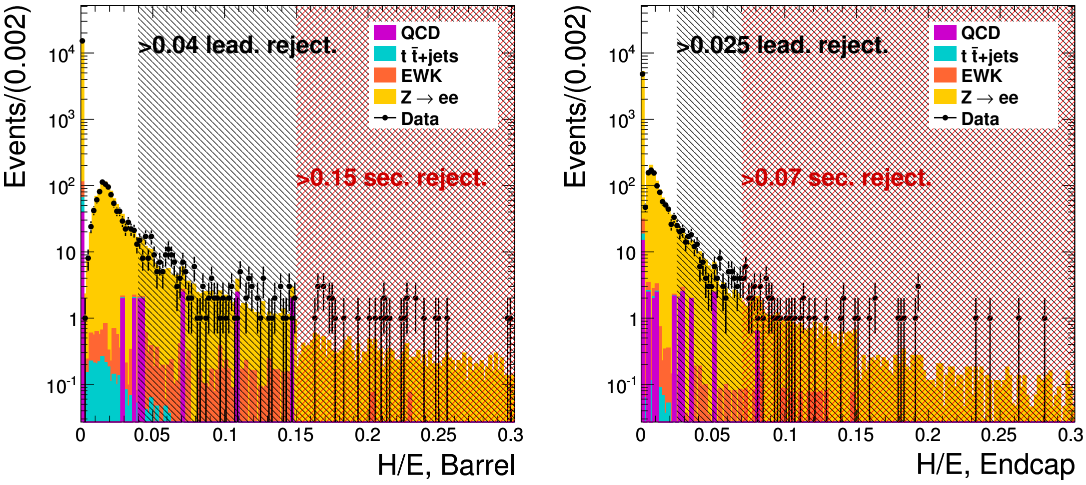
Each electron couple which passed all the selection cuts described in the previous section is considered as a Z candidate. In this section the characteristics of the electron couples selected in the 36.2 pb^{-1} data sample with all cuts applied are shown.

In Fig. 4.24 and Fig. 4.25 the p_T and η distributions are shown both for the leading and the second electrons of the selected couples, after all cuts applied. The application of the whole set of selections leaves an event sample where most of the background has been suppressed, as shown in both Fig. 4.24 and Fig. 4.25. Notwithstanding the uncertainties due to the limited luminosity of these MC samples, no QCD events are left after all selection cuts are applied. Both Fig. 4.24 and Fig. 4.25 show finally very good agreement between data and MC for what concern the selected electron kinematic variables, in all the kinematic region of interest.

In Fig. 4.26 the distribution of the invariant mass of the selected electrons is shown, while Fig. 4.27 shows the p_T and η distributions of the Z candidates. Also in this case the agreement between data and MC is good. Fig. 4.27 shows also that Z candidates are mainly boosted in the forward region. This effect is due to the fact that in pp collisions the Z boson is created from a valence quark and a quark of the

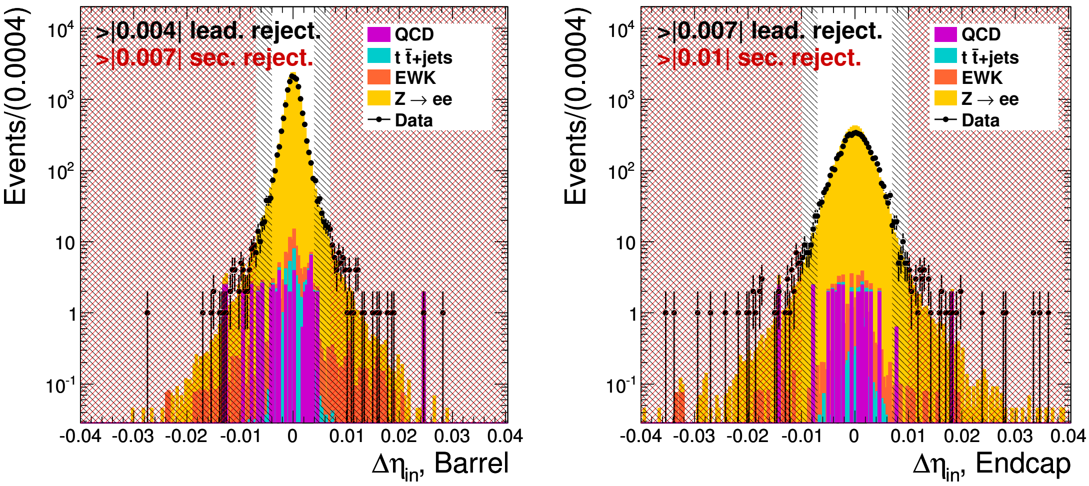
	WP80		WP95	
	Barrel	Endcap	Barrel	Endcap
H/E	0.04	0.025	0.15	0.07
$\Delta\eta_{in}$	0.004	0.007	0.007	0.01
$\Delta\phi_{in}$	0.06	0.03	-	-
$\sigma_{i\eta i\eta}$	0.01	0.03	0.01	0.03

Table 4.6: Electron Identification cuts applied on leading (WP80) and second (WP95) electrons. Barrel and Endcap cuts are shown.



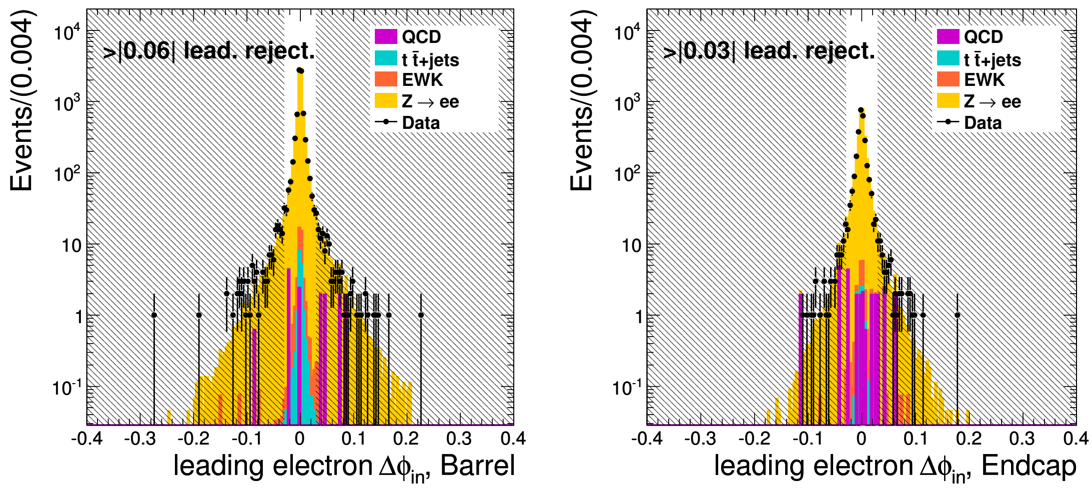
(a) (b)

Figure 4.20: Electron Identification: distributions of H/E variable, measured for leading and second electrons in barrel (a) and endcap (b). Cuts on leading (black) and second (red) electrons are also shown. All selections applied but electron identification itself.



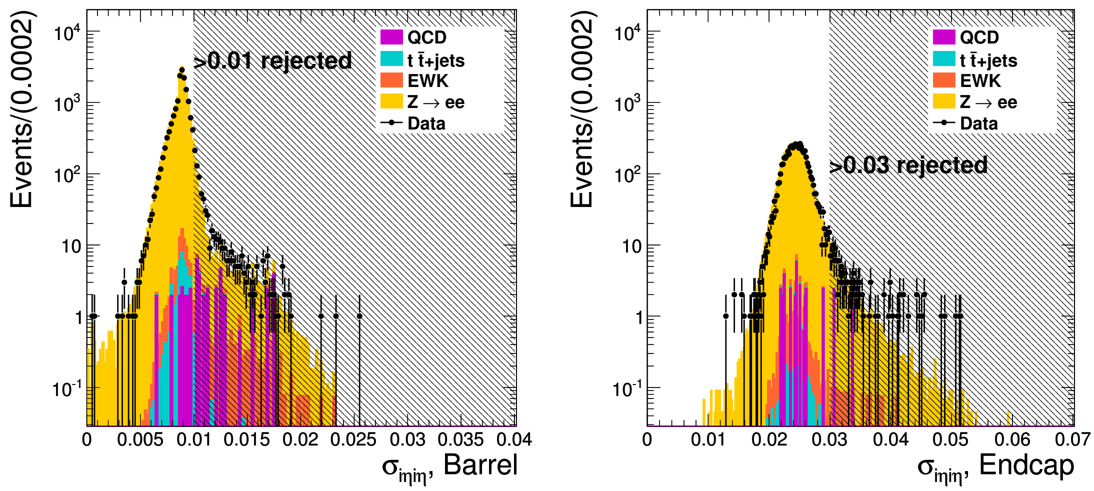
(a) (b)

Figure 4.21: Electron Identification: distributions of $\Delta\eta_{in}$ variable, measured for leading and second electrons in barrel (a) and endcap (b). Cuts on leading (black) and second (red) electrons are also shown. All selections applied but electron identification itself.



(a) (b)

Figure 4.22: Electron Identification: $\Delta\phi_{in}$ distributions for leading electrons in barrel (a) and endcap (b). All selections applied but electron identification itself.



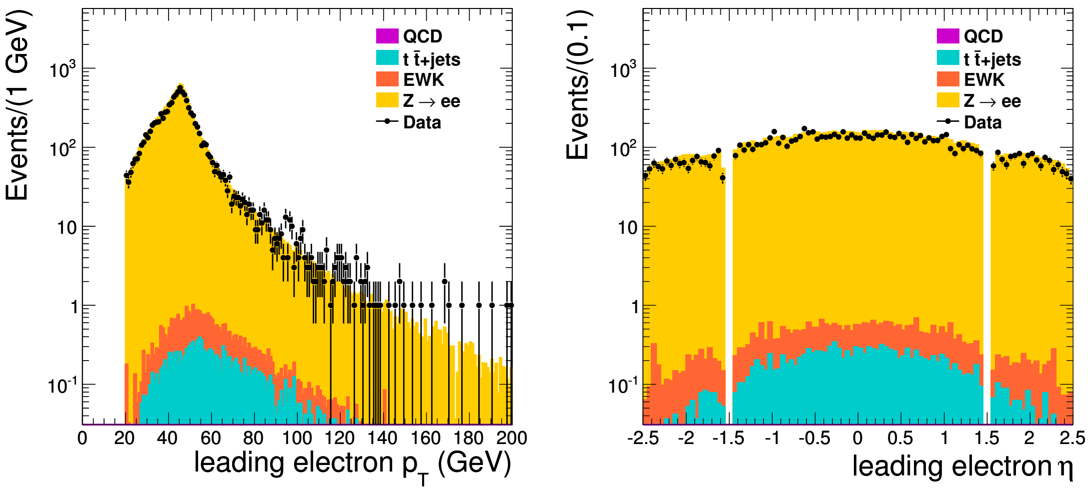
(a) (b)

Figure 4.23: Electron Identification: $\sigma_{i\eta i\eta}$ distributions in barrel (a) and endcap (b) (both leading and second electrons shown). Cuts shown are applied both on leading and second electrons (see Table 4.6). All selections applied but electron identification itself.

sea. If one or more jets are present in the event, the Z boson acquires a momentum in the central region.

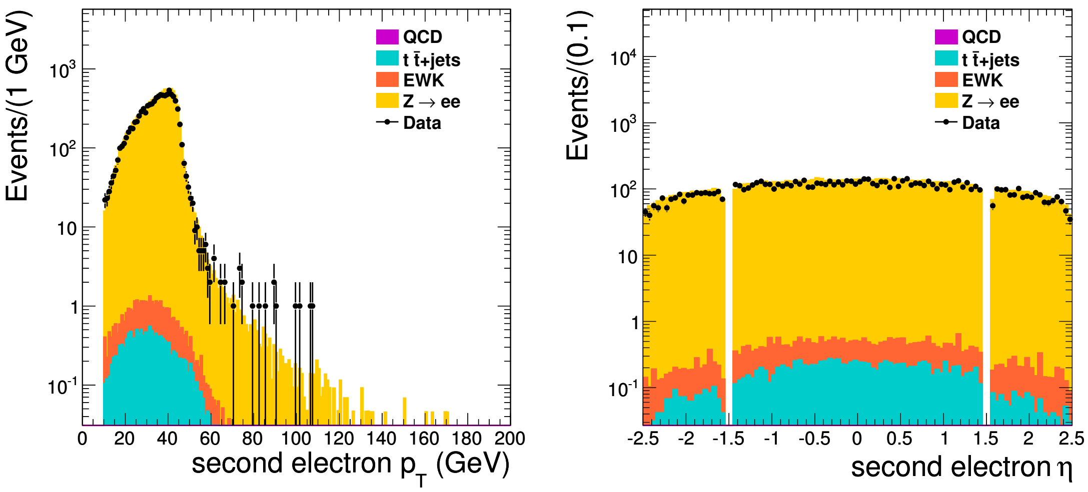
Finally Table 4.7 reports the number of Z candidates after each selection, both for data and for each MC sample normalized to the data luminosity (see Table 3.1 and Table 3.2 of Sec. 3.6 for the normalization factor of each MC sample). One can see how the application of the selection cuts gradually improves the agreement between data an MC. It is also evident the huge impact of the isolation cut on the background rejection. As shown in Table 4.7, no QCD events are left after all selection cuts are applied.

When all selections have been applied the selected Z candidates in data sample are 9717 ± 98, while the MC prediction is 10717 ± 16. This difference in the Z candidate yields must be attributed to a different efficiency of the selections between data and MC, as will be shown in Chapter 6.



(a) (b)

Figure 4.24: Distributions of p_T (a) and η (b) variables of the leading electron of the selected electron couples. All selection cuts applied.



(a) (b)

Figure 4.25: Distributions of p_T (a) and η (b) variables of the second electron of the selected electron couples. All selection cuts applied.

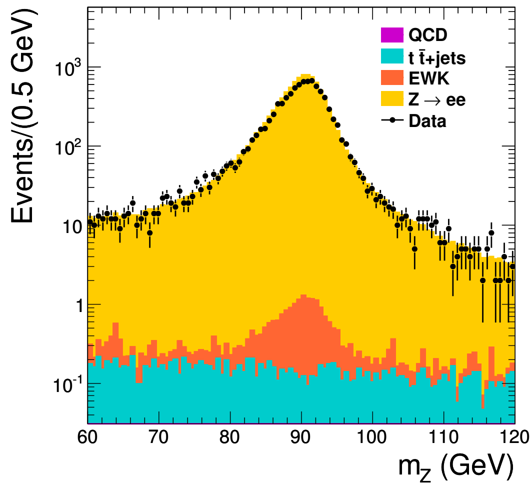


Figure 4.26: Di-electron invariant mass after full selection applied.

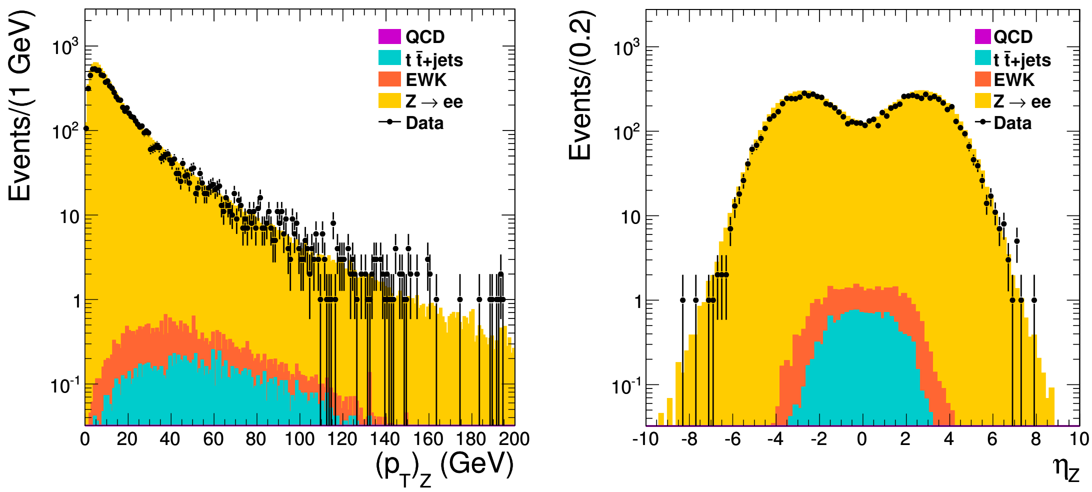


Figure 4.27: Distributions of the p_T (a) and η (b) variables evaluated for the selected Z candidates. Full selection applied.

Sample	Selections					
	Acc	**Acc** and **Trg**	**Acc**, **Trg** and **Imp**	**Acc**, **Trg**, **Imp** and **Conv**	**Acc**, **Trg**, **Imp** **Conv** and **Iso**	**Acc**, **Trg**, **Imp**, **Conv**, **Iso** and **EiD**
QCD	261833 ± 671	14591 ± 163	8880 ± 126	6833 ± 111	43.37 ± 9.5	-
$t\bar{t}$ + $jets$	218.8 ± 1.1	93.1 ± 0.7	71.7 ± 0.6	69.6 ± 0.6	17.0 ± 0.3	14.2 ± 0.3
EWK	1292.2 ± 9.7	875.0 ± 7.9	688.6 ± 7.0	665.7 ± 6.9	40.7 ± 1.2	20.6 ± 0.5
Z + $jets$	15371 ± 19	13769 ± 18	12900 ± 17	12621 ± 17	11797 ± 16	10682 ± 16
Total (MC)	278715 ± 671	29328 ± 164	22541 ± 128	20189 ± 112	11898 ± 19	10717 ± 16
Data (36 pb^{-1})	349215 ± 591	101271 ± 318	65067 ± 255	57263 ± 239	11163 ± 106	9717 ± 99

Table 4.7: Event number for each sample with its statistical uncertainty, evaluated for each selection applied. The MC samples are normalized to the integrated luminosity of the data sample, i.e. 36.2 pb^{-1}. No QCD events left after all cuts applied.

Chapter 5

Jet reconstruction and selection in $Z \rightarrow ee + jets$ events

Jets are an important tool in hadronic physics and they play a predominant role at the LHC. By defining jets one aims at accessing, from the final state particles, the hard parton level processes. Due to the large production cross section of jets at LHC, they will allow studies of new kinematic regimes, confronting predictions of perturbative QCD and probing physics processes within and beyond the Standard Model. In this chapter the jet reconstruction adopted in CMS and in this analysis is reported, together with the selection criteria used to select jets in $Z \rightarrow ee$ events.

5.1 Standard jet clustering algorithms in CMSSW

Since the definition of jet is not unique, several approaches are therefore available for jet clustering, each of them with different characteristics. Two broad classes of jet clustering algorithms exist. The first one consists in the "*conical recombination*", where jets are defined as dominant directions of energy flow. One introduces the concept of a stable cone as a circle of fixed radius R in the $\eta - \phi$ plane such that the sum of all the momenta of the particles within the cone points in the same direction as the centre of the circle. Cone algorithms attempt to identify all the stable cones. The second class is called "*sequential recombination*" and works by defining a distance between pairs of particles, performing successive recombinations of the pair of closest particles and stopping when all resulting objects are too far apart. Algorithms within that class differ in the definition of the distance.

As described in Chapter 1 Sec. 1.5.3, the main requirement for a jet clustering algorithm is the *InfraRed and Collinear safety* (IRC-safe), i.e. the algorithm must be independent of infrared and collinear corrections in perturbative QCD. The infrared corrections concern the emission of a gluon with an infinitely low energy (see Fig. 5.1), a phenomenon which happens with a probability equal to 1 in perturbative QCD. If the algorithm is IR-safe this soft emission must not affect the number of reconstructed jets. At the same time the collinear corrections concern the splitting

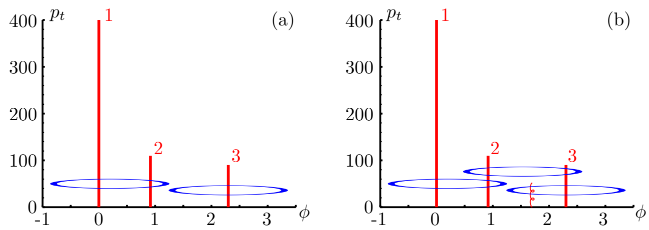


Figure 5.1: Stable cones found by an IR-unsafe algorithm for a 3-particle event (left) and for the same event with an additional infinitely soft gluon (right). The IR-unsafety points out that a different number of stable cones are found in the two cases [95].

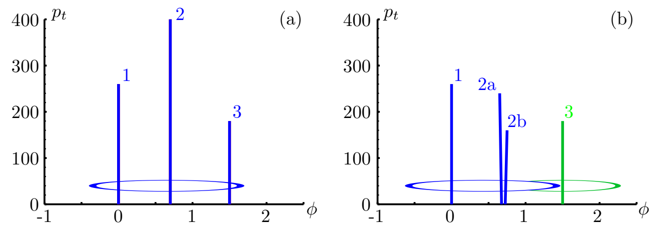


Figure 5.2: Jets found by a C-unsafe algorithm for a 3-particle event (left) and for the same event with a collinear splitting (right). The C-unsafety points out that a different number of stable cones are found in the two cases [95].

of a hard particle into two collinear particles, a process that also has a probability equal to 1 in perturbative QCD (see Fig. 5.2). Also in this case the algorithm is C-safe if this fact does not change the number of reconstructed jet. The IRC-safe algorithms present also the advantage of being less sensitive to the calorimeter noise (small energy deposits do not affect the number of reconstructed jets).

The standard IRC-safe algorithms adopted by CMS are the SISCone [95, 96], in the conical recombination class, and the k_t and anti-k_t [97] algorithms, in the sequential recombination class. In the following the different characteristics of these three algorithms are described.

5.1.1 The SISCone algorithm

The SISCone [95, 96] is a IRC-safe cone algorithm based on a seedless approach (i.e. all particles are treated as equal) to stable cones identification. The cone search follows these steps:

- for each particle i all the particles $j = 1...N$ with a distance from i lower than $2R$ are considered, where $R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ is a fixed parameter;

- for each couple of particles ij, a circle of radius R which has the two particles on the circumference is created;
- for each circle the p_T weighted centroid of all the particles which lie within the circle is calculated;
- the cone centered on the centroid of the nth circle is declared stable only if it contains all the initial particles;
- the cones with overlaps are split/joined if the scalar sum of the common particles p_T is lower/higher than a fraction f of the energy of the cone with higher momentum.

Default values used for the R and f parameters are: $R = 0.5$ and $f = 0.75$.

5.1.2 The k_t and anti-k_t algorithms

The k_T and anti-k_T [97] are IRC-safe algorithms which belong to the sequential recombination class. They introduce two definitions of distance: d_{ij}, the distance between the objects (particles or pseudojets) i and j, and d_{iB}, the distance between the object i and the beam. These distances are defined as follows:

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta_{ij}^2}{R^2}, \tag{5.1}$$

$$d_{iB} = k_{ti}^{2p}, \tag{5.2}$$

where $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, k_{ti}, y_i and ϕ_i are respectively the transverse momentum, rapidity and azimuth of particle i and p is a parameter which distinguishes among different distance definitions. The algorithm therefore proceeds as follows:

- the distances d_{ij} and d_{iB} for each particle i are calculated and the smallest distance between d_{ij} and d_{iB} is identified;
- if $d_{ij} < d_{iB}$ the objects i and j are recombined;
- if $d_{ij} > d_{iB}$ i is considered as a jet and removed from the list of objects;
- the distances are recalculated and the procedure is repeated until no objects are left.

According to the value of the parameter p present in the distance definitions we may obtain different jet clustering algorithms. The value $p = 1$ defines the inclusive k_t algorithm. It can be shown in general that for $p > 0$ the behaviour of the jet algorithm with respect to soft radiation is rather similar to that observed for the k_t algorithm, because what matters is the ordering between particles, and for finite

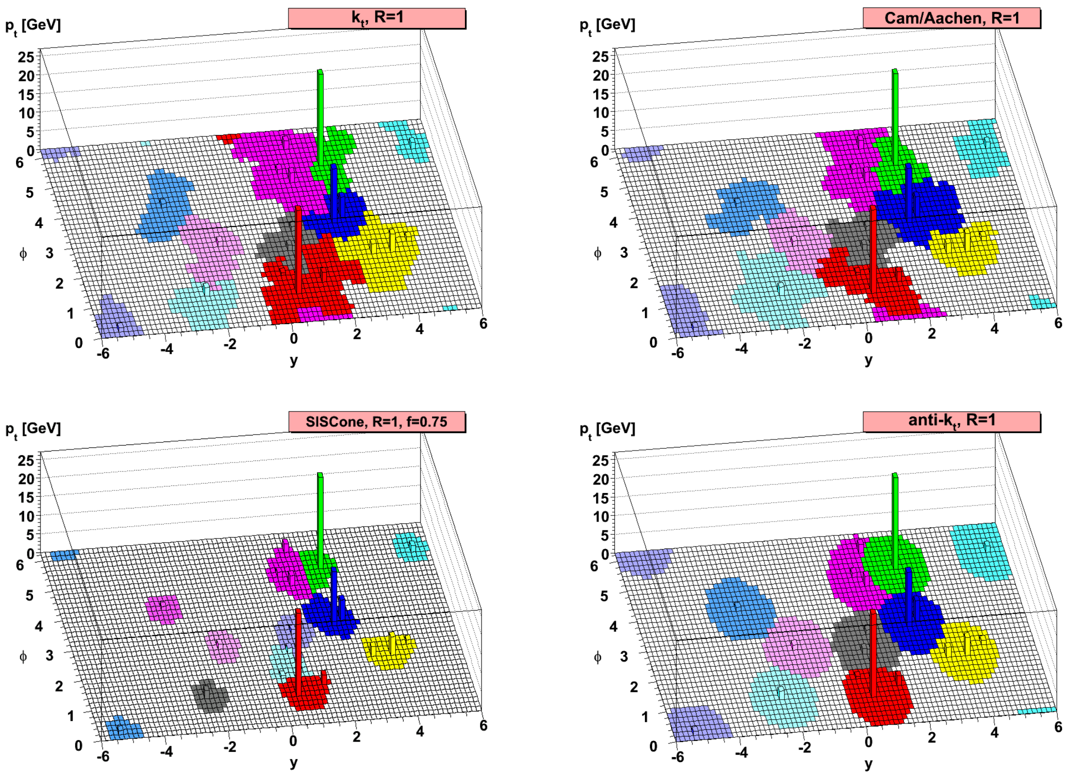


Figure 5.3: A MC sample parton-level event, together with many random soft “ghosts”, clustered with four different jets algorithms, illustrating the “active” catchment areas of the resulting hard jets. For k_t and Cam/Aachen the detailed shapes are in part determined by the specific set of ghosts used, and change when the ghosts are modified. On the contrary, the anti-k_t jet shapes are more regular and not modified by soft particles [97].

