UNIVERSITÀ DEGLI STUDI DI ROMA “LA SAPIENZA”
Facoltà di Scienze Matematiche, Fisiche e Naturali
Dottorato di Ricerca in Fisica - XVII Ciclo

Optimization of the discovery potential of the Higgs Boson in the decay channel $H \to ZZ^{(*)} \to 2e^{+}2e^{-}$ with the CMS detector

Tesi presentata da **Paolo Meridiani**
per il conseguimento del titolo di Dottore di Ricerca

Supervisore:

Chiar.mo Prof. Egidio Longo

Anno Accademico 2003-2004

Contents

Introduction

The Standard Model of the electroweak interactions [1] is one of the most successful theories of modern particle physics. It was extensively tested by recent experiments, and no discrepancy from the theoretical predictions has been found yet. The theory, however, still lacks a final test. Particles masses are introduced in the Standard Model using the so called electroweak spontaneous simmetry breaking mechanism [2], which predicts the existence of a scalar particle, the Higgs boson, which, up to now, has not been experimentally observed. There are also other aspects of the theory which induce to think the Standard Model more like a low energy effective theory, rather than a fundamental one. Several possible extensions of the Standard Model have been proposed.

The ultimate tests of the Standard Model and the possibility to search for new physics have led to the design of the Large Hadron Collider (LHC), a high-energy, high-luminosity proton-proton collider, that will be installed at the European Laboratory for Nuclear Research (CERN) in Geneva (Switzerland); the first collisions are planned for the summer of 2007.

By providing proton collisions at a center of mass energy at 14 TeV, at an expected luminosity as high as $10^{34}\mathrm{cm}^{-2}\mathrm{s}^{-1}$, LHC will allow the search of the Higgs boson in the whole expected mass range, from the present LEP limit ~ 114 GeV up to ~ 1 TeV. Four experiments will collect data at the LHC: two general-purpose ones (ATLAS and CMS), one dedicated to b-physics (LHC-b), and one dedicated to heavy ion studies (Alice). The search strategy of the Higgs boson at LHC will exploit different decay channels, according to the varying decay properties of the Higgs boson with its mass. One of the most relevant channel is the Higgs decay into 4 leptons through a $ZZ^{(*)}$ intermediate state, which can be used as a discovery channel in a wide Higgs mass range.

In this thesis I will present the work which I perfomed on the electron reconstruction and identification in the CMS detector, with the goal to optimize the discovery potential of this experiment in the $H \rightarrow ZZ^{(*)} \rightarrow 2e^{+}2e^{-}$ channel. A fundamental role in the electron reconstruction is played by the electromag-

netic calorimeter. The CMS experiment has designed the largest ever built crystal calorimeter, with the aim to reach an excellent energy and direction resolution on electrons and photons. Part of this thesis is dedicated to the evaluation of the calorimeter perfomances in the test beam conditions, and to the comparison of the test beam data with the output of detailed simulation.

The presentation of the work is organized as follows.

A general introduction to the Higgs boson properties and to the limits on its mass, coming from theoretical arguments and from experimental searches, is given in chapter 1; the LHC project is later reviewed, concluding the chapter with the Higgs boson searches at LHC.

In chapter 2 the overall design of the CMS detector is described, together with the CMS trigger system, which will play a crucial role in the event reduction and selection.

Chapter 3 and 4 are dedicated respectively to the description of the CMS electromagnetic calorimeter (ECAL) and to the analyis of the 2003 electromagnetic calorimeter test beam data. An aspect of major concern is represented by the calorimeter channel to channel intercalibration; the precision at which the intercalibration factors will be known is directly entering in the constant contribution to the energy resolution. The goal, reported in the CMS ECAL TDR [3], is to reach a precision of better than 0.5%. To this extent, it is important to understand the value at which the intercalibration factors will be known at start-up. The first part of chapter 4 is dedicated to the evaluation of the intercalibration precision which can be reached using the measurements performed during the calorimeter assembly phase, comparing these factors with those obtained by test beam intercalibration. Test beam is also a unique occasion to validate and compare with detailed simulation. I contributed to the development of a GEANT 4 [4] simulation of the electromagnetic calorimeter in the test beam configuration. In the second part of chapter 4, after introducing the main features of this simulation, a detailed comparison between test beam data and the output of the simulation is illustrated.

The electron reconstruction algorithms and identification criteria in the CMS detector are studied in detail in chapter 5. Key aspects in the $H \rightarrow ZZ^{(*)} \rightarrow 2e^{+}2e^{-}$ analysis are represented by reconstruction and identification efficiency of isolated electrons and by energy and direction resolution. All these aspects are analyzed and discussed. Particular emphasis is given to the bremsstrahlung in the tracker material, which represents the main problem in the electron reconstruc-

tion. Some variables able to select electrons with a reduced fraction of radiated energy are taken into consideration. An estimator of the electron energy and direction based on the combination between calorimeter and tracker measurements is then introduced. The last part of the chapter is dedicated to the construction of an electron identification variable based on the likelihood fraction, capable to reject fake electrons coming from jets.

Finally, in chapter 6 the discovery potential of the Higgs boson in the channel $H \to ZZ^{(*)} \to 2e^+2e^-$ is evaluated in the mass range from 115 to 550 GeV, assuming an integrated luminosity of 20 fb^{-1}, which corresponds to one year of data taking at the LHC exptected low luminosity conditions ($2 \cdot 10^{33}$ cm^{-2}s^{-1}). The main background sources are evaluated: ZZ^*, $t\bar{t}$ and $Zb\bar{b}$. Signal and background samples are all obtained from a detailed simulation of the CMS detector response. A discovery probability estimator is introduced, which is based on the future hypotheses testing between the background only hypotheses and the signal plus background one. An optimization procedure of the signal selection is proposed, in order to maximize the CMS discovery probability in the $H \to ZZ^{(*)} \to 2e^+2e^-$ channel.

The results contained in chapters 4, 5 and 6 represent my personal and original contributions to the CMS collaboration activities.

Chapter 1

The Large Hadron Collider and Higgs Boson Searches

Our current understanding of the subatomic world is summarized in the so called Standard Model of Particle Physics (SM) [1], a quantum field theory based on the $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$ simmetry group.

Extensive consistency and precision tests were performed yielding stringent constraints on the Standard Model [5] over a wide range of energies.

However, the Standard Model has not been completely tested: in particular, it requires an additional scalar field, the Higgs field, whose corresponding particle has not yet been observed. The Higgs mechanism has been introduced [2] to break the electroweak simmetry, therefore to give masses to the particles described by the Standard Model.

Direct searches for the Higgs bosons through its decays, performed at LEP and Tevatron accelerators, have not yet given positive results, setting a lower limit on its mass to about 114 GeV.

Apart from electroweak simmetry breaking mechanism, still to be confirmed, there are several reasons to think the Standard Model only as an effective description, and to foresee a more fundamental theory. Several models proposing a wider symmetry than what is in the Standard Model have been proposed, in order to solve the theoretical drawbacks affecting it.

The ultimate tests of the Standard Model and the possibility to search for new physics beyond the Standard Model has led to the design of the Large Hadron Collider (LHC), a high-energy, high-luminosity proton-proton collider, that will be installed at the European Laboratory for Nuclear Research (CERN) in Geneva (Switzerland). The first collisions are planned for the summer of 2007; at that time it will be the most powerful particle accelerator ever built.

In this chapter the basic concepts of the Standard Model are overviewed, fo-

cussing in particular on the electroweak simmetry breaking mechanism, with the aim to introduce the Higgs boson properties. The different searches conducted for the Higgs boson are then reviewed, together with its theoretical and experimental limits. The design of the LHC, the experimental environment and the requirements for LHC experiments are described in Section 1.4; an overview of the Higgs searches at the LHC are presented in Section 1.6.

1.1 The Standard Model

The Standard Model is a quantum field theory describing three of the four fundamental interactions: electromagnetic interaction, weak interaction and strong interaction.

The Standard Model is a gauge theory based on the $SU(3)_C \otimes SU(2)_T \otimes U(1)_Y$ simmetry group, obtained as direct product of the color simmtery group (C), weak isospin (T) and hypercharge (Y). Associated to the gauge simmetry groups there are 12 spin-1 gauge bosons: 8 bosons (g) for $SU(3)_C$, 3 (W_i) for $SU(2)_T$ and 1 (B) for $U(1)_Y$. The Z boson and the photon γ will be seen as linear combination of W_3 and B.

Matter fields are represented by spin-$\frac{1}{2}$ particles, fermions, divided into two categories, leptons and quarks, having the following tansformation properties with respect to the $SU(3)_C \otimes SU(2)_T \otimes U(1)_Y$ gauge group:

$$
\begin{aligned}
Q_L^i &= \begin{pmatrix} u_L^i \\ d_L^i \end{pmatrix} = \left(3, 2, \frac{1}{6}\right), \quad u_R^i = \left(3, 1, \frac{2}{3}\right), \quad d_R^i = \left(3, 1, -\frac{1}{3}\right) \\
L_L^i &= \begin{pmatrix} \nu_L^i \\ l_L^i \end{pmatrix} = \left(1, 2, -\frac{1}{2}\right), \quad l_R^i = (1, 1, -1), \quad i = 1, 2, 3
\end{aligned} \tag{1.1}
$$

where the indicated numbers are the dimension of the representation respectively for $SU(3)_C$, $SU(2)_T$ and the value of the hypercharge Y; the index i runs over the three quarks and leptons families, while the ν_R^i is not shown since it is a singlet with respect to gauge group transformation.

Fermions in the equation 1.1 are written as two components left- and right-handed Weyl spinors. Electric charge (in unit of elementary charge e=(1.602 176 462 $\times$ 10^{-19}) C can be written as $Q = T_3 + Y$, indicating with T_3 the diagonal generator of $SU(2)_T$.

Direct mass terms for fermions cannot be written since they would violate the gauge simmetry. Mass terms, instead, can be introduced after the spontaneous simmetry breaking mechanism, first introduced by Higgs [2], is exploited; the introduction of a scalar field with the proper quantum numbers allows to break the $SU(3)_C \otimes SU(2)_T \otimes U(1)_Y$ into a residual $SU(3)_C \otimes U(1)_{em}$, with the purpose to

dress weak interaction bosons with mass and at the same time keep the carrier of the electromagnetic interaction, the photon, massless. In order to accomplish this, the simplest choiche is to use a doublet of $SU(2)_T$ having hypercharge $Y = -1/2$:

$$\Phi = \begin{pmatrix} \Phi^0 \\ \Phi^- \end{pmatrix}. \tag{1.2}$$

The potential for the scalar field is quartic

$$V = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2, \tag{1.3}$$

with $\mu^2 < 0$, and a positive quartic coupling $\lambda > 0$. No higher order auto-interaction terms than the quartic $|\Phi^4|$ appear in the potential expression, in order to guarantee the renormalizability of the theory [6].

If one of the component of Φ takes a non void vacuum expectation value:

$$\langle \Phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}, \quad v = \sqrt{\frac{-\mu^2}{\lambda}}, \tag{1.4}$$

a scalar massive particle, the Higgs boson H, is predicted to have a mass $M_H^2 = -2\mu^2$, while the charged $W^\pm$ and the neutral Z^0 bosons takes respectively a mass $M_{W^\pm} = \frac{g_2 v}{2}$ and $M_{Z^0} = \frac{v\sqrt{g_1^2+g_2^2}}{2}$, where in the previous equations g_1 and g_2 are respectively the couplings associated to $U(1)_Y$ and to $SU(2)_T$. The photon remains massless.

Higgs field is also suitable to generate the masses of all the fermion fields without symmetry violating terms. This can be done introducing Yukawa couplings between the Higgs doublet and the fermion fields.

The Higgs coupling to the fermion is diagonal in the flavour and is equal to $g_{Hff} = \frac{m_f}{v}$, proportional to the fermion mass m_f.

The Higgs is coupled to W^+W^- or to ZZ with an intensity equal to:

$$g_{HWW} = \frac{g^2 v}{2} = g m_W = \frac{2m_W^2}{v} \tag{1.5}$$

$$g_{HZZ} = \frac{g^2 v}{4\cos^2\theta_W} = \frac{g m_Z}{2\cos\theta_W} = \frac{m_Z^2}{2}. \tag{1.6}$$

Even in this case, the coupling is proportional to the gauge boson mass; proportionality of the Higgs coupling to the mass is a prediction of the Standard Model.

1.2 Higgs Properties and Searches

1.2.1 Higgs Decays

The Standard Model is extremely predictive in the Higgs sector, giving all couplings, decay widths and production cross sections at given Higgs boson mass, a parameter which instead should be experimentally measured.

In figure 1.1 and 1.2 the Standard Model Higgs boson total decay width and decay branching ratios are reported as a function of the mass for a large range of values, from 50 GeV/c^2 to 1 TeV/c^2 . The curves shown are obtained using the program HDECAY [7], including also higher order radiative corrections.

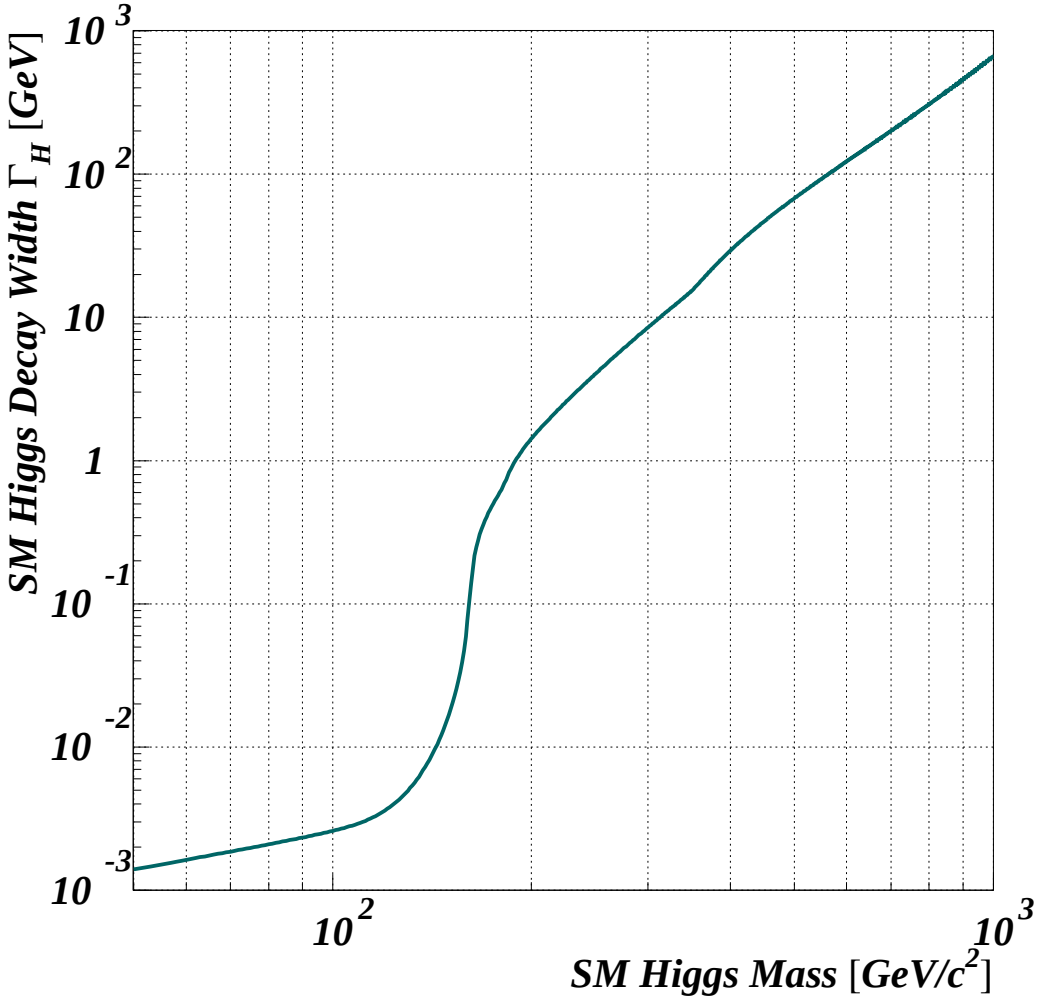


Figure 1.1: *Total decay width Γ_H of Standard Model Higgs boson as a function of the mass.*

1.2.2 Higgs mass: theoretical limits

The Higgs mass depends on the unknown coupling λ in the Higgs potential, and therefore cannot be predicted. However some constraints can be fixed on a theoretical basis [8], [9]. A lower bound can be put requiring that the theory is stable up to a certain scale: the Higgs potential is affected by radiative corrections, which depends on the renormalization scale, which can change its shape inducing a potential without an absolute minimum. At the same time, the coupling λ increases with the energy scale; the request that it remains finite up to a scale (triviality) corresponds to putting an upper bound on m_H. The theoretical bounds on m_H as a function of scale up to which the Standard Model is taken to be valid are shown in 1.3.

For the SM to remain valid up to the Planck scale (= 10^{19} GeV), the Higgs mass must be in the range 130-200 GeV/c^2. Assuming a validity of the SM only up to 1 TeV, the Higgs mass can be up to 700 GeV/c^2. In any case, the main indication is that the Higgs boson should be searched in a range of masses below 1 TeV. More stringent limits are infact coming from the direct experimental Higgs searches which will be reviewd in the following.

1.2.3 Higgs searches

The search for the Higgs boson is divided into two parallel lines: the direct searches at leptonic or hadronic colliders and the indirect searches, relying on the internal consistency of the SM and of the proposed electroweak simmetry breaking mechanism.

Indirect searches

The precision of electroweak measurements has reached such a sensitivity to be able to probe Standard Model corrections due to higher order loops to tree-level expectation values. At one loop, all electroweak parameters have at most logarithmic dependence on m_H [10, 11]; other corrections are due to top quark loops, which depend at leading order on $(m_t/m_W)^2$. Measurements of electroweak parameters coming from different experiments [5], mainly from LEP, are used in a global χ^2 fit with the aim to find the best constraints on m_H. The results of the global fit are shown in figure 1.4, where the $\chi^2 - \chi^2_{min}$ value is shown as a function of m_H.

The 95% confidence level upper limit on m_H is

$$m_H < 219 \quad \text{GeV}/c^2. \tag{1.7}$$

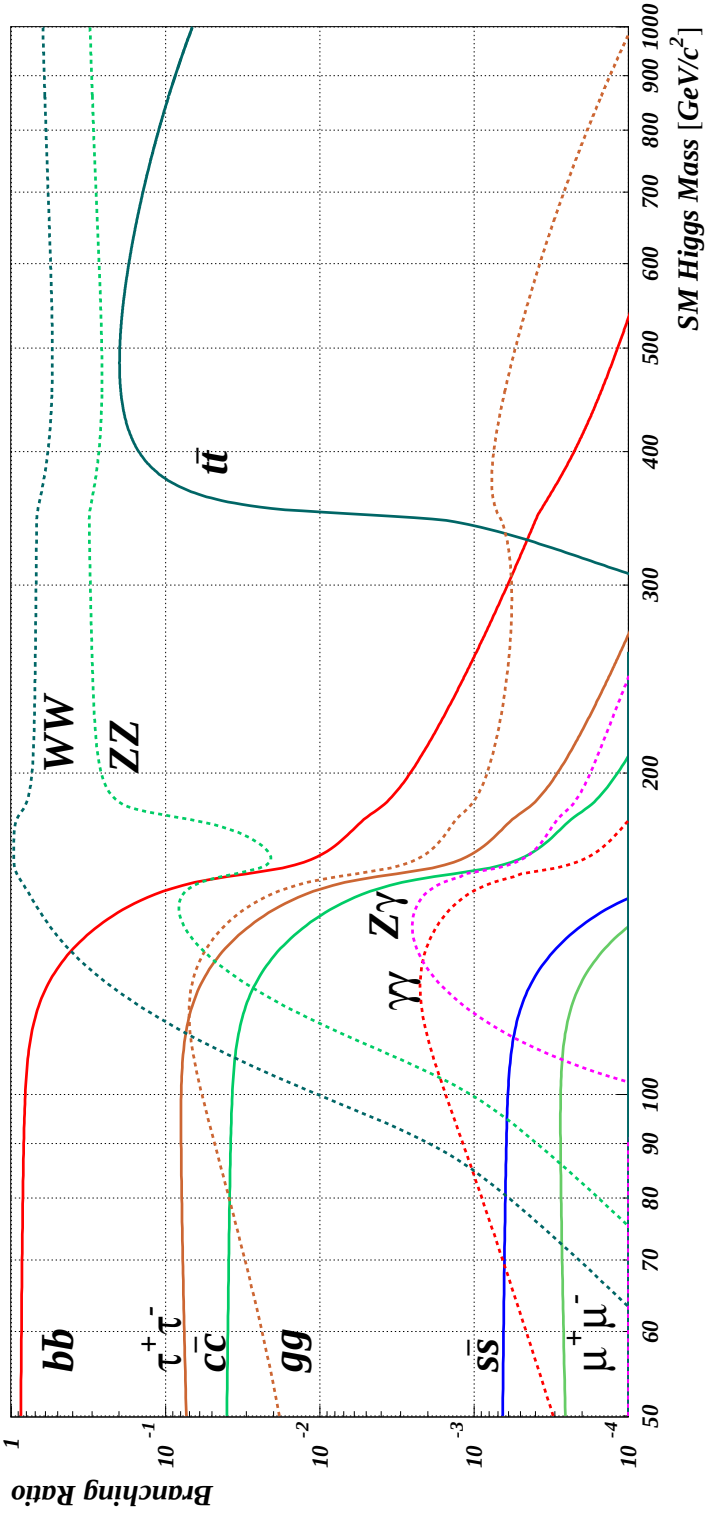


Figure 1.2: *Decay branching ratios of the Standard Model Higgs boson as a function of the mass. Decays into fermion anti-fermion pairs are represented by solid lines, decays into gauge boson pairs by dashed lines.*

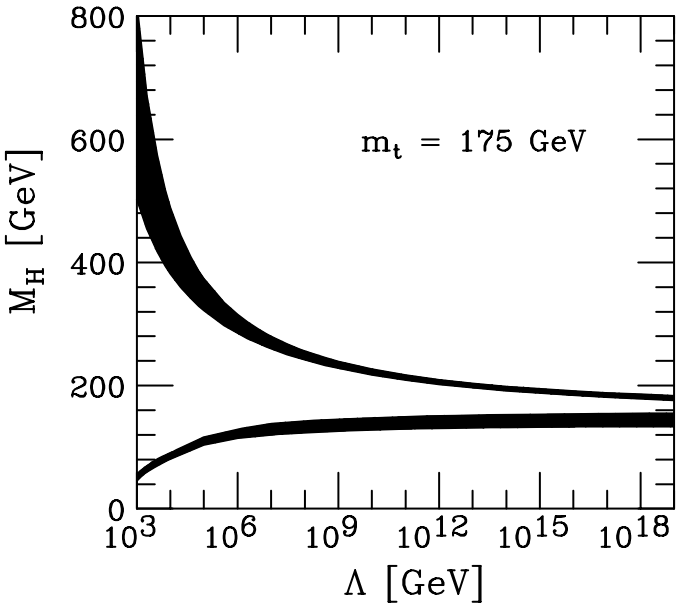


Figure 1.3: *Theoretical limits on Standard Model Higgs boson mass. The allowed region, as a function of the energy scale* Λ *at which the Standard Model breaks down, is between the two curves, obtained assuming* m_t*=175 GeV/*c^2 *and* $\alpha_s(m_Z)$*=0.118 [9].*

Direct searches at LEP

The tightest constraint on the Higgs boson mass comes from the combined results of the four LEP experiments [12]. The four collaborations ALEPH [13], DELPHI [14], L3 [15] and OPAL [16] have collected 2461 pb^{-1} of e^+e^- collision data at centre-of-mass energies $\sqrt{s}$ between 189 and 209 GeV. At LEP the Higgs boson is mainly produced in association with the Z^0 boson, through the so-called *Higgstrahlung process* ($e^+e^- \rightarrow HZ$). Inputs from the four experiments are provided for all the channels and are combined together to define a variable sensitive to the signal-to-background ratio Q: the log-likelihood test statistics $2\ln(Q)$ [12] is used. Its value is shown in figure 1.5 as a function of m_H (Higgs-like events have large Q value).

The lower bound on m_H at 95% C.L. (intersection of the solid line with the horizontal line) is

$$m_H > 114.4 GeV, \tag{1.8}$$

while the preferred mass value is m_H = 115.6 GeV corresponding to the maximum of the Likelihood $2\ln(Q) = 2.88$ (minimum of the solid line in figure 1.5). The

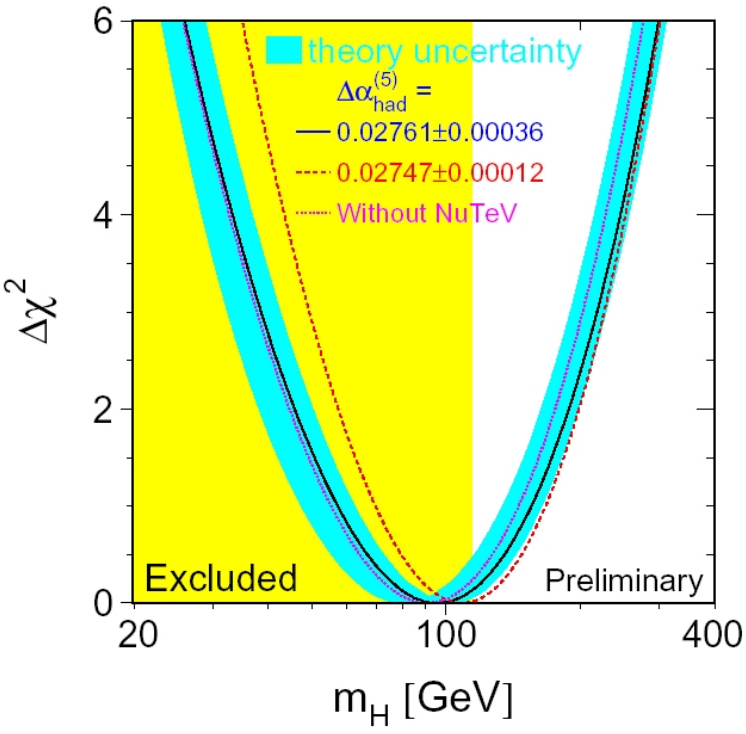


Figure 1.4: *Observed value of* $\Delta\chi^2 = \chi^2 - \chi^2_{min}$ *as a function of Higgs mass* m_H*. The line is the result of the electroweak fit using all data, the band represents the uncertainty due to neglecting higher order corrections. The vertical band is the region excluded by direct searches.*

minimum is 1.74 standard deviations away from the only background hypotheses, and it is consistent with the signal+background expectation for the same test mass. The favoured interpretation of this observation is the signal of a Standard Model Higgs boson within this mass range [12]. The signal-like behaviour mainly originates from the four-jet ALEPH data [17].

Direct searches at Tevatron

The Higgs boson can be produced via several mechanisms at the Tevatron at a centre of mass energy $\sqrt{s} = 1.96$ TeV.

The most promising discovery modes at the Tevatron are the production of the Higgs boson in association with either a W or Z boson: possible final states are $l\nu b\bar{b}$, $l^+l^-b\bar{b}$, $l^+l^-b\bar{b}$, $\nu\bar{\nu}b\bar{b}$, $q\bar{q}b\bar{b}$. In figure 1.6 [20] the luminosity thresholds as a function of m_H for the exclusion at 95% CL, and for observation of a 3σ excess, and 5σ excess are shown.

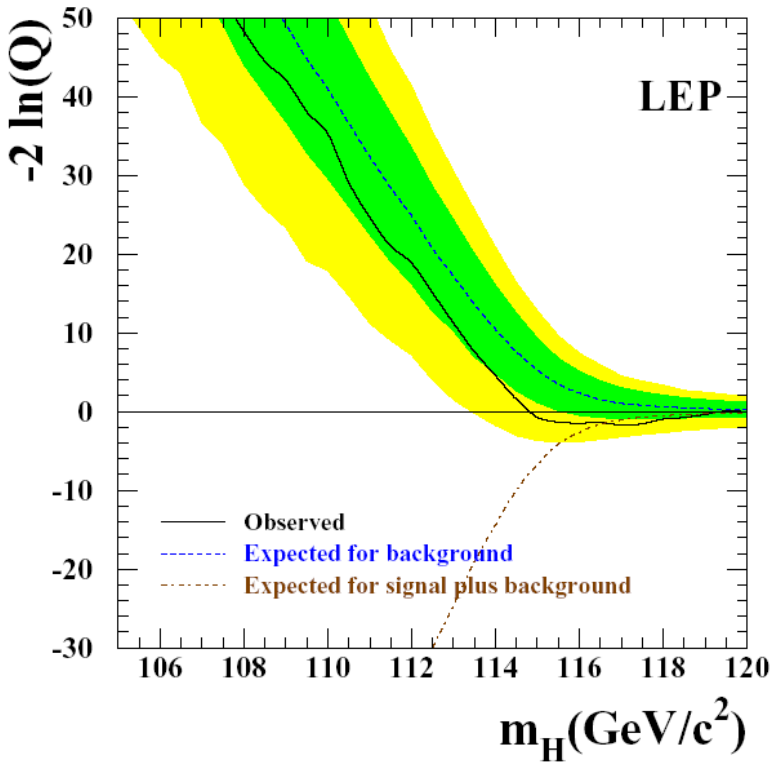


Figure 1.5: *Observed and expected behaviour of test-statistics* $2\ln(Q)$ *as a function of the test-mass* m_H*, obtained combining the data of the four LEP experiments. The solid line is the observed curve, the dashed (dot-dashed) is the median expectation in the hypotheses of background only (signal+background). The two shaded areas are the 68% and 95% probability bands around the median background expectation.*

The integrated luminosity required for each experiment (CDF[18] and D0 [19]) to exclude a 115 GeV SM Higgs boson at 95% C.L. is 1.5 fb^{-1}, while an observation at 3σ requires instead about 2 fb^{-1}. The effective luminosity delivered by the Tevatron Collider RUNII to each experiment is plotted in figure 1.7, showing an integrated luminosity of 650 pb^{-1} from 2001 up to now, a factor 4 lower than the planned 2 fb^{-1}. If the Tevatron Collider will not succeed in a significant ramping up of the luminosity, the direct Higgs search at the Tevatron Collider seems to be difficult.

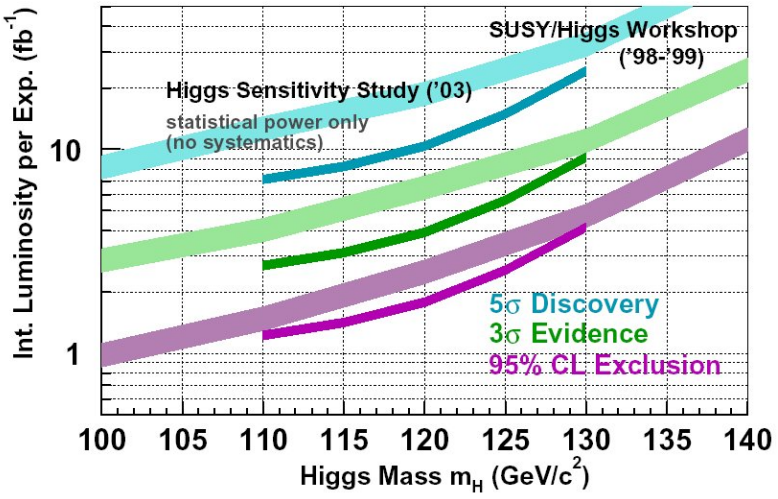


Figure 1.6: *Integrated luminosities per experiment corresponding to the median expectations for 95% CL exclusion, 3 σ evidence and 5 σ discovery for m_H = 110 - 130 GeV/c^2. The thicker curves correspond to a previous study (1999) [21].*

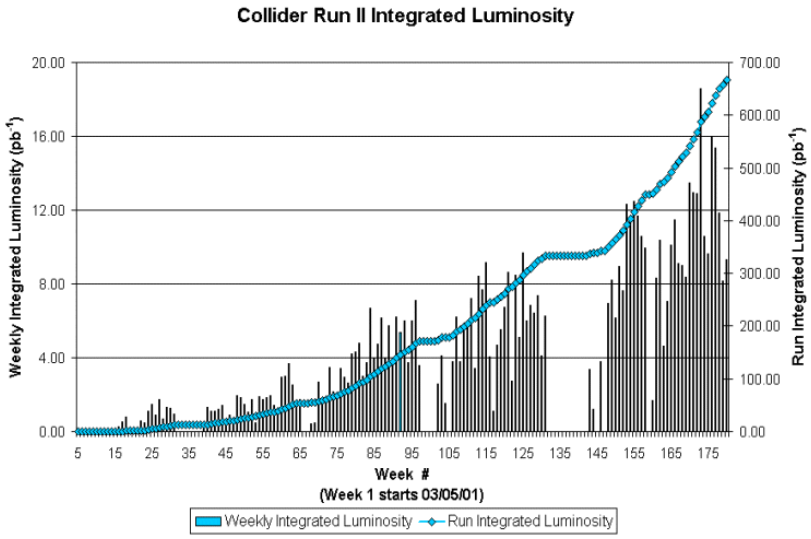


Figure 1.7: *Integrated Luminosity delivered by the Tevatron Collider during RUNII.*

1.3 Beyond the Standard Model

Despite the remarkable agreement between the precision measurements of electroweak observables and Standard Model predictions, there are strong theoretical

arguments that the Standard Model is not the ultimate theory of the fundamental particles and their interactions. It has 17 free arbitrary parameters, which may seem too many for a fundamental theory and leaves several unanswered questions. Some of them concern the problems of unification of interactions, number of fermions families, neutrino masses, naturalness/hierarchy problem [22]. Therefore, the Standard Model is generally considered as an effective field theory, valid up to some energy scale.

Among these problems, the naturalness-hierarchy problem is considered to be one of the most serious theoretical drawbacks of the Standard Model. There are two ways that propose to solve it: one is to avoidi the scalar field and construct a new strong force with new vector bosons, the other is to introduce a new symmetry and new particles that cancel exactly and naturally the divergences. Each of these proposals would have some experimental observables at the energy scale of 1 TeV.

In the case of a new strong force, the electroweak symmetry could be broken by a condensates of new fermions that are attracted by the new strong force like in the technicolour theories [23]. Such a mechanism for the symmetry breaking is also offered for example in the BESS (Breaking Electroweak Symmetry Strongly) model [24]. It would result in three new vector bosons and the Higgs would not be a physical particle. Composite models where the vector bosons and the Higgs are not elementary particles would result in a spectrum of new particles.

Models proposing a new symmetry are extensions of the Standard Model. In E_6 (from the symmetry group E_6) gauge models [25], there is an additional $U(1)$ symmetry arising from the superstring theories. This could result in new heavy gauge bosons $Z^{'}$ and $W^{'}$ at the TeV scale. The most popular theory extending the Standard Model is Supersymmetry (SUSY) [26] that introduces a symmetry between bosons and fermions. Each particle should have a SUSY partner, a sparticle, with a spin differing by $1/2$. At least two Higgs doublets are required resulting in five observable Higgs particles. The naturalness problem is solved by an exact cancellation between the particle and sparticle contributions.
There is no experimental evidence of any of these models and their validity can only be confirmed, or ruled out, with experiments.

1.4 The Large Hadron Collider

It has been shown how the main target of particle physics for the years to come will be the comprehension of the electroweak simmetry breaking mechanism and the search for possible new physics. These are the reasons that led the particle

physics community to design and build a new and more powerful accelator, the Large Hadron Collider (LHC); in this section the physics requirements and feasibility will be reviewed.

In a circular collider of radius R, the energy loss per turn due to synchrotron radiation is proportional to $(\frac{E}{M})^4 \cdot \frac{1}{R}$, where E and M are respectively the energy and mass of the particles accelerated; it then follows that a circular electron colliders would need enormous dimensions to reach energies of the order of 500 GeV per beam, therefore the natural choice for a collider with current technologies is to use beams of protons, which are almost 2000 times heavier than electrons. In a proton-proton collider the interactions involve the proton constituents (quarks, anti-quarks and gluons), which carry only a fraction of the proton momentum. A drawback of this is that the centre-of-mass energy and the rest frame for the hard scattering are unknown, but an advantage is that a wider range of energies can be explored with respect to with fixed-energy beams.

The event rate R_i of a physics channel i can be defined as the number of events per unit of time occurring with cross section σ_i

$$\frac{dN_i}{dt} = R_i = \sigma_i \cdot L\,. \tag{1.9}$$

It is proportional to cross section σ_i via the constant L, luminosity, which depends only on the machine parameters. Assuming a small crossing angle between the beams and gaussian-shaped beam bunches, the luminosity L can be expressed as [28]

$$L = f\frac{n_b N_1 N_2}{4\pi\sigma_x\sigma_y} \tag{1.10}$$

where f is the revolution frequency of the n_b bunches, N_1 and N_2 number of protons in the two colliding bunches, σ_x and σ_y the beam profiles in horizontal (bend) and vertical directions at the interaction point.

In the figure 1.8 cross-sections for different processes are given as a function of the centre-of-mass energy in $p-p$ collisions; in particular it can be noted how the inclusive Higgs production cross-section steeply increases with the centre-of-mass energy while the background, the total inelastic $p-p$ cross-section, remains approximatively constant over a wide range of energies. Therefore, it can be argued that, to raise the Higgs statistics, the highest possible centre-of-mass energy should be used.

One of the basic ideas behind the LHC design is to install a new hadron collider into the existing 27 km long tunnel previously occupied by the LEP (sited

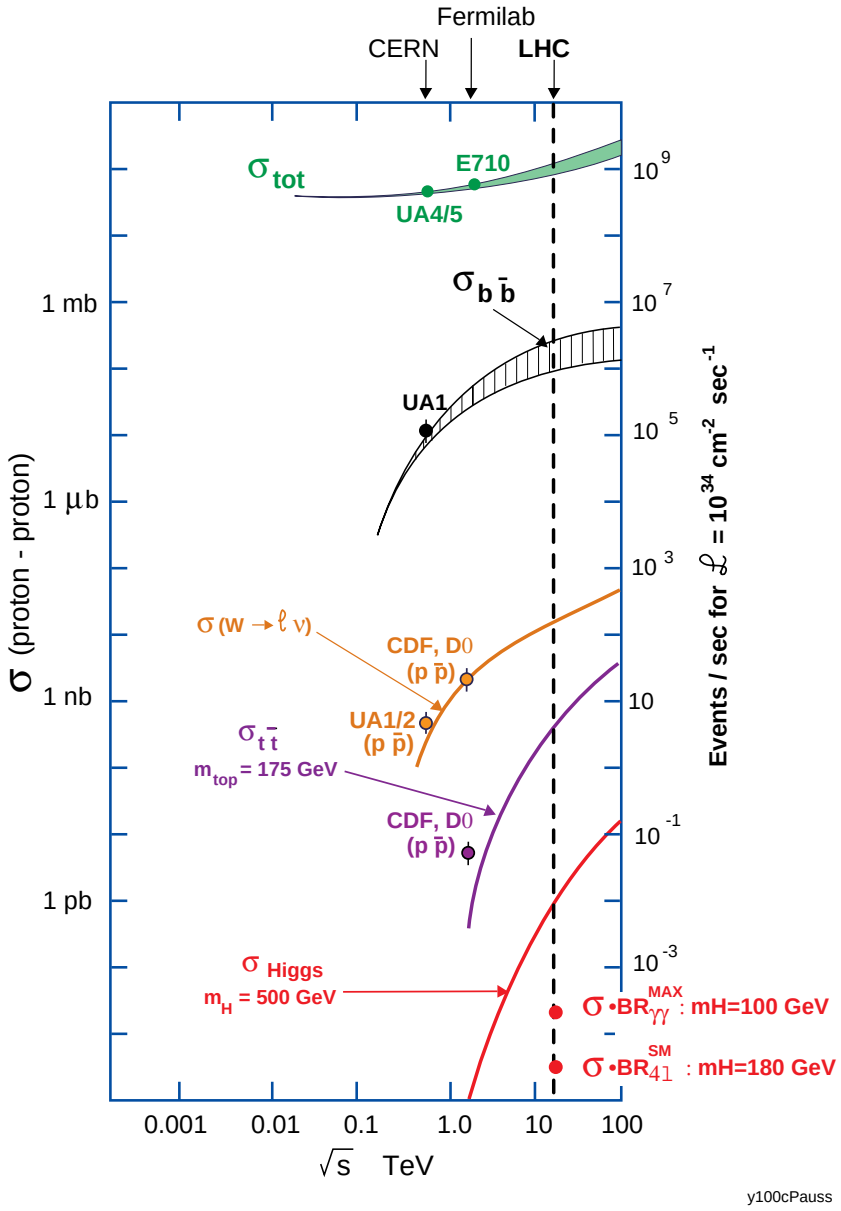


Figure 1.8: *Cross-sections for different processes as a function of the centre-of-mass energy in $p-p$ collisions*

100 m depth at CERN laboratories in Geneva). This gives also the possibility to reuse several infrastructures, including preaccelerators. In the LHC design, 1232 main dipoles operating at 1.9 K and generating a magnetic field up to 8.33 T will be used to steer the particles into curvilinear trajectories together with 386

quadrupoles, 360 sextupoles and 336 octupoles for stability control.

The other important characteristic to maximize the rate for a particular process is the luminosity; this has the drawback, however, that the total event rate can become so high that several interactions overlap in the same bunch crossing (*pile up*). The LHC will operate at a bunch crossing rate of 40 MHz and at a design LHC luminosity of $10^{34}\text{cm}^{-2}\text{s}^{-1}$ = 10 $\text{nb}^{-1}\text{s}^{-1}$. The bunch structure is such that only about 80% of the bunches will be full [28]; since the total non-diffractive inelastic $p-p$ cross section predicted by PYTHIA [29] is 55 mb, on average 17.3 events will occur at every bunch crossing. With about 50 charged tracks per interaction, this pile-up poses several experimental problems.
In the first three years of operation, the LHC will run at a reduced luminosity of $2 \times 10^{33}\text{cm}^{-2}\text{s}^{-1}$; only after that it will run at its design luminosity. The two luminosity regimes are commonly called *High luminosity* and *Low Luminosity*, respectively.

The LHC will also be able to accelerate and collide beams of heavy ions such as Pb at 2.76 ATeV to study the deconfined state of matter, the quark-gluon plasma. The parameters of the LHC are summarised in table 1.4.

LHC is planned to produce the first collision on April 2007 and start the physics programme from August 2007.