Δ this is maintained for all positive values of p. The case of $p = 0$ is special and it corresponds to the inclusive Cambridge/Aachen algorithm [98].

For $p = -1$ we obtain the anti-k_t algorithm. The functionality of the anti-k_t algorithm can be understood by considering an event with a few well separated hard particles with transverse momenta $k_{t1}, k_{t2}, \ldots$ and many soft particles. The $d_{1i} = min(1/k_{t1}^2, 1/k_{ti}^2)\Delta_{1i}^2/R^2$ between a hard particle 1 and a soft particle i is exclusively determined by the transverse momentum of the hard particle and the Δ_{1i} separation. The d_{ij} between similarly separated soft particles will instead be much larger. Therefore soft particles will tend to cluster with hard ones long before they cluster among themselves. If a hard particle has no hard neighbours within a distance 2R, then it will simply accumulate all the soft particles within a circle of radius R, resulting in a perfectly conical jet. If we have two hard particles separated by a distance $R < \Delta_{12} < 2R$, then there will be two hard jets, and it is not possible for both to be perfectly conical. If $k_{t1} \gg k_{t2}$ jet 1 will result conical, while jet 2 will be partly conical, since it will miss the part overlapping with jet 1. Otherwise, if $k_{t1} \sim k_{t2}$ both cones will be clipped. Finally if we have two hard particles within a distance $\Delta_{12} < R$ they are joined to form a single jet, which will be conical and centered on k_1 if $k_{t1} \gg k_{t2}$, and will be instead more complex if $k_{t1} \sim k_{t2}$. The key feature above is that the soft particles do not modify the shape of the jet, while hard particles do. It results in a more regular shape of the reconstructed jets with respect to the other algorithms, as shown in Fig. 5.3.

Due to this fact, the anti-k_t algorithm is more robust than the other ones with respect to non-perturbative effects like hadronization and underlying event contamination, improving in this way the momentum resolution and therefore the calorimeter performance. Finally the clusterization performed by the anti-k_t is faster than the one obtained by using the other algorithms, e.g. the SISCone [95]. For all these reasons the anti-k_t algorithm was used for the present work. The distance parameter was fixed at the value $R = 0.5$ and the reconstructed jets are therefore denoted as "ak5" jets.

5.2 Jet reconstruction in CMS

The CMS subdetectors involved in the jet reconstruction are the hadronic calorimeter HCAL, the electromagnetic calorimeter ECAL, which identifies photons and electrons belonging to the jet (in particular the ones coming from π_0 decays), and the silicon Tracker, which adds track information in order to improve the p_T response and resolution of calorimeter jets. Three types of jets are reconstructed in CMS, combining in different ways the individual contributions from subdetectors to form the inputs to the jet clustering algorithms previously described: Calorimeter jets (CALO), Jet Plus Track (JPT) jets and Particle Flow (PF) jets [99].

For the present analysis the Particle Flow jets are used, since they are better reconstructed and have a better jet p_T resolution as described in the following. The jet clustering algorithm used was the anti-k_t described in the previous section and the jet collections are therefore named as "ak5PF" jets.

5.2.1 Particle Flow jets (PF jets)

The Particle Flow algorithm [87–89], as already stated in Sec. 4.1.4, combines the information from all CMS subdetectors to identify and reconstruct all particles in the event as shown in Fig. 5.4, namely muons, electrons, photons, charged hadrons and neutral hadrons. Charged hadrons, in particular, are reconstructed from tracks in the central tracker. Photons and neutral hadrons are reconstructed from the energy clusters in the electromagnetic and hadron calorimeters which do not have a corresponding track. A neutral particle overlapping with charged particles in the calorimeters can be detected as a calorimetric energy excess with respect to the sum of the associated track momenta. PF jets are then reconstructed from the

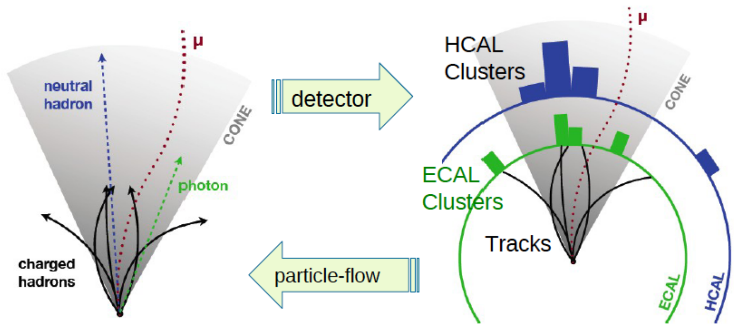


Figure 5.4: The Particle Flow algorithm. Particles in the CMS detector are seen as tracks and energy deposits. The PF algorithm attempt to fully reconstruct an event by combining information from all CMS subdetectors.

resulting list of particles. The jet momentum and spatial resolutions is improved with respect to calorimetric jets since the use of the tracking detectors and the excellent granularity of the ECAL allow to resolve and precisely measure charged hadrons and photons inside jets, which constitute about 90% of the jet energy.

5.2.2 Jet position and momentum measurements

Once an ensemble of pseudojets (i.e. identified particle for PF jets) clustered in phase space is chosen, for example, within an area of $\eta - \phi$ space defined by a cone of radius R, then a jet axis is defined such that the total momentum of the ensemble perpendicular to that axis is zero [100]. With that choice of axis the jet four vector is the sum of the momentum and energy of the pseudojets. For PF jets the sum is therefore performed on the identified particles that belong to the jets. Jet four momentum and axis direction are defined as follows [100, 101]:

$$\vec{P}_J = \sum_i \vec{p}_i, \quad E_J = \sum_i e_i \,, \tag{5.3}$$

$$\tan\phi_J = \frac{P_{yJ}}{P_{xJ}}, \quad \tanh\eta_J = \cos\theta = \frac{P_{zJ}}{P_J}\,, \tag{5.4}$$

where jet quantities are uppercase and pseudojets quantities are lowercase.

5.3 Jet energy corrections and p_T resolution

A detailed understanding of the energy calibration and resolution of jets is of crucial importance and is a leading source of systematic uncertainty for many analyses with jets in the final state.

The jet energy measured in the detector is typically different from the corresponding particle jet energy. The latter is obtained in the simulation by clustering, with the same jet algorithm, the stable particles produced during the hadronization process that follows the hard interaction. The main cause for this energy mismatch is the non uniform and non linear response of the CMS calorimeters to the jet showers. Furthermore, electronics noise and additional pp interactions in the same bunch crossing (event *pile-up*) can lead to extra unwanted energy. The purpose of the jet energy correction is to relate, on average, the energy measured in the detector to the energy of the corresponding particle jet.

CMS has developed a factorized multistep procedure for the jet energy calibration (JEC) [102]. The following three subsequent corrections are devised to correct PF jets to the corresponding particle jet level:

- **L1 Offset Correction** (E_{offset}), which aims to correct the jet energy for the excess unwanted energy due to electronics noise and pile-up. In principle this correction removes any dataset dependence on luminosity, so that the following corrections are applied upon a luminosity independent sample. The data-driven L1FastJet [103] pile-up subtraction algorithm has been used. It is based on estimating the pile-up and underlying event transverse momentum density using the calculated jet area.

- **L2 Relative Correction** (C_{Rel}), that removes variations in jet response versus jet η through a balance, in two jet events, with respect to a central control region chosen as a reference ($|\eta| < 1.3$).

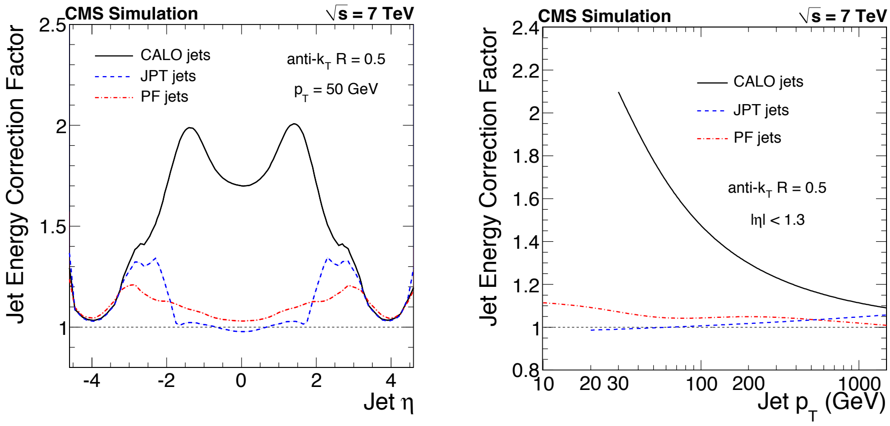
- **L3 Absolute Correction** (C_{Abs}), which removes variations in jet response versus jet p_T.

The default sequence for the jet energy corrections is expressed mathematically as [99]:

$$E_{Corrected} = (E_{Uncorrected} - E_{offset}) \times C_{Rel}(\eta, p_T^{''}) \times C_{Abs}(p_T^{'})\,, \tag{5.5}$$

where $p_T^{''}$ is the transverse momentum of the jet corrected for offset and $p_T^{'} = p_T^{''} \times C_{Rel}(\eta, p_T^{''})$ is the transverse momentum of the jet corrected for offset and pseudorapidity dependence.

All the correction factors described above are determined first by using the MC true jets. The MC calibration is based on the simulation and corrects the energy of the reconstructed jets such that it is equal on average to the energy of the generated MC particle jets. Calorimeter jets require a large correction factor due to the non-linear response of the CMS calorimeters. The track-based jet types (JPT and PF) require much smaller correction factors because the charged component of the jet shower is measured accurately in the CMS tracker, as shown in Fig. 5.5. The low jet p_T threshold indicates the minimum recommended p_T for each jet type: 30 GeV , 20 GeV , and 10 GeV for CALO, JPT, and PF jets respectively. A wider p_T spectrum



(a) (b)

Figure 5.5: Monte Carlo jet energy correction factors derived from simulation for CALO, JPT, and PF jets at $\sqrt{s} = 7$ TeV, as a function of η ($p_T = 50\,\mathrm{GeV}$) (a) and of corrected jet transverse momentum (b). Jets are reconstructed by using the anti-k_t algorithm with distance parameter $R = 0.5$ [104].

is therefore available using PF jets. PF jets require also much smaller corrections with respect CALO jets as these jets rely heavily on the tracking information, and correction dependence versus η is more stable for PF jets than for CALO and JPT jets.

For the 2010 data taken period CMS adopted also a hybrid approach to determine the JEC, i.e. both MC truth information and physics processes from pp collisions are used for in-situ jet calibration. For data analysis indeed small residual L2 and L3 corrections need to be applied on top of MC truth, in order to make the data look like the MC. Dijet, γ+jets and Z+jets data samples were used to determine residual corrections, producing a further correction factor that must be applied only on jets from data. This fourth factor is

- **L2L3 Residual Correction**, small residual correction to be applied on top of MC truth:
  - relative scale: η dependent, constant in p_T residual bewteen 0.92 and 1.03,
  - absolute scale: a constant factor of 1.007.

In Fig. 5.6 and Fig. 5.7 the total jet energy correction factor is shown as a function of η and p_T respectively, for the three different jet types. The overall uncertainties on the jet energy correction factors are also shown in the plots. For the combined MC and residual calibration, the residual corrections for the relative and the absolute response are multiplied with the generator level MC correction, while the corresponding uncertainties are added in quadrature. Because of the

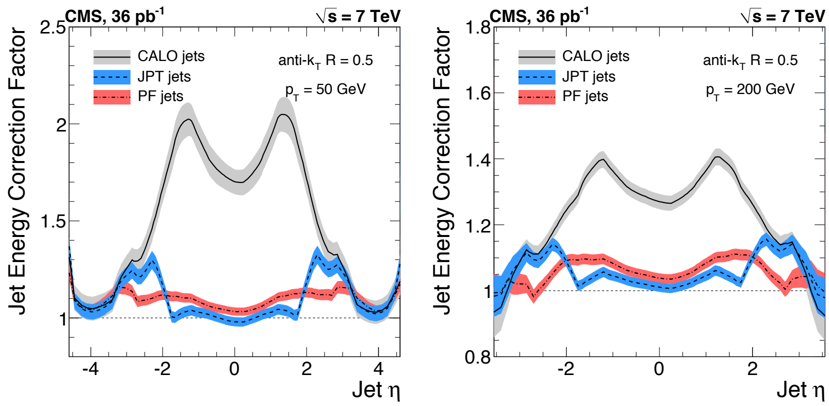
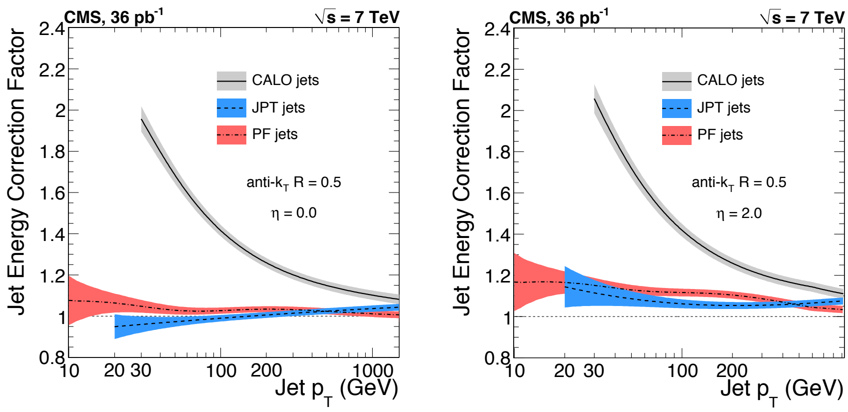
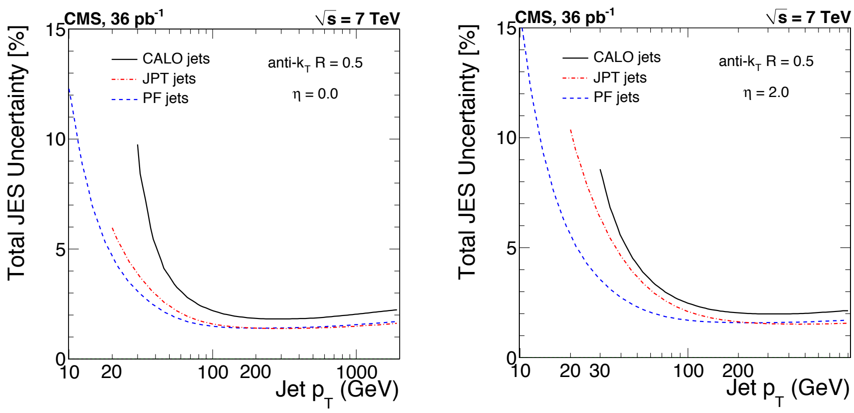


Figure 5.6: Total jet energy correction factor, as a function of jet η for $p_T = 50$ GeV (left) and $p_T = 200$ GeV (right). The bands indicate the corresponding uncertainty [104].



(a) (b)

Figure 5.7: Total jet energy correction factor, as a function jet p_T for $\eta = 0$ (a) and $\eta = 2$ (b). The bands indicate the corresponding uncertainty [104].

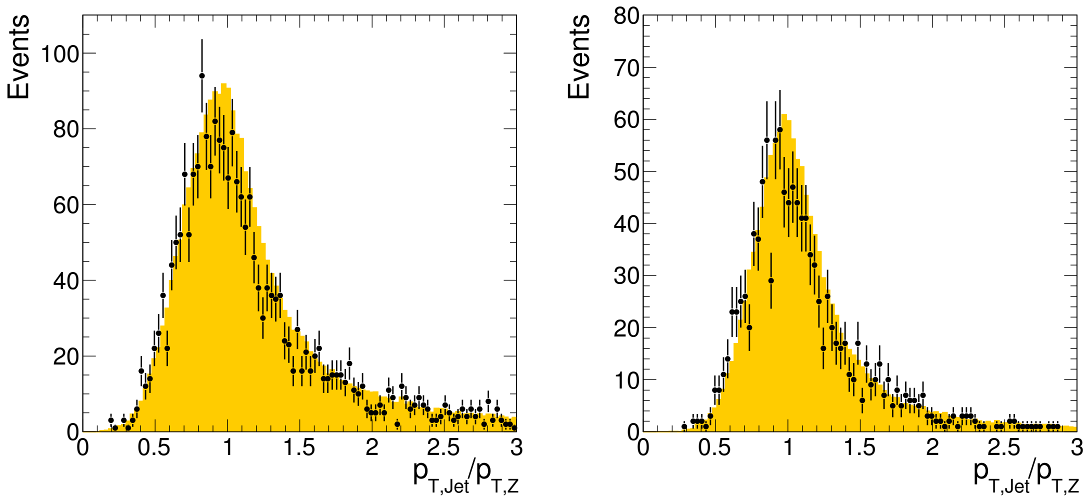


(a) (b)

Figure 5.8: Total uncertainty on jet energy calibration as a function of jet p_T, for $\eta = 0$ (a) and $\eta = 2$ (b) [104].

smallness of the residual corrections, the combined correction has the same shape of the MC component. In Fig. 5.8 the total uncertainty on jet energy correction is shown as a function of jet p_T, for $\eta = 0$ (a) and jet $\eta = 2$. In general the PF jets have the smallest systematic uncertainty, while CALO jets have the largest one. As shown in Fig. 5.8, for PF jets with $p_T = 15\,\mathrm{GeV}$ the total JEC uncertainties is about 8% while for jet $p_T = 30\,\mathrm{GeV}$ the total uncertainty is of the order of 3% [104].

In order to check the jet calibration for the present analysis, the ratio $p_{T,Jet}/p_{T,Z}$ has been measured for Z + 1 jet events. In such events indeed the $p_{T,Jet}/p_{T,Z}$ ratio should be equal to 1 because of the transverse momentum balance between the Z boson and the jet. In Fig. 5.9 the $p_{T,Z}/p_{T,Jet}$ ratio for data and for $Z + jets$ MC sample with Z2 tune is shown both for jet $p_T > 15$ GeV and jet $p_T > 30$ GeV. The events shown are required to have a reconstructed $Z \to ee$ event with all the selection cuts applied (see Sec. 4.8). The background is not considered, since it results absolutely negligible as already shown in Sec. 4.8. For these plots the L1, L2 and L3 jet energy corrections have been applied on MC jets, while on data jets, together with the L1, L2 and L3 corrections, also the L2L3Residual correction has been applied. In Fig. 5.10 the same measurement has been performed comparing data and D6T $Z + jets$ MC sample. These plots show for data a clear peak around 1, within the calibration uncertainties described above, which confirms the correct jet energy calibration for the present work. The agreement between data and MC is better for Z2 tune, while Fig. 5.10 shows worse performances of D6T tune simulation



(a) (b)

Figure 5.9: Ratio between $p_{T,Jet}$ and $p_{T,Z}$ in Z + 1 jet events. Data (black) and $Z + jets$ MC sample with Z2 tune (yellow) are shown. The background is not considered because negligible. The ratio is evaluated by using PF jets with $p_T > 15$ GeV (a) and PF jets with $p_T > 30$ GeV (b). In both cases jets are required to have $\eta < 2.5$.

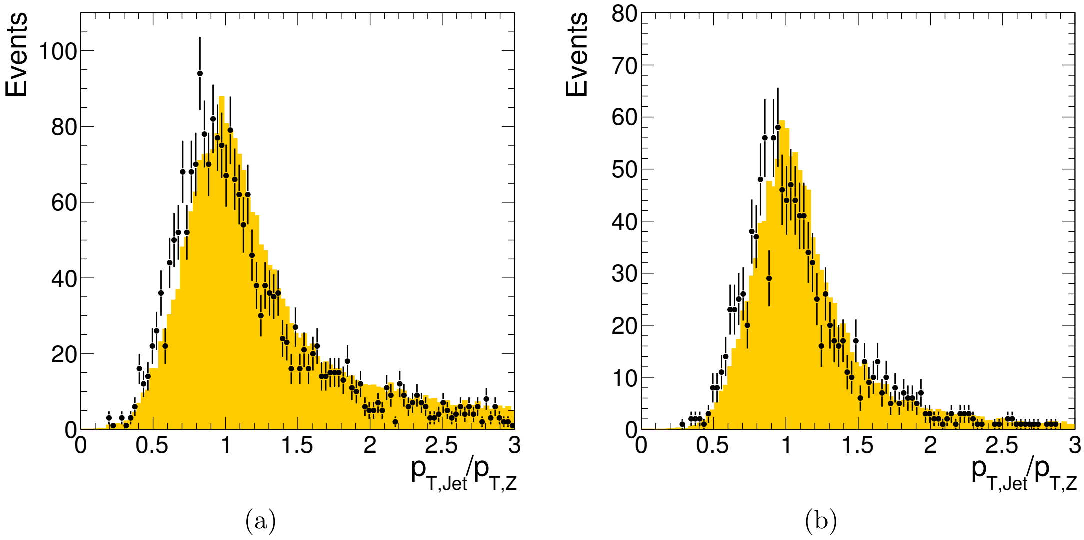


Figure 5.10: Ratio between $p_{T,Jet}$ and $p_{T,Z}$ in Z + 1 jet events. Data (black) and $Z + jets$ MC sample with D6T tune (yellow) are shown. The background is not considered because negligible. The ratio is evaluated by using PF jets with $p_T > 15$ GeV (a) and PF jets with $p_T > 30$ GeV (b). In both cases jets are required to have $\eta < 2.5$.

in properly reproducing jet energy.

A jet p_T resolution measurement was performed using dijet events in both data and MC samples [104]. The dijet resolution is defined for events with at least two jets through the variable A defined as

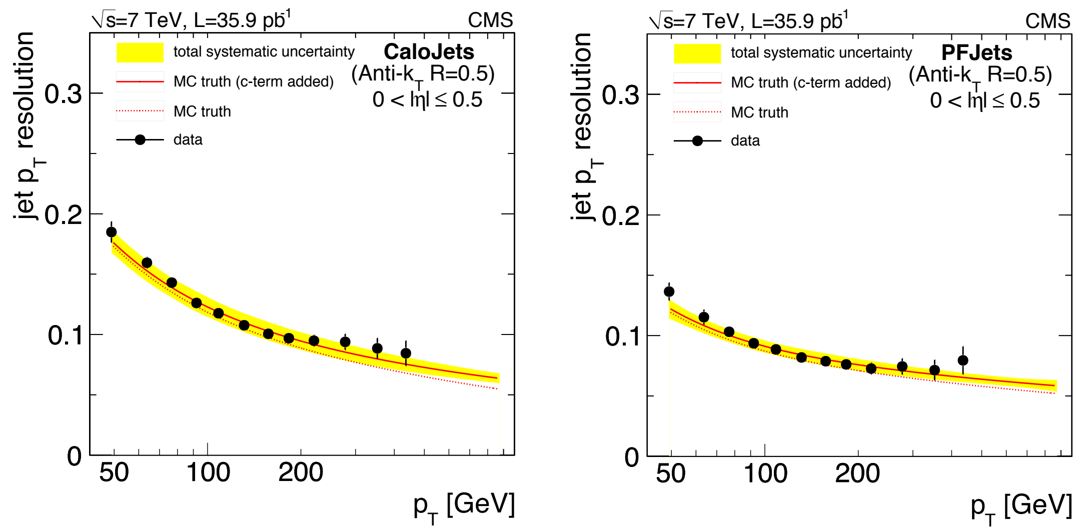
$$A = \frac{p_T^{Jet1} - p_T^{Jet2}}{p_T^{Jet1} + p_T^{Jet2}}\,, \tag{5.6}$$

where p_T^{Jet1} and p_T^{Jet2} refer to the randomly ordered transverse momenta of the two leading jets. In the limit $p_T \equiv < p_T^{Jet1} > = < p_T^{Jet2} >$ and $\sigma\left(p_T\right) \equiv \sigma\left(p_T^{Jet1}\right) = \sigma\left(p_T^{Jet2}\right)$, the resolution is calculated to be

$$\left(\frac{\sigma(p_T)}{p_T}\right) = \sqrt{2}\sigma_A\,, \tag{5.7}$$

where σ_A is the variance of the A variable [104].

Fig. 5.11 shows the dijet resolution measurement results obtained for CALO and PF jets in the central region. In each plot, the solid red line depicts the resolution from generator level MC, corrected for the measured discrepancy between data and simulation (constant term), and represents the best estimate of the jet p_T resolution in data. The yellow band represents the total systematic uncertainty.



(a) (b)

Figure 5.11: Jet p_T resolution measurement for CALO (a) and PF (b) jets in $|\eta| < 0.5$. Data are compared to the generator level (MC truth) p_T resolution before (red dashed line) and after (red solid line) correction for the measured discrepancy between data and simulation [104].

The uncorrected generator level MC resolution is shown as a red dashed line, while black dots represent data, which are in good agreement with the corrected MC resolution within the statistical and systematic uncertainties. As shown in Fig. 5.11 PF jets have better performances than CALO jets in terms of transverse momentum resolution. In particular in the region $|\eta| < 0.5$ with a p_T of 100 GeV the measured PF jet resolution in the data is better than 10%.

5.4 The `PATJet` object

Like the electrons (Chapter 4 Sec. 4.6), also jets are reconstructed by the PAT algorithm (Chapter 2 Sec. 2.4.2) and the `PATJet` object is defined, which makes use of the anti-k_t algorithm. In particular the `PATJet` object stores the following information:

- jet axis coordinates and jet momentum;
- jet energy corrections, as described in the previous section;
- jet energy correction uncertainties;
- tracks associated to the jet. Track matching is based on spatial separation in $\eta - \phi$ between the jet axis and the track momentum measured at the interaction vertex;

For the present analysis a `PATJet` collection of PF jets was used. PF jets were chosen because of their lower and more precise jet energy corrections and for their better p_T resolution, as stated in the previous sections. After the application of the jet energy correction the following very loose preselection cuts are applied on the `PATJet` collections:

- jet $(p_T)_{Corrected} > 10\,\mathrm{GeV}$,

- jet $|\eta| < 10$.