	$p-p$	${}^{82}_{208}Pb-{}^{82}_{208}Pb$
Beam energy at injection	450 GeV	73.8 TeV
Beam energy at collision	7 TeV	574 TeV(2.76 ATeV)
Maximum Luminosity	$1\times10^{34}\text{cm}^{-2}\text{s}-1$	2×10^{27} $\text{cm}^{-2}\text{s}^{-1}$
Number of Bunches	2808	1608
Bunch spacing	7.48 cm	5.3 cm
Bunch separation	24.95 ns	124.75 ns
Number of particles per bunch	1.1×10^{11}	8×10^{7}
Total crossing angle	300 μrad	<100 μrad
Bunch Length (r.m.s.)	7.5 cm	7.5 cm
Transverse beam size at Impact Point	15 μm	15 μm
Luminosity lifetime	10 h	4.2 h
Filling time per ring	4.3 min	9.8 min

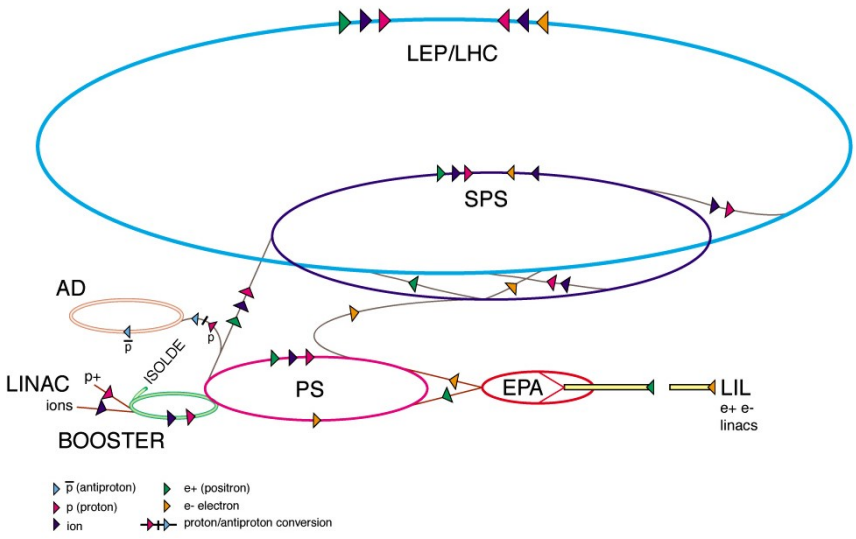


Figure 1.9: *Overview of the accelerator complex at CERN. The LEP e^+ and e^- trajectories and LHC $p-p$ and $Pb-Pb$ trajectories are indicated. Protons will be accelerated and brought up to 50 MeV by a linear accelerator LINAC. A Booster raises the beam energy up to 1.4 GeV injecting proton beams into the old circular accelerator PS. The 25 GeV energy beams extracted from PS are injected to a bigger circular accelerator SPS, which introduces 450 GeV proton beams into the LHC ring. The tunnel of LHC is a 26.659 km circumference, composed with 8 curvilinear sections (2.840 km) and 8 rectilinear sections, where the beams collide.*

1.5 Basic phenomenology of proton-proton collisions

When two protons collide at energies higher than their mass, the interaction involves their constituents, since the proton is resolved into its "partons" (quarks and gluons), carrying only a fraction x of the total momentum of the proton. The distributions of the x variable for the different constituents are called *Parton Density Functions* $f(x,Q^2)$. They depend on x and on Q^2, the exchanged four-momentum during the interaction: at low Q^2 the major contribution comes from the valence quarks, while at high Q^2, the PDFs are shifted towards lower values of x, uniformizing the contribution of valence and sea quarks.

In figure 1.10 the CTEQ6L PDFs [30] at two different values of Q^2 are presented.

The general schema of a $p-p$ interaction is shown in figure 1.11.

The energy available for each interaction of the di-parton system is $\sqrt{\hat{s}} =$

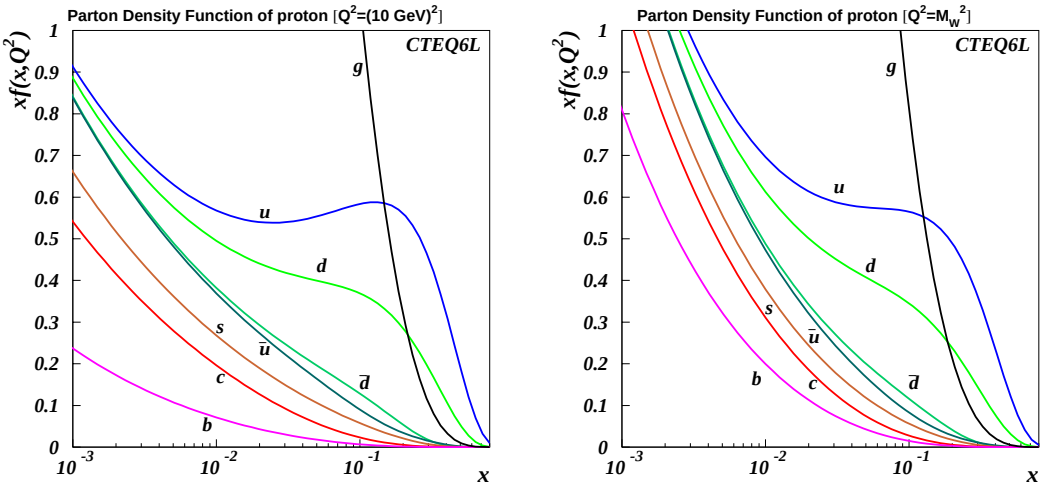


Figure 1.10: *Parton Density Function for a proton with Q^2=10 GeV^2 (left) and $Q^2=m_W^2$ (right).*

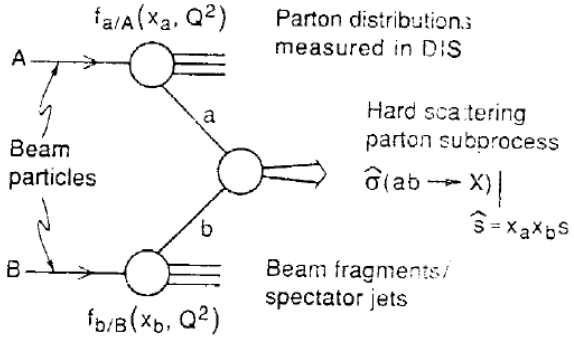


Figure 1.11: *The general schema of a $p-p$ interaction.*

$\sqrt{x_a x_b s}$, and in general the two fractions x_a and x_b are different. The cross-section of a generic $p-p$ interaction can be written as

$$\sigma = \sum_{a,b} \int dx_a dx_b \, f_a(x_a, Q^2) \, f_b(x_b, Q^2) \hat{\sigma}_{ab}(x_a, x_b), \tag{1.11}$$

where $\hat{\sigma}_{ab}$ is the cross-section for the elementary interaction between partons a and b, and $f_a(x_a, Q^2)(f_b(x_b, Q^2))$ represents the PDF for fraction $x_a(x_b)$.

The $p-p$ total cross-section, estimated from the results of the experiments UA4, UA5 and E710, is equal to [31]:

$$\sigma_{tot} = (100 \pm 20)\text{mb},$$

where the purely elastic one has been included (30%). The inelastic interactions, around 60-70 mb, belong to two classes:

- Large distance collisions between the two incoming protons, where only a small momentum is transferred during the interaction. They are soft collisions with production of particles with large longitudinal momentum and small transverse momentum (p_T around 500 MeV/c). The scattering at large angle is suppressed, most of the particles escaping detection along the beam pipe. This kind of processes is referred to as *Minimum Bias* and represents the large majority of $p-p$ collisions.

- Head-on collisions at small distances between parton *a* from one proton and parton *b* from the other. In this hard scattering there is a transferred momentum larger than in *Minimum Bias* and massive particles could be created, with higher p_T and large angles with respect to the beam line. These are the interesting physics events, but unfortunately they are rare. For example, the inclusive W (Z) production cross section is 140 nb (43 nb), which results in an interesting event every about 2 millions (8 millions) $p-p$ interactions.

In table 1.1 the rates ($R_i = L \cdot \sigma_i$) are reported for some important processes at LHC (*Low Luminosity*).

Process	Events/s	Events/year
$W \rightarrow e\nu$	20	$5 \cdot 10^8$
$Z \rightarrow l^+l^-$	2	$5 \cdot 10^7$
$t\bar{t}$	4	10^8
$b\bar{b}$	10^5	10^{12}
H ($M_H \simeq 800$ GeV)	0.002	10^4
QCD jet ($p_T > 200$ GeV)	10^2	10^9

Table 1.1: *Expected rates for some processes at LHC at* Low Luminosity *[31]*

It is convenient to introduce boost invariant quantities to define the kinematics of the process:

- transverse momentum, p_T, the projection of the momentum p on a plane perpendicular to the beam axis

- rapidity, y, defined as

$$y = \frac{1}{2}\ln\frac{E+p_z}{E-p_z} = tanh^{-1}(\frac{p_z}{E}) \tag{1.12}$$

indicating with E the energy, and with p_Z the projection of momentum p on the beam axis [1]

Thus the invariant differential cross section is conveniently expressed as

$$E\frac{d^3\sigma}{d^3\vec{p}} = \frac{d^2\sigma}{\pi dy d(p_T^2)} \simeq \frac{d^2\sigma}{\pi d\eta d(p_T^2)}, \tag{1.14}$$

having used the relation $\frac{dy}{dp_z} = \frac{1}{E}$ and integrating over the azimuthal angle ϕ.

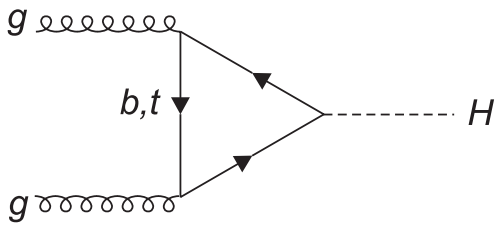
1.6 Standard Model Higgs searches at LHC

1.6.1 Standard Model Higgs Production at LHC

In the following the different production processes for the Standard Model Higgs at LHC will be reviewed.

Gluon fusion: $gg \rightarrow H$

The gluon fusion process is the dominant Higgs production mode at the LHC over the entire mass range, that is up to about 1 TeV. The diagram at leading order is presented on the figure on the right, and the complete calculations are given in the reference [32].



A more precise calculation should take into account the 2-loop QCD radiative corrections.

Usually the higher order correction is expressed in the K-factor defined as the ratio of the Higher-Order cross-section over the Leading-Order one, which considering only NLO is written as

$$K = \frac{\sigma_{NLO}}{\sigma_{LO}}. \tag{1.15}$$

[1] Under a boost in the z direction with velocity β, $y \rightarrow y - tanh^{-1}\beta$; quantities like the rapidity distribution $\frac{dN}{dy}$ are invariant under a boost along the z direction. In the ultrarelativistic approximation $\frac{m}{p} \ll 1$, the rapidity may be expanded to obtain

$$y = \frac{1}{2}\ln\frac{1+cos\theta+\frac{1}{2}(\frac{m}{p})^2+o((\frac{m}{p})^2)}{1-cos\theta+\frac{1}{2}(\frac{m}{p})^2+o((\frac{m}{p})^2)} \simeq -\ln\tan(\frac{\theta}{2}) = \eta \tag{1.13}$$

with $\cos\theta = \frac{p_z}{p}$. The equation 1.13 defines the quantity η pseudorapidity, approximatively equal to y if $\frac{m}{p} \ll 1$ and $\theta \gg \frac{1}{\gamma}$ and in any case measurable when either the mass or the momentum of a particle are unknown.

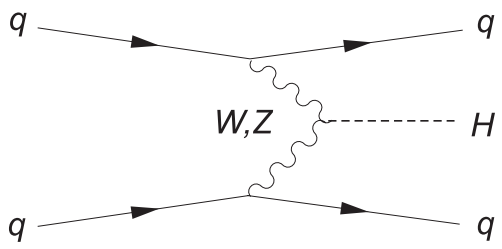
Complete calculations of the total NLO cross-section σ_{NLO} in this channel can be found in the reference [33].

The total correction K in this channel is large, ranging from 1.6 to 1.9 for different m_H. The most important theoretical uncertainties come from the parametrization of the parton distribution functions, especially the gluon ones, and from the contributions of higher orders, still unknown.

An insight to parton density functions uncertainties can be obtained by calculating the cross section with different structure functions. In the reference [34], using different proton structure functions (CTEQ4M, MRS(R1), GRV(92)), variations of about 10% of the cross section over the entire mass range were found. The variations of σ_{NLO} with the renormalization and factorization scales are small with respect to the variations of σ_{LO} and contribute to less than 15%.

W and Z fusion: $qq \rightarrow Hqq$

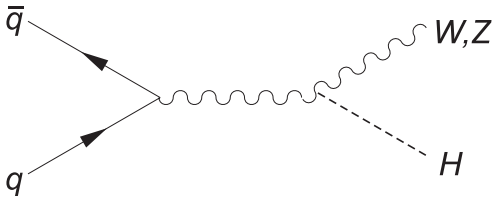
The cross section of the Higgs boson production through the fusion of virtual W or Z bosons is one order of magnitude smaller than then the gluon fusion in the intermediate mass range, only becoming competitive for Higgs masses around 1 TeV [35], [36]. The process at leading order is shown on the figure on the right.



The QCD corrections are well known [36]; the K factors for these processes are smaller than the gluon fusion one, in the interval 1.08-1.1. The most important characteristics of these production processes are the presence of two forward jets, a high invariant mass of these jets and a suppression of the hadronic production in the central region. Despite the lower cross-sections, the particular features of this channel can be exploited to increase the signal-to-background ratio in the search for an intermediate mass Higgs at LHC.

Higgs-strahlung $qq \rightarrow VH$

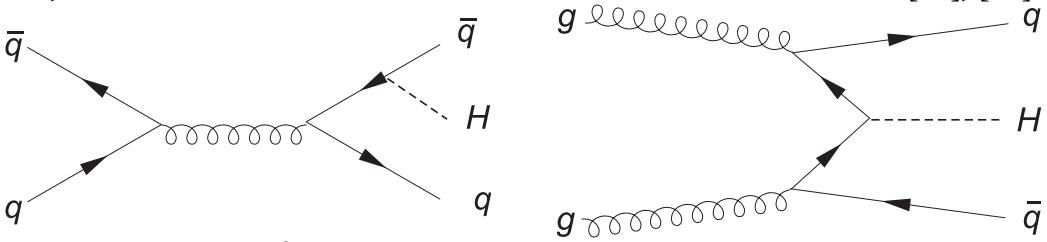
The Higgs production through the *Higgs-strahlung*, associated Higgs production with a W or Z vector boson, presents the interesting feature to observe and tag on the vector boson decay products. Cross section for these processes is about one to two orders of magnitude smaller the gluon fusion process one in the range m_H<200 GeV. The *Higgs-strahlung* process at leading order is presented in the figure on the right (complete calculation given in [37]).



QCD corrections are identical to the one for the Drell-Yan process [38]. K factor for this process ranges to 1.25 to 1.40.

Associated production with a $t\bar{t}$ pair

In the intermediate mass range, the cross section for the Higgs production in association with top quarks becomes similar to the *Higgs-strahlung* cross section. With the detection of the associated pair and in the decay channel, this process gives an additional possibility to search for the Higgs in the mass region below 130 GeV [39]. The process at leading order is presented on the diagrams below, and the cross section calculations can be found in the references [39], [40].



The calculation of QCD corrections is rather involved and only recently has been made available [42]; K factors are around 1.2.

Summary

The cross sections of the various Higgs production mechanisms at the LHC are presented in figure 1.12 as a function of the Higgs mass [34]. All known QCD and QED corrections are included in the calculations, apart from the ones for the $gg, qq \rightarrow t\bar{t}H$ channel.

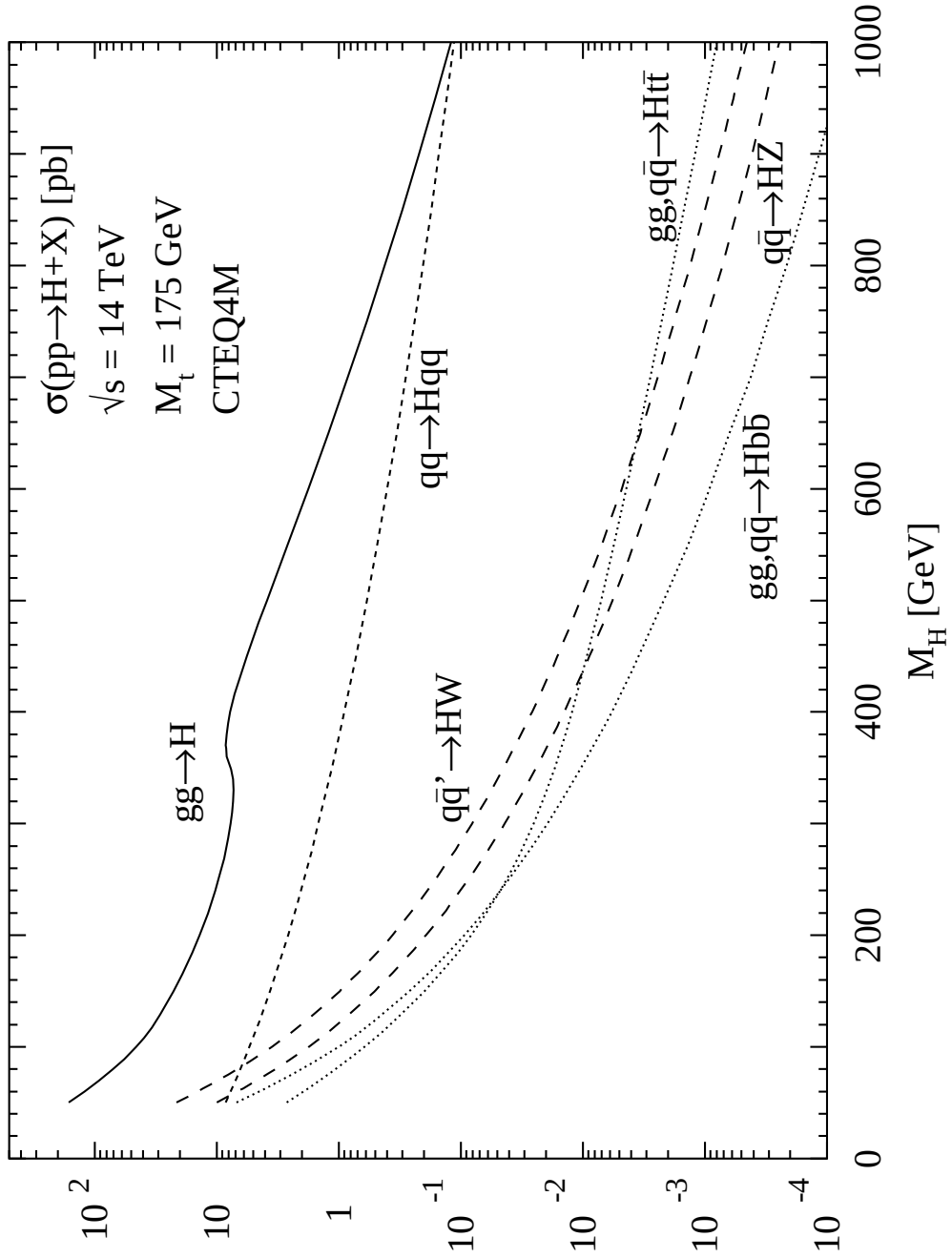


Figure 1.12: *Higgs production cross sections at the LHC as a function of the Higgs mass.*

Over the entire mass range, the gluon fusion is the dominant production mode, vector boson fusion becoming competitive only in the region around 1 TeV. In the intermediate mass range, $100 < m_H < 200$ GeV, several combinations of produc-

tion and decay channels are accessible at LHC, giving the possibility to measure the Higgs couplings [41].

1.6.2 Strategy for the Standard Model Higgs search at LHC

The Higgs decays into fully hadronic final states are the most copious at LHC, however it would be very difficult to resolve them when merged in the higher QCD background. Therefore topologies with leptons or photons are preferred, even if disfavoured by their smaller branching ratios. The associated production with a leptonically decaying particle or with forward jets can be exploited.

The different search strategies at LHC depends on the Higgs mass since the Higgs decay channels branching ratios (see figure 1.2 on page 6); it is possible to define according to the Higgs decay properties three regions:

- Low Mass Region $m_H < 130$ GeV where the $b\bar{b}$ decay mode dominates
- Intermediate Mass Region 130 GeV $< m_H < 2\ m_Z$ where the $b\bar{b}$ decay decreases with increasing $H \to VV^{(*)}$ (V=$W^{\pm}$ or Z).
- High Mass Region $m_H > 2\ m_Z$ where the Higgs boson decays mainly into on-shell W^+W^- or ZZ pairs.

In what follows the Higgs searches at LHC into these three different regions will be discussed, introducing the subject to which the last part of this thesis will be devoted, the Higgs search in the channel $H \to ZZ^{(*)} \to e^+e^-e^+e^-$.

Low Mass Region

The dominant decay channel in this mass region is $H \to b\bar{b}$. Difficulties arise for this channel considering the overwhelming background due to QCD di-jet rate and the fact that the inclusive $H \to b\bar{b}$ decay lacks any useful trigger for CMS, since neither the jet trigger nor the leptonic trigger can be used. A more favourable situation can be obtained restricting the search into the associate production channels, where the decay products of the top quark pair or of the vector boson produced together the Higgs, allow either to trigger the events, searching for an high energy lepton (also mixed triggers can be used, e.g. lepton+b-jets) or to enhance the S/B ratio, adopting tagging techniques. The HZ channel seems to be of low interest, being already suppressed in comparison with the HW, and taking into account the leptonic branching ratio of the Z (an addition factor three). So with a lepton, missing energy and two/four tagged b-jets, the Higgs search in the $H \to b\bar{b}$ decay channel can be possible.

Another possibility is to select rare decays, with a favourable S/B ratio: the "golden channel" in this region is $H \to \gamma\gamma$. The requirements for a good $\gamma\gamma$ invariant mass recontruction are an excellent energy and direction resolution, hence an excellent electromagnetic calorimeter is required. To have a good direction resolution, the primary vertex associated to the Higgs production should be identified and reconstructed: at high luminosity algorithms for vertex reconstruction using tracks are under study, but a good knowledge of the high luminosity pile-up is needed. The background for this channel has an irreducible component due to $pp \to \gamma\gamma + X$ and to $pp \to \gamma +$ jet $+X$, with a hard bremsstrahlung coming from a quark jet [3], and a reducible component due to QCD multi-jets production or γ+jet; the latter background is about 40% of the irreducible one. A good π^0/γ discrimination is needed to reduce this contribution.

This decay channel can also be studied in the exclusive production process $pp \to H + jet + X$, where the Higgs boson is produced at large p_T, and in the associated WH production channel with one isolated lepton from W [43]. The search in these processes is less sensitive to the $\gamma\gamma$ mass resolution, and the backgrounds can be significantly reduced with the requirement of a lepton or a jet yielding a $S/B \simeq 1$.

Another promising channel in this region is the $H \to \tau^+\tau^-$ decay, with a branching ratio of about 8%, using in particular the vector boson fusion process, where the energetic quark jets in the forward and backward direction allow to suppress the brackground processes, coming mainly from QCD and Zjj.

Intermediate Mass Region

In this mass region the Higgs boson starts decaying into couples of vector bosons WW or ZZ. The most promising channels are $pp \to H \to WW \to l^+\nu l^{'-}\bar{\nu}$ or $pp \to H \to ZZ \to l^+l^-l^{'+}l^{'-}$ with $l, l^{'}$ = e or μ. The WW decay mode has to be extracted from a background mainly due to $qq \to WW$ continuum or $t\bar{t} \to bW^+\bar{b}W^-$ and $W^{\pm}t(b)$ associated production.

The fully leptonic decay $H \to ZZ^* \to 4l$ has a very clean experimental signature. In particular a good lepton identification and reconstruction is required in this mass range, where the invariant mass resolution is dominated by the experimental resolution, being the natural Higgs Boson width negligible. The signal selection is based on the identification of two opposite charged lepton pairs coming from a common vertex. The invariant mass of one of the two pairs should be compatible with m_Z. The main irreducible background is continuum ZZ^* production together with reducible background $t\bar{t} \to 4l + X$ and $Zb\bar{b} \to 4l + X$. In the first case leptons come from $t \to Wb$ decay followed by $W \to l\nu$ and semileptonic b decay, in the

second case two leptons are from the real $Z \rightarrow ll$ and the other two from b quark decay chains.

A sharp decrease of the branching ratio $H \rightarrow ZZ$ (see figure 1.2) can be noted around 160-170 GeV due to a threshold effect when the decay to two W bosons on the mass shell becomes possible.

High Mass Region

In this region the predominant decay channels are $H \rightarrow W^+W^-$ and $H \rightarrow ZZ$ with both vector bosons *on-shell*. The $H \rightarrow ZZ \rightarrow 4l$ channel has a smaller ZZ irreducible background than in the intermediate mass region, requiring both pair of invariant masses close to m_Z. Furthermore, the intrinsic Higgs width Γ_H is larger than the achievable experimental mass resolution, therefore the detector performance is less critical. For all these reasons, the $H \rightarrow ZZ \rightarrow 4l$ channel is a gold-plated Higgs boson signature at LHC in this mass region. For very large masses, $m_H > 600$ GeV/c^2, other decay modes are used to supplement $H \rightarrow ZZ \rightarrow 4l$, because the production cross section decreases significantly and the resonance peak of the four leptons become, due to very large Γ_H, no longer visible: $H \rightarrow Z(l^+l^-)Z(\nu\bar{\nu})$ or $H \rightarrow Z(l^+l^-)Z(q\bar{q})$.

Chapter 2

The CMS Experiment

LHC detectors will operate in a very difficult environment: the high bunch crossing frequency, the high event rate and the pile-up of several events in the same bunch crossing dictate strict requirements on the design of detectors. To cope with a bunch crossing rate of 25 ns and a pile-up of about 20 events per crossing, the detectors should have a very fast time response and readout electronics. Due to the presence of pile-up, high granularity is also required to avoid the overlap of particles in the same sensitive elements. High granularity means a large number of electronics channels, and therefore high costs. LHC detectors will also have to stand an extremely high radiation dose; special radiation-hard electronics must be used. Additional requirements apply to the online trigger selection, that has to deal with a background rate several orders of magnitude higher than the signal rate.

Four experiments will be installed at the LHC. A map of them is shown in figure 2.1.

Two of them are devoted to specific topics: ALICE to heavy ions collisions and LHC-b to b-physics. The other two are the general-purpose experiments ATLAS and CMS. Their design differs significantly, since two very different solutions were chosen for the configuration of magnetic field: CMS uses a solenoidal field generated by a big superconducting solenoid, while ATLAS uses a toroidal field produced by three sets of air-core toroids complemented by a small inner solenoid.

This chapter describes the general design of CMS and of its subdetectors. A general description of the CMS trigger system is given in the last section of this chapter.

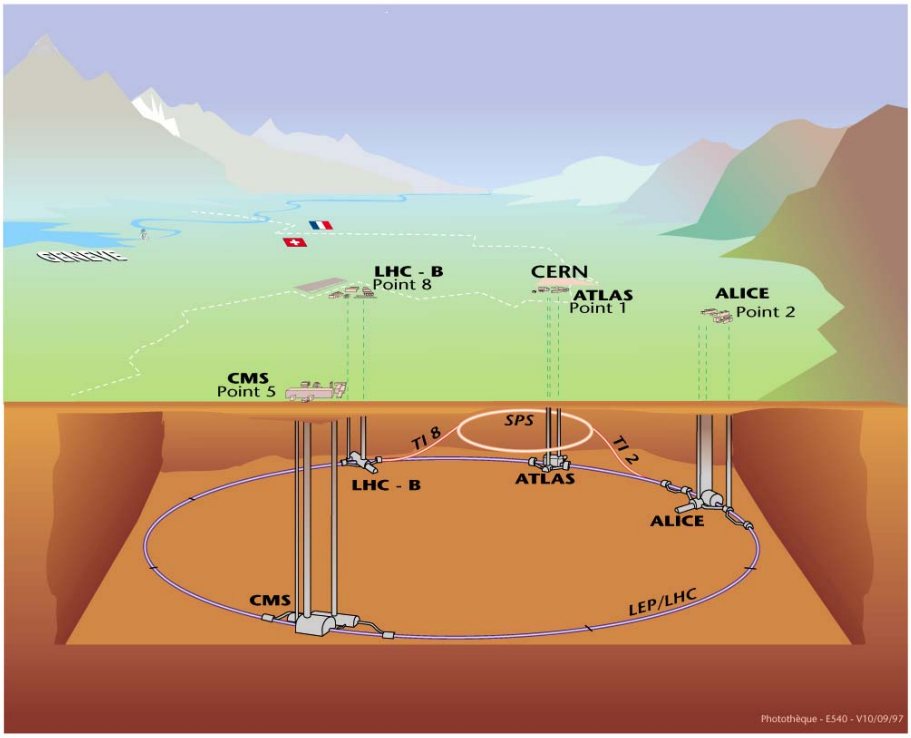


Figure 2.1: *Map of LHC and related experiments.*

2.1 CMS Overall Design

The Compact Muon Solenoid experiment, CMS [44] has as its main feature the compactness of design, obtained thanks to a very strong magnetic field of 4 T, generated by a superconducting solenoid. The design priorities expressed in the the CMS project [45] are a redundant muon system, a good electromagnetic calorimeter and a high quality tracking system.

CMS design has a cylindrical simmetry around the beam axis and has a typical structure of collider based physics experiment: several cylindrical layers coaxial to the beam direction, referred to as *barrel* layers, closed at both ends by detector disks orthogonal to the beam pipe, the *endcaps*, to ensure detector hermeticity. In figure 2.2 a schematic view of CMS: the full length is 21.6 m, diameter is 15 m for a total weight of $\simeq$ 12500 t.

The natural coordinate frame used to describe the detector geometry is a right-handed cartesian system with the x axis pointing to the centre of LHC ring, the

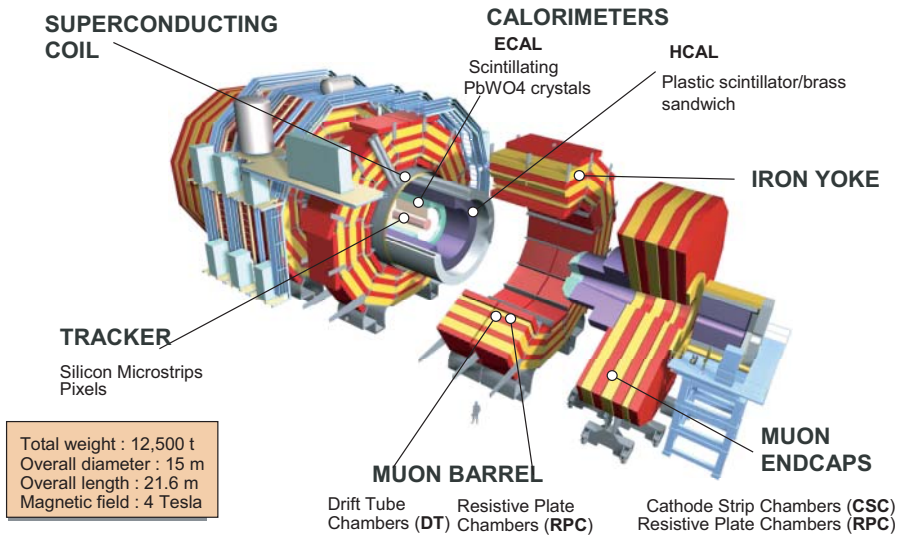


Figure 2.2: *Schematic picture of CMS experiment at LHC.*

z axis coincident with the CMS cylinder axis and the y axis directed almost upwards along the vertical. Cylindrical symmetry of CMS design drives the use of a pseudo-angular reference frame, given by the triplet (r, ϕ, η), with r distance from z axis, ϕ azimuthal coordinate with respect to x axis and pseudorapidity η defined by equation 1.13.

The longitudinal view of one quarter of the CMS detector is shown in figure 2.3. The transversal view of the barrel region is shown in figure 2.4. Detectors and non-sensitive volumes are indicated with a standard defined by two-letter code.

The CMS design is driven by the choice of its magnet (**CB**), a 13 m long superconducting solenoid [46] with a diameter of 5.9 m. Cooled with liquid helium, it will generate a magnetic field of 4 T, which is kept uniform by a massive iron return yoke (**YB**,**YE**). The yoke will also host the muon system (**MB**,**ME**), composed by drift tube detectors in the barrel region and cathode strip chambers in the endcaps (up to $|\eta| < 2.4$), complemented by a system of resistive plate chambers with a coverage of $|\eta| < 2.1$.
The calorimeters and the inner tracker are installed inside the coil. Very fine segmentation is crucial for the innermost detector to deal with a very high track density; therefore a silicon pixel detector was chosen. In the baseline design it consists of 3 barrel layers and 2 forward disks. Outside the pixel detector, a silicon strip detector is installed, extending up to a radius of about 1.2 m. The full silicon tracker allows charged tracks reconstruction in the acceptance region of

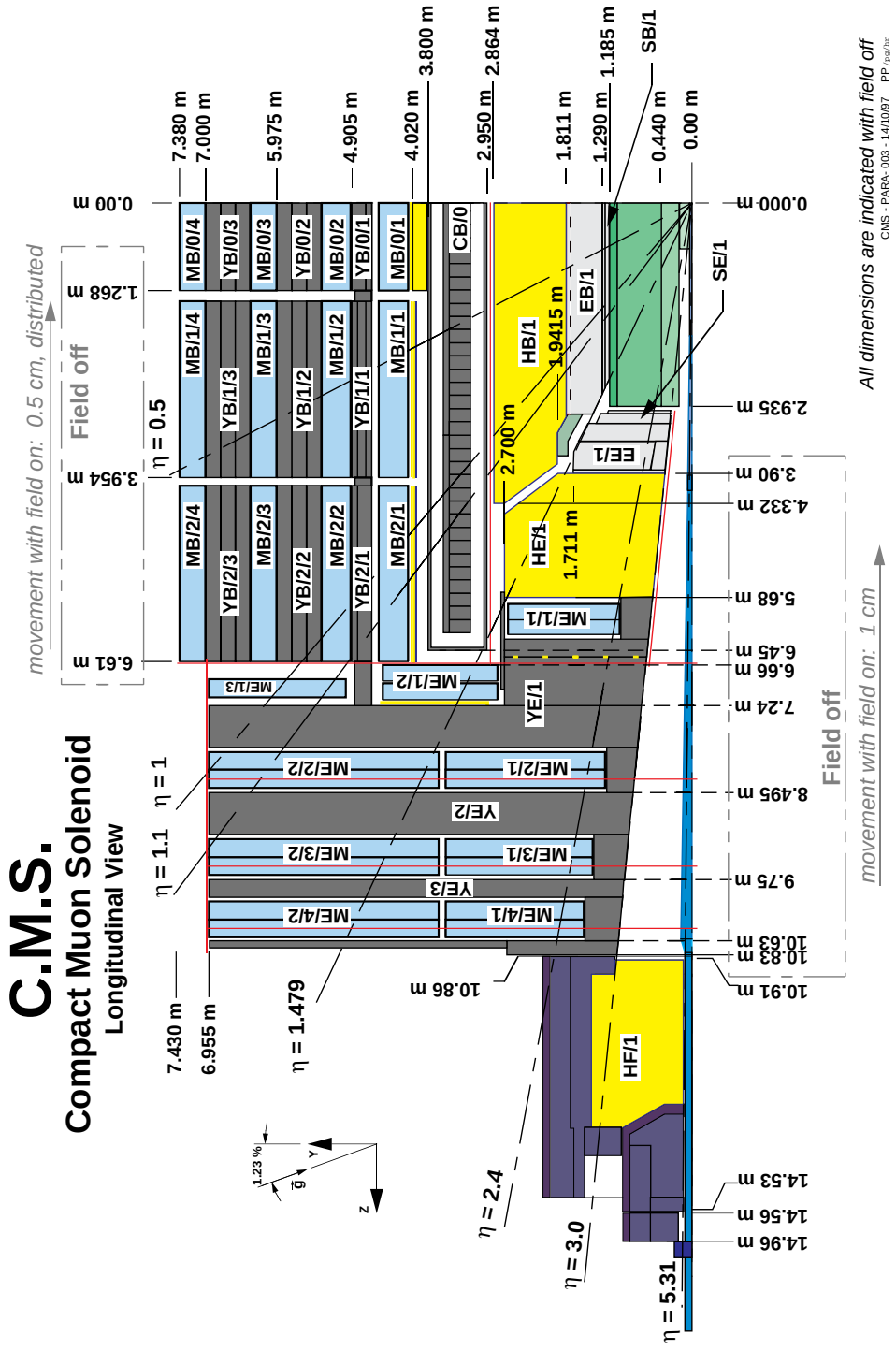


Figure 2.3: *Longitudinal view of a quarter of CMS experiment. Detectors and non-sensitive volumes are indicated by two-letter code: the first letter indicates the subdetector (S=Silicon tracker, E=Electromagnetic calorimeter, H=Hadron calorimeter, C=magnet Coil, Y=magnet iron Yoke, M=Muon chambers), the second letter refers to the position (B=Barrel, E=Endcap, F=Forward region).*

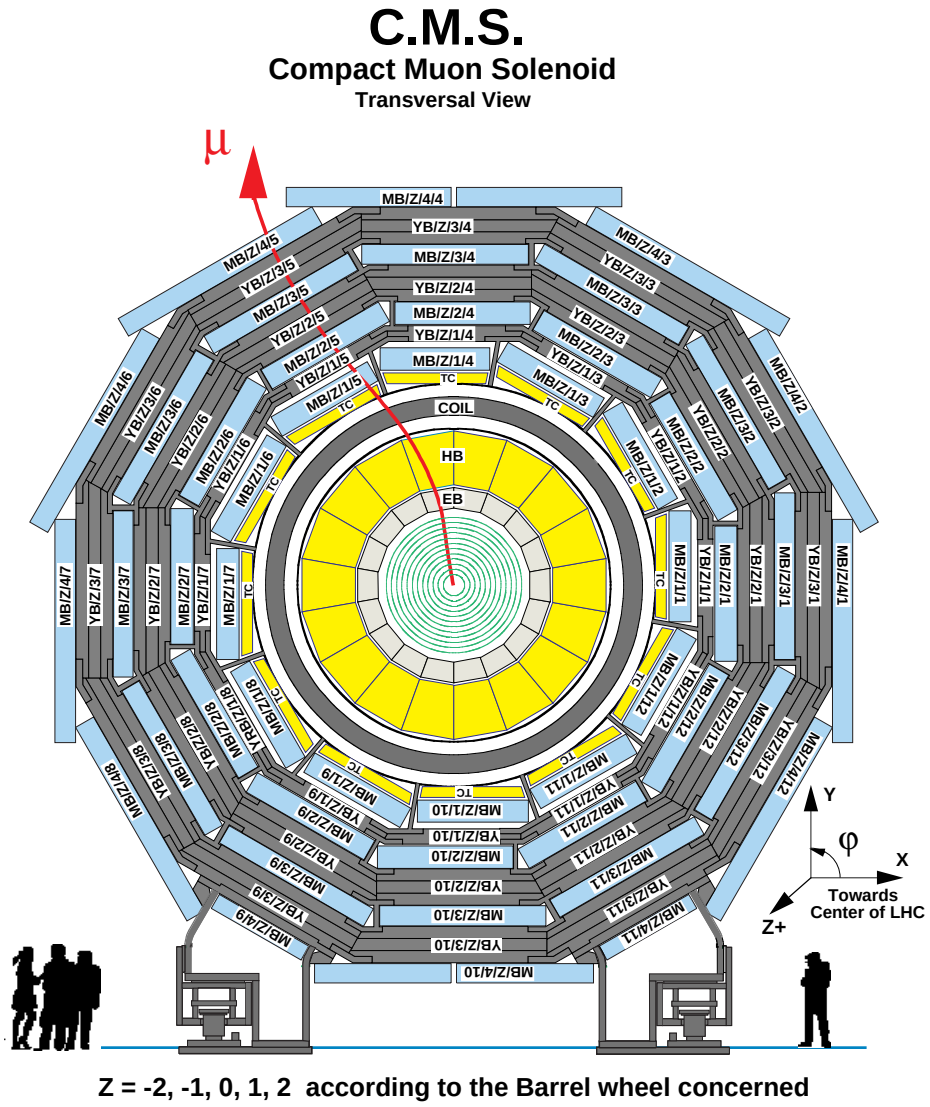


Figure 2.4: *Transversal view of the barrel region of CMS. Barrel wheels are identified with the same two-letter code of figure 2.3 and numbered along z direction as -2, -1, 0, +1, +2.*

$|\eta| < 2.5$.
Photons and electrons are measured by a homogeneous electromagnetic calorimeter (ECAL), composed by lead tungstate $PbWO_4$ scintillating crystals covering the region $|\eta| < 3.0$ (**EB**,**EE**). In the endcaps, it will be supplemented by a lead/silicon preshower detector, to improve the resolution in the determination of electron and photon direction and to help pion rejection.
Jets and energy imbalance are measured by a sampling hadronic calorimeter (HCAL) installed just before the coil. It is composed of a copper alloy and stainless steel instrumented with plastic scintillators. The barrel and endcap parts (**HB**, **HE**) have the same η coverage of the ECAL, and are complemented by a very forward calorimeter (**HF**), which extends the coverage up to $|\eta| < 5.3$.

2.2 The Tracker

Track and vertex finding at the LHC will be an important tool for identifying signal events and rejecting background. Efficient track reconstruction provides electron-photon separation, and is helpful in identifying W and Z bosons which will be involved in many new physics signatures at the LHC. Track isolation is another important tool because it can be used to suppress jet backgrounds to isolated high energy photons and electrons.

The design of the tracking system has changed substantially since the Technical Proposal and an updated description is available in the tracker TDR [47]. The tracking system consists of a number of silicon pixel layers close to the interaction point, surrounded by a large silicon tracking detector. A major constraint on the design of the tracking system is reduce as much as possible the material budget in front of the calorimeters. Early conversion of photons degrades sensitivity to $H \rightarrow \gamma\gamma$, and bremsstrahlung in the tracker material and high magnetic field impairs the energy resolution for electrons.

In addition to accurate track reconstruction, the tracker will provide vertex identification. In the case of $H \rightarrow \gamma\gamma$, charged recoil tracks can be used to identify the Higgs vertex, for example. Massive particle decays at the LHC will frequently involve B mesons, with markedly displaced secondary vertices. An important role of the tracking system will be the identification of these secondary vertices in order to tag b jets in CMS.

The tracker extends in the region $|\eta| < 2.5$, r<120 cm, |z|<270 cm and it is completely based on semiconductor detectors made of silicon covering the largest ever-designed *Si* detector surface of 198 m^2.

To better solve the pattern recognition problem, the tracker is designed to fulfil two basic properties: low cell occupancy and large hit redundancy. The low occupancy is obtained by working with high granularity detectors, mainly the

ones closer to the interaction point because they have to cope with higher particle fluxes, and fast primary charge collection, obtained with thin detectors and overdepleting the silicon bulks. The redundancy is guaranteed by the overall design of figure 2.5, which allows many measured points per track within an acceptable material budget.