5.5 Jet - lepton isolation

All energy deposits in the electromagnetic calorimeter are used by the jet clustering algorithm. Thus all the electrons present in the event are classified as jets, since they release almost all their energy in ECAL. Since we are interested in jet counting, the Z decay electrons must be separated from the real jets present in the event. To this end the distance in the $\eta - \phi$ plane between the jet axis and the electrons coming from Z decay is used. If a jet is separated by a distance $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ higher than a certain value from both the Z electrons it is therefore considered as *lepton-isolated*, otherwise it is removed from the jet collection (this operation is also named "*jet cleaning*"). Fig. 5.12 shows for the $Z + jets$ MC sample the values of the ΔR separation variable evaluated between each jet and the electron coming from Z decay nearest to it. All the jets with a jet-lepton separation $\Delta R < 0.5$ are rejected as not isolated jets. The value $\Delta R = 0.5$ is the distance parameter that has been used for the anti-k_t jet clustering algorithm and it is the minimum in the ΔR distribution of Fig. 5.12.

5.6 Z+jet events selection

Within the selected $Z \rightarrow ee$ events (see Chapter 4 Sec. 4.8) a jet energy corrected PF jet collection is reconstructed with the preselection cuts described in Sec. 5.4. Further cuts are needed in order to discard electrons identified as originating from the Z boson decay and badly reconstructed jets. To discard these unwanted jets, while retaining most of the real ones, a set of loose jet identification criteria has been defined which keep $> 99\%$ of the real jets while removing a significant fraction of unwanted ones [105]. The applied requirements are:

- jet-lepton separation $\Delta R > 0.5$,
- $|\eta| < 2.5$,
- charged hadron fraction > 0.0,

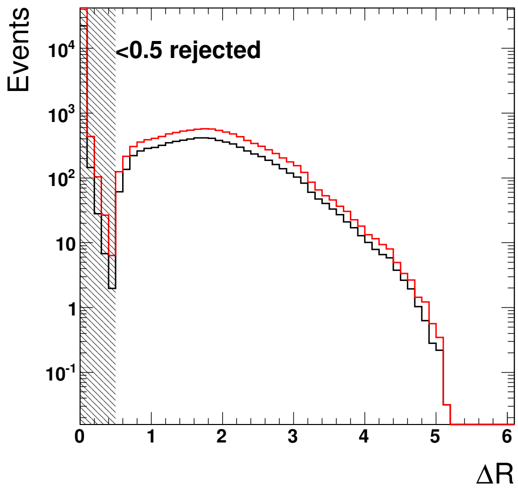


Figure 5.12: Separation between each jet and the Z electron nearest to it, evaluated for the $Z + Jets$ MC sample by using GEN jets (red) and PF RECO jets (black), with $p_T > 15\,\mathrm{GeV}$ and $|\eta| < 2.5$. The separation is expressed in terms of $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$. The anti-$k_t$ algorithm is used for jet reconstruction. The cut applied for jet-lepton isolation is $\Delta R > 0.5$. For this plot the MC sample is normalized to a reference luminosity of 50 pb^{-1}.

- neutral hadron fraction < 0.99,
- charged multiplicity > 0.0,
- charged electromagnetic fraction < 0.99,
- neutral electromagnetic fraction < 0.99,
- two different p_T cuts studied: $p_T > 15\,\mathrm{GeV}$ and $p_T > 30\,\mathrm{GeV}$.

The jet lepton separation requirement $\Delta R > 0.5$ leads to a PF jet collections cleaned from electrons coming from Z decay, while the cut on η limits the reconstruction region to the tracker η region, since PF jets rely on tracking information.

For the Z+jets cross section measurement two different PF jet classes have been used: one with $p_T > 15\,\mathrm{GeV}$ and the other with $p_T > 30\,\mathrm{GeV}$. In fact, although PF jets have better resolution and more precise energy corrections for $p_T > 30\,\mathrm{GeV}$, their excellent performances make feasible also a Z+jets cross section measurement for $p_T > 15\,\mathrm{GeV}$ jets.

5.6.1 Jet p_T and η distributions after cuts

In Fig. 5.13 the transverse momentum of the leading jets found in reconstructed $Z \to ee$ events are shown for $p_T > 15\,\mathrm{GeV}$. Selected events satisfy all of the selection cuts defined in Chapter 4 Sec. 4.8 to identify a $Z \rightarrow ee$ event, plus the PF jets selection criteria defined in Sec. 5.6. The plot shows a good agreement between data and MC, especially for $p_T > 30\,\text{GeV}$. As found in Chapter 4 the background is completely negligible after the application of the selection cuts. In particular, as already shown in Chapter 4, no QCD events are left after all $Z \rightarrow ee$ selection cuts applied due to the limited luminosity of these MC samples.

The η variable of the PF leading jets is shown in Fig. 5.14, both for $p_T > 15\,\text{GeV}$ and $p_T > 30\,\text{GeV}$. Also in this case the agreement between data and MC is good, for both p_T cuts. The jet multiplicity found in the selected $Z \rightarrow ee$ events will be discussed in detail in Chapter 7.

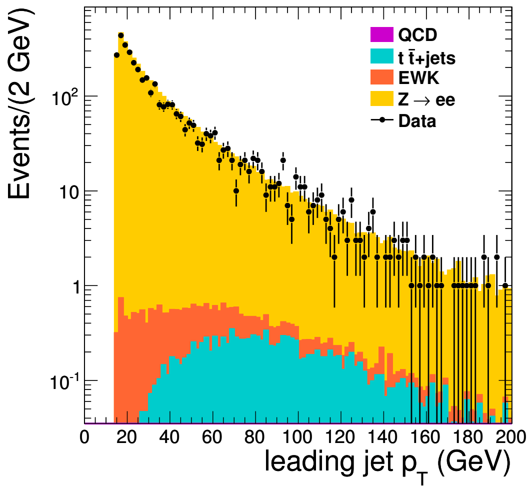


Figure 5.13: Distributions of the p_T variables of the leading PF jets found in reconstructed $Z \to ee$ events, with the cuts $p_T > 15\,\mathrm{GeV}$ and $|\eta| < 2.5$.

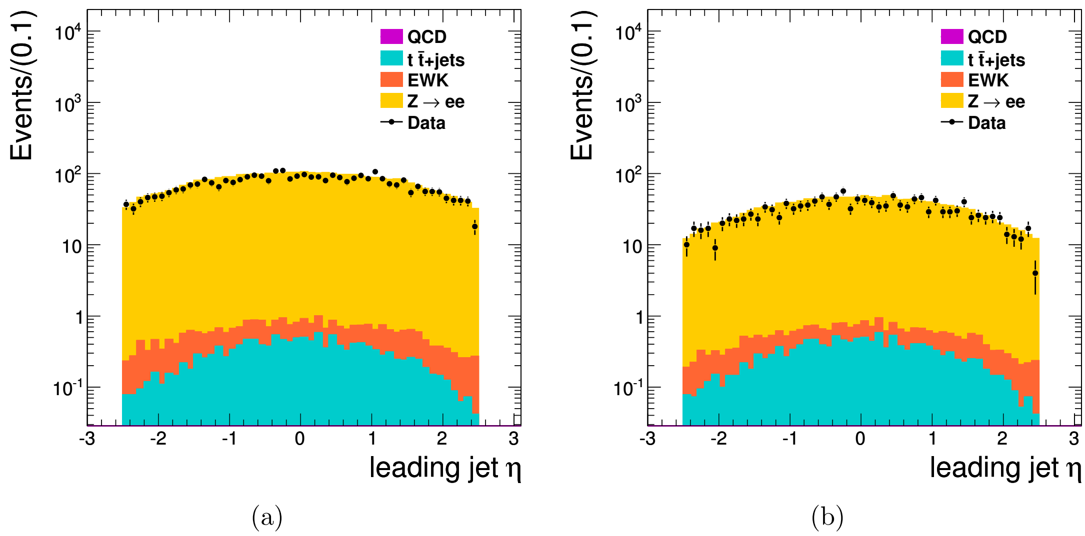


Figure 5.14: Distributions of the η variable of the leading PF jets found in reconstructed $Z \to ee$ events, with $p_T > 15\,\mathrm{GeV}$ (a) and $p_T > 30\,\mathrm{GeV}$ (b).

Chapter 6

Selection efficiency measurement

In order to properly calculate the cross section of Z+jets events as a function of jet multiplicity, the event yields must be corrected for the measurement efficiency. In this Chapter the event selection efficiency measurement is shown, following both a MC and a data driven approach. The final efficiency estimation is evaluated as a function of the jet multiplicity and takes advantage of both MC and data driven methods.

6.1 The $Z \rightarrow ee + jets$ event selection efficiency

The global efficiency for selecting $Z \rightarrow ee$ events can be factorized as follows:

$$\epsilon = \epsilon_{Acc} \times \epsilon_{Trg} \times \epsilon_{Imp} \times \epsilon_{Conv} \times \epsilon_{Iso} \times \epsilon_{EiD} \, , \tag{6.1}$$

where each factor represents the relative efficiency of the correspondent selection step described in Chapter 4 Sec. 4.8. In more detail,

- ϵ_{Acc} is the efficiency to reconstruct an electron couple within the acceptance, i.e. within the kinematic cuts described in Sec. 4.8; in particular $p_T > 20\,\text{GeV}$ for the leading electron and $p_T > 10\,\text{GeV}$ for the second electron are required,

- ϵ_{Trg} is the trigger efficiency for the leading electron,

- ϵ_{Imp} is the efficiency to have an impact parameter < 0.02 cm both for the leading and the second electrons of the couple,

- ϵ_{Conv}, ϵ_{Iso} and ϵ_{EiD} are the efficiencies for the working points WP80 (leading electron) and WP95 (second electron) requirements, concerning the conversion rejection, isolation and electron identification respectively.

All the factors in Eq. (6.1) represent therefore double electron efficiencies, except for ϵ_{Trg} which is a single electron trigger efficiency. Selection cuts applied on the electron couples are asymmetric, with the exception of the impact parameter requirement which is the same both for the leading and the second electrons. In the following sections, the efficiency of each particular cut will be evaluated by requiring for both the leading and the second electrons of $Z \to ee$ event candidates their own selection cuts, i.e. it will be given by the product of the two single electron efficiencies relative to the leading and the second electron respectively.

The increased hadronic activity in events with extra jets is expected to affect the electron efficiency, in particular the electron isolation requirement. Each one of the efficiencies in Eq. (6.1) must be therefore evaluated as a function of the jet multiplicity. In the following sections the event selection efficiency is evaluated as a function of jet inclusive multiplicity bins, considering two different p_T^{jet} cuts: $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$. For $p_T^{jet} > 15\,\mathrm{GeV}$ 5 multiplicity bins are studied, i.e. ≥ 0, ≥ 1, ≥ 2, ≥ 3 and ≥ 4 jets. For $p_T^{jet} > 30\,\mathrm{GeV}$ instead only 4 multiplicity bins are taken into account, i.e. ≥ 0, ≥ 1, ≥ 2, and ≥ 3 jets, since for data sample the bin with ≥ 4 jets is affected by lack of statistics, which do not allow to apply in this case the data driven method used to evaluate the corresponding efficiency, as will be described in Sec. 6.3.

6.2 MC estimation of selection efficiency

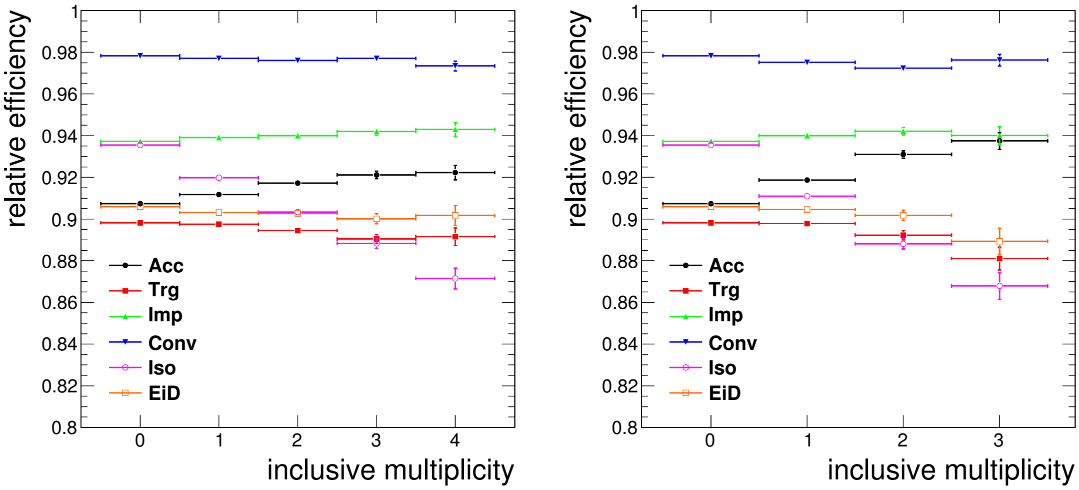
In order to check the effect of each cut on the $Z \to ee$ event selection its relative efficiency was evaluated, making use of the $Z+jets$ `MADGRAPH` sample with Z2 tune. The relative efficiency is defined as the efficiency of each cut evaluated with respect to the previous one. The relative efficiency of the first cut is calculated with respect to the total number of events generated in acceptance. The relative efficiency can be therefore expressed as follows:

$$\epsilon_{Rel}^{MC}(\mathrm{Acc}) = \frac{N^{Reco}(\mathrm{Acc})}{N^{Gen}(\mathrm{Acc})}\,, \tag{6.2}$$

$$\epsilon_{Rel}^{MC}(n) = \frac{N^{Reco}(\mathrm{cut}\,n)}{N^{Reco}(\mathrm{cut}\,n-1)}\,, \quad n = 2, \ldots, 6\,, \tag{6.3}$$

where $N^{Reco}(\mathrm{Acc})$ is the number of events reconstructed within the acceptance, $N^{Reco}(\mathrm{cut}\,n)$ ($N^{Reco}(\mathrm{cut}\,n-1)$) is the number of reconstructed events which pass the n-th (($n-1$)-th) cut and $N^{Gen}(\mathrm{Acc})$ represents the total number of events generated within the acceptance. Eq. (6.2) is the efficiency of the first cut, while Eq. (6.3) is the efficiency of the added cuts. In Fig. 6.1 the relative efficiency of each cut is shown as a function of jet inclusive multiplicity, both for $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$.

As shown in Fig. 6.1 the relative efficiencies appear to be quite flat, except for the isolation cut which becomes less efficient as the number of jets increases, as expected because of the enhanced hadronic activity. Also the first applied cut (Acceptance) seems to show a dependence on jet multiplicity, opposite to the one



(a) (b)

Figure 6.1: Relative efficiency of the selection cuts as a function of jet inclusive multiplicity, for $p_T^{jet} > 15\,\mathrm{GeV}$ (a) and for $p_T^{jet} > 30\,\mathrm{GeV}$ (b). The efficiency of each cut is calculated with respect to the number of events passing the previous cut. For this efficiency estimation the $Z + jets$ `MADGRAPH` Z2 sample was used.

found for the isolation cut. In this case the increase of the acceptance cut efficiency with the multiplicity is due to the presence of fake $Z \to ee$ electron couples in this very first step of the selection procedure, i.e. couples in which one or both the electrons belong to the hadronic activity present in the event and not to actual Z decays. The number of fake $Z \to ee$ electron couples indeed increases as the jet multiplicity increases. This effect disappears with the application of further cuts, as will be shown in the following.

The global efficiency needed to the cross section measurement is defined as the efficiency of the whole selection calculated with respect to the total number of events generated within the acceptance. We can also define the "cascade efficiency" of each cut as the efficiency of the selection cut cascade evaluated at that particular cut. The cascade efficiency formula is therefore:

$$\epsilon_{Casc}^{MC}(n) = \frac{N^{Reco}(\mathrm{cut}\,1 + \ldots + \mathrm{cut}\,n)}{N^{Gen}(\mathrm{Acc})}\,, \quad n = 1, \ldots, 6\,, \tag{6.4}$$

where $N^{Reco}(\mathrm{cut}\,1 + \ldots + \mathrm{cut}\,n)$ is the number of reconstructed events passing the first n cuts and $N^{Gen}(\mathrm{Acc})$ is the total number of events generated within the acceptance. The global efficiency of the whole selection sequence is therefore given by $\epsilon_{Casc}^{MC}(6)$, i.e. with all the cuts applied. In the following $\epsilon_{Casc}^{MC}(6)$ will be simply called as global efficiency of the $Z \to ee + jets$ event selection and will be denoted as ϵ_{Glob}^{MC}.

The cascade efficiency calculated at each step of the selection sequence is shown in Fig. 6.2 as a function of jet inclusive multiplicity, both for $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$. As shown in the plots of Fig. 6.2, the global efficiency of the whole selection expressed as a function of the jet inclusive multiplicity is roughly included for $p_T^{jet} > 15\,\mathrm{GeV}$ between 63% for the bin 0 and 59% for the bin 4, while for $p_T^{jet} > 30\,\mathrm{GeV}$ the minimum values is about 58%, reached at the bin 3 of inclusive multiplicity. As previously shown in Fig. 6.1 the isolation cut dependence on the jet multiplicity is still recognizable in Fig. 6.2, while the same dependence of the acceptance cut is canceled by the further cuts as expected.

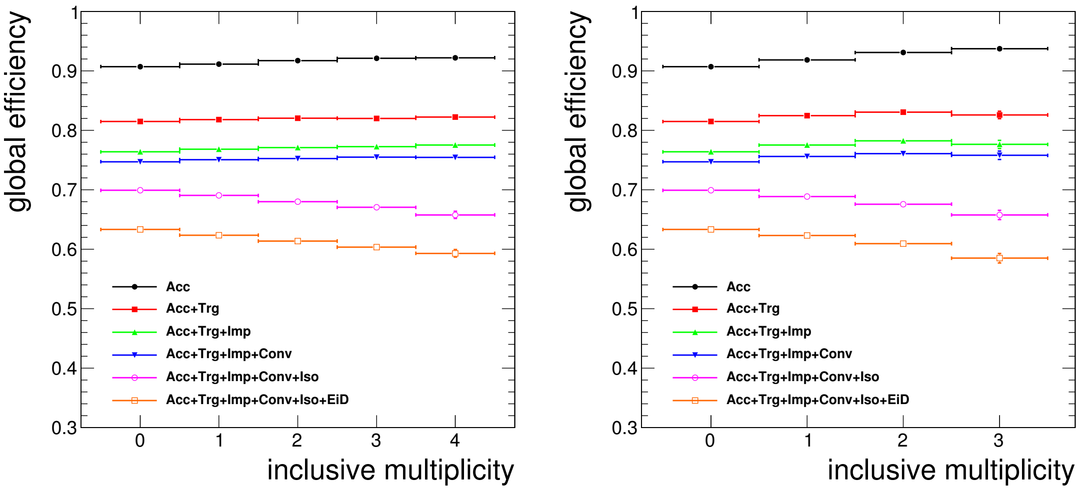
Also the $Z + jets$ `MADGRAPH` sample with D6T tune was used for the cascade and global efficiency calculation in order to check possible differences in efficiency calculation with respect to Z2 tune, due to a possible different data and detector response modeling of the two tunes. The cascade efficiency obtained at each step of the selection sequence making use of the D6T sample is shown in Fig. 6.3, both for $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$.

The global efficiency obtained with D6T $Z + jets$ sample is quite lower than the one obtained with the Z2 sample, being included between 62% and 56%. In particular the difference is recognizable starting from the isolation cut and denotes a difference in the underlying event modeling between Z2 and D6T. In Table 6.1, Table 6.2 and Table 6.3 the selection global efficiency values are shown, evaluated as a function of jet inclusive multiplicity both for Z2 and D6T samples.

As shown in the tables the efficiencies calculated by using the Z2 and D6T MC simulations differ by about $3-6\%$. The efficiency calculation made by using MC simulations leads therefore to large systematic errors due to differences in data and detector simulation. Moreover imperfections in data and detector modeling may lead to large discrepancy between the MC efficiency and the actual data efficiency, as will be shown in the following. This fact makes necessary to use a data driven method for properly evaluating the global efficiency.

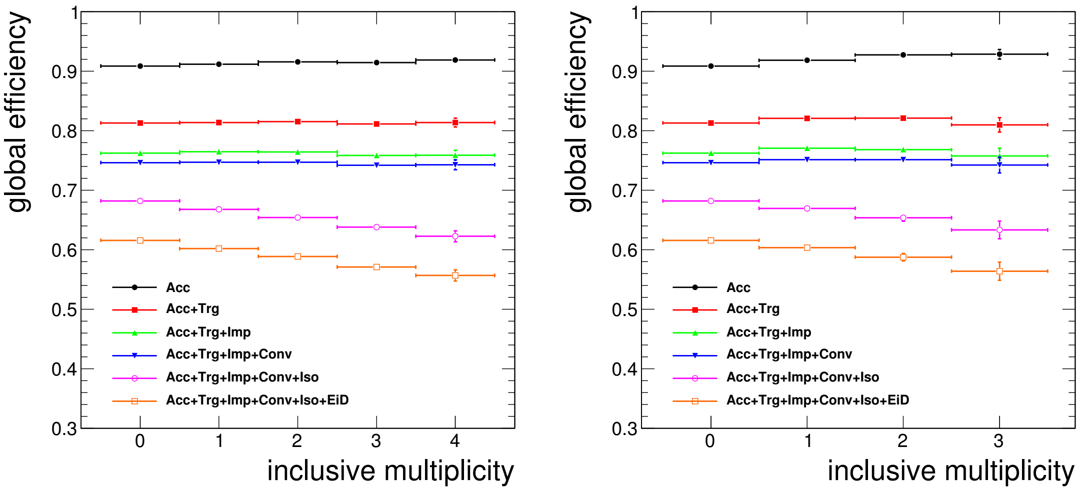
Selection Global Efficiency - MC Z2		
jet multiplicity	$\mathbf{p_T > 15}\,\mathrm{GeV}$	$\mathbf{p_T > 30}\,\mathrm{GeV}$
≥ 0	0.633 ± 0.001	0.633 ± 0.001
≥ 1	0.624 ± 0.001	0.623 ± 0.001
≥ 2	0.614 ± 0.002	0.609 ± 0.003
≥ 3	0.604 ± 0.003	0.585 ± 0.008
≥ 4	0.593 ± 0.006	

Table 6.1: Global Efficiency of the $Z \to ee + jets$ event selection, calculated using the MC Z2 sample. The efficiency is shown as a function of jet inclusive multiplicity, both for $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$.



(a) (b)

Figure 6.2: Cascade efficiency of the selection sequence as a function of jet inclusive multiplicity, for $p_T^{jet} > 15\,\mathrm{GeV}$ (a) and for $p_T^{jet} > 30\,\mathrm{GeV}$ (b). For this efficiency estimation the $Z + jets$ `MADGRAPH` Z2 sample was used.



(a) (b)

Figure 6.3: Cascade efficiency of the selection sequence as a function of jet inclusive multiplicity, for $p_T^{jet} > 15\,\mathrm{GeV}$ (a) and for $p_T^{jet} > 30\,\mathrm{GeV}$ (b). For this efficiency estimation the $Z + jets$ `MADGRAPH` D6T sample was used.

Selection Global Efficiency - MC D6T		
jet multiplicity	$\mathbf{p_T > 15}$ GeV	**Diff.% Z2-D6T**
≥ 0	0.616 ± 0.001	2.74%
≥ 1	0.602 ± 0.002	3.49%
≥ 2	0.589 ± 0.003	4.07%
≥ 3	0.571 ± 0.005	5.39%
≥ 4	0.557 ± 0.009	6.12%

Table 6.2: Global Efficiency of the $Z \rightarrow ee{+}jets$ event selection, calculated using the MC D6T sample. The efficiency is shown as a function of jet inclusive multiplicity, for $p_T^{jet} > 15\,\text{GeV}$. The difference between the Z2 and D6T calculated efficiencies is also shown.

Selection Global Efficiency - MC D6T		
jet multiplicity	$\mathbf{p_T > 30}$ GeV	**Diff.% Z2-D6T**
≥ 0	0.616 ± 0.001	2.74%
≥ 1	0.604 ± 0.003	3.12%
≥ 2	0.588 ± 0.006	3.59%
≥ 3	0.564 ± 0.015	3.59%

Table 6.3: Global Efficiency of the $Z \rightarrow ee{+}jets$ event selection, calculated using the MC D6T sample. The efficiency is shown as a function of jet inclusive multiplicity, for $p_T^{jet} > 30\,\text{GeV}$. The difference between the Z2 and D6T calculated efficiencies is also shown.

6.3 Tag & Probe estimation of selection efficiency

The efficiency calculation based on MC predictions leads to large discrepancies between different MC modeled samples (as shown in the previous Section) and between data and MC, due to the imperfections in the simulation. These discrepancies would produce large systematic errors in the efficiency estimation. It is therefore preferable to measure the selection efficiency from the data itself, with no reference to simulation. To perform such a measurement a data driven method was used, which is called the "Tag & Probe" (T&P) method. Taking advantage of the fact that $Z \rightarrow ee$ is a well defined resonance, the T&P selects electron couples of a well defined type and probes the efficiency of the selection criteria on those couples. In the following the T&P tool will be described in detail and the global efficiency measurement obtained by using T&P on the data sample will be shown, as usual as a function of the jet inclusive multiplicity both for $p_T^{jet} > 15\,\text{GeV}$ and $p_T^{jet} > 30\,\text{GeV}$.

6.3.1 Description of Tag & Probe tool

In the T&P method known resonances, like the Z resonance in the present case, are used to measure efficiencies. This method allows to measure the single electron efficiency for both the hard and soft selection cuts applied in the present analysis on the leading and the second Z electrons respectively. Since asymmetric cuts have been adopted, the T&P measurement of single electron efficiency will produce different efficiency values. The actual T&P efficiency for $Z \to ee$ event selection will be given therefore by the product of the two T&P single electron efficiencies.

In order to evaluate the efficiency for each one of the two legs of the $Z \to ee$ resonance, events are identified in which one of the two decay products passes tight selections (Tag selection), that aims at rejecting most of the background. The second decay product, the Probe, is used to measure the desired efficiency. The definition of the Probe object depends therefore on the specifics of the selection criteria being examined.

Tag electrons are often referred to as "golden" electrons and the fake rate for passing Tag selection criteria should be very small. Nevertheless too tight cuts on the Tags may introduce a bias in the T&P efficiency measurement, selecting T&P pairs where the Probe electron is polarized towards higher quality values of the selection parameters. The cuts used for selecting Tag electrons are the following:

- $p_T > 20\,\mathrm{GeV}$,
- $|\eta| < 1.44 \cup 1.57 < |\eta| < 2.50$,
- $d_0 < 0.05$,
- Combined Relative Isolation < 0.1,

where d_0 is the impact parameter (see Chapter 4 Sec. 4.8) and the *Combined Relative Isolation* is defined in Chapter 4 Sec. 4.5. Each Tag-Probe pair is also required to have an invariant mass between 60 GeV and 120 GeV.

Cuts on the Probes are applied in the same order specified in Chapter 4 Sec. 4.8. The T&P single electron efficiencies for each cut are therefore obtained from the Tag-Probe pairs as follows:

$$\epsilon_{lead}^{Single} = \frac{Probe(\mathrm{cut}_{lead}\,1 + ... + \mathrm{cut}_{lead}\,n)}{Probe(\mathrm{Acc}_{lead})}\,, \tag{6.5}$$

$$\epsilon_{sec}^{Single} = \frac{Probe(\mathrm{cut}_{sec}\,1 + ... + \mathrm{cut}_{sec}\,n)}{Probe(\mathrm{Acc}_{sec})}\,, \tag{6.6}$$

where $Probe(\mathrm{cut}_{lead/sec}\,1 + ... + \mathrm{cut}_{lead/sec}\,n)$ is the number of Probes passing the selection criteria up to cut number n, distinguished between *leading* (Eq. (6.5)) and *second* (Eq. (6.6)) electron criteria due to the asymmetric cuts, $Probe(\mathrm{Acc}_{lead/sec})$ is the number of Probes passing the acceptance requirement (for leading and second electron respectively) and ϵ_{lead}^{Single} and ϵ_{sec}^{Single} are the single electron efficiencies (for leading and second electron respectively). The T&P selection efficiency is therefore given by:

$$\epsilon^{TP} = \epsilon_{lead}^{Single} \cdot \epsilon_{sec}^{Single} \,. \tag{6.7}$$

Despite the background rejection operated by the Tag requirements, some background still survives the Tag cuts and the mass constraint and must therefore be subtracted from the true signal yields. The residual background subtraction is operated by a fitting procedure. In a nutshell, the efficiency estimation strategy proceeds according to the following steps. For each Probe efficiency to be evaluated, the shape of the invariant mass of Tag-Passing Probe and of Tag-Failing Probe pairs is fitted on a MC training sample of signal events only. On another MC training sample of background events only the invariant mass is plotted, both for Tag-Passing Probe and of Tag-Failing Probe pairs and its shape is fitted. The fits are performed for each jet inclusive multiplicity bin. The shapes of these distributions are the only input from Monte Carlo on which the procedure relies. The MC samples used for the training fit are the $Z+jets$ `MADGRAPH` sample with Z2 tune for the signal, and all the background samples listed in Table 3.2 of Chapter 3, except for the WZ and ZZ samples. All the background samples used are normalized to a reference luminosity of 50 pb^{-1} and merged in an unique background sample. The WZ and ZZ samples were excluded because, for the purpose of the efficiency measurement, they may be considered exactly as $Z \to ee$ events.

On the data sample used, for each multiplicity bin a simultaneous fit is performed on the Tag-Passing Probe and the Tag-Failing Probe pairs, where the signal and the background shapes are fixed in the training step. The normalization of the two distributions, for each bin, is related to the efficiency ϵ^{Single} (leading or second) as follows:

$$s_{t-pp} = s \times \epsilon^{Single} \,, \tag{6.8}$$

$$s_{t-fp} = s \times (1 - \epsilon^{Single}) \,, \tag{6.9}$$

where s is the total number of Tag-Probe pairs, s_{t-pp} is the number of signal Tag-Passing Probe pairs and s_{t-fp} is the number of Tag-Failing Probe pairs. The estimation of ϵ_{lead}^{Single} and ϵ_{sec}^{Single} for each bin is obtained from such fits. The signal shape used for the fit is a Breit Wigner convoluted with a Crystal Ball function, which is defined as

$$f(x; m, s, a, n) = \begin{cases} \dfrac{\left(\frac{n}{|a|}\right)^n e^{-\frac{1}{2}a^2}}{\left(\frac{n}{|a|} - |a| - x\right)^n} & , \text{for } x < -|a| \\ \exp\left(-\frac{1}{2}\left(\frac{x-m}{s}\right)^2\right) & , \text{for } x > -|a| \end{cases} \tag{6.10}$$

where m, s, a and n are the fit parameters. The background was instead fitted by using an exponential function. Fits were performed with the RooFit [106] toolkit, using an unbinned likelihood fit.

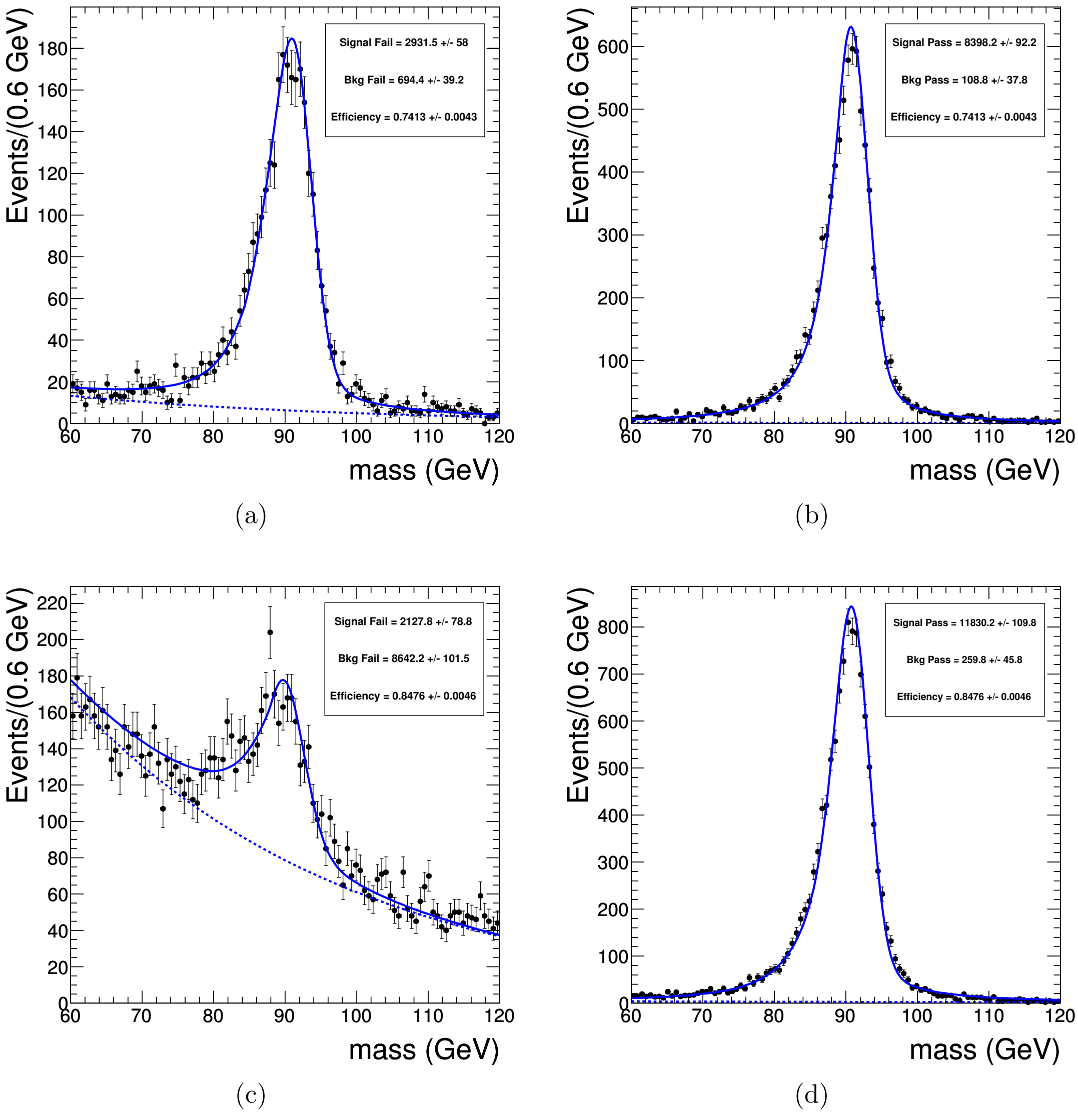


Figure 6.4: T&P fits for the inclusive bin ($\geq$ 0 jets). The fits shown are: Tag-Failing Probe (a) and Tag-Passing Probe (b) for leading electron cuts (WP80), Tag-Failing Probe (c) and Tag-Passing Probe (d) for second electron cuts (WP95). Single electron efficiencies for leading (a-b) and second (c-d) electron cuts are also shown in the plots. Data (black points), signal fit (solid line) and background fit (dashed line) are shown.

In Fig. 6.4 the T&P fits obtained for the Tag-Failing Probe pairs and the Tag-Passing Probe pairs of the inclusive bin ($\geq$ 0 jets) are shown, both for the leading (WP80) and the second (WP95) electron global selections. Single electron efficiencies of leading and second electron selections are also shown in the plots respectively. In the figures from Fig. 6.7 to Fig. 6.13 all the other fits used for performing the global T&P efficiency measurement are shown, both for $p_T^{jet} > 15\,\mathrm{GeV}$ (≥ 1 ≥ 2 ≥ 3 and ≥ 4 bins) and for $p_T^{jet} > 30\,\mathrm{GeV}$ (≥ 1 ≥ 2 and ≥ 3 bins). In Table 6.11, Table 6.12 and Table 6.13 at the end of this Chapter, the fit results, the single electron efficiencies and the global T&P efficiencies are summarized, for the inclusive bin (≥ 0 jets), the $p_T^{jet} > 15\,\mathrm{GeV}$ multiplicity bins and the $p_T^{jet} > 30\,\mathrm{GeV}$ multiplicity bins respectively.

Due to the lack of statistics for the bin 4 of inclusive multiplicity in case of $p_T^{jet} > 30\,\mathrm{GeV}$, the fit procedure does not converge. This fact does not allow to properly evaluate the background contribution, which is particularly relevant for the Tag-Failing Probe pairs. For this reason the bin 4 has not been studied in the present work for $p_T^{jet} > 30\,\mathrm{GeV}$.

6.3.2 Global efficiency measurement with Tag & Probe

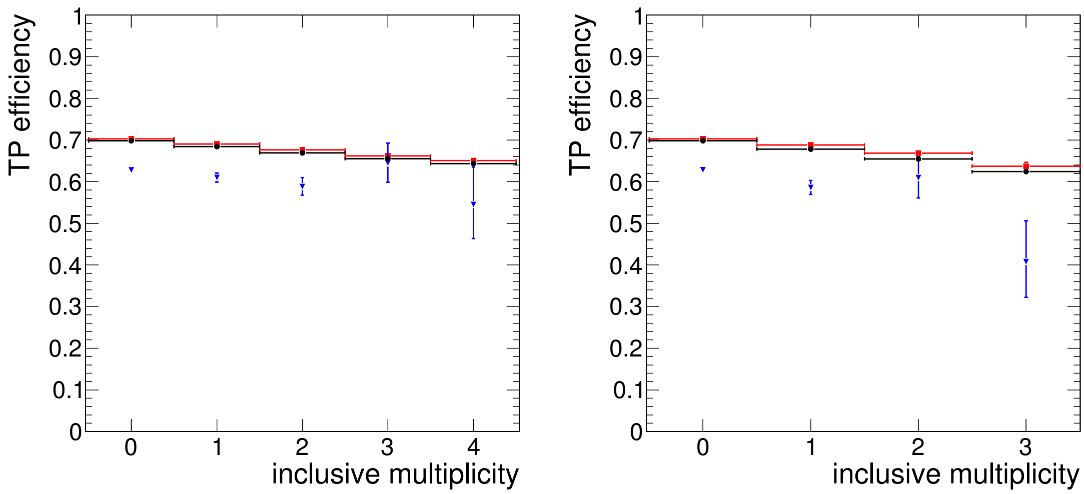
After the measurement of ϵ_{lead}^{Single} and ϵ_{sec}^{Single}, the T&P estimation of $Z \to ee$ event selection efficiency ϵ^{TP} will be given by Eq. (6.7), while its relative uncertainty will be equal to the quadratic sum of the relative uncertainties of the measured single electron efficiencies. The T&P efficiency can not be compared with the MC global efficiency defined in Eq. (6.4), since ϵ_{Glob}^{MC} is an estimation of the number of events passing the selection criteria with respect to the events generated in acceptance while T&P can not rely on generator information being a data driven method. The MC global efficiency of Eq. (6.4) can be broken down as follows:

$$\begin{aligned}\epsilon_{Glob}^{MC} = \epsilon_{Casc}^{MC}(6) &= \frac{N^{Reco}(\mathrm{Acc})}{N^{Gen}(\mathrm{Acc})} \cdot \frac{N^{Reco}(\mathrm{cut\,1} + ... + \mathrm{cut\,6})}{N^{Reco}(\mathrm{Acc})} \\ &= \epsilon_{Rel}^{MC}(\mathrm{Acc}) \cdot \frac{N^{Reco}(\mathrm{AllCuts})}{N^{Reco}(\mathrm{Acc})}\,.\end{aligned} \tag{6.11}$$

The $\epsilon_{Rel}^{MC}(\mathrm{Acc})$ factor in Eq. (6.11) can be evaluated only via MC simulations, while the second factor can be obtained by using the T&P method.

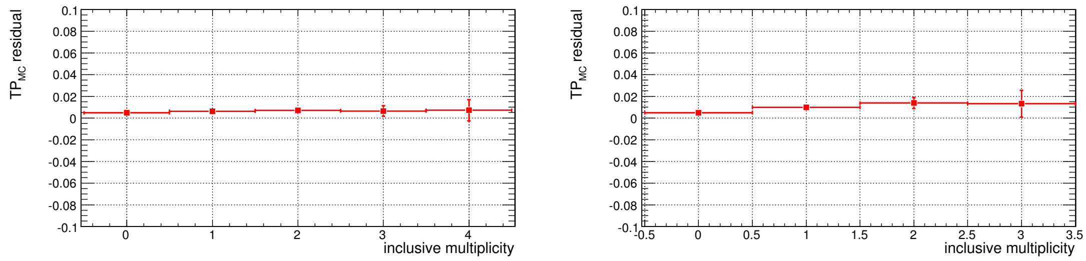
In Fig. 6.5 the T&P efficiencies measured as a function of the jet inclusive multiplicity are shown (blue points), both for $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$ respectively. The T&P efficiencies are compared with the ratio $N^{Reco}(\mathrm{AllCuts})/N^{Reco}(\mathrm{Acc})$ evaluated by using MC simulations (black points). As a closure test also the T&P efficiencies measured on the signal only sample (the same used for the MC calculation) are shown (red points), in order to check the consistency of the T&P procedure and the possible presence of a bias. In the following the T&P efficiency measured on data sample will be denoted as ϵ_{Data}^{TP}, while T&P efficiency measured on the MC signal only sample will be named ϵ_{MC}^{TP}. In the limit of no bias introduced by the T&P procedure, the ratio $N^{Reco}(\mathrm{AllCuts})/N^{Reco}(\mathrm{Acc})$ (black points) and ϵ_{MC}^{TP} (red points) would be exactly equal.

As shown in Fig. 6.5 the consistency check confirms the validity of the T&P procedure with a negligible bias, which however will be taken into account as shown in



(a) (b)

Figure 6.5: T&P efficiency as a function of jet inclusive multiplicity, for $p_T^{jet} > 15\,\mathrm{GeV}$ (a) and $p_T^{jet} > 30\,\mathrm{GeV}$ (b). T&P efficiency measured on data ϵ_{Data}^{TP} (blue) and on MC signal only sample (closure test) ϵ_{MC}^{TP} (red) are shown, compared with the ratio $N^{Reco}(\mathrm{AllCuts})/N^{Reco}(\mathrm{Acc})$ calculated from MC simulation (black).



(a) (b)

Figure 6.6: Residuals between ϵ_{MC}^{TP} and the ratio $N^{Reco}(\mathrm{AllCuts})/N^{Reco}(\mathrm{Acc})$ calculated from MC simulation. Both $p_T^{jet} > 15\,\mathrm{GeV}$ (a) and $p_T^{jet} > 30\,\mathrm{GeV}$ (b) inclusive multiplicity are shown.

the next Section. In Fig. 6.6 the residuals between ϵ_{MC}^{TP} and the ratio MC calculated $N^{Reco}(\mathrm{AllCuts})/N^{Reco}(\mathrm{Acc})$ are shown. As shown in Fig. 6.5, the T&P efficiency measured on data results to be lower than the one predicted from MC simulations. As stated in the previous Section this discrepancy is due to MC imperfections in modeling the detector response.

In Tables 6.4 and 6.5 the ϵ_{MC}^{TP} and ϵ_{Data}^{TP} measured values are shown for each bin of jet inclusive multiplicity, for $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$ respectively.

T&P Efficiency - $p_T^{jet} > 15\,\mathrm{GeV}$		
jet multiplicity	$\epsilon^{\mathbf{TP}}_{\mathbf{MC}}$	$\epsilon^{\mathbf{TP}}_{\mathbf{Data}}$
≥ 0	0.703 ± 0.001	0.628 ± 0.006
≥ 1	0.690 ± 0.001	0.610 ± 0.011
≥ 2	0.676 ± 0.002	0.588 ± 0.021
≥ 3	0.662 ± 0.004	$0.645^{+0.048}_{-0.046}$
≥ 4	0.650 ± 0.007	$0.545^{+0.089}_{-0.081}$

Table 6.4: T&P efficiency measured as a function of jet inclusive multiplicity, for $p_T^{jet} > 15\,\mathrm{GeV}$. Both the T&P efficiencies ϵ^{TP}_{MC}, measured on the MC signal only sample, and ϵ^{TP}_{Data} measured on data sample are shown.

T&P Efficiency - $p_T^{jet} > 30\,\mathrm{GeV}$		
jet multiplicity	$\epsilon^{\mathbf{TP}}_{\mathbf{MC}}$	$\epsilon^{\mathbf{TP}}_{\mathbf{Data}}$
≥ 0	0.703 ± 0.001	0.628 ± 0.006
≥ 1	0.688 ± 0.002	0.586 ± 0.017
≥ 2	0.669 ± 0.004	$0.610^{+0.051}_{-0.049}$
≥ 3	0.637 ± 0.009	$0.408^{+0.098}_{-0.086}$

Table 6.5: T&P efficiency measured as a function of jet inclusive multiplicity, for $p_T^{jet} > 30\,\mathrm{GeV}$. Both the T&P efficiencies ϵ^{TP}_{MC}, measured on the MC signal only sample, and ϵ^{TP}_{Data} measured on data sample are shown.

6.4 Final efficiency estimation

The selection efficiency measurement used for the cross section calculation takes advantage of both the T&P measurement and the MC estimation (necessary for the $\epsilon^{MC}_{Rel}(\mathrm{Acc})$ calculation). As stated before, in the ideal case of no bias in the T&P measurement the ratio $N^{Reco}(\mathrm{AllCuts})/N^{Reco}(\mathrm{Acc})$ calculated by using MC signal sample and the ϵ^{TP}_{MC} efficiency described in the previous section would be exactly the same. Therefore, in order to correct for a possible residual bias in the T&P measurement, the following $Z \to ee$ selection efficiency estimation ϵ_Z is used:

$$\begin{aligned}\epsilon_Z = \epsilon^{MC}_{Rel}(\mathrm{Acc}) \cdot \frac{N^{Reco}(\mathrm{AllCuts})}{N^{Reco}(\mathrm{Acc})} \cdot \frac{\epsilon^{TP}_{Data}}{\epsilon^{TP}_{MC}} \\ = \epsilon^{MC}_{Glob} \cdot \frac{\epsilon^{TP}_{Data}}{\epsilon^{TP}_{MC}} = \epsilon^{MC}_{Glob} \cdot \rho\,,\end{aligned} \tag{6.12}$$

where ϵ^{MC}_{Glob} is the selection global efficiency calculated by using the MC method described in Sec. 6.2, while ϵ^{TP}_{Data} and ϵ^{TP}_{MC} are the T&P efficiencies described in Sec. 6.3.2, measured on data sample and on the MC signal only sample respectively.

$\epsilon_{\mathbf{Z}} = \epsilon^{\mathbf{MC}}_{\mathbf{Glob}} \cdot \rho$ **(MC Z2)**				
jet multiplicity	$\mathbf{p_T} > \mathbf{15}\,\mathrm{GeV}$	**Diff.%** $\epsilon^{\mathbf{MC}}_{\mathbf{Glob}} - \epsilon_{\mathbf{Z}}$	$\mathbf{p_T} > \mathbf{30}\,\mathrm{GeV}$	**Diff.%** $\epsilon^{\mathbf{MC}}_{\mathbf{Glob}} - \epsilon_{\mathbf{Z}}$
≥ 0	0.566 ± 0.006	-11%	0.566 ± 0.006	-11%
≥ 1	0.551 ± 0.010	-12%	0.531 ± 0.015	-15%
≥ 2	0.534 ± 0.019	-13%	$0.556^{+0.046}_{-0.045}$	-9%
≥ 3	$0.588^{+0.044}_{-0.043}$	-3%	$0.375^{+0.090}_{-0.079}$	-36%
≥ 4	$0.497^{+0.082}_{-0.074}$	-16%		

Table 6.6: Final efficiency estimation as a function of jet inclusive multiplicity, evaluated by using T&P data driven efficiency and the MC $Z + jets$ simulation with tune Z2 for the MC part of Eq. (6.12). The percentage difference with respect the MC efficiencies evaluated in Sec. 6.2 is also shown.

In Eq. (6.12) the variable ρ defined as

$$\rho = \frac{\epsilon^{TP}_{Data}}{\epsilon^{TP}_{MC}} , \qquad (6.13)$$

has been introduced, and the corrected efficiency for each multiplicity bin i can be finally expressed as:

$$(\epsilon_Z)_i = \left(\epsilon^{MC}_{Glob} \cdot \rho\right)_i . \qquad (6.14)$$

In Table 6.6 the final selection efficiency estimation is shown, evaluated as described in Eq. (6.12) as a function of jet inclusive multiplicity both for $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$. As shown in Table 6.6 the final estimated efficiency is lower than the one obtained from the MC only estimation of about 11% for the inclusive bin, while for the other bins the difference is included between 3% for the bin 3 of $p_T^{jet} > 15\,\mathrm{GeV}$ and 36% for the bin 3 of $p_T^{jet} > 30\,\mathrm{GeV}$ (which is however affected by a large statistical error). This fact gives reason for the discrepancy between data and MC yields after all selections applied, which has been already shown for the

$\epsilon_{\mathbf{Z}} = \epsilon^{\mathbf{MC}}_{\mathbf{Glob}} \cdot \rho$ **(MC D6T)**		
jet multiplicity	$\mathbf{p_T} > \mathbf{15}\,\mathrm{GeV}$	**Diff.%** $\epsilon_{\mathbf{Z}}$**(D6T)-**$\epsilon_{\mathbf{Z}}$**(Z2)**
≥ 0	0.567 ± 0.006	+0.18%
≥ 1	0.553 ± 0.010	+0.26%
≥ 2	0.540 ± 0.019	+1.12%
≥ 3	$0.596^{+0.044}_{-0.042}$	+1.37%
≥ 4	$0.497^{+0.083}_{-0.075}$	-0.02%

Table 6.7: Final efficiency estimation as a function of jet inclusive multiplicity, evaluated by using T&P data driven efficiency and the MC $Z + jets$ simulation with tune D6T for the MC part of Eq. (6.12).

$\epsilon_{\mathbf{Z}} = \epsilon_{\mathbf{Glob}}^{\mathbf{MC}} \cdot \rho$ (MC D6T)		
jet multiplicity	$\mathbf{p_T} > \mathbf{30}\,\mathrm{GeV}$	**Diff.%** $\epsilon_{\mathbf{Z}}$**(D6T)-**$\epsilon_{\mathbf{Z}}$**(Z2)**
≥ 0	0.567 ± 0.006	+0.18%
≥ 1	0.528 ± 0.016	-0.60%
≥ 2	$0.566^{+0.048}_{-0.046}$	+1.76%
≥ 3	$0.356^{+0.085}_{-0.076}$	-5.10%

Table 6.8: Final efficiency estimation as a function of jet inclusive multiplicity, evaluated by using T&P data driven efficiency and the MC $Z + jets$ simulation with tune D6T for the MC part of Eq. (6.12).

inclusive bin in Table 4.7 of Chapter 4 Sec. 4.9. An analogous effect is present also for the other multiplicity bins, as will be shown in the next Chapter.

In order to check the stability of the procedure adopted for the efficiency measurement, the same measurement has been redone by using the $Z + jets$ MC sample simulated with tune D6T for the calculation of the parts of Eq. (6.12) where MC simulations are involved. The $Z + jets$ D6T sample was also used as signal training in order to check the stability of the fitting procedure. In Table 6.7 and in Table 6.8 the values of the efficiency $\epsilon_{Glob}^{MC} \cdot \rho$ obtained by using D6T simulations for the MC parts of the efficiency calculation are shown. While for the MC global efficiency the difference between the Z2 and the D6T estimations shown in Table 6.2 and Table 6.3 is about $3 - 6\%$, in case of the $\epsilon_{Glob}^{MC} \cdot \rho$ measurements shown in Table 6.6, Table 6.7 and Table 6.8 the discrepancy becomes lower than 1.8%, except for the bin 3 of $p_T^{jet} > 30\,\mathrm{GeV}$ where the difference is about 5% due to the lower available statistics which produces larger uncertainties in the fitting step.

The differences between the Z2 and D6T estimations of $\epsilon_{Glob}^{MC} \cdot \rho$ will be considered as a systematic uncertainty on the estimated efficiencies. Finally in Table 6.9 and Table 6.10 the efficiencies used in the next Chapter for the $Z \rightarrow ee + jets$ analysis are summarized, for $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$ respectively.