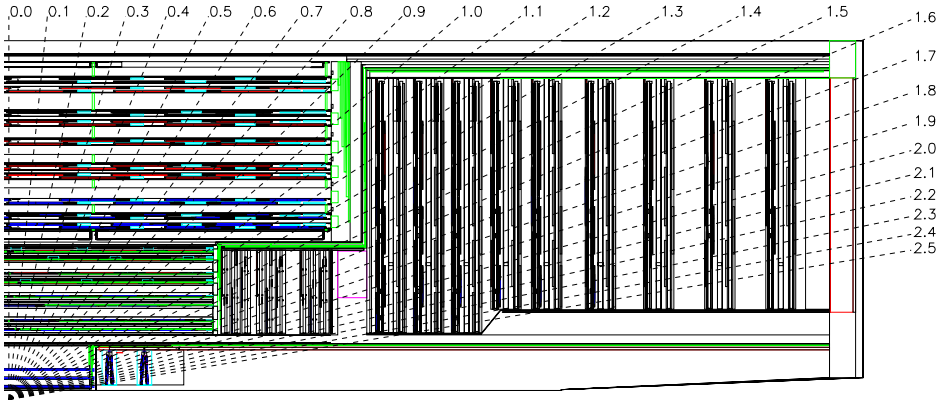


Figure 2.5: *Schematic view of a quarter of the CMS silicon tracker comprehensive of the supporting structures, cables and services.*

An average of 12-14 points (hits) per track are guaranteed to permit a high tracking efficiency and a low rate (10^{-3} or less) of fake tracks (reconstructed tracks not corresponding to any real track). To contrast the radiation damage, both pixel and microstrips detectors have to be kept cold at a working temperature of -10^o C for the whole tracker volume.

The tracker will reconstruct high energy electrons (from W or Z decays, for example) with an efficiency better than 90%. Simulations of single muons within tracker show that an efficiency close to 100% is reachable in the range $|\eta|$ <2.0 (see figure 2.6). The standalone tracker performance for isolated muons over a range of pseudorapidities and for several different muon transverse momenta is shown in figure 2.7.

Leptons and charged hadrons which are produced in the central region are reconstructed with the momentum precision given by equation 2.1.

$$\frac{\Delta p_T}{p_T} \approx 0.005 + 0.15 p_T \qquad p_T \text{ in TeV} \tag{2.1}$$

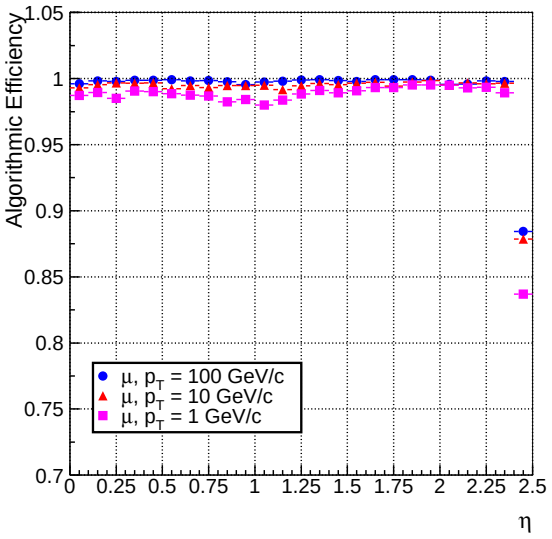


Figure 2.6: *Global track reconstruction efficiency for single muons.*

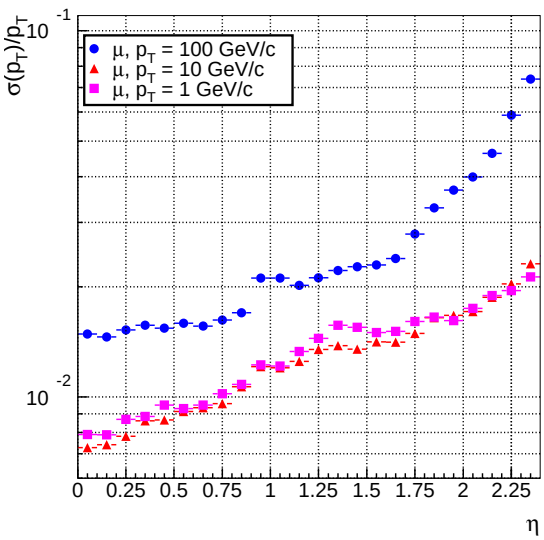


Figure 2.7: *Single muon track transverse momentum resolution.*

2.3 The Electromagnetic Calorimeter (ECAL)

A high performance electromagnetic calorimeter is a fundamental requirement for any general purpose LHC experiment, in order to have precise measurements on electrons and photons. The design of CMS ECAL [3] has been prompted by the possibility to observe the decay of a light Higgs boson into a couple of photons. Since in the region m_H<140 GeV/c^2 the intrinsic Higgs width Γ_H is less than 30 MeV, the $\gamma\gamma$ invariant mass resolution is dominated by experimental resolution, which should be of the order of 1% to enhance the significance of a possible signal.

The CMS collaboration has chosen a homogeneous calorimeter composed with finely segmented crystals of lead tungstate ($PbWO_4$), which is a radiation resistant and chemically inert scintillator suited to work in the LHC high dose environment (from 0.18 Gy/h at $|\eta| = 0$ to 6.5 Gy/h at $|\eta| = 2.6$ at high luminosity). Figure 2.8 shows a longitudinal view of a quarter of the electromagnetic calorimeter: it is organized in a barrel region $|\eta| <1.479$ and a forward region to cover the pseudorapidity area below 3.0. Precision energy measurements will be made only in the region $|\eta| \leq 2.6$, matching the coverage of the tracker.

A detailed description of the electromagnetic calorimeter and of its performances in a test beam configuration will follow in chapters 3 and 4.

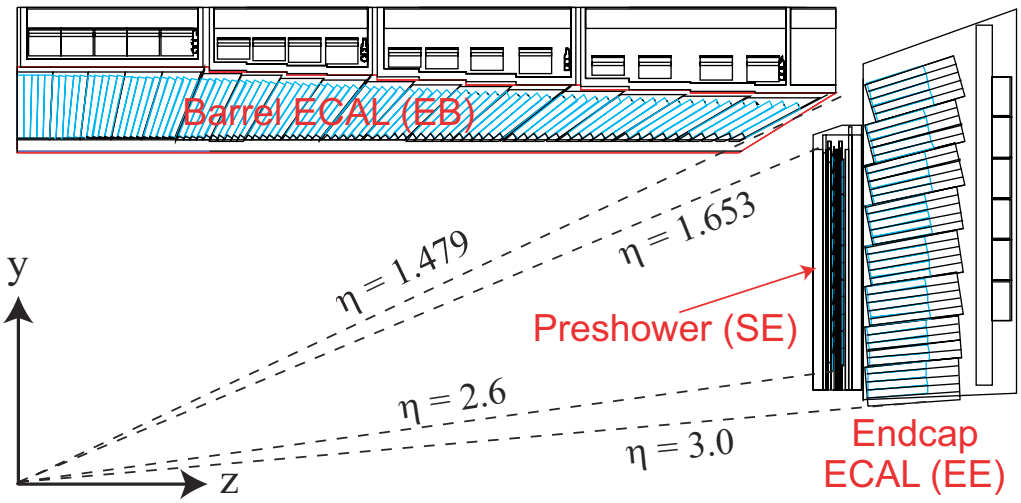


Figure 2.8: *Longitudinal view of a quadrant of CMS electromagnetic calorimeter.*

2.4 The Hadron Calorimeter (HCAL)

The hadron calorimeter is used together with the electromagnetic one to measure the energy and direction of jets, the transverse energy E_T and the imbalance of transverse energy, or missing transverse energy, E_T^{miss} . High hermeticity is required, combined to a material thickness sufficient to contain the whole hadron shower.

The CMS HCAL [48] is a sampling calorimeter with 3.7 mm thick active layers of plastic scintillators alternated with 5 cm thick brass plate absorbers. The signal is readout with wavelength-shift fibres. The barrel granularity $\Delta\eta \times \Delta\phi$=0.087×0.087, matching a 5×5 crystals ECAL tower, is fine enough to allow an effcient di-jet separation. The longitudinal view of HCAL is represented in figure 2.9: the barrel ($|\eta|$<1.4) and endcap (1.4< $|\eta|$ < 3.0) with an overall thickness from 8.9 to 10 interaction lengths λ_0 respectively. Since the barrel part of the calorimeter is not sufficiently thick to contain all the energy of highly energetic showers, an additional tail-catcher composed of scintillators tiles is placed outside the magnet.

To improve the hermeticty a very forward calorimeter (HF) is placed outside the magnet yoke, ±11 m away along the beam direction from the nominal interaction point, covering from $|\eta|$ = 3 to $|\eta|$ = 5. It has, as active elements, quartz fibres parallel to the beam, interleaved into steel plate absorbers, which constitute the passive material. With this configuration the complex of CMS hadron calorime-

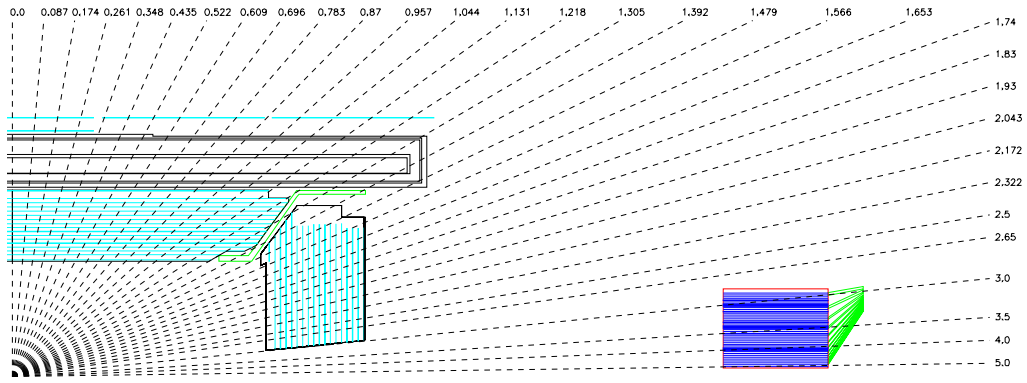


Figure 2.9: *Longitudinal view of a quarter of CMS hadron calorimeter, subdivided into barrel and endcap HCAL, placed inside the magnetic coil, the outer barrel tail-catcher and the very forward calorimeter HF.*

ter has an overall depth of more than 11 λ_0 over the full coverage. The hadronic energy resolution combined with ECAL measurements is

$$\frac{\sigma(E)}{E} = \frac{100\%}{\sqrt{E[GeV]}} \oplus 4.5\% \tag{2.2}$$

and it is expected to sensibly degrade around $|\eta|$ = 1.4, where there will be installed services and cables resulting in a higher amount of inactive material.
The performance of the very forward calorimeter

$$\frac{\sigma(E)}{E} = \frac{182\%}{\sqrt{E[GeV]}} \oplus 9\%(hadrons) \qquad \frac{\sigma(E)}{E} = \frac{138\%}{\sqrt{E[GeV]}} \oplus 5\%(electrons) \tag{2.3}$$

is sufficient to improve the missing transverse energy resolution to the desired level.

2.5 The Muon System

A huge muon detection system [49] is placed outside the magnet coil. Its purposes are multiple: muon reconstruction and identification, trigger for events with muons as well as precise time measurement of the bunch crossings.
The layout of the muon detection system is sketched in figure 2.10.

The muon detectors are integrated in the iron return yoke of the magnet. Both barrel and end-caps are made out of four active layers and three planes of absorber. The barrel region extends up to $|\eta| < 1.3$. It is divided into five segments. Each

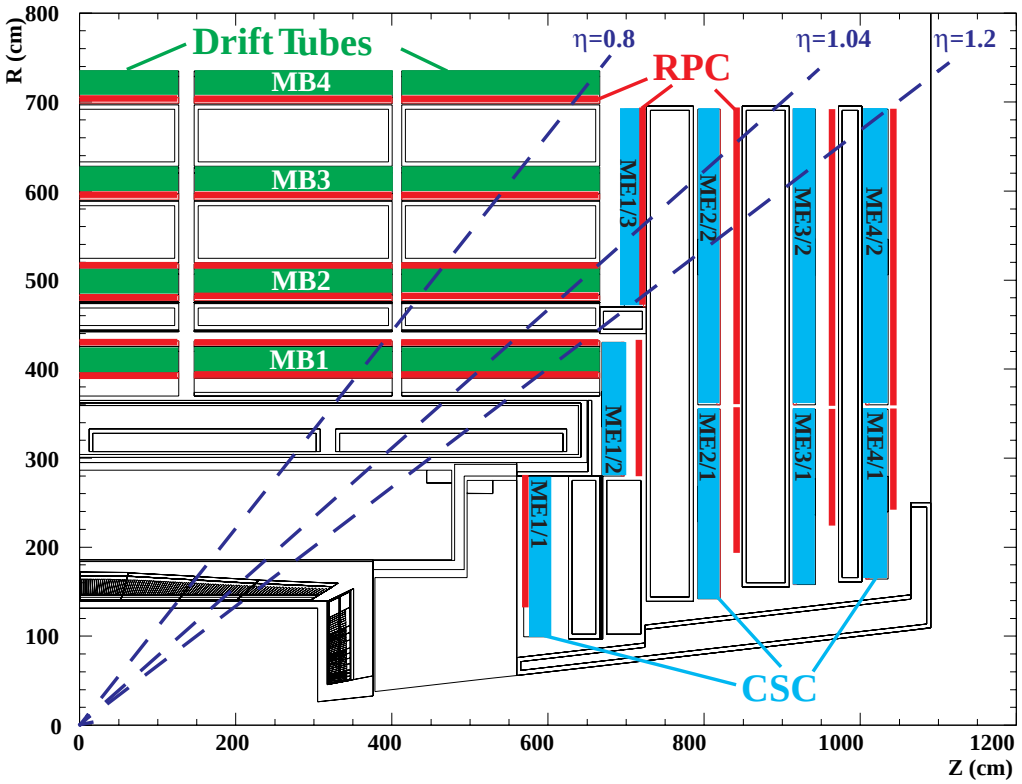


Figure 2.10: *Longitudinal view of a quarter of the muon system, subdivided into barrel, with drift tubes (DT) and resistive plate chambers (RPC), and endcap with cathode strip chambers (CSC) and RPCs.*

detection unit houses 12 layers of Drift Tube (DT) with approximatively 400 ns drift time and a time resolution of 5 ns. The DT spatial resolution is 250 μm per tube and an overall resolution of 100 μm in $R-\phi$ and 150 μm in z is expected. Endcaps extend the coverage up to $|\eta| < 2.4$. The active layers are equipped with trapezoidal shaped Cathode Strip Chamber (CSC) detectors. Each chamber is made of 6 sandwiches of cathode strips and wires which provide 3-dimensional reconstruction. CSC are designed to operate in non uniform magnetic field ranging from 1 to 3 T. The spatial resolution varies from 75 μm, for the first two inner layers, to 150 μm for the outer ones.

To complement DT and CSC measuremtents at trigger level, the information provided by the Resistive Plate Chambers (RPC) is used. RPC detectors have a very prompt time response and excellent time resolution (σ < 1-2 ns). There is a plane of RPC detectors for each layer of CSC detectors in the endcaps and the first, second and fourth layers of DT detectors in the barrel. Each RPC chamber in the barrel is made of two phenolic resin planes separated by a gap of a few mm filled with gas. Planes are coated by a conductive graphite paint in the shape of electrodes. Readout is made by plastic insulated aluminium strips outside the resin plates. The spatial resolution of RPC is of the order of the strip size (10-40 mm in $R-\phi$) and 100-1300 mm in z. These devices operate in avalanche mode, instead of the more common streamer mode, to cope with the LHC high rate.

2.6 The CMS Trigger

At the LHC nominal luminosity, the total event rate is of the order of 10^9 Hz; as it has been noticed, however, the rate for interesting events is very small, as shown in table 1.1. A large fraction of the corresponding selection has to be performed online, since the raw event size is of the order of 1 MB and storing and processing the resulting amount of data for a subsequent *off-line* analysis would be prohibitive. This online selection will be performed by the trigger system, reducing the event rate to the order of 100 Hz, enough to accommodate the interesting signal channels produced at LHC. This task is quite difficult not only due to the high rejection factors it requires (10^7), but also because the output rate is almost saturated already by *standard* processes like Z and W production. Therefore the trigger, in order to make its decision, should have a level of sophistication comparable to offline reconstruction, even if the time available to perform this selection is limited: bunch crossings will occur at a rate of 40 MHz, so that a decision must be taken every 25 ns, a time which is too small even to read out all raw data from the detector.

The *accept/reject* decision will be taken in several steps (**levels**) of increasing refinement, where each one takes a decision using only a subsample of the available data. The foreseen schema is presented in figure 2.11, where it is possible to identify two main decision step: **Level-1** Trigger and High Level Trigger(**HLT**).

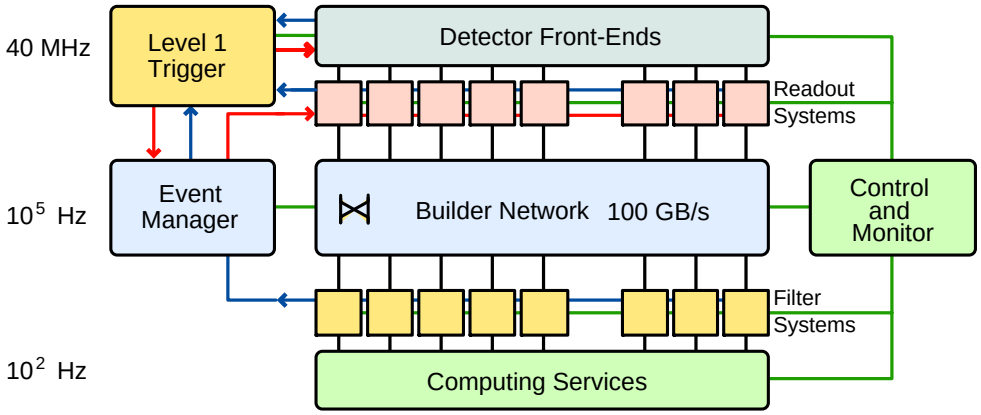


Figure 2.11: *Data flow in the CMS Trigger/DAQ system. The software-based High-Level Trigger (HLT) filters via the Data Acquisition system (DAQ) the events passing hardware-based Level-1 trigger. Time axis goes from upside down.*

2.6.1 Level-1 Trigger

The Level-1 trigger [50] is implemented on custom-built programmable hardware. It runs dead-time free and has to take an accept/reject decision for each bunch crossing, i.e. every 25 ns. This is achieved with a synchronous pipelined structure of processing elements, each taking less than 25 ns to complete. At every bunch crossing, each processing element passes its results to the next element and receives a new event to analyse. During this process, the complete detector data are stored in pipeline memories, whose depth is technically limited to 128 bunch crossings. The Level-1 decision is therefore taken after a fixed time of 3.2 μs. This time must include also the transmission time between the detector and the counting room (a cable path of up to 90 m each way) and, in the case of Drift Tube detectors, the electron drift times (up to 400 ns). The time available for calculations can therefore be as low as 1 μs. The Level-1 trigger is divided into three subsystems: the Calorimeter Trigger, the Muon Trigger and the Global Trigger.

The Calorimeter and Muon Triggers identify trigger objects of different types: isolated and nonisolated electrons/photons; forward, central and τ-jets; and muons. The four best candidates of each type are selected and sent to the Global Trigger,

together with the measurement of their position, transverse energy or momentum and a quality word. The Global Trigger also receives the total and missing transverse energy measurement from the Calorimeter Trigger. The Global Trigger selects the events according to programmable trigger conditions, that can include requirements on the presence of several different objects with energies or momenta above predefined thresholds. Topological conditions and correlations between objects can be required as well.

2.6.2 High Level Trigger

The High-Level trigger [51] selection (HLT) is realized with a software running on a farm of commercial processors running in parallel. The goal of HLT is to reduce the Level-1 output rate to 100 Hz mass storage with dedicated fast algorithms. The Level-1 measurement of jets, leptons and photons are refined through intermediate steps, divided into logical levels (Level-2, Level-2.5, Level-3) with somewhat arbitrary classifications depending on the peculiar algorithms of each subdetector. The pixel hits are available together with tracker signals after zero-suppression, hence primary vertex reconstruction and track finding should be possible and algorithms similar to the offline ones will run online too. The output rate of 100 Hz data to be stored on disks is subdivided into different topologies, listed in table 2.6.2 for the initial period at low luminosity.

The HLT system will receive, on average, an event every 10 μs from Level-1 selection and has to reduce by a factor 1000 the amount of data. Each recorded event has an average size of about 1 MB. The DAQ system must provide the means to feed data from the front-ends to the PC farm at a sustained bandwidth up to 100 kHz$\times$1 MB=100 GB s^{-1}. The time spent on average to process one event passing Level-1 trigger during HLT is roughly 300 ms on a Intel Pentium III processor. Assuming a total time of 20 hours data taking per day, a total disk space of 10 TB per day will be stored at full luminosity for subsequent off-line analysis.