$\epsilon_{\mathbf{Z}} = \epsilon_{\mathbf{Glob}}^{\mathbf{MC}} \cdot \rho$ **- jet** $\mathbf{p_T} > \mathbf{15}\,\mathrm{GeV}$			
jet multiplicity	$\epsilon_{\mathbf{Z}}$	**Statistical Unc.**	**Systematic Unc.**
≥ 0	0.566	± 0.006	+0.18%
≥ 1	0.551	± 0.010	+0.26%
≥ 2	0.534	± 0.019	+1.12%
≥ 3	0.588	$^{+0.044}_{-0.043}$	+1.37%
≥ 4	0.497	$^{+0.082}_{-0.074}$	−0.02%

Table 6.9: Final estimation of efficiency ϵ_Z as a function of jet inclusive multiplicity, for $p_T^{jet} > 15\,\mathrm{GeV}$. Statistical and systematic uncertainties are shown.

$\epsilon_{\mathbf{Z}} = \epsilon^{\mathbf{MC}}_{\mathbf{Glob}} \cdot \rho$ - **jet** $\mathbf{p_T} > \mathbf{30}$ GeV			
jet multiplicity	$\epsilon_{\mathbf{Z}}$	**Statistical Unc.**	**Systematic Unc.**
≥ 0	0.566	± 0.006	$+0.18\%$
≥ 1	0.531	± 0.015	-0.60%
≥ 2	0.556	$^{+0.046}_{-0.045}$	$+1.76\%$
≥ 3	0.375	$^{+0.090}_{-0.079}$	-5.10%

Table 6.10: Final estimation of efficiency ϵ_Z as a function of jet inclusive multiplicity, for $p_T^{jet} > 30$ GeV . Statistical and systematic uncertainties are shown.

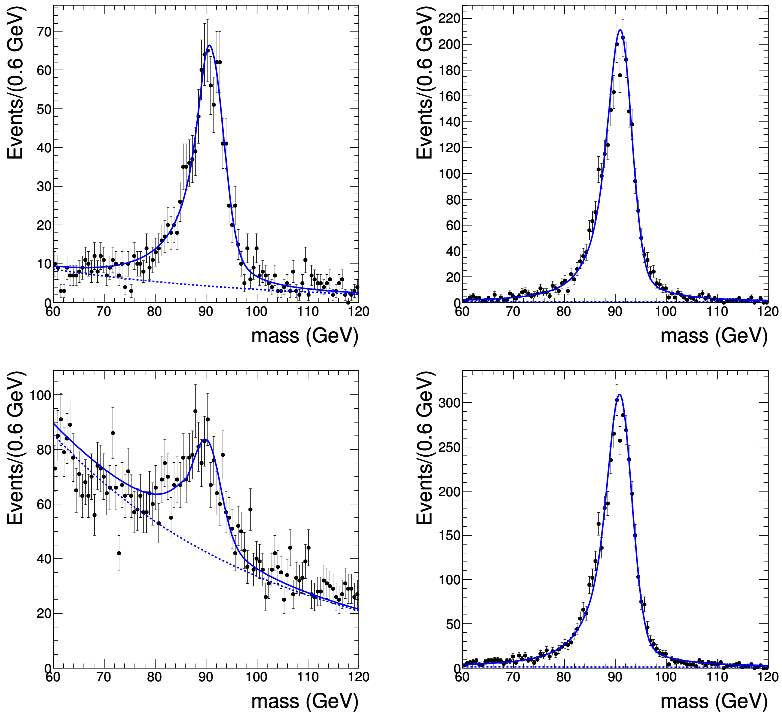


Figure 6.7: T&P fits: multiplicity bin ≥ 1, $p_T^{jet} > 15\,\text{GeV}$. Failing - Pass fits for leading electron cuts (top) and Failing - Pass fits for second electron cuts (bottom) are shown. Data (black points), signal fit (solid line) and background fit (dashed line) are shown.

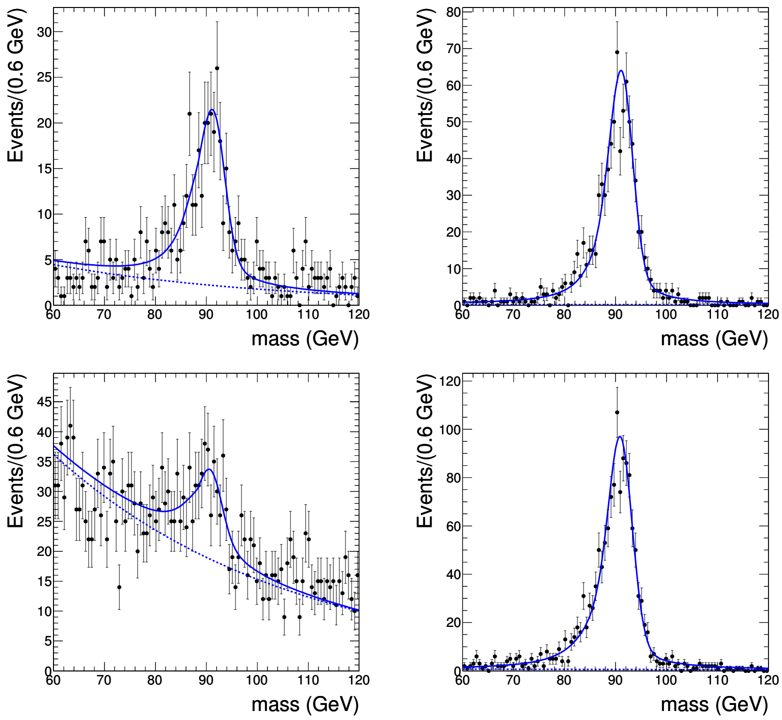


Figure 6.8: T&P fits: multiplicity bin ≥ 2, $p_T^{jet} > 15\,\text{GeV}$. Failing - Pass fits for leading electron cuts (top) and Failing - Pass fits for second electron cuts (bottom) are shown. Data (black points), signal fit (solid line) and background fit (dashed line) are shown.

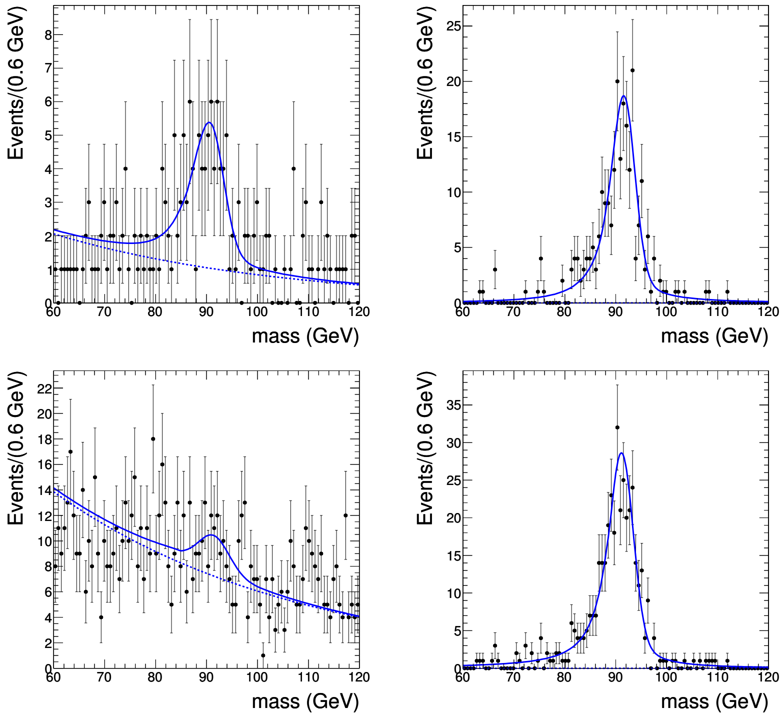


Figure 6.9: T&P fits: multiplicity bin ≥ 3, $p_T^{jet} > 15\,\mathrm{GeV}$. Failing - Pass fits for leading electron cuts (top) and Failing - Pass fits for second electron cuts (bottom) are shown. Data (black points), signal fit (solid line) and background fit (dashed line) are shown.

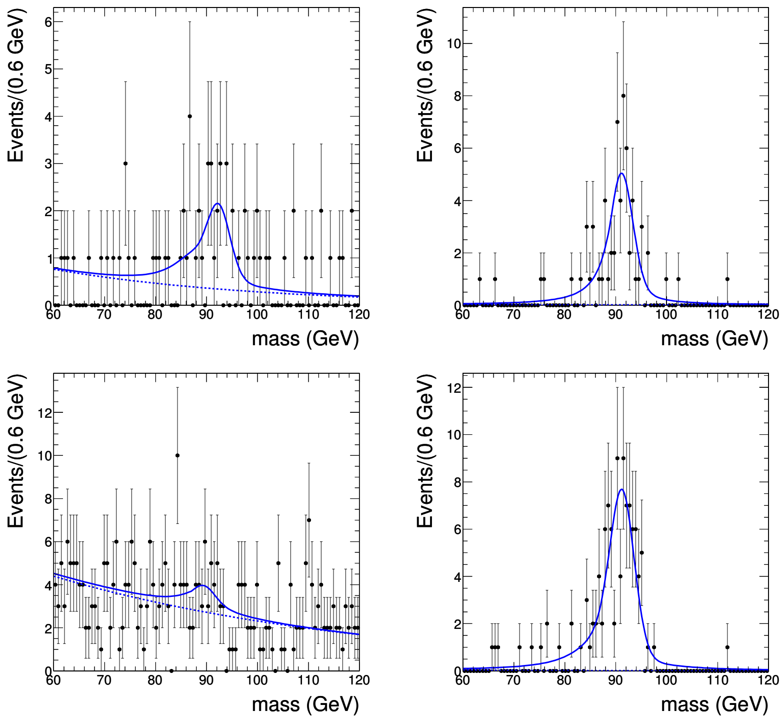


Figure 6.10: T&P fits: multiplicity bin ≥ 4, $p_T^{jet} > 15\,\mathrm{GeV}$. Failing - Pass fits for leading electron cuts (top) and Failing - Pass fits for second electron cuts (bottom) are shown. Data (black points), signal fit (solid line) and background fit (dashed line) are shown.

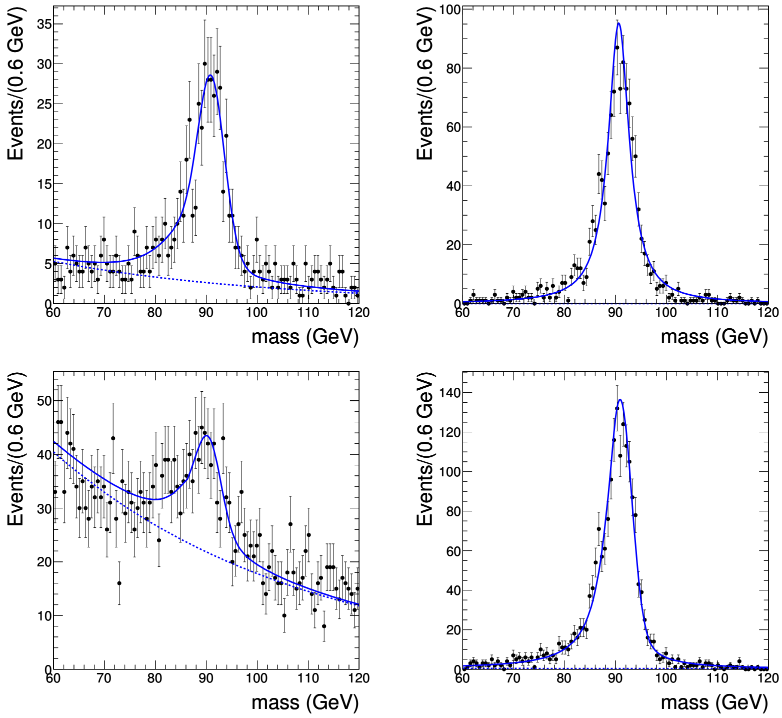


Figure 6.11: T&P fits: multiplicity bin ≥ 1, $p_T^{jet} > 30\,\mathrm{GeV}$. Failing - Pass fits for leading electron cuts (top) and Failing - Pass fits for second electron cuts (bottom) are shown. Data (black points), signal fit (solid line) and background fit (dashed line) are shown.

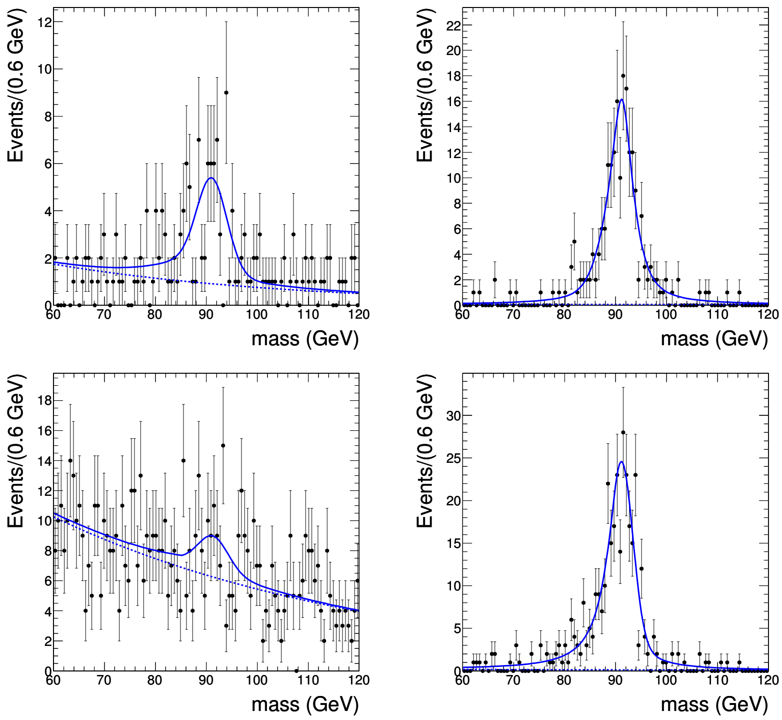


Figure 6.12: T&P fits: multiplicity bin ≥ 2, $p_T^{jet} > 30\,\mathrm{GeV}$. Failing - Pass fits for leading electron cuts (top) and Failing - Pass fits for second electron cuts (bottom) are shown. Data (black points), signal fit (solid line) and background fit (dashed line) are shown.

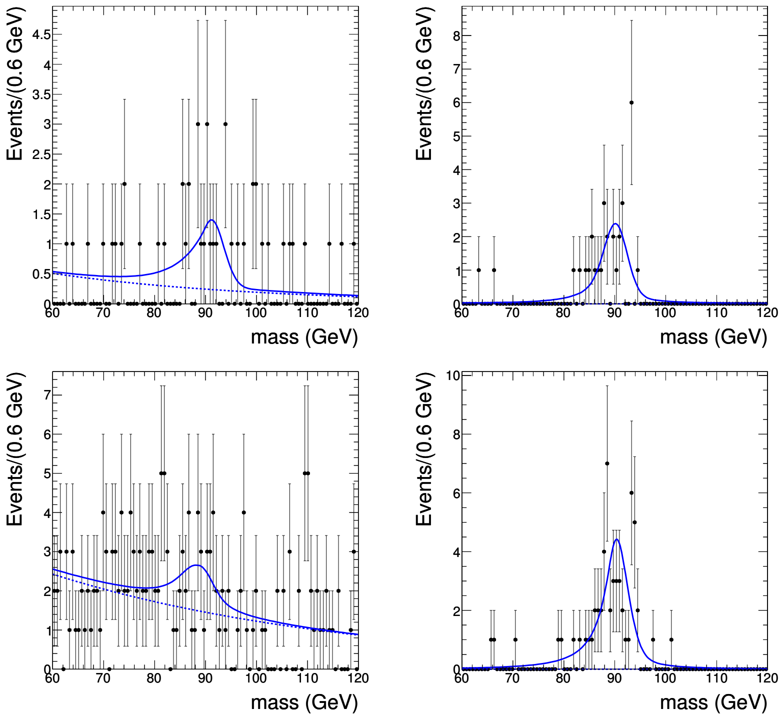


Figure 6.13: T&P fits: multiplicity bin ≥ 3, $p_T^{jet} > 30\,\text{GeV}$. Failing - Pass fits for leading electron cuts (top) and Failing - Pass fits for second electron cuts (bottom) are shown. Data (black points), signal fit (solid line) and background fit (dashed line) are shown.

T&P Efficiency - Inclusive Bin									
Multiplicity		**Leading Electron**			**Second Electron**			$\epsilon^{\mathrm{TP}}_{\mathrm{Data}}$
		Fail	**Pass**	**Single Eff.**	**Fail**	**Pass**	**Single Eff.**	
≥ 0	Signal	2931 ± 58	8398 ± 92	0.741 ± 0.004	2128 ± 79	11830 ± 110	0.848 ± 0.005	0.628 ± 0.006
	Background	694 ± 39	109 ± 38		8642 ± 101	260 ± 46		

Table 6.11: T&P signal and background fit results for the inclusive bin ≥ 0 jets. Single electron efficiencies are shown for both leading and second electron selections. T&P global efficiency ϵ^{TP}_{Data} is also shown.

T&P Efficiency - jet $\mathbf{p_T > 15\,GeV}$								
Multiplicity		**Leading Electron**			**Second Electron**			$\epsilon^{\mathrm{TP}}_{\mathrm{Data}}$
		Fail	**Pass**	**Single Eff.**	**Fail**	**Pass**	**Single Eff.**	
≥ 1	Signal	1012 ± 36	2808 ± 53	0.735 ± 0.008	883 ± 55	4318 ± 66	0.830 ± 0.006	0.610 ± 0.011
	Background	466 ± 29	44 ± 22		4598 ± 73	106 ± 29		
≥ 2	Signal	319 ± 21	849 ± 29	0.727 ± 0.015	316 ± 35	1342 ± 37	0.810 ± 0.015	0.588 ± 0.021
	Background	242 ± 19	17 ± 13		2033 ± 48	60 ± 17		
≥ 3	Signal	76 ± 11	245 ± 16	0.763 ± 0.028	72 ± 21	394 ± 20	0.845 ± 0.031	$0.645^{+0.048}_{-0.046}$
	Background	113 ± 12	0.1 ± 38		794 ± 29	0 ± 6		
≥ 4	Signal	30 ± 7	63 ± 8	0.676 ± 0.056	25 ± 8	105 ± 21	0.806 ± 0.060	$0.545^{+0.089}_{-0.081}$
	Background	40 ± 8	1.7 ± 2.5		283 ± 17	1.5 ± 2.6		

Table 6.12: T&P signal and background fit results for the $p_T^{jet} > 15\,\mathrm{GeV}$ inclusive multiplicity. Single electron efficiencies are shown for both leading and second electron selections. T&P global efficiency ϵ^{TP}_{Data} is also shown.

T&P Efficiency - jet $\mathbf{p_T} > 30\,\mathrm{GeV}$									
Multiplicity		**Leading Electron**			**Second Electron**			$\epsilon^{\mathbf{TP}}_{\mathbf{Data}}$
		Fail	**Pass**	**Single Eff.**	**Fail**	**Pass**	**Single Eff.**	
≥ 1	Signal	436 ± 24	1181 ± 34	0.730 ± 0.012	452 ± 39	1836 ± 43	0.802 ± 0.013	0.586 ± 0.017
	Background	284 ± 21	0 ± 32		2325 ± 52	31 ± 19		
≥ 2	Signal	79 ± 11	204 ± 14	0.721 ± 0.031	60 ± 14	328 ± 34	0.845 ± 0.033	$0.610^{+0.051}_{-0.049}$
	Background	99 ± 12	6.1 ± 5.9		662 ± 27	12 ± 8		
≥ 3	Signal	20 ± 6	31 ± 6	0.603 ± 0.077	27 ± 10	57 ± 7	0.676 ± 0.071	$0.408^{+0.098}_{-0.086}$
	Background	27 ± 6	0 ± 8		152 ± 13	0 ± 8		

Table 6.13: T&P signal and background fit results for the $p_T^{jet} > 30\,\mathrm{GeV}$ inclusive multiplicity. Single electron efficiencies are shown for both leading and second electron selections. T&P global efficiency ϵ^{TP}_{Data} is also shown.

Chapter 7

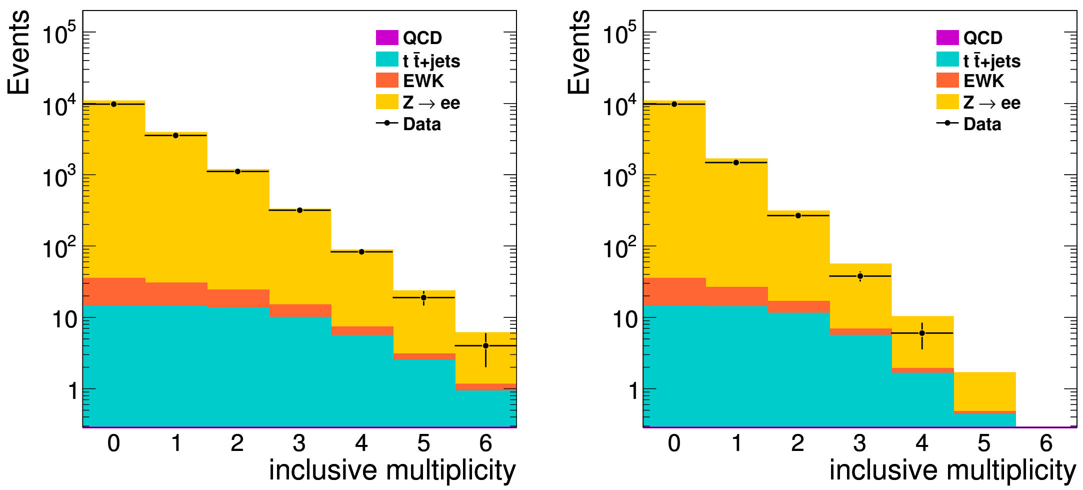
Cross section measurement

In this Chapter, after studying the reconstructed inclusive jet multiplicities, the measurement of the cross section $\sigma(\mathrm{Z} + \geq n \text{ jets})$ is evaluated both for $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$. In fact jet multiplicity needs first to be corrected for the reconstruction efficiencies shown in the previous Chapter and for the effects of the finite detector resolution, in order to go back to the true multiplicity distribution. In the following the analysis of the jet multiplicity will be shown together with the study of the systematic uncertainties that affect the measurement. Finally the $\sigma(\mathrm{Z} + \geq n \text{ jets})$ cross section will be obtained from the corrected multiplicity distributions.

7.1 Raw jet multiplicity

In Fig. 7.1 the inclusive multiplicity distributions of the reconstructed jets are shown, both for $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$. In these plots the number of data, MC signal and MC background yields are shown for each multiplicity bin, after having applied all the $Z \to ee$ selection cuts described in Chapter 4 Sec. 4.8. The MC samples are normalized to the data luminosity. For the signal simulation the $Z + jets$ `MADGRAPH` sample with Z2 tune has been used (see Chapter 3 Sec. 3.6 for a complete MC dataset description).

This kind of multiplicity distribution is named "raw multiplicity", since no corrections are applied. The plots in Fig. 7.1 show the inclusive multiplicity up to the bin ≥ 6 jets, in order to compare reconstructed data and MC at this first step (for $p_T^{jet} > 30\,\mathrm{GeV}$ the bins ≥ 5 and ≥ 6 have no data yields). Nevertheless only the bins 0, 1, 2, 3, 4 for $p_T^{jet} > 15\,\mathrm{GeV}$ and 0, 1, 2, 3 for $p_T^{jet} > 30\,\mathrm{GeV}$ are taken into account for the analysis, since for higher multiplicity values the lack of statistics do not allow a proper evaluation of the selection efficiency, as described in the previous Chapter. In Tables 7.1 and 7.2 data and MC yields are shown as a function of the multiplicity bins, for $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$ respectively. As shown in Fig. 7.1 and in Tables 7.1 and 7.2 no QCD background events are left after all $Z \to ee$ selection cuts applied.



(a) (b)

Figure 7.1: Raw multiplicity distributions after all $Z \to ee$ selections applied, for $p_T^{jet} > 15\,\text{GeV}$ (a) and for $p_T^{jet} > 30\,\text{GeV}$ (b). Data and MC signal (MADGRAPH Z2) and background yields are shown. No QCD events left after all cuts applied.

raw multiplicity yields - jet $p_T > 15\,\text{GeV}$							
Sample	≥ 0	≥ 1	≥ 2	≥ 3	≥ 4	≥ 5	≥ 6
$t\bar{t}$+jets	14.2 ± 0.3	14.2 ± 0.3	13.5 ± 0.3	9.9 ± 0.2	5.5 ± 0.2	2.47 ± 0.11	0.92 ± 0.06
EWK	20.6 ± 0.5	15.8 ± 0.4	10.4 ± 0.3	5.0 ± 0.2	1.8 ± 0.1	0.57 ± 0.07	0.21 ± 0.07
Z+jets	10682 ± 16	3841 ± 9	1151 ± 5	312 ± 2	80.5 ± 1.4	20.4 ± 0.7	4.9 ± 0.3
Total MC	10717 ± 16	3871 ± 9	1175 ± 5	327 ± 3	87.7 ± 1.4	23.4 ± 0.7	6.04 ± 0.34
Data	9717 ± 99	3543 ± 59	1111 ± 33	318 ± 18	83 ± 9	19 ± 4	4 ± 2

Table 7.1: Data and MC yields as a function of jet inclusive multiplicity for $p_T^{jet} > 15\,\text{GeV}$, after all $Z \to ee$ selections applied. MC samples are normalized to data luminosity. No QCD events left after all cuts applied.

raw multiplicity yields - jet $p_T > 30\,\text{GeV}$						
Sample	≥ 1	≥ 2	≥ 3	≥ 4	≥ 5	≥ 6
$t\bar{t}$+jets	14.0 ± 0.3	11.3 ± 0.2	5.5 ± 0.2	1.6 ± 0.1	0.42 ± 0.04	0.06 ± 0.01
EWK	12.1 ± 0.2	5.3 ± 0.1	1.26 ± 0.01	0.257 ± 0.008	0.042 ± 0.003	0.008 ± 0.001
Z+jets	1628 ± 6	291 ± 3	49.1 ± 1.1	8.3 ± 0.4	1.20 ± 0.16	0.14 ± 0.06
Total MC	1654 ± 6	308 ± 3	55.8 ± 1.1	10.2 ± 0.4	1.67 ± 0.17	0.21 ± 0.06
Data	1483 ± 38	267 ± 16	38 ± 6	6 ± 2	0	0

Table 7.2: Data and MC yields as a function of jet inclusive multiplicity for $p_T^{jet} > 30\,\text{GeV}$, after all $Z \to ee$ selections applied (bin $\geq$ 0 is shown in Table 7.1). MC samples normalized to data luminosity. No QCD events left after all cuts applied.

The Tables also show a discrepancy between the data and MC yields after all cuts are applied, of the order of 10% for the inclusive bin ($\geq$ 0 jets) and lower for the others. As described in the previous Chapter (Sec. 6.4) this discrepancy is related to the different efficiency between data and MC, which is about 10% for the inclusive bin. For the bins of multiplicity higher than 0 also the bin by bin migration must be taken into account, as described in the following.

7.2 Signal and background extraction

In order to properly evaluate signal and background contributions to the content of each multiplicity bin, a data driven method was used which takes advantage of the T&P machinery described in the previous Chapter. By fitting the Z mass plots for each multiplicity bin it is possible to extract signal and background contributions, except for the WZ and ZZ backgrounds which are considered as $Z \rightarrow ee$ events by the fitting procedure. These two background contributions must be therefore evaluated from the MC predictions. The WZ and ZZ yields (included in EWK in Tables 7.1 and 7.2) evaluated by using the MC `PYTHIA` samples, described in Chapter 3 Sec. 3.6, are shown in Table 7.3 for each analyzed multiplicity bin, both for $p_T^{jet} > 15\,\text{GeV}$ and $p_T^{jet} > 30\,\text{GeV}$.

The mass plots to be fitted are obtained for each multiplicity bin from data that have passed all the selection cuts for both the leading and the second electrons. The number of entries of each plot must be therefore equal to the data yields shown in Tables 7.1 and 7.2 for each multiplicity bin. The fit functions used are the convolution of a Crystal Ball with a Breit Wigner for the signal and an exponential function for the background, as already described for the T&P procedure in Chapter 6 Sec. 6.3.1. Thus the signal shape is first fitted for each multiplicity bin on a MC training sample of signal events only, while the background shape is fitted on the invariant mass plot obtained from a MC training sample of background events only. The signal and background shapes are then fitted to the data Z mass plot, keeping their shapes fixed (only the signal peak is free to shift). The sum of the integrals of the two fitted curves must be equal to the number of entries, i.e. to the number of yields shown in Tables 7.1 and 7.2 for each multiplicity bin. In

WZ - ZZ background yields ($\mathbf{N_{WZ-ZZ}}$) from MC datasets					
	≥ 0	≥ 1	≥ 2	≥ 3	≥ 4
$p_T^{jet} > 15\,\text{GeV}$	13.98 ± 0.06	12.32 ± 0.05	8.96 ± 0.04	4.35 ± 0.03	1.59 ± 0.02
$p_T^{jet} > 30\,\text{GeV}$	13.98 ± 0.06	10.41 ± 0.05	5.01 ± 0.03	1.24 ± 0.02	-

Table 7.3: WZ and ZZ background yields as a function of jet inclusive multiplicity after all $Z \rightarrow ee$ selections applied, both for $p_T^{jet} > 15\,\text{GeV}$ and $p_T^{jet} > 30\,\text{GeV}$. MC `PYTHIA` samples normalized to data luminosity are used for the yields evaluation.

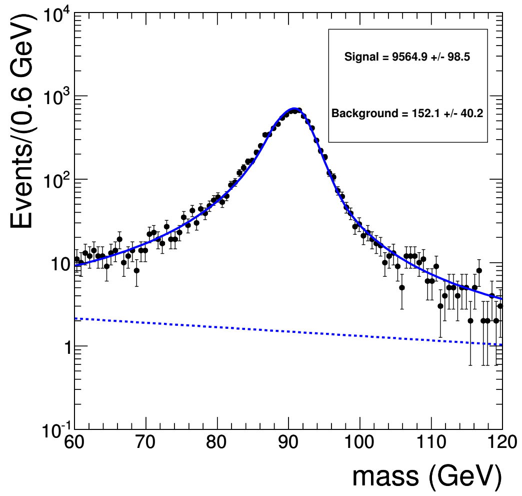


Figure 7.2: Z invariant mass fit for the inclusive bin (≥ 0 jets). Signal and background estimations are shown in the plots.

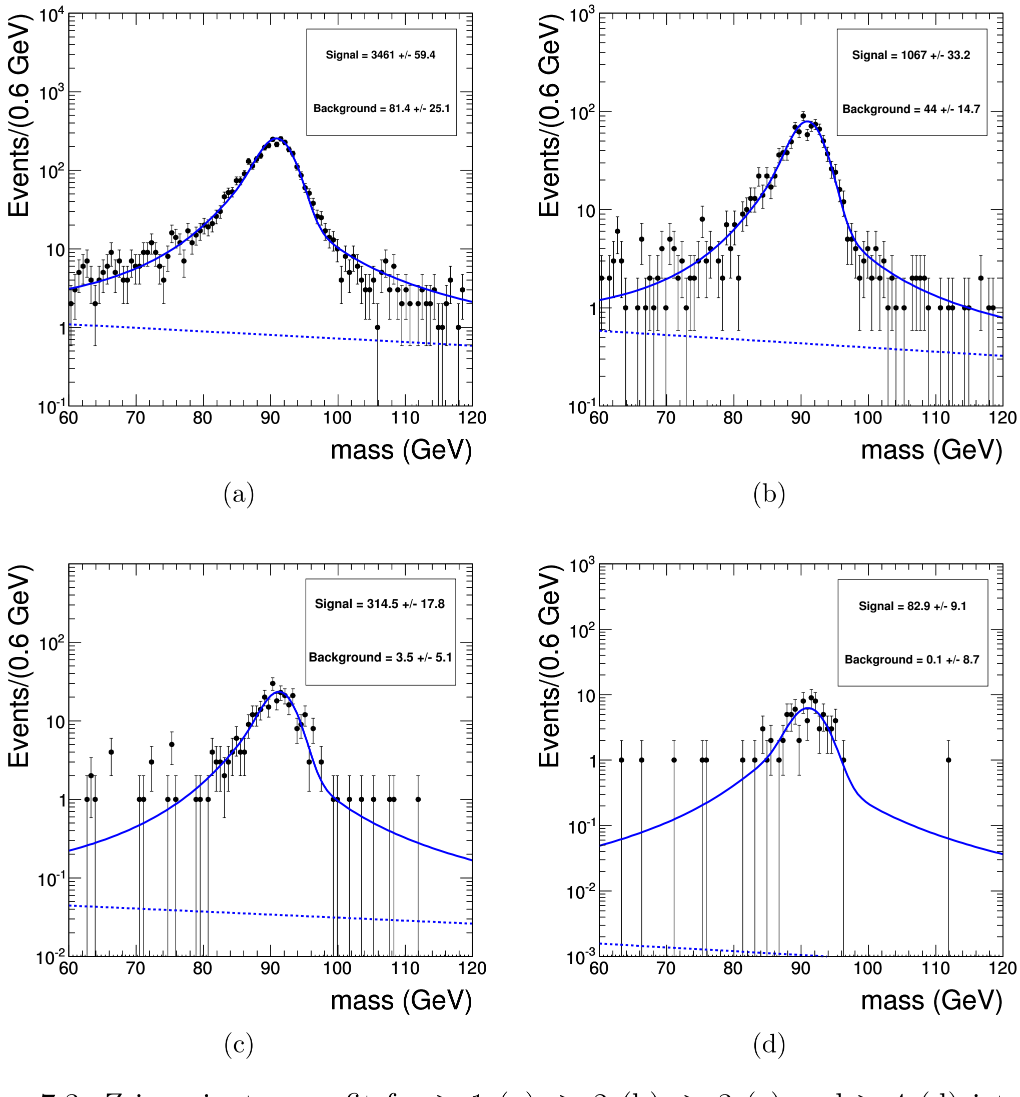
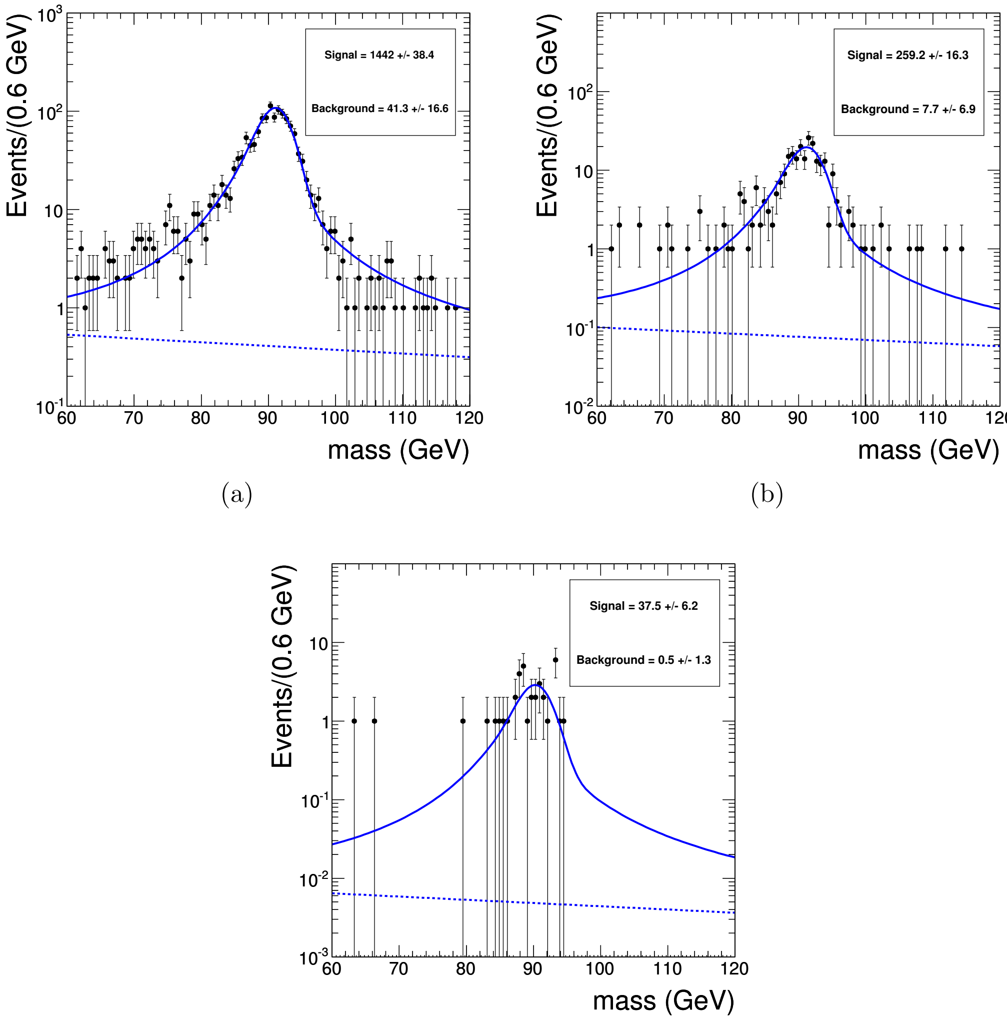


Figure 7.3: Z invariant mass fit for ≥ 1 (a), ≥ 2 (b), ≥ 3 (c) and ≥ 4 (d) jets for $p_T^{jet} > 15\,\mathrm{GeV}$. Signal and background estimations are shown in the plots.



(a) (b)

(c)

Figure 7.4: Z invariant mass fit for $\geq$ 1 (a), $\geq$ 2 (b), and $\geq$ 3 (c) jets for $p_T^{jet} > 30\,\mathrm{GeV}$. Signal and background estimations are shown in the plots.

Fig. 7.2 the signal and background fit obtained for the inclusive bin is shown, while in Fig. 7.3 and Fig. 7.4 the fits obtained for the other multiplicity bins are shown, both for $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$. In Tables 7.4 and 7.5 the signal and background estimations extracted by using the fitting procedure are summarized.

For the present analysis the signal yields N_{Sig} estimated for each multiplicity bin i are therefore:

$$(N_{Sig})_i = \left(N_{Sig}^{Fit} - N_{WZ-ZZ}\right)_i \,, \tag{7.1}$$

where N_{Sig}^{Fit} is the signal estimation extracted from the Z invariant mass fits (shown in Tables 7.4 and 7.5), while N_{WZ-ZZ} is the MC estimation for WZ-ZZ background, as summarized in Table 7.3. The signal yields N_{Sig} of Eq. (7.1) are affected by the

$\mathbf{p_T^{jet}} > \mathbf{15}\,\mathrm{GeV}$	$\mathbf{N_{Sig}^{Fit}}$	$\mathbf{N_{Bkg}^{Fit}}$
≥ 0	9565 ± 98	152 ± 40
≥ 1	3461 ± 59	81 ± 25
≥ 2	1067 ± 33	44 ± 15
≥ 3	315 ± 18	3.5 ± 5.1
≥ 4	83 ± 9	0.1 ± 8.7

Table 7.4: Signal and background yields extracted from the Z invariant mass fits, as a function of jet inclusive multiplicity for $p_T^{jet} > 15\,\mathrm{GeV}$.

$\mathbf{p_T^{jet}} > \mathbf{30}\,\mathrm{GeV}$	$\mathbf{N_{Sig}^{Fit}}$	$\mathbf{N_{Bkg}^{Fit}}$
≥ 0	9565 ± 98	152 ± 40
≥ 1	1442 ± 38	41 ± 17
≥ 2	259 ± 16	7.7 ± 6.9
≥ 3	37 ± 6	0.5 ± 1.3

Table 7.5: Signal and background yields extracted from the Z invariant mass fits, as a function of jet inclusive multiplicity for $p_T^{jet} > 30\,\mathrm{GeV}$.

Signal estimation $\mathbf{N_{Sig}}$		
$\mathbf{p_T^{jet}} > \mathbf{15}\,\mathrm{GeV}$	$\mathbf{N_{Sig}} = \left(\mathbf{N_{Sig}^{Fit}} - \mathbf{N_{WZ-ZZ}}\right)$	**Statistical Unc.**
≥ 0	9551	± 98
≥ 1	3449	± 59
≥ 2	1058	± 33
≥ 3	311	± 18
≥ 4	81	± 9

Table 7.6: Signal estimations evaluated as shown in Eq. (7.1) as a function of jet inclusive multiplicity for $p_T^{jet} > 15\,\mathrm{GeV}$. The statistical uncertainties are shown, while the systematic ones are negligible.

Signal estimation $\mathbf{N_{Sig}}$		
$\mathbf{p_T^{jet}} > \mathbf{30}\,\mathrm{GeV}$	$\mathbf{N_{Sig}} = \left(\mathbf{N_{Sig}^{Fit}} - \mathbf{N_{WZ-ZZ}}\right)$	**Statistical Unc.**
≥ 0	9551	± 98
≥ 1	1432	± 38
≥ 2	254	± 16
≥ 3	36	± 6

Table 7.7: Signal estimations evaluated as shown in Eq. (7.1) as a function of jet inclusive multiplicity for $p_T^{jet} > 30\,\mathrm{GeV}$. The statistical uncertainties are shown, while the systematic ones are negligible.

statistical uncertainty of N_{Sig}^{Fit} (shown in Tables 7.4 and 7.5), while the uncertainty on N_{WZ-ZZ} should be considered as a contribution to the systematics. Nevertheless the relative contribution of the N_{WZ-ZZ} systematic uncertainty to the total uncertainty on N_{Sig} is of the order of $5 \cdot 10^{-6} \div 4 \cdot 10^{-4}$, being therefore completely negligible with respect to the statistical uncertainty as well as to the other sources of systematic uncertainty described in the following. In Tables 7.6 and 7.7 the signal yields estimated for each multiplicity bin are shown, both for $p_T^{jet} > 15\,\text{GeV}$ and $p_T^{jet} > 30\,\text{GeV}$, together with their uncertainties.

7.3 Efficiency correction

For the measurement of $\sigma(\text{Z} + \geq \text{n jets})$, the number of signal events must be corrected for detector efficiency, in order to extrapolate the true number of events produced in pp collisions for each multiplicity bin. The results are not corrected for detector acceptance in order to keep them independent of MC generators that would be used to make such correction and of the added uncertainties this would introduce.

Making use of the efficiency definition introduced in Chapter 6, the number of events corrected for detector efficiency $N_{Sig}^{EffCorr}$ is defined for each multiplicity bin as

$$\left(N_{Sig}^{EffCorr}\right)_i = \left(\frac{N_{Sig}^{Fit} - N_{WZ-ZZ}}{\epsilon_{Glob}^{MC} \cdot \rho}\right)_i = \left(\frac{N_{Sig}}{\epsilon_Z}\right)_i , \tag{7.2}$$

where i is the multiplicity bin and Eqs. (6.14) and (7.1) have been used. The measurement of ϵ_Z has been previously described in Chapter 6 and its measured values with their uncertainties have been summarized in Tables 6.9 and 6.10 of Chapter 6. The efficiency uncertainty should be considered as a systematic uncertainty for the measurement that we are interested in. Its contribution to the overall systematic uncertainty of the measurement will be therefore discussed in the following, together with the other sources of systematics.

7.4 Unfolding multiplicity distributions

In order to extrapolate the true multiplicity distribution, one more correction must be applied because of the limited detector momentum resolution. The finite resolution in jet momentum reconstruction can indeed modify the true multiplicity distribution in the sense that for each event it can affect the number of jets that pass the p_T thresholds, causing a possible migration of the event to neighboring multiplicity bins. Without such a correction the results produced by two different experiments can not be compared, and $\sigma(\text{Z} + \geq \text{n jets})$ can not be correctly measured for $n > 0$.

To correct for this effect, an *unfolding* procedure has been applied, in order to retrieve the true multiplicity distributions both for $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$. The method adopted to perform this correction is to compare reconstructed and generated multiplicity distributions of the MC sample $Z + jets$ `MADGRAPH` with Z2 tune, calculating bin by bin the correction factors which allow to retrieve the generated distribution from the reconstructed one. The correction factor used are therefore defined as

$$C_i = \left(\frac{N_{Jet}^{Gen}}{N_{Jet}^{Reco}} \right)_i , \tag{7.3}$$

where N_{Jet}^{Gen} is the number of generated jets, N_{Jet}^{Reco} is the number of reconstructed jets and i is the inclusive multiplicity bin. For the correction factor calculation of Eq. (7.3), the presence of a Z boson generated in the acceptance is also required. In fact, in order to compare generated and reconstructed multiplicity distributions coming from the same event sample, obtaining therefore a correction factor equal to 1 for the inclusive bin (≥ 0 jets), the events with a Z generated out of the acceptance must be rejected. In Fig. 7.5 the Z2 `MADGRAPH` generated and reconstructed multiplicity distributions are compared, while in Fig. 7.6 the calculated correction factors are shown, both for $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$. In Table 7.8 the calculated correction factors are summarized with their statistical uncertainties. As shown in Table 7.8 the statistical uncertainties on the correction factors go from about 0.07% at bin 1 of both $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$, up to 0.6% for the bin 4 of $p_T^{jet} > 15\,\mathrm{GeV}$. These uncertainties are therefore completely negligible with respect to the statistical and systematic uncertainties of the measurement, as shown in the following.

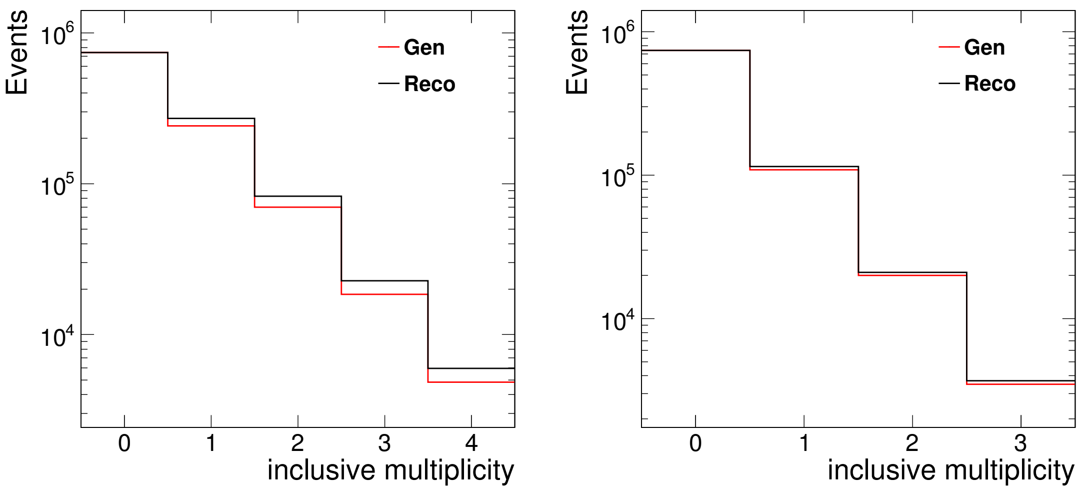
Once the unfolding correction factors have been calculated, the corrected multiplicity distributions may be obtained. Starting from Eqs. (7.1) and (7.2) the true number of events for each multiplicity bin i is given by:

$$\left(N_{Sig}^{True}\right)_i = \left(N_{Sig}^{EffCorr}\right)_i \cdot C_i = \left(\frac{N_{Sig}^{Fit} - N_{WZ-ZZ}}{\epsilon_Z} \right)_i \cdot C_i \,, \tag{7.4}$$

where $\left(N_{Sig}^{Fit}\right)_i$ is the number of signal yields in bin i extracted from the Z invariant

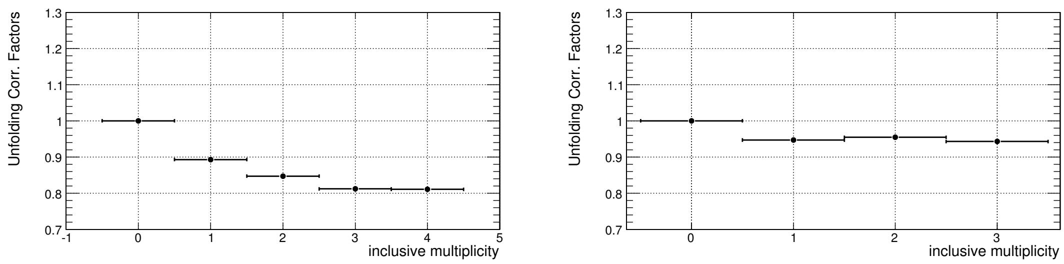
Unfolding Correction Factors - `MADGRAPH` Z2 sample					
	≥ 0	≥ 1	≥ 2	≥ 3	≥ 4
$p_T^{jet} > 15\,\mathrm{GeV}$	1	0.8927 ± 0.0006	0.847 ± 0.001	0.812 ± 0.003	0.811 ± 0.005
$p_T^{jet} > 30\,\mathrm{GeV}$	1	0.9469 ± 0.0007	0.954 ± 0.001	0.943 ± 0.004	-

Table 7.8: Unfolding Correction Factors, evaluated for each inclusive multiplicity bin as shown in Eq. (7.3) by using $Z + jets$ `MADGRAPH` Z2 sample, both for $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$. Statistical uncertainties are also shown.



(a) (b)

Figure 7.5: Generated (red) and reconstructed (black) jet inclusive multiplicity distributions for $p_T^{jet} > 15\,\mathrm{GeV}$ (a) and $p_T^{jet} > 30\,\mathrm{GeV}$ (b), obtained from the MC sample $Z + jets$ `MADGRAPH` Z2. A Z boson generated in the acceptance is required. The MC sample is not normalized (corresponding to a luminosity of 1593 pb^{-1}).



(a) (b)

Figure 7.6: Unfolding correction factors evaluated by using the $Z + jets$ `MADGRAPH` Z2 sample, for $p_T^{jet} > 15\,\mathrm{GeV}$ (a) and $p_T^{jet} > 30\,\mathrm{GeV}$ (b).

mass fits, N_{WZ-ZZ} is the WZ-ZZ background in bin i evaluated from MC, ϵ_Z is the measured efficiency for the bin i and C_i is the unfolding correction factor evaluated for bin i by using Eq. (7.3). The unfolded multiplicity distribution obtained from Eq. (7.4) is therefore corrected for the effects of the detector, in terms of efficiency and momentum resolution. It can be therefore compared with the MC predictions and then used for the cross section measurement. In Fig. 7.7 the results of the unfolded inclusive multiplicity are shown, both for $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$, together with a comparison with the MC generated multiplicity distribution and a ratio between the unfolded multiplicity and the MC predictions. Only statistical uncertainties are considered for these plots.

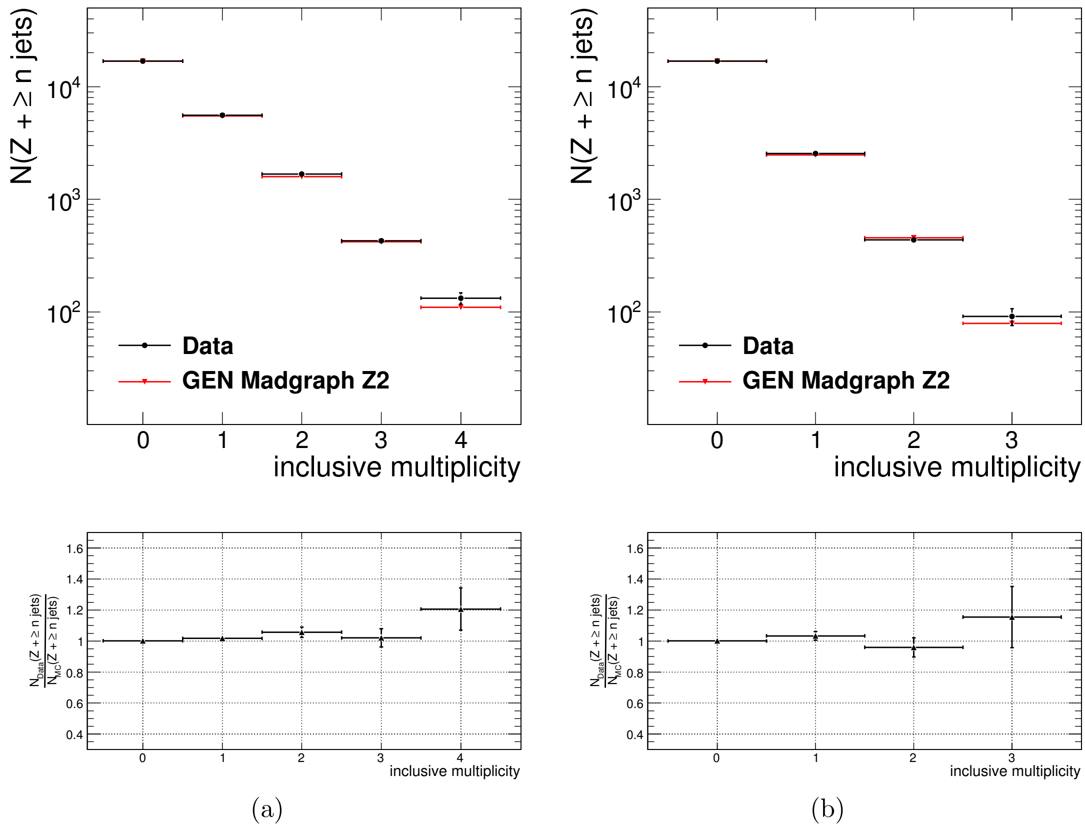


Figure 7.7: Data inclusive multiplicity distributions corrected for efficiency and unfolding (black), for $p_T^{jet} > 15\,\text{GeV}$ (a) and $p_T^{jet} > 30\,\text{GeV}$ (b). Inclusive multiplicity distribution of $Z + jets$ `MADGRAPH` Z2 generated jets is also shown (red). The ratios between data corrected multiplicity and the MC predictions are shown at the bottom.