2.7 Status of CMS Construction

The CMS construction is now in its final phase: the various modules of the different subdetectors are in production. Afterwards, an integration phase is planned where different parts from all the subdetectors will be assembled together before the final commissioning of CMS, planned for the beginning of the year 2007. For what concerns the Magnet, the Coil Barrel modules (CB-1 and CB-2) are delivered to CERN. From November 2004 it will be possible to start the cooling tests and further Magnet tests are planned until September 2005. The construction of

Trigger	Threshold [GeV]	Expected Rate [Hz]	Cumulative Rate [Hz]
Inclusive e	29	33	33
ee	17	1	34
Inclusive γ	80	4	38
$\gamma\gamma$	$40\otimes25$	5	43
Inclusive μ	19	25	68
$\mu\mu$	7	4	72
Inclusive τ-jets	86	3	75
Two τ-jets	59	1	76
1 jet and E_T^{miss}	$180\otimes123$	5	81
1 jet or 3 jet or 4 jet	657,247,113	9	89
e and jet	$19\otimes52$	1	90
Inclusive b-jets	237	5	95
Calibration and other events (10%)		10	105
Total			105

Table 2.1: *High-Level trigger table at low luminosity. The thresholds correspond to the values of E_T or p_T with 95% efficiency (90% efficiency for muons) [51].*

ECAL Barrel is ongoing with the serial assembly of supermodules. For what concerns HCAL, the majority of HF (Hadronic Forward) wedges have been completed and a sample was tested in 2003. The HE (Hadronic Endcap) plus absorber was completed and megatiles have been installed in it. The HO (Hadronic Outer) trays have been delivered and tested at CERN and HB (Hadronic Barrel) was completed and assembled. More than 50% of the DT and RPC chambers, needed for the muon system, have been produced and the full lot will be completed at middle 2005. CSC chambers are on schedule; about 303 are done and 284 chambers are at CERN on a total of 482. Up to now 90 chambers are installed on CMS and 63 of them are under test.

Chapter 3

The CMS Electromagnetic Calorimeter

The benchmark physics channel which has been used to set the requests on the performance of the CMS electromagnetic calorimeter is $H \rightarrow \gamma\gamma$. The signal significance for this decay mode is critically dependent on the di-photon mass resolution of the calorimeter, and therefore on the energy resolution of the ECAL for single photons.

The di-photon mass resolution can be expressed as

$$\frac{\sigma_M}{M} = \frac{1}{2}\left(\frac{\sigma_{E_1}}{E_1} \oplus \frac{\sigma_{E_2}}{E_2} \oplus \frac{\sigma_\theta}{\tan(\theta/2)}\right) \tag{3.1}$$

where $\oplus$ denotes addition in quadrature, and θ is the angular separation between the two photons measured in radians, hence excellent energy and angular resolution are required from the electromagnetic calorimeter.

A well designed homogeneous crystal calorimeter guarantees a very high energy resolution. Because of the high crossing rate and the high luminosity further requirements for the calorimeter are a very fast detector response, high granularity and radiation hardness in order to withstand the hazardous environment.

The above considerations have led the CMS collaboration to opt for a hermetic homogeneous electromagnetic calorimeter built out of 75848 Lead Tungstate ($PbWO_4$) crystals.

In this chapter a detailed description of the design and capabilities of the CMS electromagnetic calorimeter is given. A particular attention will be devoted to the calorimeter calibration, one of the main factors that will set the limit on the ultimate ECAL performances.

3.1 $PbWO_4$ Crystals

In table 3.1 the main $PbWO_4$ properties are listed and compared to other scintillators.

	$PbWO_4$	NaI(Tl)	BGO
Density [g/cm^3]	8.28	3.67	7.13
Radiation length [cm]	0.89	2.59	1.12
Molière radius [cm]	2.2	4.5	2.4
Emission peak [nm]	440	410	480
dLY/dT$\times$1/LY a T=20 ^{o}C [%/^{o}C]	-2	~ 0	-1.6
LY relative to NaI(Tl) [%]	1.3	100	15
Light decay time [ns]	5-15	250	300

Table 3.1: Some $PbWO_4$ properties compared to other scintillators (LY indicates light yield).

The choice of $PbWO_4$ was motivated by its short radiation length X_0 (0.89 cm) and its small Molière radius (2.19 cm). This allows the calorimeter to be sufficiently compact to be inserted within the magnet coil and sufficiently fine-grained to have a good $\pi^0 - \gamma$ separation, to reduce the *Pile-up* effects and to obtain a good angular resolution.
Other important aspects are that it is a radiation resistant and chemically inert scintillator, suited to work in the LHC high dose environment (from 0.18 Gy/h at $|\eta|$ = 0 to 6.5 Gy/h at $|\eta|$ = 2.6 at high luminosity), and it also has a short scintillation decay time τ = 10 ns that allows to collect 85% of the light in the 25 ns interval between two pp collisions.

The main drawback of the $PbWO_4$ choice is the low light yield, about two order of magnitude less than NaI. This is due to the $PbWO_4$ scintillation mechanism, which is strongly limited by the thermal quenching; this same mechanism explains the short decay time and the strong light yield dependence on the temperature ($-2\%/^{o}$C at room temperature). To overcome these difficulties a new photodetector has been developed, the Avalanche Photo Diode (APD) and a complex cooling system has been designed in order to stabilize the system temperature. Both of them will be described later in the text.

The $PbWO_4$ emission spectrum is represented in fig. 3.1 [52]; a peak is found in the region around 420 nm.

An intensive R&D has been done also to improve the $PbWO_4$ crystal radiation resistance, one of the most critical issue at LHC. The environment for the CMS

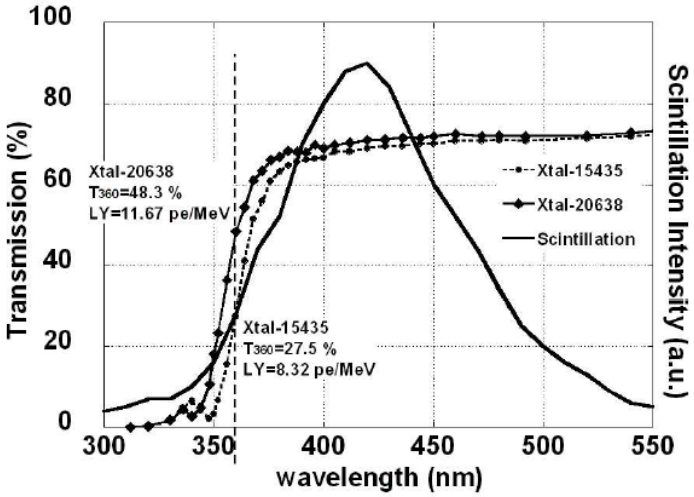


Figure 3.1: $PbWO_4$ *emission spectrum and transmission [52].*

ECAL is very hostile in this respect, since at high luminosity one expects maximum dose rates for electromagnetic radiation of around 0.25 Gy/h in the ECAL barrel, values of around 0.30 Gy/h at the barrel-EE interface and 15 Gy/h at $\eta = 3.0$. The neutron flux in the EB calorimeter is expected to reach 2×10^{13} n/cm^2 in 10 years of running at LHC, reaching in the EE $\sim 10^{15}$ n/cm^2. However, no damage by neutron irradiation has been seen so far, while crystal transparency is affected by ionizing radiation. It has been shown that the scintillation mechanism is not damaged, nor is the scintillation emission spectrum changed [3]. The small decrease in light output is due to radiation-induced absorption, i.e. to the formation of color centers, which reduce the crystal transparency. Stochiometric fine tuning, doping and optimized growth conditions have improved the radiation hardness and the light transmission of the crystals. Production crystals show a light yield loss which is, at saturation, on average less than 3 % if irradiated with γs at 0.15 Gy/h.

Because the transparency only is affected, it is possible to monitor the loss in transmission due to irradiation using a light injection system in the calorimeter, and to apply corrections for it. This is the fundamental reason for the implementation of the transparency monitoring system which will described later in this chapter.

The R&D has also succeeded in reducing light collection non-uniformity (by depolishing and shading) under the limit of $0.35\%/X_0$ (in section 3.5 the implications of a non-uniform light yield will be discussed).

3.2 Mechanical Structure

In the CMS electromagnetic calorimeter, the support structure of the crystals will be attached to the surrounding hadron calorimeter and the weight of the detector is transmitted to the hadron calorimeter front face. Therefore, all the heavy support structure and cabling lies behind the crystals and only a light-weight alveolar structure holds the crystals and readout electronics.

The crystals are arranged in barrel and endcaps. A three-dimensional view of the calorimeter is shown in figure 3.2.

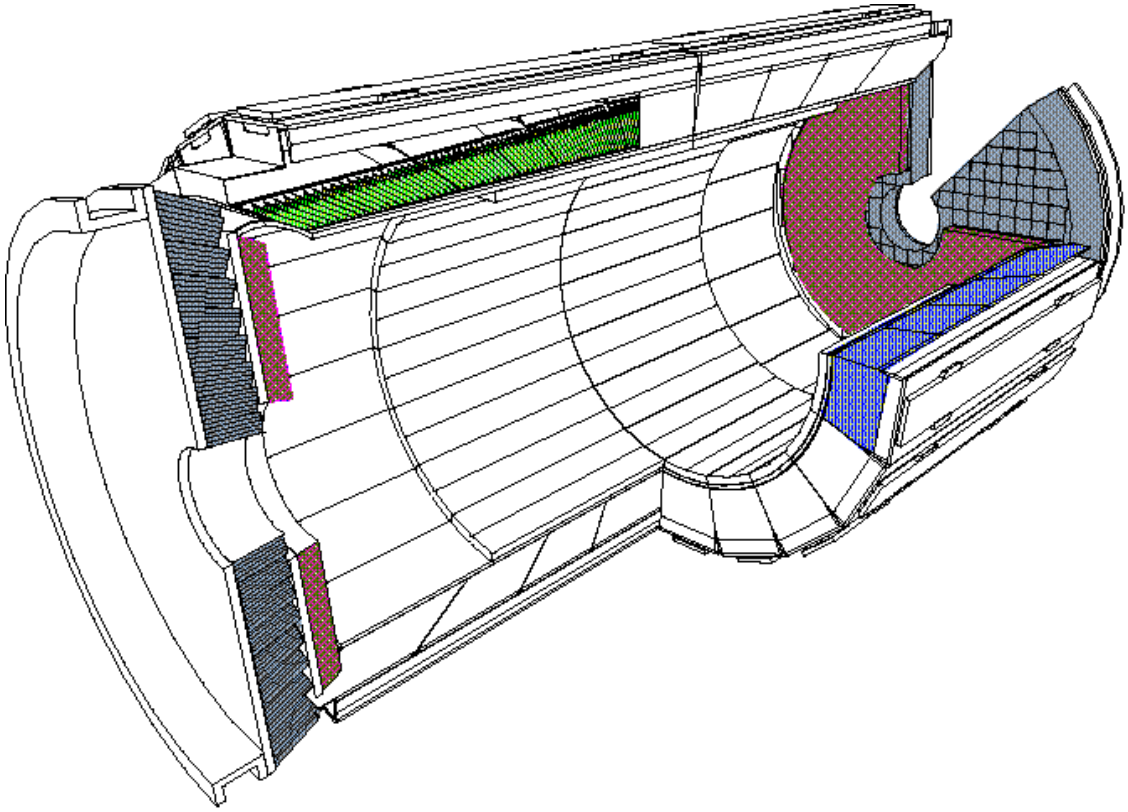


Figure 3.2: *A three-dimensional view of the calorimeter.*

The barrel consists of 2×18 supermodules each of which contains 20 crystals in ϕ and 85 crystals in η. In both detector halves in η the supermodules are divided in 4 support structures called modules. There are 17 different crystal types in the barrel: each crystal has a length of 230 mm (25.8 X0) and a front face area close to 22×22 mm^2, exact dimensions depending on crystal type, giving a granularity of $\Delta\eta \times \Delta\Phi = 0.0175 \times 0.0175$. The crystals are tapered and their axes have a

constant off-pointing angle of 3^o with respect to the nominal vertex position in both η and ϕ to avoid gaps that are pointing to the interaction region, increasing the hermeticity of the structure.

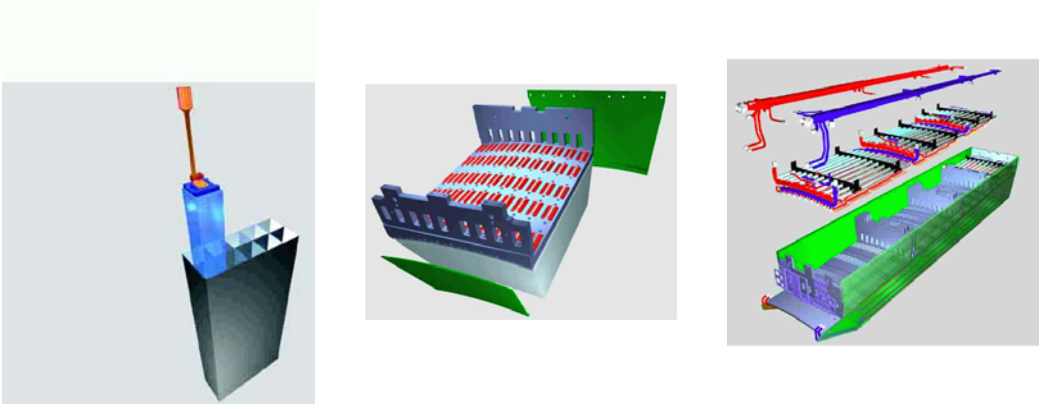


Figure 3.3: *Stages of an ECAL super-module assembly. (left) Alveolar submodule during insertion of the first crystal; (center) Complete module; (right) Expanded view of a super-module with cooling circuits.*

The barrel coverage extends over the pseudorapidity range $\eta < 1.479$. The endcap is arranged in two halves (Dees). The ECAL endcaps are built up of structures composed of identical 5×5 crystals, called super-crystals, arranged in array in (x-y) direction as shown in figure 3.4. Endcap crystals are 220 mm long (owing to the presence of the $3X_0$ thick pre-shower detector in front) and with 24.7×24.7 mm^2 front face dimensions: they are slightly larger in cross-section, shorter, and almost parallelepipedic, compared to the barrel crystals. The front face of the endcap shall be at a distance of 3170 mm from the interaction point. To ensure a hermetic design, the crystal axis will be oriented to a point 1300 mm beyond the interaction point, off-pointing to a similar extent as the barrel crystals. The endcap crystals are supported by a back plate which also provides a shield to the electronics equipment which is located behind the back plate. The endcaps are bolted to the hadron calorimeter endcap front face.

3.3 Photodetectors

Because of the relatively low light yield of $PbWO_4$, photo-detectors with intrinsic amplification are needed. In addition, they have to be fast, radiation hard and able to operate in a strong magnetic field. For the barrel these conditions are met by

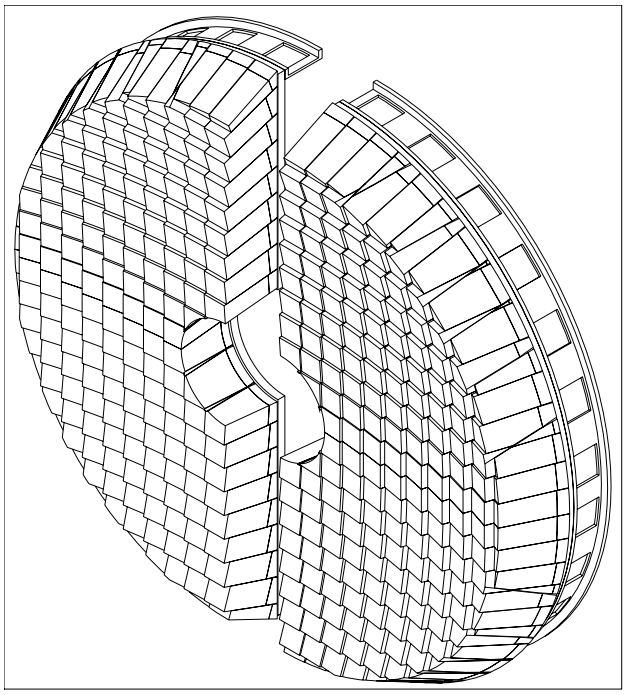


Figure 3.4: *Mechanical structure of the ECAL Endcap.*

avalanche photodiodes (APDs), an outcome of a 10-year R&D effort between the CMS collaboration and Hamamatsu Photonics. In the ECAL endcaps the radiation levels can be order of magnitudes higher than in the barrel. In particular, the expected neutron flux would be too large for APDs. Therefore the CMS collaboration has chosen to use vacuum phototriodes (VPTs) in this region, profiting from favourable orientation of the magnetic field ($8.5^o < \theta < 25.5^o$) between the field and the photodetector axis. A brief description of these photodetectors will follow in next two sections.

3.3.1 APD

The Avalanche Photo Diode principle of operation is shown in figure 3.5. The primary photon generates a photoelectron which is then accelerated in an electric field where it generates an avalanche of electrons. Table 3.2 summarises the parameters of the APDs as developed. Two APDs are integrated and connected in parallel within a capsule (3.5), which provides an active area per crystal of 2×25 mm^2.

This photo-detection devices has an impact on the the calorimeter perfor-

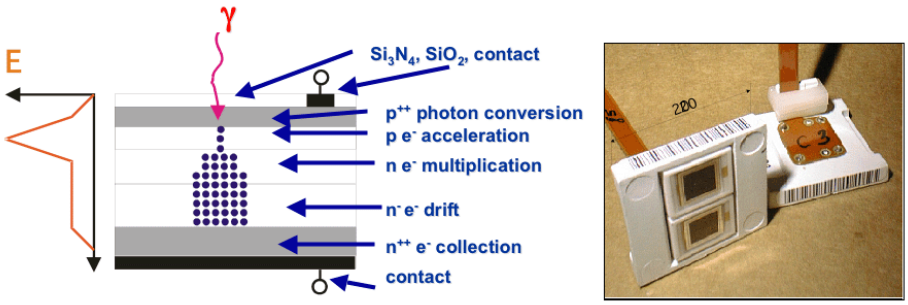


Figure 3.5: *Left : Principle of operation of an APD. Right : Two APDs per capsule.*

mances by contributing to all the terms of the energy resolution. The effect on the stochastic energy resolution term can be quantified using the excess-noise factor, which is due to the probabilistic nature of multiplication process. In an ideal gain mechanism, the excess-noise factor equals 1. Another issue is represented by the voltage and temperature sensitivity of the gain. Third, the calorimeter noise is affected by the capacitance of the APD and leakage currents flowing on the surface of the device (I_s) or, more important, flowing through the device ("bulk current" I_b).

Radiation may cause damage which increases the leakage currents. The main damage comes from the neutrons which create defects in the silicon lattice. Gain seems to be not affected up to absorbed fluxes of 10^{13} n/cm^2, while the leakage current increases linearly with the flux up to values of ~ 10 μA for the maximum dose expected.

Sensitivity to traversing radiation is also an issue: in the APD this effect is reduced by the amplification that occurs just behind the photo-conversion layer, meaning that only the energy deposited by particles in this thin 5 μm layer is amplified, while signal from ionising particles traversing the bulk of the silicon is not amplified.

More detailed description of the characteristics can be found in [54]

3.3.2 VPT

A vacuum phototriode (VPT) is a device capable of working in an axial or quasi-axial magnetic field. A typical VPT is illustrated in figure 3.6.

Photoelectrons are generated on the photocathode and part of them passes through the anode mesh and impacts on the dynode where the secondary electron cascade arises. The secondary electrons are attracted back to the anode mesh which captures a large fraction of them. The quantum efficiency of these devices for the PWO peak wavelength is 15% and their sensitive area of roughly 300

Constructor	Hamamatsu
Active Area	25 mm^2
Quantum efficiency	75% (at 430nm)
Capacitance	80 pF
Excess Noise Factor (F) @ G=50	2.0
Operating Voltage @ G=50	$\sim$380 V
Initial Dark Current @ G=50	< 50 nA (<10 nA typical)
Gain (G)	50 (max >1000)
Voltage Sensitivity dG/dV$\times$1/G @ G=50	3.15 (%/V)
Temperature Sensitivity dG/dT$\times$1/G @ G=50	-2.2 (%/^{o}C)

Table 3.2: *Prinicipal properties of Hamamatsu APD. G stands for gain.*

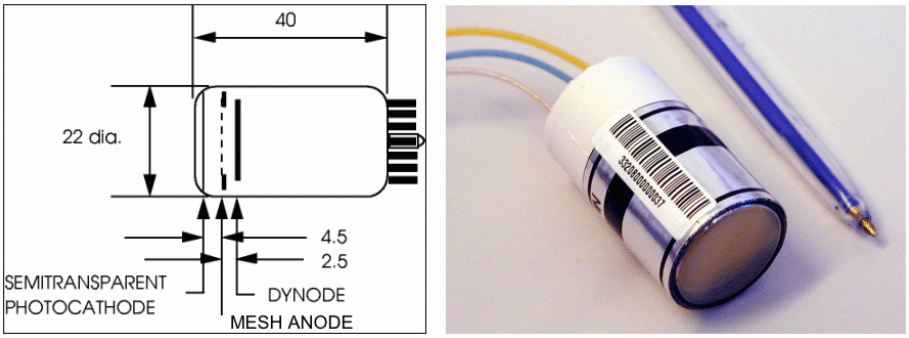


Figure 3.6: *Schematics and view of a VPT.*

mm^2. Thus the total light collection efficiency is of the same order of the APDs. The gain of the VPTs in a 4 T magnetic field is of order of 8-10 compared to 50 of APDs, however they are less sensitive to the temperature (<1%/^{o}C) and they are very insensitive to the bias voltage variation (< 0.1%/V). VPTs have a low capacitance (few pF) and a leakage current <2 nA. The excess noise factor of the currently available VPTs is 2.5-3.0. Particular attention is posed in order to obtain photodetectors with a uniform response for different rates and radiation conditions: it is shown that anode response and the excess noise factor of VPT are independent of radiation dose and neutron fluence [55].

3.4 Electronics chain

The objectives for the front-end (FE) electronics are extremely challenging. The electronics must be extremely fast to match the 25 ns LHC crossing rate, and it

must provide a very precise energy measurement on a dynamic range as large as ∼95 dB (50 MeV - 1.5 TeV). The noise has to be kept below ∼50 MeV per crystal in the barrel region. Since it is placed on the detector, the FE electronics must be radiation-hard and reliable.

An ECAL electronics system design review held during the first quarter of 2002 concluded that some of the criteria which were originally driving the system layout [3] were not valid anymore. In particular, there was the possibility to profit from a low cost radiation hard 0.25μm CMOS technology, allowing the use of ASICs (Application Specific Integrated Circuit) inside the detector volume. This allows to move some tasks like the Trigger Primitives Generation (TPG) from the off-detector electronics onto the detector, decreasing the number of data links and the quantity of off-detector electronics by about a factor of eight (the number of datalinks reduced from 91000 to ∼ 12500 while the number of off detector boards diminished from 800 to ∼220). This concept is not new and had been considered earlier in the development of the ECAL project [44] but it was not adopted at the as a baseline due to a lack of proven, low cost radiation-hard electronics technology.

The new electronics scheme is presented in figure 3.7 [56].

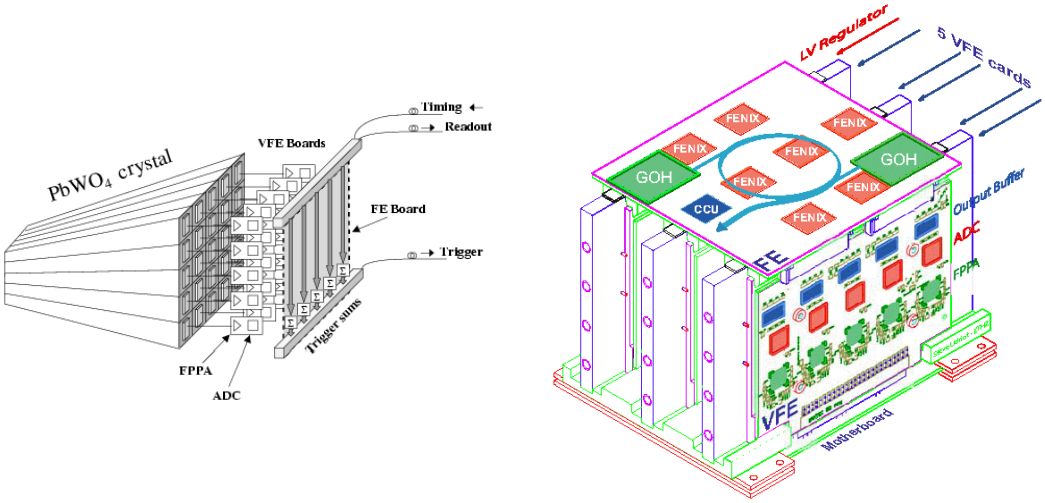


Figure 3.7: *Left : Schematic view of the ECAL readout electronics for 5×5 crystals. Right : The arrangement of the front-end electronics into VFE and FE boards.*

The crystals are organized in matrices of 5×5 crystals forming a so called trigger tower. The small electrical signals coming from each photodetector are sent to a motherboard, housing five Very Front End (VFE) cards and a Low Voltage Regulator (LVR) card [57], which distributes regulated voltages to the VFE cards.

The motherboard is also used to filter and distribute high voltage to the photodectors. The output of the 5 VFE cards is feed into the Front End (FE) card which processes the digitized data of one trigger tower.
Each VFE cards houses five identical electronics channels. Also the scheme of the VFE card was subject of a major redesign. In the previous version the signals from the photodetectors were pre-amplified and shaped by a trans-impedance amplifier with internal shaping, followed by a four-range amplification and logic stage in order to adapt the output to the dynamic range of a 12-bit ADC from Analog Devices. This functionality was integrated into a unique radiation-hard integrated circuit named FPPA (Floating Point Pre-Amplifier). An important noise due to parasitic resistance and capacitance was affecting the first version of the FPPA. The problem was partially corrected in the last iteration but, at that time, it was considered safer to develop an alternative to FPPA: the MGPA (MultiGain PreAmplifier) [58] based on the 0.25μm CMOS technology, which provides three analogue output signals at three different gains. These three signals are digitized in parallel by a new four-channel, 12-bit 40 MHz ADC [59] with an integrated digital gain switching logic, developed in 0.25μm technology. Such a solution would also simplify the low voltage distribution, requiring only two instead of four voltages.

In the current design each channel of the VFE card consists of 3 different ASIC's, a multi gain pre amplifier (MGPA), an ADC and a buffer, all designed in radiation tolerant 0.25μm technology. In addition, in each VFE card, a Detector Control Unit (DCU) is implemented, measuring the APD leakage currents and the crystal temperature.

The FE board stores and processes the digitized data from VFE cards during the Level 1 trigger latency of 3 μsec. It utilizes 7 radiation tolerant FENIX ASIC's performing three main tasks: the data from a row of five crystals are stored until arrival of a L1 accept, the pulse shape data from a row of five crystals are corrected and summed to produce a strip energy. Finally in the case of the ECAL barrel, the sum of all strip energies is computed to produce the energy deposited in the trigger tower. The trigger data are transmitted to the off-detector electronics through a serial digital data link consisting of a GOL (Gigabit Optical Link) serializer chip, a link system based on the CMS tracker link technology, a digital receiver module and a serializer. The link is operated at 800 Mb/s, which is sufficient to complete the transfer of trigger data every 25 ns. On receipt of a L1 accept, the data from the triggered event, stored in the memory of the five FENIX chips, are transmitted to the off-detector electronics through a separate serial data link also operated at 800 Mb/s. For every L1 accept, 10 time-slices of data are transmitted off the detector in 7.5 μs. This system uses a digital optical link system, controlled by the off-detector front-end control boards.

3.5 Energy Resolution

The energy resolution of a homogenous calorimeter can be written as:

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c, \tag{3.2}$$

where E is the energy and a, b, c respectively represent the *stochastic*, *noise* and *constant* term of the energy resolution. This section will be devoted to analyze in detail the different contributions to each term.

The stochastic term a is associated to statistical Poisson-like fluctuations in the number of the primary processes that generate the signal in the measuring volume, hence the fluctuations of the light production and collection. These are poissonian contributions, so the stochastic term a is:

$$a = \frac{1}{\sqrt{N_{pe}}}, \tag{3.3}$$

where N_{pe} is the number of photoelectrons emitted per energy unit, which are equal to

$$N_{pe} = LY_{scint} \cdot \varepsilon_{lc} \cdot \varepsilon_{geom} \cdot \varepsilon_{pc}. \tag{3.4}$$

LY_{scint} is the amount of photons produced by the scintillator per energy unit, and the product $\varepsilon_{lc} \cdot \varepsilon_{geom} \cdot \varepsilon_{pc}$ measures the light collection efficiency, taking into account the fraction of light produced which escapes the rear face of the scintillator ε_{lc} and therefore can be detected, the ratio of the photodector active area over the rear face surface ε_{geom} and the photodector quantum efficiency ε_{qe}. There can be other contributions from the multiplication process inside the photodetector: these contributions can taken into account using the *excess noise factor*, F, already defined in section 3.3.1:

$$a = \sqrt{\frac{F}{N_{pe}}}. \tag{3.5}$$

Lateral shower fluctuactions and material in front of the calorimeter also contribute to a. In the CMS ECAL TDR [3] a has been evaluated to be 2.7%, estimating the energy in a 5×5 crystal matrix.

Noise term b includes the electronics noise and the physical noise due to energy released by particles coming from *Pile-up* events. Electronics noise can be due both to the photodetector and to the preamplifier. The photodetector contributes basically via two components: one is proportional to its capacitance, the other is connected to the fluctuactions of the leakage current; the latter accounts to a contribution of 30 MeV and 110 MeV respectively for low and high luminosity

scenarios. Preamplifier noise instead can be parametrized as

$$\sigma_{noise} = \frac{ENC}{e \cdot N_{pe} \cdot G}, \tag{3.6}$$

where e is the electron charge, G is the photodetector gain while ENC is the noise equivalent charge, depending on the input capacitance, preamplifier shaping time and gain; for a single crystal this term is estimated to be around 30-40 MeV. This value is particularly high if compared to other electromagnetic calorimeters (due to the short shaping time), but it is comparable with the physical noise from pile-up events (estimated to be 30 MeV and 95 MeV for low and high luminosity regime).

Constant term c is the dominant one at high energies. Several effects are contributing to it, so the main challenge will be to master all the different contributions in order to maintain the constant term below the design goal 0.5%. The different contributions to the constant term are listed below:

- Longitudinal non-uniformity of the crystal light yield. Because of the tronco-pyramidal shape of the crystals, a strong focusing effect of the light takes place and so the crystal light collection efficieny is not uniform. In order to have a uniform light yield, one of the lateral faces of the crystals is depolished by the producer. The non-uniformity contribution has been studied to be less than 0.3% if the non-uniformity slope in the shower maximum region is required to be less than 0.35%/X_0 [60].

- Temperature stability. Both scintillation mechanism and APD gain are particularly sensitive to the temperature. A temperature stability within 18.00 $\pm$ 0.05 ^{o}C should be achieved by the cooling system in order to mantain the temperature stability contribution to c below 0.1%. Temperature measurements in the testbeam for a period longer than two months show that this limit can be reached

- High voltage stability. APDs gain strongly depends on bias voltage. The power supply system has to ensure stability better than 30 mV (including long term stability, regulation, noise, ripple and reproducibility) in order to give a contribution to the constant term smaller than 0.1%. During the 2002 test-beam the operation of the HV system was very successfull and the stability of the bias voltage was within the specifications. [61]

- Longitudinal containment. Test beam and simulations showed that this contribution to the constant term is below 0.2%.

- Intercalibration errors. Crystal intercalibration precision goes directly into the constant term with very little scaling, due to the fact that most of the energy of an electromagnetic cluster is contained within a single crystal (around 80%). Therefore crystals' calibration will represent a key aspect in order to reach the design performance of the electromagnetic calorimeter.

In table 3.3 the design goal for the different contributions to the barrel energy resolution is reported, estimating the energy with a 5×5 crystal matrix. The values are extracted from the ECAL TDR [3]. A comparison of the the contributions estimated with a detailed MonteCarlo simulation and real test beam data will be presentend in the section 4.7.3.

Stochastic term *a*	
Shower containment	1.5%
Photostatistics	2.3%
Total	2.7%
Noise term *b*	
Preamplifier	150 MeV
Leakage current (low L)	30 MeV
Pile-up (low L)	30 MeV
Total Noise (low L)	155 MeV
Leakage current (high L)	110 MeV
Pile-up (high L)	95 MeV
Total Noise (high L)	210 MeV
Constant term *c*	
Shower containment	0.16%
LY Non-uniformity	0.3%
Intercalibration	0.4%
Total	0.5%

Table 3.3: *Design goal for the different energy resolution contributions (5×5 crystal matrix).*

3.6 Calibration procedure

Careful monitoring and calibration of the ECAL will be necessary to achieve the target constant term of 0.5%.

Prior to installation in CMS, it is foreseen that at least some supermodules (each supermodule=1700 crystals) will be precalibrated in an electron beam at CERN. This will allow verification and correction of Monte-Carlo models, and provide an initial set of crystal calibration coefficients at startup. Initially it was foreseen to calibrate all crystals of the 36 super-modules using a high energy electron beam with an expected accuracy of 0.5-1%. Due to the tight CMS schedule only few super-modules will be calibrated in this way. For the ECAL crystals that will not be calibrated with the beam, it has been suggested to use the laboratory measurements performed in the construction phase [62]. It is very important to give an estimate of the intercalibration precision that can be reached using only laboratory measurements. In the next chapter it will be demonstrated that a precision better than 4.5% can be reached combining all available laboratory measurements used in production qualification; a brief description of the qualification measurements is given later in the text (section 3.7.

When installed in CMS, it will be necessary to perform continuous calibration *in situ* using physics events. The *in situ* calibration process has two distinct parts, *inter*calibration, and *global* calibration: Intercalibration relates the response of different crystals to the same energy deposition, but does not calibrate the absolute response of the crystals, which is the task of the global calibration. The intercalibration process can be further broken down into the following steps:

- Intercalibration of rings of crystals in the same η interval. Intercalibration in ϕ relies on the (assumed) azimuthal symmetry of crystal activity. It is expected [63] to be possible to perform this intercalibration very quickly at startup.

- Intercalibration of ϕ rings at different η using $Z \rightarrow e^+e^-$ decays. Fitting the Z mass will also calibrate the absolute energy scale of the ECAL.

- Regional intercalibration, which will intercalibrate crystals in $\eta - \phi$ regions, using isolated electrons mainly from $W \rightarrow e\nu$.

Another aspect which should be considered is the fact that radiation damage reduces the transparency of the crystals. Because of this effect, the response of the crystals must be constantly monitored, and a laser monitoring system will be used to track the evolution of each channel in the time between two successive calibrations with physics events.

3.6.1 Monitoring System

The light monitoring system will be crucially important for tracking changes in crystal transparency. It will inject laser light into the ECAL crystals regularly during normal running of the machine, and during shutdown periods in order to monitor crystal recovery. It distributes laser light of three frequencies, blue, green, red [64], through a two-level fanout system with cross-verification by precision radiation-hard PN diodes at each level, and eventually to each crystal via a radiation-hard optical fiber. The logical regional unit for intercalibration is a group of crystals all sharing the same reference photodiodes (typically 400 or 500 crystals in the barrel). A diagram of the fanout and PN diode configuration in one barrel module is shown in figure 3.8.

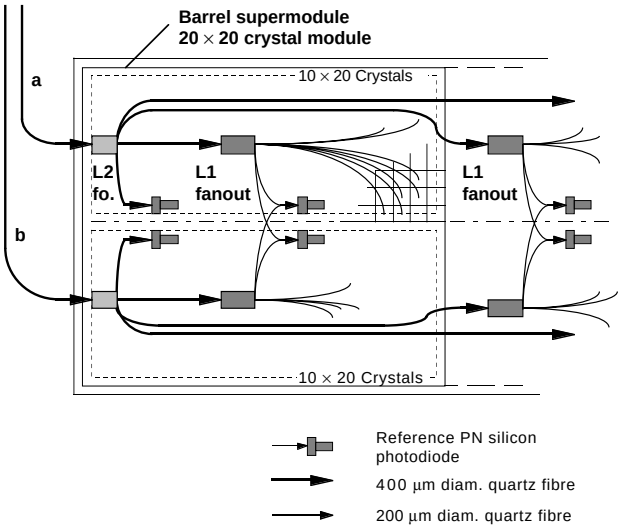


Figure 3.8: *Light distribution system for a single 20 × 20 crystal module. Monitoring light is injected into either fibre (a) or fibre (b) during a monitoring cycle.*

Under irradiation the losses of the signal from the laser light (ΔL) and of the signal from beam particles (ΔB) are related by [65]:

$$\frac{\Delta B}{B} = \left(\frac{\Delta L}{L}\right)^{\alpha}. \tag{3.7}$$

The principle of the laser monitoring is based on the universality of α: α should be independent of time and the same for all the crystals. As a first approximation,

the energy resolution contribution from the dispersion of α is:

$$\frac{\sigma(E)}{E} = \frac{\Delta L}{L}\sigma(\alpha) \tag{3.8}$$

The complete monitoring system has been tested during the 2002 test beam on a calorimeter module with 100 channels equipped with the FPPA FE electronics. A stability of 0.1% was achieved by the laser system. Some promising results concerning the universality of α, altough obtained on a limited statistics, are presented in [66]. Given the expected transparecy loss (see section 3.1), the observed dispersion of 6% on α is compatible with the limit fixed at 0.3% to the contribution of the laser monitoring system to the constant term of the energy resolution.

3.7 Production qualification

The CMS electromagnetic calorimeter will be composed of many parts, each of them having an impact on its final performances. Each part should be carefully tested to verify that its quality would respect the design calorimeter performances. In the following the qualification processes of crystal, APD and electronics will be briefly discussed.

3.7.1 Crystal qualification

All ECAL crystals are characterized in two regional Centres: CERN and Rome (INFN-ENEA at Casaccia) before assembly. The crystal characterization consists of the measurement of the longitudinal (LT) and transversal (TT) optical transmission, the scintillation light yield (LY) and its uniformity and the dimensions. These measurements are performed in an automatic way with special machines called ACCOS (Automatic Crystal Control System) [67],[68]. Cross calibration of the Cern and Rome machines is described in [69]. The two Centers are dedicated to the assembly and test of different modules corresponding to different sets of crystal shapes: modules of so-called type 2 and 3 (Rome) and modules of type 1 and 4 (CERN). The LY is measured using different techniques described in [70], [71]. A radioactive source along the crystal length in steps of 1 cm with a step motor, allowing measurements of the light collection uniformity. The $PbWO_4$ light is collected by a photomultiplier (PMT).
Crystals are accepted if their LY is greater than 8 photoelectrons per MeV (pe/MeV). This limit is set for crystals measured in a standard configuration to which the automatic machines are calibrated.

The LY versus position measurements are fitted with two straight lines from 3.5 to 11.5 cm from the crystal front and from 11.5 to 19.5 cm. The slope of these fits in units of radiation lengths (X_0) corresponds to the so-called Front (FNUF) and Rear (RNUF) Non-Uniformity. From the fit results the LY at 8 X_0 (distance corresponding to the shower-maximum for a 50 GeV electrons) is also deduced. This is what is usually referred to as "the LY" of the crystal. The crystals for which the FNUF is found outside the specifications ($\pm 0.35\%/X_0$) are further depolished at CERN.

The longitudinal transmission (LT) and transversal transmission (TT) are measured using special spectrophotometers [70], [68]. This measurement is particularly important since transmission is correlated to the radiation hardness of the crystal; but as it will be shown in section 4.4 it is also correlated to the LY of the crystal, giving the possibility to improve the direct LY estimate.

3.7.2 APD qualification

Since the capsules, once mounted, will be inaccessible for the whole life-time of the experiment, very strict Q&A measures are being applied in order to guarantee an APD failure rate below 1/1000 in whole experiment life. All APDs are subjected to a screening programme which consists in a 500Krad preirradiation with a Co^{60} source and a subsequent annealing and accelerated ageing under bias. Only those devices are retained which fulfil certain pre-defined selection criteria for critical parameters, such as the dark current [54].

3.7.3 Electronics qualification

All the FE electronics components will be characterized before installing them. Several tests will be performed, in particular noise and calibration measurements of the preamplifier and of the ADC [56].

Chapter 4

The 2003 Electromagnetic Calorimeter Test Beam

In the year 2003 a very extensive test beam programme on a quasi final version of the whole ECAL system has been conducted. Two partially equipped supermodules were tested and exposed to electrons with energies ranging from 20 to 200 GeV. The first super-module (SM0) was equipped with 100 channels of readout electronics using the old FPPA2001 and the AD9042 ADC, including the newly designed final cooling system, the mother-boards, LVR boards and FE boards. It has been studied for about two months in the H4 beam line at CERN. During the last week of the test beam programme, the second super-module (SM1) was tested, equipped with the new electronics. Only in the first period, the amount of data was sufficient to allow the detailed studies presented in this chapter, even if the old version of the electronics was found to be too noisy.

The objectives of the test beam were many. Apart from a full system test, the main interest was in evaluating the calorimeter performances and in demonstrating the capability to carry out a precalibration of a supermodule. The comparison of the intercalibration constants with those obtained from laboratory measurements is another issue of particular importance, since not all the supermodules will be precalibrated and for most of the crystals the only available calibration at the startup will be the one obtained in the laboratory.

The efficiency of the precalibration data collection is an issue. To this aim, an automatic procedure capable of validating quasi on line the data collected has been developed, in order to obtain a prompt identification of potential problems and, in case, take the appropriate actions.

Another aspect which has been considered is the fact that during the year 2003

a major change in the CMS simulation software took place, which now is based on the C++ object oriented version of GEANT (GEometry ANd Tracking), GEANT 4 [4], instead of the FORTRAN based one, GEANT 3 [72]. Since most of the studies on the reconstruction software are based on simulated data, it is extremely important to demonstrate the validity of the simulation comparing it to test beam data.

In this chapter the results in which the author have directly contributed will be described. At first, after illustrating the test beam setup and the crystal signal amplitude reconstruction, the intercalibration procedure and the comparison with the intercalibration obtained with laboratory measurements will be discussed. Secondly, the data validation architecture and the precalibration data validation task will be described. The comparison between test beam data and a GEANT 4 simulation will conclude this chapter.

4.1 Test Beam Setup

The ECAL test beam set-up is installed in the H4 line at CERN. It allows the test of an entire *Supermodule*; the supermodule is fixed onto a rotating table, allowing to set the angle between the incoming beam particles and the crystal axis as if they were originated in the LHC interaction point. As described in the previous chapter, CMS ECAL barrel has a quasi-pointing geometry with respect to the mean position of the primary vertex, with a 3 degrees tilt in both the η and ϕ directions.

Along the beam line, 6 scintillator counters, named S1 to S6, are available with various logic patterns to construct the trigger.

A precise determination of the impact position of the incident beam particle on the crystal surface is obtained with a hodoscope system, which measures the trajectory of the beam particles and allows the extrapolation of the coordinates of the incident point on the crystal front face. The system is composed of 2 fibre planes with orthogonal orientation, separated by 2.5 m, constructed as staggered two layered set of 1mm scintillating fibres read by multi-anode photomultipliers. The resolution attainable on the position of the measured point per plane is around 200 μm and combining the two measurements in the two different orientation 145 μm is reached.