A close agreement is observed in Fig. 7.7 between the data unfolded multiplicity and the MC truth. As shown in both the ratio plots, unfolded data and MC predictions are indeed in agreement within the statistical uncertainty for all the multiplicity bins, except for the bins 3 and 4 of $p_T^{jet} > 15\,\text{GeV}$. Once the systematic uncertainty that affects the measurement is taken into account, these last two bins also show agreement between data and MC, as will be shown in the following.

7.5 Systematic uncertainties

In order to properly evaluate the measurement uncertainty, several sources of systematic uncertainties must be taken into account. For the present analysis the sources of systematics are relative to the efficiency measurement, to the unfolding procedure, to the jet energy corrections and to the measured luminosity of the data

Efficiency Systematics					
	≥ 0	≥ 1	≥ 2	≥ 3	≥ 4
$p_T^{jet} > 15\,\mathrm{GeV}$	1.0%	1.8%	$^{+3.8\%}_{-3.6\%}$	$^{+7.3\%}_{-7.4\%}$	$^{+15\%}_{-16\%}$
$p_T^{jet} > 30\,\mathrm{GeV}$	1.0%	$^{+2.9\%}_{-3.0\%}$	$^{+8.1\%}_{-8.3\%}$	$^{+21\%}_{-25\%}$	-

Table 7.9: Systematic uncertainties due to the efficiency measurement as a function of inclusive multiplicity, both for $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$.

sample. In the following each one of these sources of systematic uncertainties will be described. The total systematic uncertainty on the measurement will be given as the quadratic sum of each one of the following contributions.

7.5.1 Systematics of the efficiency

As described in Chapter 6 Sec. 6.4, the measured efficiency is affected by a statistical uncertainty, relative to the measurement method and to the available statistics, and by a systematic uncertainty. The latter takes into account the stability of the Tag & Probe procedure used for performing the measurement, and was evaluated as a comparison with the efficiency obtained by using the D6T tune MC sample, in place of the Z2 sample, for the signal training in the fitting step and for calculating the MC efficiency that has to be rescaled to the data driven estimated one (see Chapter 6 Sec. 6.4). Both the statistical and systematic uncertainties of the efficiency measurement are added in quadrature in order to provide the efficiency contribution to the total systematic uncertainty of the final measurement.

As shown in Tables 6.9 and 6.10 of Chapter 6 Sec. 6.4, the larger contribution to the uncertainty on the efficiency is the statistical one, which is of the order of 1% for the inclusive bin (≥ 0 jets) up to 25% for the bin 3 of $p_T^{jet} > 30\,\mathrm{GeV}$ multiplicity distribution, while the systematic is about 0.2% for the inclusive bin and it is below the 2% for the others, except for the bin 3 of $p_T^{jet} > 30\,\mathrm{GeV}$ where it is about 5%. The large statistical uncertainty is due to the lack of statistics for high multiplicity bins (in particular for $p_T^{jet} > 30\,\mathrm{GeV}$).

In Table 7.9 the efficiency contribution to the systematic uncertainty of the measurement is shown, both for $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$.

7.5.2 Systematics of the unfolding procedure

The unfolding procedure described in Sec. 7.4 is affected by the particular choice of the MC simulation used for calculating the correction factors, i.e. the $Z + jets$ `MADGRAPH` with tune Z2. In order to evaluate the systematic uncertainty due to the simulation chosen for the correction factors calculation, the unfolding was also performed using another MC simulation, i.e. the D6T $Z + jets$ `MADGRAPH`. In Fig. 7.8

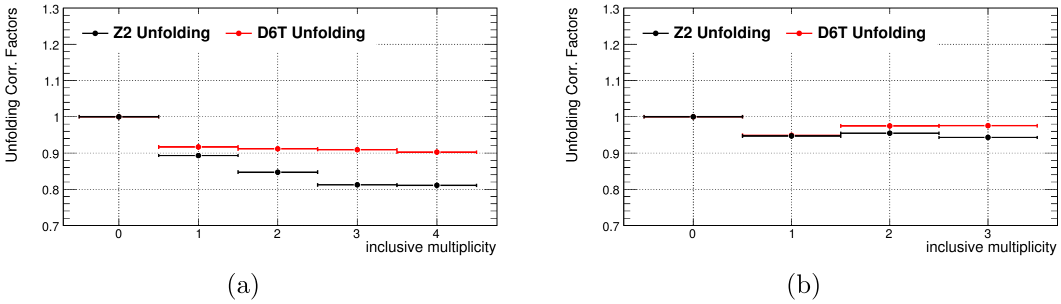


Figure 7.8: Unfolding correction factors obtained from Z2 MADGRAPH simulation (black) and D6T MADGRAPH simulation (red), for $p_T^{jet} > 15\,\mathrm{GeV}$ (a) and $p_T^{jet} > 30\,\mathrm{GeV}$ multiplicity distributions.

Unfolding Systematics					
	≥ 0	≥ 1	≥ 2	≥ 3	≥ 4
$p_T^{jet} > 15\,\mathrm{GeV}$	-	+2.3%	+7.6%	+12%	+11%
$p_T^{jet} > 30\,\mathrm{GeV}$	-	+0.2%	+2.1%	+3.4%	-

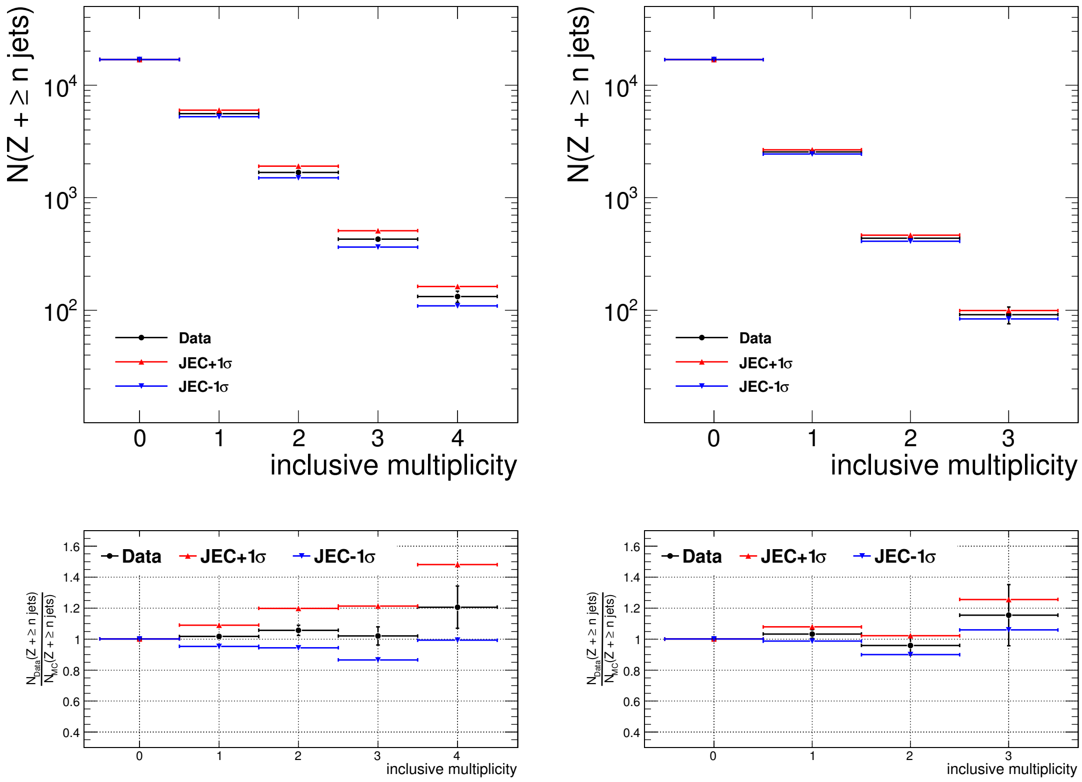
Table 7.10: Systematic uncertainties due to the unfolding procedure evaluated as a function of inclusive multiplicity, both for $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$.

the correction factors obtained by using the D6T simulation are compared with the ones shown in Sec. 7.4, evaluated by using the Z2 simulation, both for $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$. The relative difference between the Z2 and D6T correction factors is therefore considered as a systematic uncertainty of the measurement for each multiplicity bin, except for the inclusive bin which is not unfolded. The contributions of this source of systematic uncertainty are shown in Table 7.10 as a function of the inclusive multiplicity, both for $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$. Since this uncertainty have been evaluated as a difference between two different sets of unfolding correction factors, its contribution to the total systematic uncertainty results to be asymmetric.

7.5.3 Systematics for the jet energy correction

The uncertainty in jet energy calibration directly affects the jet multiplicity spectrum and therefore is an important source of systematic uncertainty for the present measurement. This contribution to the systematic uncertainty is evaluated as the change in the multiplicity distribution induced by shifting the p_T^{jet} by $\pm\sigma_{JEC}$. It is estimated with pure MC samples because of the limited data statistics in the high multiplicity jet bins.

To calculate σ_{JEC} the uncertainties of all the jet energy correction factors L1,



(a) (b)

Figure 7.9: Data inclusive multiplicity distributions corrected for efficiency and unfolding (black) and systematic uncertainties due to jet energy calibration, evaluated by shifting p_T^{jet} by $+\sigma_{JEC}$ (red) and $-\sigma_{JEC}$ (blue). Both the multiplicity distributions obtained for $p_T^{jet} > 15\,\mathrm{GeV}$ (a) and $p_T^{jet} > 30\,\mathrm{GeV}$ (b) are shown. The ratios between data corrected multiplicity and the MC predictions are shown at the bottom, together with the JEC systematics.

L2, L3 and L2L3Residual (see Chapter 5 Sec. 5.3) are added in quadrature. As described in Sec. 5.3 the total uncertainty on p_T^{jet} for PF jets with $p_T = 15\,\mathrm{GeV}$ is about 8% and it decreases to about 3% for $p_T = 30\,\mathrm{GeV}$.

In Fig. 7.9 the unfolded multiplicity evaluated in Sec. 7.4 is shown both for $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$, together with the systematic uncertainties obtained by shifting the p_T^{jet} by $\pm\sigma_{JEC}$. The systematics on the ratios between data and MC `MADGRAPH` generated jets are also shown. As expected the jet energy calibration contribution to the systematic uncertainty on the measured multiplicity distribution obtained for $p_T^{jet} > 15\,\mathrm{GeV}$ is larger than for the one at $p_T^{jet} > 30\,\mathrm{GeV}$. The systematic uncertainties from this source of error are shown in Table 7.11 both for $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$ multiplicity bins.

JEC Systematics					
	≥ 0	≥ 1	≥ 2	≥ 3	≥ 4
$p_T^{jet} > 15\,\mathrm{GeV}$	-	$^{+7.2\%}_{-6.2\%}$	$^{+13\%}_{-11\%}$	$^{+19\%}_{-15\%}$	$^{+23\%}_{-18\%}$
$p_T^{jet} > 30\,\mathrm{GeV}$	-	$^{+4.4\%}_{-4.5\%}$	$^{+6.5\%}_{-6.2\%}$	$^{+8.7\%}_{-8.3\%}$	-

Table 7.11: Systematic uncertainties due to the jet energy calibration uncertainty evaluated as a function of inclusive multiplicity, both for $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$.

7.5.4 Systematic from the luminosity measurement

The last contribution to the systematic uncertainty is due to the error on the luminosity measurement. The luminosity uncertainty affects the cross section measurement, which is given by

$$\sigma\,(\mathrm{Z}+\geq i\,\mathrm{jets}) = \frac{\left(N_{Sig}^{True}\right)_i}{\mathscr{L}}\,, \tag{7.5}$$

where $\left(N_{Sig}^{True}\right)_i$ is the final estimation of signal events (corrected for the efficiency and unfolding) given by Eq. (7.4) for the i-th multiplicity bin and $\mathscr{L}$ is the data luminosity. This uncertainty affects equally all the multiplicity bins both for $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$ since it is relative to the knowledge of the actual size of the data sample collected. As described in Chapter 2 Sec. 2.5 the measured luminosity is [66]

$$\mathscr{L} = (36.2\,\pm\,1.4)\,\mathrm{pb}^{-1}\,. \tag{7.6}$$

The contribution of the luminosity uncertainty to the total systematic uncertainty of the measurement is therefore equal to 4% for each multiplicity bin, both for $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$.

Also MC predicted events for each multiplicity bin are normalized to the data luminosity and thus the luminosity uncertainty affects the data-MC comparisons. In the following, the systematic uncertainties shown in the data-MC comparison plots of signal yields will include therefore also the luminosity systematics.

7.6 Measurement of $\sigma(\mathrm{Z}+\geq n$ jets)

After the study of all the systematic uncertainties, the final estimation of the multiplicity distributions can be made. As stated above, all the examined sources of systematic uncertainties are added in quadrature, in order to obtain the total systematic uncertainty. The results are compared with the MC predictions obtained from the $Z+jets$ `MADGRAPH` simulation with tune Z2 and D6T. In order to compare the measurement with other MC generators, also a $Z+jets$ `PYTHIA` sample with Z2 tune has been used to check the results. In Figs. 7.10 and 7.12 the final measured number of events, corrected for efficiency and unfolding, are shown as a function of the inclusive multiplicity, respectively for $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$, together with `MADGRAPH` and `PYTHIA` predictions. Both statistical and overall systematic uncertainties are shown. In order to highlight the different contributions to the systematic uncertainty, the ratios of the measured multiplicity distribution with respect to the MC `MADGRAPH` Z2 predictions are shown in Figs. 7.11 and 7.13. In these plots the different sources of systematic uncertainties are stacked in the following order: luminosity, efficiency, unfolding and JEC. Although the systematic due to the luminosity uncertainty does not affect the signal yield measurement, it must be considered when comparing data and MC predictions due to the normalization of the MC samples to the data luminosity.

The agreement of the measured multiplicity distributions with the `MADGRAPH` Z2 predictions is very good both for $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$, as shown in Figs. 7.10 and 7.12. In contrast, increasing disagreement is observed with `PYTHIA` Z2 as the number of jets increases, since `PYTHIA` is a fixed-order tree-level matrix element calculator, and therefore does not handle the presence of extra jets. For what concerns the MC tune dependence of the simulated multiplicity distributions one can see a large discrepancy between the measurements and the D6T tune predictions for $p_T^{jet} > 15\,\mathrm{GeV}$, while for $p_T^{jet} > 30\,\mathrm{GeV}$ the Z2 and D6T distributions are almost the same and they both are in agreement with the measurements.

As shown in Figs. 7.11 and 7.13, the contributions to the total systematic uncertainty due to the jet energy calibration and to the unfolding are more important for the $p_T^{jet} > 15\,\mathrm{GeV}$ multiplicity distribution, while in the case of $p_T^{jet} > 30\,\mathrm{GeV}$ distribution the dominant source of systematic uncertainty is due to the efficiency measurement. The efficiency systematic contribution is strictly related to the available statistics, since it is due to the data driven procedure adopted for determining the efficiency (see Chapter 6). Therefore the contribution related to this systematic source as well as the one related to the luminosity should strongly decrease with a larger size of the analyzed data sample.

In Tables 7.12 and 7.13 the final measured yields are shown with their statistical and systematic uncertainties, for $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$ respectively. The inclusive bin (≥ 0 jets) is shown only in Table 7.12.

Finally, the measurement of the cross section $\sigma(\mathrm{Z} + \geq \mathrm{n}$ jets) is given for each multiplicity bin by dividing the measured yields summarized in Tables 7.12 and 7.13 and their uncertainties by the measured luminosity, as shown in Eq. (7.5). In Figs. 7.14 and 7.15 the obtained values of $\sigma(\mathrm{Z} + \geq \mathrm{n}$ jets) expressed in pb are shown versus the inclusive multiplicity bins, for $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$ respectively. In Tables 7.14 and 7.15 the cross section values measured for each multiplicity bin and their uncertainties are summarized. The luminosity uncertainty is the most important contribution for the systematic uncertainty of the inclusive bin. As stated above, an increase in the size of the analyzed data sample would

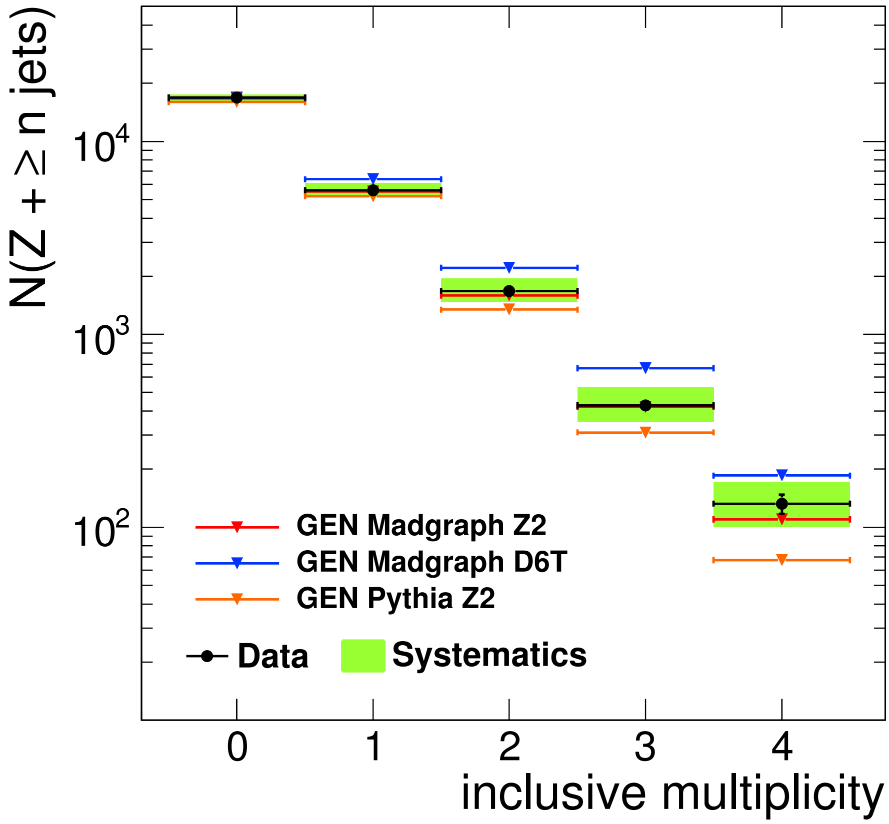


Figure 7.10: Measurement of N(Z+≥n jets) for $p_T^{jet} > 15\,\text{GeV}$ (black points). Statistical (black bars) and total systematic (green area) uncertainties are shown. MC `MADGRAPH` Z2 (red triangles), `MADGRAPH` D6T (blue triangles) and `PYTHIA` Z2 (orange triangles) predictions are also shown.

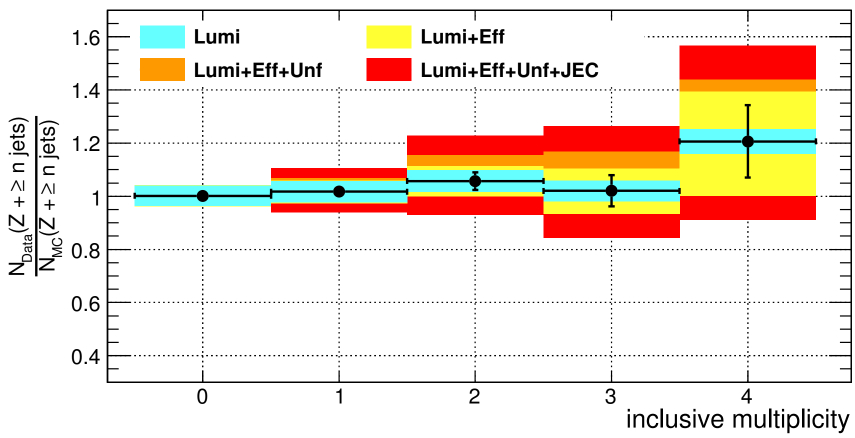


Figure 7.11: Ratio between the measured multiplicity distribution for $p_T^{jet} > 15\,\text{GeV}$ and the MC `MADGRAPH` Z2 predictions. Systematic uncertainties are shown separately, while black error bars represent the statistical uncertainty. Different systematic contributions are added in quadrature.

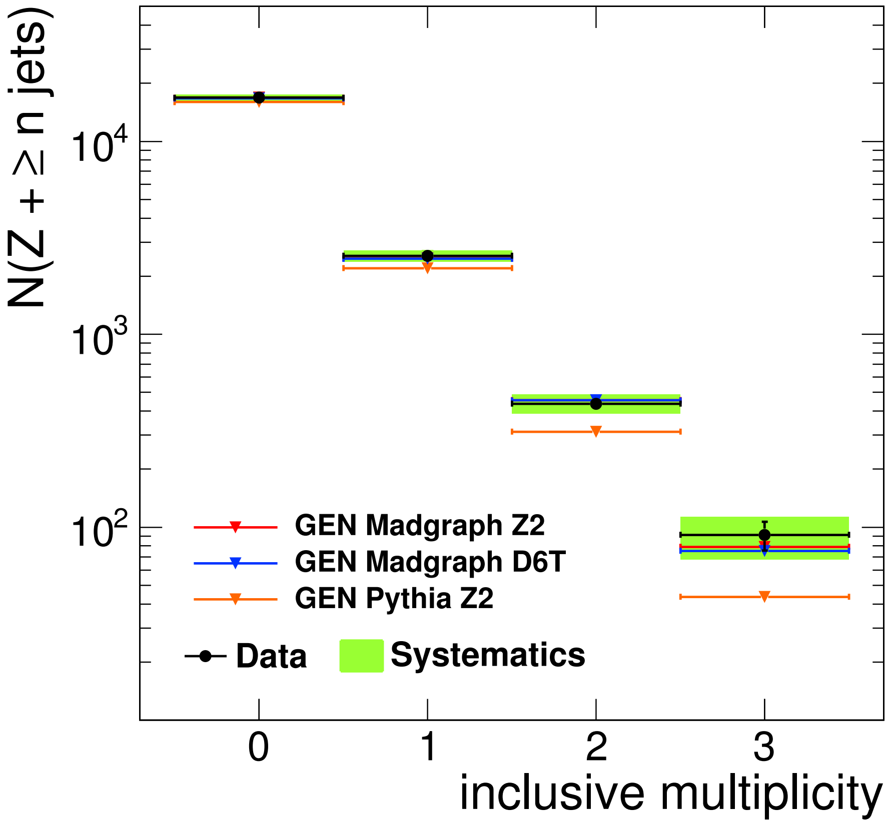


Figure 7.12: Measurement of N(Z+≥n jets) for $p_T^{jet} > 30\,\mathrm{GeV}$ (black points). Statistical (black bars) and total systematic (green area) uncertainties are shown. MC `MADGRAPH` Z2 (red triangles), `MADGRAPH` D6T (blue triangles) and `PYTHIA` Z2 (orange triangles) predictions are also shown.

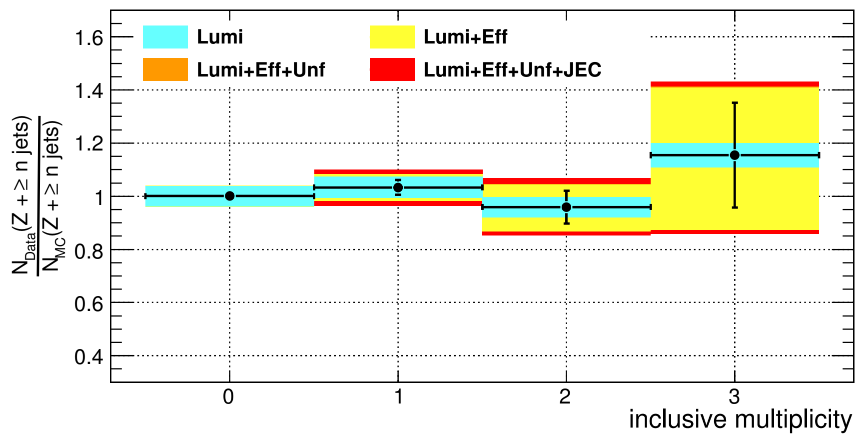


Figure 7.13: Ratio between the measured multiplicity distribution for $p_T^{jet} > 30\,\mathrm{GeV}$ and the MC `MADGRAPH` Z2 predictions. Systematic uncertainties are shown separately, while black error bars represent the statistical uncertainty. Different systematic contributions are added in quadrature.

Final Measured Yields - $p_T^{jet} > 15\,\mathrm{GeV}$			
	Measured Yields	**Statistical Unc.**	**Systematic Unc.**
≥ 0	16871	± 174	± 166
≥ 1	5585	± 96	$^{+441}_{-363}$
≥ 2	1678	± 53	$^{+265}_{-189}$
≥ 3	428	± 25	$^{+101}_{-72}$
≥ 4	132	± 15	$^{+39}_{-32}$

Table 7.12: Final measured yields as a function of inclusive multiplicity for $p_T^{jet} > 15\,\mathrm{GeV}$. Statistical and systematic uncertainties are shown.

Final Measured Yields - $p_T^{jet} > 30\,\mathrm{GeV}$			
	Measured Yields	**Statistical Unc.**	**Systematic Unc.**
≥ 1	2554	± 69	± 136
≥ 2	436	± 28	± 46
≥ 3	91	± 15	$^{+22}_{-23}$

Table 7.13: Final measured yields as a function of inclusive multiplicity for $p_T^{jet} > 30\,\mathrm{GeV}$. Statistical and systematic uncertainties are shown. The inclusive bin ≥ 0 jets is shown in Table 7.12.

strongly reduce the luminosity contribution to the overall systematic uncertainty as well as the one due to the efficiency measurement, which affects in particular the $p_T^{jet} > 30\,\mathrm{GeV}$ multiplicity distribution and the bin 4 of the $p_T^{jet} > 15\,\mathrm{GeV}$ multiplicity distribution.

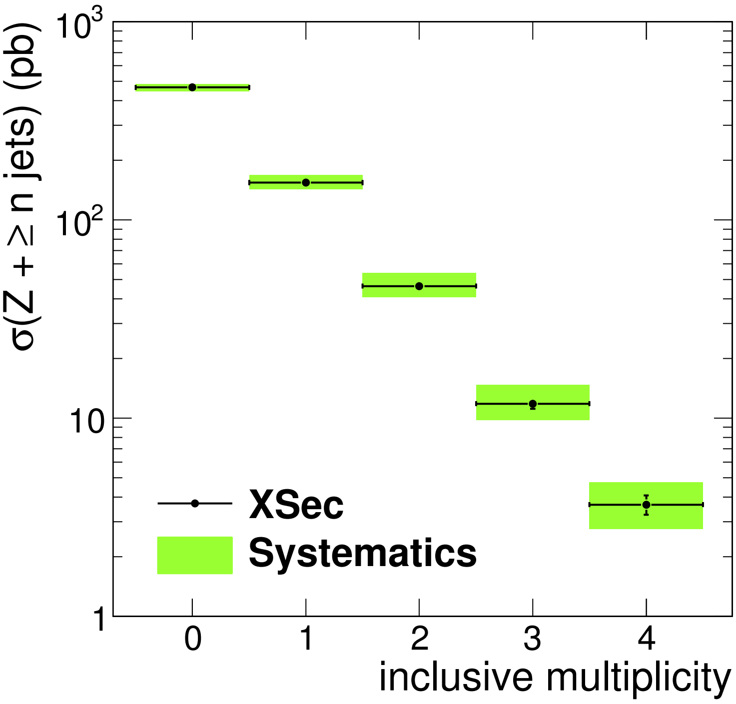


Figure 7.14: Measurement of $\sigma(\mathrm{Z}+\geq\mathrm{n}$ jets) for $p_T^{jet} > 15\,\mathrm{GeV}$. Statistical (black bars) and total systematic (green area) uncertainties are shown.

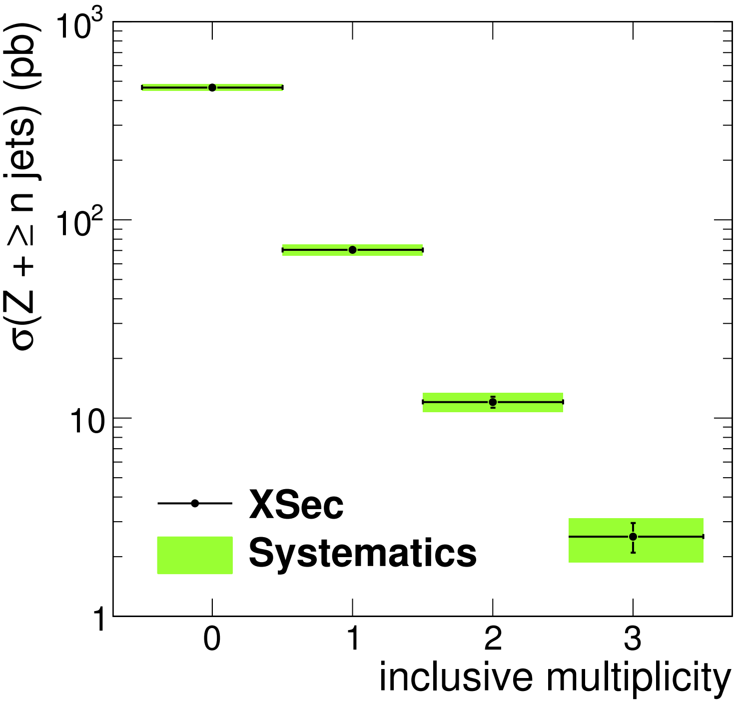


Figure 7.15: Measurement of $\sigma(\mathrm{Z}+\geq\mathrm{n}$ jets) for $p_T^{jet} > 30\,\mathrm{GeV}$. Statistical (black bars) and total systematic (green area) uncertainties are shown.

Cross Section Measurement - $\mathbf{p_T^{jet}} > \mathbf{15\,GeV}$			
	σ(**Z +** $\geq$ **n jets) (pb)**	**Statistical Unc.**	**Systematic Unc.**
≥ 0	466	± 5	± 19
≥ 1	154	± 3	$^{+14}_{-12}$
≥ 2	46.4	± 1.5	$^{+7.5}_{-5.5}$
≥ 3	11.8	± 0.7	$^{+2.8}_{-2.0}$
≥ 4	3.7	± 0.4	$^{+1.1}_{-0.9}$

Table 7.14: Measurement of σ(Z+$\geq$n jets) for $p_T^{jet} > 15\,\text{GeV}$. Statistical and systematic uncertainties are shown.

Cross Section Measurement - $\mathbf{p_T^{jet}} > \mathbf{30\,GeV}$			
	σ(**Z +** $\geq$ **n jets) (pb)**	**Statistical Unc.**	**Systematic Unc.**
≥ 1	70.6	± 1.9	± 4.7
≥ 2	12.1	± 0.8	± 1.4
≥ 3	2.5	± 0.4	± 0.6

Table 7.15: Measurement of σ(Z+$\geq$n jets) for $p_T^{jet} > 30\,\text{GeV}$. Statistical and systematic uncertainties are shown. The inclusive bin ≥ 0 jets is shown in Table 7.14.

Conclusions

In the previous Chapters the full $Z \to e^+e^- + jets$ analysis was described, presenting first the CMS detector, i.e. the experimental apparatus that made this analysis possible, and the reconstruction algorithms that give us access to higher level objects built from the raw energy deposits recorded by the various subdetectors. The analysis was performed by using the CMS official selection for the 2010 data taken, corresponding to a total integrated luminosity of $\mathscr{L} = (36.2 \pm 1.4)\,\mathrm{pb}^{-1}$.

In order to maximize the signal yields an asymmetric selection approach was adopted, requiring a well identified and isolated electron as the primary leg of the Z boson and keeping a looser second leg. One more reason for using this particular selection for the tight leg was that this analysis was also part of a W/Z ratio measurement, and therefore the tight leg had to be kept synchronized with the W electron selection. Despite the much looser requirements on the second leg, an almost background-free sample was obtained after applied the whole set of selections, thanks to the clear Z production signature.

The inclusive jet rates were measured for two different p_T^{jet} cuts: $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$. For $p_T^{jet} > 15\,\mathrm{GeV}$ 5 multiplicity bins have been studied, i.e. ≥ 0, ≥ 1, ≥ 2, ≥ 3 and ≥ 4 jets. For $p_T^{jet} > 30\,\mathrm{GeV}$ only 4 multiplicity bins have been taken into account, i.e. ≥ 0, ≥ 1, ≥ 2, and ≥ 3 jets, since for the higher p_T^{jet}multiplicity distribution the bin with ≥ 4 jets was affected by lack of statistics. The measured yields have been corrected for the detector efficiency, measured using data driven techniques (Tag & Probe), and for the bin migrations due to the finite detector resolution in p_T (unfolding).

The final measurements obtained for the multiplicity distributions were compared with predictions of different MC generators, i.e. `MADGRAPH` with Z2 and D6T tunes and `PYTHIA` with Z2 tune. A good agreement is found between data and the `MADGRAPH` simulation with Z2 tune, both for $p_T^{jet} > 15\,\mathrm{GeV}$ and $p_T^{jet} > 30\,\mathrm{GeV}$ distributions. The D6T tune `MADGRAPH` predictions are almost equal to the Z2 predictions for the high p_T^{jet} distribution, whereas they do not reproduce well the low p_T^{jet} distribution. The `PYTHIA` simulation does not model well the multiplicity distributions in both cases, with increasing disagreement as the number of jets increases. The fact that the `PYTHIA` sample does not model the data well is expected as it does not handle multiple jets in the matrix element.

The most important contributions to the total systematic uncertainty are the ones related to the jet energy calibration and to the unfolding for the $p_T^{jet} > 15\,\mathrm{GeV}$

multiplicity distribution, while in the case of $p_T^{jet} > 30\,\text{GeV}$ distribution the dominant source of systematic uncertainty is due to the efficiency measurement.

The increment of the analyzed data sample size will strongly reduce the contributions to the systematic uncertainty due to the luminosity and to the efficiency. The efficiency systematic uncertainty indeed is strictly related to the available statistics, since it is due to the data driven procedure adopted for determining the efficiency. As more statistics also mean a more accurate p_T^{jet} measurement, also the systematic uncertainty related to the jet energy calibration should decrease. Finally, the increase of the analyzed statistics will allow therefore to extend the measurement to higher jet multiplicity bins, both for $p_T^{jet} > 15\,\text{GeV}$ and $p_T^{jet} > 30\,\text{GeV}$.

Acknowledgements

I would like to thank my supervisor Prof. Raffaello D'Alessandro for his support, encouragement and precious help during all the PhD.

I would also like to thank Dr. Vitaliano Ciulli and Dr. Elisabetta Gallo for their guidance and for the constant dedication that they showed in following my work.

Finally I would like to thank Dr. Piergiulio Lenzi, Dr. Sandro Gonzi, Dr. Antonio Tropiano and all the other members of the CMS group of Firenze for their useful help and their suggestions.

Bibliography

[1] F. Halzen and A. Martin, *Quarks and leptons: an introductory course in modern Particle Physics*. Wiley, New York USA, 1984.

[2] S. L. Glashow, "Partial symmetries of weak interactions," *Phys. Rev.* **22** (1961) 579–588.

[3] A. Salam, *Elementary Particle Theory*. Almquist and Wiksells, N. Svartholm, Stockholm, 1968.

[4] S. Weinberg, "A model of leptons," *Phys. Rev. Lett.* **19** (1967) 1264–1266.

[5] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, "Experimental test of Parity Conservation in Beta Decay," *Phys. Rev.* **105** (1957) 1413–1414.

[6] R. L. Garwin, L. M. Lederman, and M. Weinrich, "Observations of the Failure of Conservation of Parity and Charge Conjugation in Meson Decays: The Magnetic Moment of the Free Muon," *Phys. Rev.* **105** (1957) 1415–1417.

[7] M. Goldhaber, L. Grodzins, and A. W. Sunyar, "Helicity of neutrinos," *Phys. Rev.* **109** (1958) 1015–1017.

[8] E. Fermi, "*Trends to a theory of Beta radiation*," *Nuovo Cimento* **11** (1952) 1–19.

[9] Lep Injector Study Group, "Lep design report vol. 1: The lep injector chain." Cern-lep-th-83-29, cern-ps-dl-83-81, cern-sps-83-26, lal-rt-83-09, 1983.

[10] Lep Injector Study Group, "Lep design report vol. 2: The lep main ring." Cern-lep/8-01, 1984.

[11] Lep Injector Study Group, "Lep design report vol. 3: Lep2." Cern-ac-96-01-lep2, 1996.

[12] T. C. Collaboration, "The CDF detector: an overview," *Nucl. Inst.* **271** (1988) 387–403.

[13] T. D. Collaboration., "The D0 detector," *Nucl. Inst.* **338** (1994) 185.

[14] K. Nakamura *et al.*, "Particle Data Group," *J. Phys.* **G 37** (2010) .

[15] Alpeh, Delphi, L3, Opal and SLD Collaborations, "Precision Electroweak Measurements on the Z Resonance," *CERN-PH-EP/2005-041* (2005) .

[16] B. Kilminster, "Higgs search at Tevatron," *ICHEP2010 Proceedings* (2010) .

[17] M. Gell-Mann, "A Schematic Model of Baryons and Mesons," *Phys. Lett.* **8** (1964) 214–215.

[18] M. Y. Han and Y. Nambu, "Three-triplet model with double SU(3) symmetry," *Phys. Rev.* **139** (1965) B1006–B1010.

[19] H. Fritzsch, M. Gell-Mann, and H. Leutwyler, "Advantages of the Color Octet Gluon Picture," *Phys. Lett.* **B47** (1973) 365–368.

[20] D. J. Gross and F. Wilczek, "Asymptotically Free Gauge Theories. 1," *Phys. Rev.* **D8** (1973) 3633–3652.

[21] S. Weinberg, "Nonabelian Gauge Theories of the Strong Interactions," *Phys. Rev. Lett.* **31** (1973) 494–497.

[22] Gavin Salam, "Elements of qcd for hadron colliders." Cern yellow report cern-2010-002, 2010.

[23] **CTEQ** Collaboration, R. Brock *et al.*, "Handbook of perturbative QCD: Version 1.0," *Rev. Mod. Phys.* **67** (1995) 157–248.

[24] F. Ambroglini *et al.*, *Proceedings of the Workshop on Monte Carlos, Physics and Simulations at the LHC - Part I.* arXiv:0902.0293v1 [hep-ph], 2009.

[25] G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language," *Nucl. Phys.* **B126** (1977) 298.

[26] H. L. Lai *et al.*, "Global QCD analysis and the CTEQ parton distributions," *Phys. Rev.* **D51** (1995) 4763–4782, `arXiv:hep-ph/9410404`.

[27] A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne, "MRST2001: Partons and alpha(s) from precise deep inelastic scattering and Tevatron jet data," *Eur. Phys. J.* **C23** (2002) 73–87, `arXiv:hep-ph/0110215`.

[28] R. K. Ellis, W. J. Stirling, and B. R. Webber, *QCD and Collider Physics.* Cambridge University Press, N. Svartholm, Stockholm, 1996.

[29] S.Bolognesi *et al.*, "Measurement of the $w \to e\nu$ + jets to $z \to ee$ + jets ratio in proton-proton collisions at $\sqrt{s} = 7$ tev,".

[30] S.D.Drell and T.M.Yan, "Massive Lepton-Pair Production in Hadron-Hadron Collisions at High Energies," *Phys. Rev. Lett.* **25** (1970) 316.

[31] A. M. Cooper-Sarkar, "Low-x physics and W and Z production at the LHC," `arXiv:hep-ph/0512228`.

[32] R. Gavin, Y. Li, F. Petriello, and S. Quackenbush, "FEWZ 2.0: A code for hadronic Z production at next-to- next-to-leading order," *Comput. Phys. Commun.* **182** (2011) 2388–2403, `arXiv:1011.3540 [hep-ph]`.

[33] A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne, "Parton distributions and the LHC: W and Z production," *Eur. Phys. J.* **C14** (2000) 133–145, `arXiv:hep-ph/9907231`.

[34] J. Alwall, P. Demin, S. de Visscher, R. Frederix, M. Herquet, F. Maltoni, T. Plehn, D. L. Rainwater, and T. Stelzer, "Madgraph/madevent v4: the new web generation," *Journal of High Energy Physics* **2007** (2007) no. 09, 028. `http://stacks.iop.org/1126-6708/2007/i=09/a=028`.

[35] C. F. Berger *et al.*, "Vector Boson + Jets with BlackHat and Sherpa," *Nucl. Phys. Proc. Suppl.* **205-206** (2010) 92–97, `arXiv:1005.3728 [hep-ph]`.

[36] The LHC Study Group, "The large hadron collider conceptual design report." Cern/ac 95-05, Geneva, CERN, 1995.

[37] O. S. Bruning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. B. Proudlock, "LHC Design Report Vol.1: The LHC Main Ring," *CERN-2004-003-V1* (2004) .

[38] O. S. Bruning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. B. Proudlock, "LHC Design Report Vol.1: The LHC Infrastructure and General Services," *CERN-2004-003-V2* (2004) .

[39] M. Benedikt, P. Coolier, V. Mertens, J. Poole, and K. Schindl, "LHC Design Report Vol.1: The LHC Injector Chain," *CERN-2004-003-V3* (2004) .

[40] The CMS Collaboration, "The CMS experiment at LHC." Jinst 08, 2008.

[41] "Design report Tevatron 1 project," *FERMILAB-DESIGN-1983-01* (1983) .

[42] The ATLAS Collaboration, "Atlas: A toroidal lhc apparatus technical proposal.." CERN/LHCC 94-18, LHCC/P1, 1994.

[43] The LHCb Collaboration, "A large hadron collider beauty experiment for precision measurements of cp-violation and rare decays lhcb technical proposal.." CERN/LHCC 98-04, LHCC/P4, 1998.

[44] The ALICE Collaboration, "A large ion collider experiment technical proposal design.." CERN/LHCC 95-71, LHCC/P3, 1995.

[45] The CMS Collaboration, "The magnet project: Technical design report." CERN/LHCC 97-10, CMS TDR Vol.1, 1997.

[46] V. Karimki, *The CMS tracker system project: Technical Design Report.* Technical Design Report CMS. CERN, Geneva, 1997.

[47] *The CMS tracker: addendum to the Technical Design Report.* Technical Design Report CMS. CERN, Geneva, 2000.

[48] V. Karimki, "The CMS tracker system project: Technical Design Report." Technical Design Report CMS. CERN, Geneva, 1997.

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

[50] P. Bloch, R. Brown, P. Lecoq and H. Rykaczewski, *Changes to CMS ECAL electronics: addendum to the Technical Design Report.* Technical Design Report CMS. CERN, Geneva, 2002.

[51] K. Deiters *et al.*, "Investigation of the avalanche photodiodes for the cms electromagnetic calorimeter operated at high gain," *Nucl. Inst. and Meth. A* **461** (2001) no. 1-3, 574 – 576. `http://www.sciencedirect.com/science/article/pii/S0168900200013036`. 8th Pisa Meeting on Advanced Detectors.

[52] K. W. Bell *et al.*, "Vacuum phototriodes for the cms electromagnetic calorimeter endcap," *Nuclear Science IEEE Transaction* **51** (2004) 2284 – 2287.

[53] The CMS Collaboration, "Electromagnetic calorimeter calibration with 7 TeV data," *CMS Physics Analysis Summary, CMS PAS EGM-10-003* (2010) .

[54] *The CMS hadron calorimeter project: Technical Design Report.* Technical Design Report CMS, CERN-LHCC-97-031. CERN, Geneva, 1997.

[55] *The CMS muon project: Technical Design Report.* Technical Design Report CMS, CERN-LHCC-97-032. CERN, Geneva, 1997.

[56] *CMS TriDAS project: Technical Design Report; 1, the trigger systems.* Technical Design Report CMS, CERN-LHCC-2000-038. CERN, Geneva, 2000.

[57] *CMS trigger and data-acquisition project: Technical Design Report.* Technical Design Report CMS, CERN-LHCC-2002-026. CERN, Geneva, 2002.

[58] A. Oh, "The CMS DAQ and Run Control System," *J.Phys.: Conf. Ser.* **110** (2008) 1099–1116.

[59] **CMS** Collaboration, "CMSSW, CMS reconstruction software," 2008. `http://cms.cern.ch/iCMS/jsp/page.jsp?mode=cms&action=url&urlkey=CMS_OFFLINE`.

[60] B. Stroustrup, *The C++ Programming Language (Third Edition).* Addison-Wesley, 1997.

[61] M. Lutz, *Programming Python.* Sebastopol, CA: O'Reilly, 1996.

[62] R. Brun and F. Rademakers, "ROOT: An object oriented data analysis framework," *Nucl. Instrum. Meth.* **A389** (1997) 81–86.

[63] M. Dobbs and J. B. Hansen, "The HepMC C++ Monte Carlo event record for High Energy Physics," *Comput. Phys. Commun.* **134** (2001) 41–46.

[64] **GEANT4** Collaboration, S. Agostinelli *et al.*, "GEANT4: A simulation toolkit," *Nucl. Instrum. Meth.* **A506** (2003) 250–303.

[65] A. Adam *et al.*, "PAT: the CMS Physics Analysis Toolkit." CMS-CR 2009/083, 2009.

[66] CMS Collaboration, "Absolute luminosity normalization," *CMS Detector Performance Summary DP-2011-002* (2011) .

[67] T. Sjostrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 physics and manual," *JHEP* **05** (2006) 026, `arXiv:hep-ph/0603175`.

[68] T. Stelzer and W. F. Long, "Automatic generation of tree level helicity amplitudes," *Comput. Phys. Commun.* **81** (1994) 357–371, `arXiv:hep-ph/9401258`.

[69] F. Maltoni and T. Stelzer, "MadEvent: Automatic event generation with MadGraph," *JHEP* **02** (2003) 027, `arXiv:hep-ph/0208156`.

[70] T. Gleisberg *et al.*, "SHERPA 1.alpha, a proof-of-concept version," *JHEP* **02** (2004) 056, `arXiv:hep-ph/0311263`.

[71] M. Bengtsson and T. Sjostrand, "Coherent Parton Showers Versus Matrix Elements: Implications of PETRA - PEP Data," *Phys. Lett.* **B185** (1987) 435.

[72] M. H. Seymour, "Matrix element corrections to parton shower algorithms," *Comp. Phys. Commun.* **90** (1995) 95–101, `arXiv:hep-ph/9410414`.

[73] G. Miu and T. Sjostrand, "W production in an improved parton shower approach," *Phys. Lett.* **B449** (1999) 313–320, `arXiv:hep-ph/9812455`.

[74] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, "Parton fragmentation and string dynamics," *Phys. Rep.* **97** (1983) 31–145.

[75] The CMS Collaboration, "Measurement of the Underlying Event Activity at the LHC with $\sqrt{s} = 7$ TeV and Comparison with $\sqrt{s} = 0.9$ TeV." CMS PAS, QCD-2010-010, 2010.

[76] R. Field, "Early LHC Underlying Event Data - Findings and Surprises," *High Energy Physics - Phenomenology (hep-ph)* (2010) .

[77] S. Baffioni *et al.*, "Electron reconstruction in CMS," *Eur. Phys. J. C* **49** (2007) 1099–1116.

[78] M. Anderson *et al.*, "Review of clustering algorithms and energy corrections in ECAL." CMS-IN 2010/008, 2010.

[79] P. Vanlaer, "Electron and photon reconstruction in CMS." CMS-CR 2009/201, 2009.

[80] E. Meschi *et al.*, "Electron reconstruction in the cms electromagnetic calorimeter." CMS Note 2001/034, 2001.

[81] W. Adam *et al.*, "Electron reconstruction in CMS." CMS-AN 2009/164, 2009.

[82] R. Fruhwirth *et al.*, "Application of Kalman filtering to track and vertex fitting," *Nucl.Instrum.Meth. A* **262** (1987) 444–450.

[83] C. Rovelli, "The CMS electromagnetic calorimeter and the search for the Higgs boson in the decay channel $H \to WW^{(*)} \to 2e2\nu$." PhD Thesis, Ecole Polytechnique De Palaiseau, 2006.

[84] H. Bethe and W. Heitler, "On the Stopping of Fast Particles and on the Creation of Positive Electrons," *Proc. R. Soc. Lond. A* **146** (1934) 83–112.

[85] W. Adam *et al.*, "Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC." CMS Note 2005/001, 2005.

[86] F. Beaudette *et al.*, "Electron Reconstruction within the Particle Flow Algorithm." CMS-AN 2010/034, 2010.

[87] The CMS Collaboration, “Particle-Flow Event reconstruction in CMS and Performance for Jets, Taus, and E_T^{miss}.” CMS PAS PFT 09/001, 2009.

[88] D. Benedetti, “Performance of the particle flow algorithm in CMS.” CMS-CR 2010/127, 2010.

[89] The CMS Collaboration, “Commissioning of the Particle-Flow Reconstruction in Minimum-Bias and Jet Events from pp Collisions at $\sqrt{s} = 7$ TeV.” CMS PAS PFT 10/002, 2010.

[90] A. Vartak *et al.*, “Lepton isolation in the CMS Tracker, ECAL and HCAL.” CMS-AN 2010/106, 2010.

[91] S. Baffioni *et al.*, “Electron Identification in CMS.” CMS-CR 2009/178, 2009.

[92] N. Ackhurin *et al.*, “Study of Production of Vector Bosons and Jets at 7 TeV for ICHEP2010 Conference,” *CMS Analysis Note, CMS AN-10-136* (2010) .

[93] D. Barge *et al.*, “Study of photon conversion rejection at CMS,” *CMS Analysis Note, CMS AN-2009/159* (2009) .

[94] S. Baffioni *et al.*, “Electron Identification in CMS,” *CMS Analysis Note, CMS AN-2009/178* (2009) .

[95] G. Soyez, “The SISCone and anti-k_t jet algorithm.” DIS2008, 2008.

[96] G. P. Salam and G. Soyez, “SISCone.” http://projects.hepforge.org/siscone.

[97] M. Cacciari, G. Salam, and G. Soyez, “The anti-k_t jet clustering algorithm,” *JHEP 0804* **063** (2008) .

[98] Y. L. Dokshitzer *et al.*, “Better Jet Clustering Algorithms,” *JHEP 9708* **001** (1997) .

[99] The CMS Collaboration, “Jet Performance in pp Collisions at $\sqrt{s} = 7$ TeV.” CMS PAS JME-10-003, 2010.

[100] D. Green, “Dijet mass resolution at the LHC,” `arXiv:hep-ex/0401036`.

[101] W. Busza, “Multiparticle Production in pp, pA, dA, and AA Collisions,” *Acta Physica Hungarica A) Heavy Ion Physics* (2005) .

[102] The CMS Collaboration, “Plans for Jet Energy Corrections at CMS.” CMS PAS JME-07-002, 2008.

[103] M. Cacciari and G. P. Salam, “Pileup subtraction using jet areas,” *Phys. Lett.,* **B659** (2008) 119–126.

[104] The CMS Collaboration, "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS." CMS JME-10-011, 2011.

[105] N. Saoulidou, "Particle Flow Jet Identification Criteria," *CMS Analysis Note, CMS AN 2010-003* (2010) .

[106] W. Verkerke and D. Kirkby, "The RooFit toolkit for data modeling," `arXiv:physics/0306116`. `http://arxiv.org/abs/physics/0306116`.