The reference system to which the positions in the rest of the chapter will be referred is described in the following. The z axis is parallel to the beam axis. The zero position along the beam axis is the pivot point, the point used as a reference to align and rotate the support table. The other two coordinates, x and y, which are those measured by the hodoscope for each particle composing the beam, are defined in the plane perpendicular to the beam axis The origin in the transversal

plane is chosen in such a way that it is coincident with the average beam center.

The table is configured to put the center of the beam in the position where a 120 GeV electron releases the largest fraction of its energy for each crystal, the so called "maximum containment point". Its position is different from the center of the crystal front face, due to the 3^o tilt angle between the beam direction and the crystal longitudinal axis. The angles used to rotate the table are calculated from a GEANT 3 simulation of the supermodule.

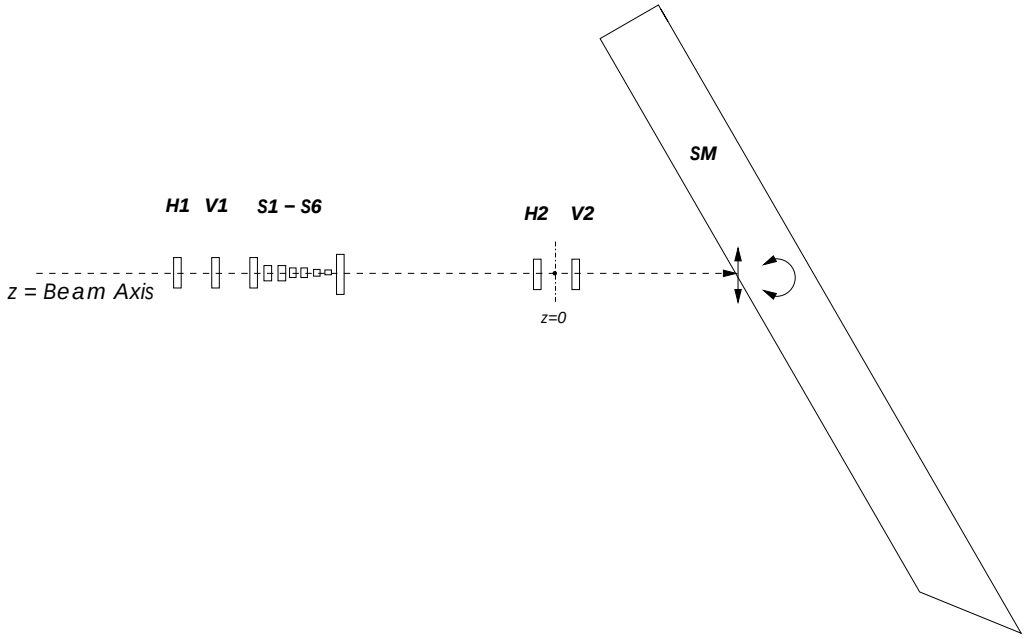


Figure 4.1: *Drawing of the beam line indicating the relative position of the four scintillating fibre planes (H1 H2 V1 V2) composing the hodoscope system, of the six scintillator counters (S1-S6) used as triggers and of the supermodule (SM) mounted onto the rotaing table. The zero position along the beam axis is the pivot point, the point used as a reference to align and rotate the support table.*

The main differences between this setup and the final CMS one are:

- the lack of material in front of the crystals
- the absence of a magnetic field.

The map of the crystals tested in the first period of data taking along with their numbers is given in figure 4.2. Two different barrel crystal types were tested: type 15 and type 16 contained in module of type 4 placed at highest η in the supermodule.

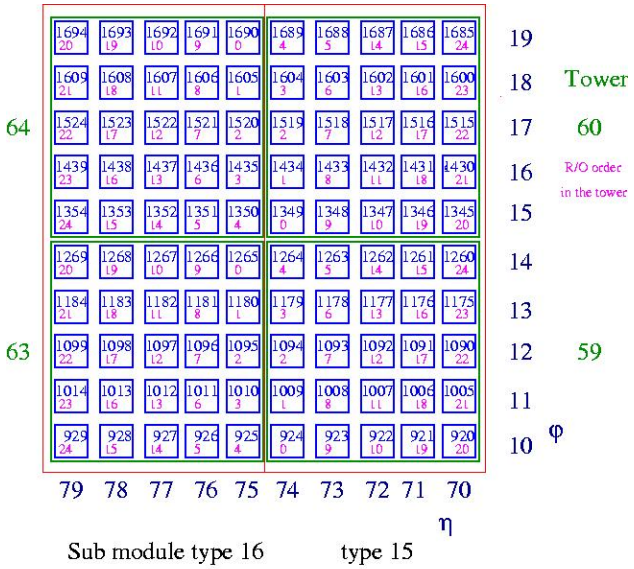


Figure 4.2: *Numbering of the 100 crystals tested during the 2003 summer ECAL test beam campaign. For each crystal, it is displayed its representative number (0-1699), its position inside the trigger tower (0-24), and an index identifying its* η *(0-84) and* ϕ *(0-19) position.*

4.2 Amplitude reconstruction

In this section the method that is used to reconstruct the amplitude of the signal due to the energy deposited in each crystal is described. This is an issue of particular importance, since this is the first step to obtain an unbiased and precise estimate of the energy contained in a single crystal over the full energy range.

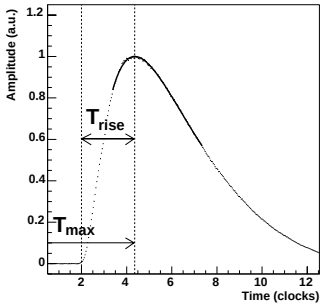
The ADC digitize the signal at 40 Mhz, storing the digitized value of 14 time samples for each clock (every 25 ns). The algorithm used to reconstruct the amplitude from the individual ADC samples is based on the technique of digital filtering. The amplitude is obtained from a linear weighting of the individual time samples:

$$\tilde{A} = \sum_i w_i \, S_i \ , \tag{4.1}$$

where $\tilde{A}$ is the estimate for the amplitude, the S_i are the individual digitized samples. The set of optimal weights is obtained by minimizing the expected variance of the reconstructed amplitude.

In the 2003 test beam, when calculating the optimal weights, an analytic description of the pulse shape is used. The functional form used is:

$$\left(\frac{t - T_{\max}}{T_{\rm rise}}\right)^{\alpha} e^{-\alpha\frac{(t-T_{\max})-T_{\rm rise}}{T_{\rm rise}}} \quad \text{if t} > (\mathrm{T_{max}}\text{-}\mathrm{T_{rise}})$$



In this analytic description, $T_{\rm rise}$ and α define the shape of the pulse, while $T_{\max}$ simply defines the absolute position of the peak with respect to the trigger.

When optimizing the weights, another issue specific to the test beam setup should be considered: in CMS final installation the ADC will be clocked by the LHC, keeping the phase between the signal and the clock constant, but in the test

beam, as the signal is asynchronous to the ADC clock, the phase can assume any value in the interval 0-25 ns. In the test beam a TDC is used to measure this phase.

4.3 Crystals intercalibration

The first step in the analysis of test beam data is intercalibration, aiming to determine the relative response of different channels (crystals) to impinging electrons of the same energy. To calibrate the array, data were recorded at two known energies, 50 and 120 GeV, for all the crystals composing the array. In each calibration run 30000 events were recorded per crystal. The procedure adopted here to extract the relative calibration of the single crystal response is based on a method first introduced in the 2000 test beam [74].

The reconstructed amplitude Ã measured in a single crystal varies as a function of the electron impact position (x,y), as measured by the hodoscope. For our purposes the 2D function of response $h(x,y)$ can be adequately approximated by the product of two 1D functions $h(x,y) = f(x)g(y)$. Both $f(x)$ and $g(y)$ can be parametrized as fourth degree polynomial, and are obtained from an iterative fit with a fourth degree polynomial to the response of the crystal. The function $h(x,y)$ is now used to produce an event by event position correction, renormalizing the signal to the magnitude it would have had if it had struck the point of maximum response.

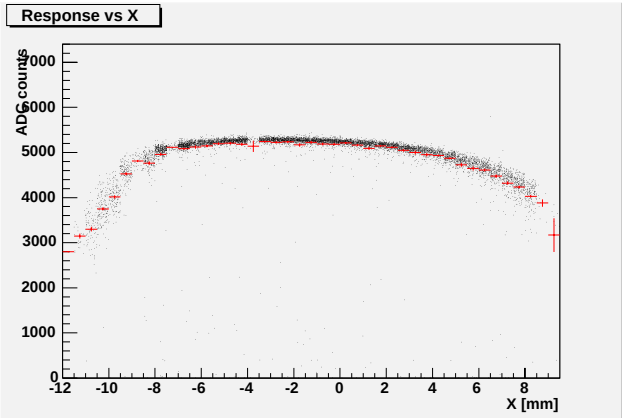


Figure 4.3: *Response of the hit crystal versus the impact point of the electron on the crystal front face.*

Figure 4.4 shows a crystal response to the 120 GeV beam. Notwithstanding

the correction, the response is not gaussian, because of the leakage. The fit is performed with an iterative gaussian fit in an interval [1.5σ,3σ], from which it is possible to extract the $\tilde{A}_i^{max}$, the maximum response for the crystal *i*. Intercalibration for crystal *i* can be obtained relating $\tilde{A}_i^{max}$ to the average maximum response over a sample of crystals.

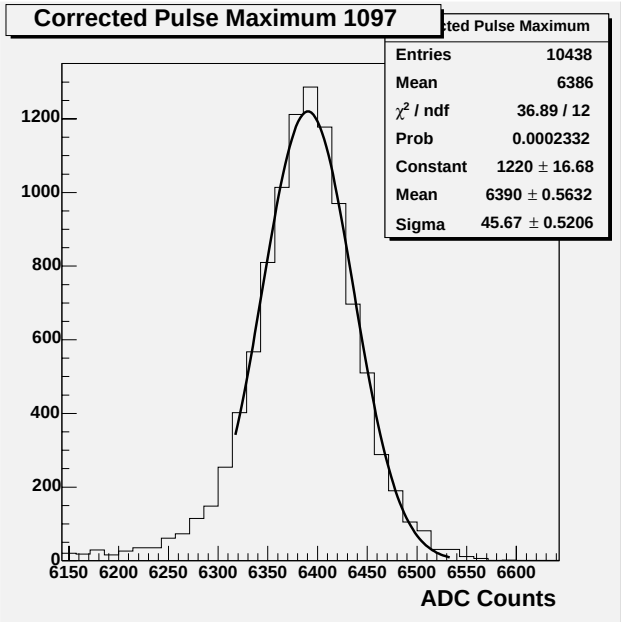


Figure 4.4: *Corrected crystal response to the 120 GeV beam. Notwithstanding the correction, the response is not gaussian, because of the leakage. The fit is performed with an iterative gaussian fit in an interval [1.5σ,3σ], from which it is possible to extract the* $\tilde{A}_i^{max}$*, the maximum response for the crystal i.*

This method allows the evaluation of the intercalibration constants from the single crystal response only and it also gives a quite high event selection efficiency (about 20-30% with the current setup). The method has been preferred to alternative iterative methods using the sum of nine crystals or the sum of 25 crystals around the hitted crystals, which are less dependent to the electron impact point position, due to the high noise levels which were found in the 2003 test beam data.

This procedure is repeated for all crystals composing the array with the 50 GeV beam and 120 GeV beam. Consistency checks were made in order to extract the precision reached in determining the intercalibration factors: 4.5 shows the distribution of the relative difference between the intercalibration factors determined with 50 GeV and 120 GeV runs.

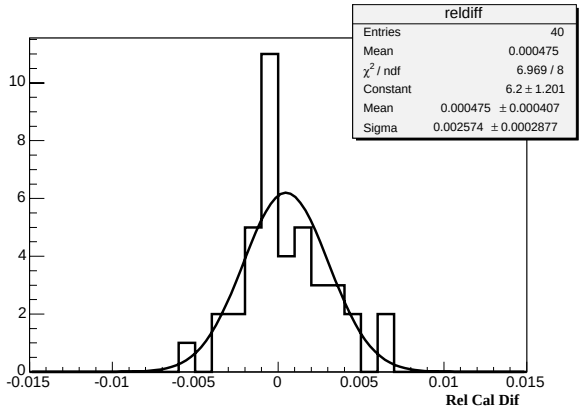


Figure 4.5: *Relative difference between the intercalibration factors determined with 50 GeV and 120 GeV runs.*

A precision of around 0.3% has been reached including statistical fluctuactions and possible reproducibility systematic effects.

4.4 Comparison with intercalibration from laboratory measurements

Since not all the supermodules will be precalibrated and for most of the crystals the only available calibration will be the one obtained from laboratory measurements, it will be crucial to understand which is the level of precision that can be reached, by comparing the results with the beam intercalibration.

The value measured in the laboratory in the process of crystal qualification is light yield which is determined with a precision of about 3%. To perform this comparison the crystal light yield should be extracted using the electron beam data:

$$LY = \frac{\tilde{\mathrm{A}}_i^{max}}{\varepsilon_{cont} E_{beam} G_{APD} G_{el}\ e}, \tag{4.2}$$

where $\tilde{\mathrm{A}}_i^{max}$ is the maximum response of the single crystal, ε_{cont} the containment factor in the single crystal in a narrow region around the maximum containment point, E_{beam} is the energy of the beam, G_{APD} is the gain of the photodector, G_{el} is the gain of the electronics and e is the electron charge.

Considering the different experimental setup in the laboratory and the test beam, Instead of absolute light yield, it is worth to compare the relative light

yield (RLY), defined as the light yield coefficient of each crystal, normalized to the crystals average:

$$RLY_{Lab(TB)} = \frac{LY_{Lab(TB)}}{LY^{avg}_{Lab(TB)}} \tag{4.3}$$

and the difference:

$$\Delta RLY = \frac{RLY_{Lab} - RLY_{TB}}{RLY_{Lab}}, \tag{4.4}$$

This quantity, apart from the absolute normalization, still gives information about the relative precision achieved.
In Figure 4.6 the distribution of ΔRLY for the SM0 intercalibrated crystals is given; the test beam intercalibrations obtained from the 120 GeV data are used. In plot (a) the ΔRLY has been calculated using the raw LY_{Lab} corrected for the variation of the automatic machines reference crystals light yield as proposed in [71], neglecting the APD and the electronics correction in formula 4.2. A gaussian fit is superimposed to the data. The standard deviation of the distribution is (4.6±0.2)%.

Another correction has been suggested ([75]). This one is based on the correlation that was noticed between the crystal measured longitudinal transmission at 360 nm, LT_{360}, and the LY of the crystal. Since the longitudinal transmission measurement is simpler and more precise, it is proposed to use this correlation to improve the precision of the laboratory measurement of the LY. The dependence between the LY and the LT_{360} can be fitted by a straight line

$$LY = a + b \cdot LT_{360}, \tag{4.5}$$

and the parameters a and b can be used to derive a new value for the light yield ($LY_{from\ LT}$). The values of a and b are obtained from a linear fit and are equal to: $a = 4.64$ p.e./MeV, $b = 0.128$ p.e./MeV % ([75]). Figure 4.6 (b) shows the ΔRLY where LY_{Lab} is prediticted from LT_{360}. The gaussian fit gives a σ=(4.3±0.2)%. The two indipendent measurement of the light yield LY_{Lab} and $LY_{from\ LT}$ can be combined. Since the values of the gaussian width of (a) and (b) are comparable, it is possible to use the a simple mean:

$$MEANLY = \frac{LY_{Lab} + LY_{from\ LT}}{2}. \tag{4.6}$$

Plot 4.6 (c) shows the ΔRLY obtained with the MEAN LY. The standard deviation of this distribution is σ=(4.1±0.2)%.

The improvement of the resolution obtained with the MEAN LY method is particularly interesting for the strategy of initial calibration of the calorimeter.

A correction is finally applied for the real gain of the APDs and of the preamplifier of the VFE board: the results are shown in plots 4.6(d), showing no particular improvements. A further correction could come from the ADC calibrations

which were not available for this module. Figure 4.7 compares the laboratory and the test beam LY before and after the described corrections.

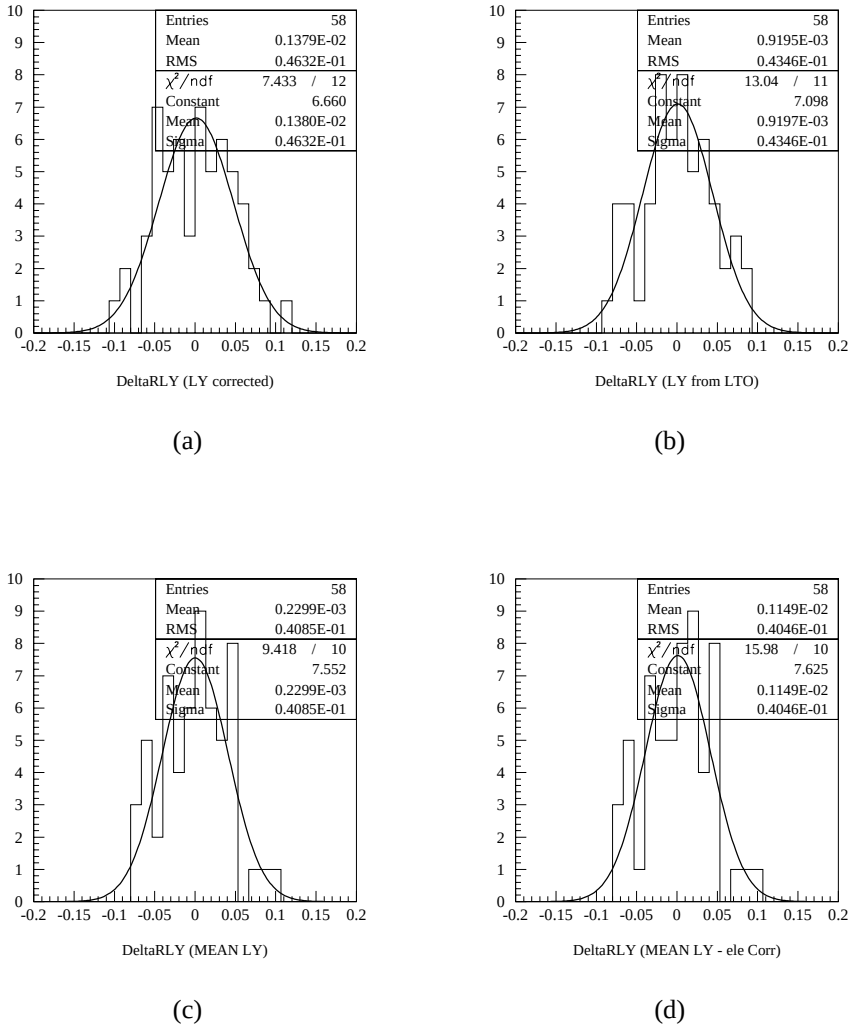
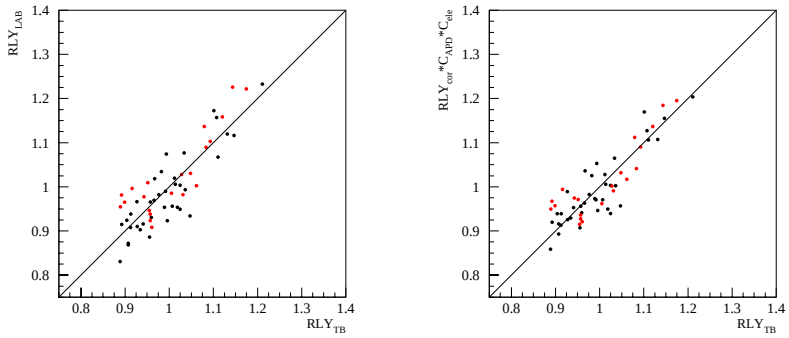


Figure 4.6: *Distribution of ΔRLY for the 58 SM0 crystals calibrated with the beam. The different plots show the effect of all corrections described in the text.*



(a) (b)

Figure 4.7: *Comparison between the laboratory and the test beam LY before and after the corrections described in the text.*

4.4.1 Other corrections

In order to deduce start-up crystal inter-calibrations from the laboratory measurement with the maximum precision, all the possible fine effects should be taken into account. So far only the effect of the variation of the containment factor (ε_{cont} in formula 4.2) has been investigated

Containment factor

Since crystals have 17 different shapes, in principle the containment factor could vary from type to type. This effect has been investigated using a detailed Monte Carlo simulation of the test beam. In particular, for each crystal type a simulated beam of 120 GeV electrons was shot on the crystal front face. The beam was pointing to the position of the shower maximum for electrons of 120 GeV, as done in the test beam.
Figure 4.8 shows the energy contained in the central crystal (1×1) and in the matrix of 9 (3×3) and 25 crystals (5×5) around it, for electrons hitting the crystal in a 4×4 mm^2 region around the maximum containment point. It is clear that, excluding type 1 which has a peculiar shape and requires to be corrected, the single crystal containment is rather uniform for different crystal types (systematic effect being much smaller than the precision of the intercalibration estimated in section 4.3, 0.3%). Relative variation of the containment is also shown in figure 4.8. It is interesting to note that variation of the cointainment with different crystal

types is less for the single crystal than for 3×3 and 5×5 crystals array. This effect is due to the front leakage of the shower to the relative positions of the crystal composing the array, more than to crystal dimensions. The crystals are infact gradually staggered towards higher η, and the front faces of any adjacent crystal are no longer aligned in the forward region. Within a staggered array of crystals, showered photons spreading out from one side of a crystal near the front face leak out to an empty space in front of the adjacent crystal. The effect is higher when considering an increasing number of crystals.

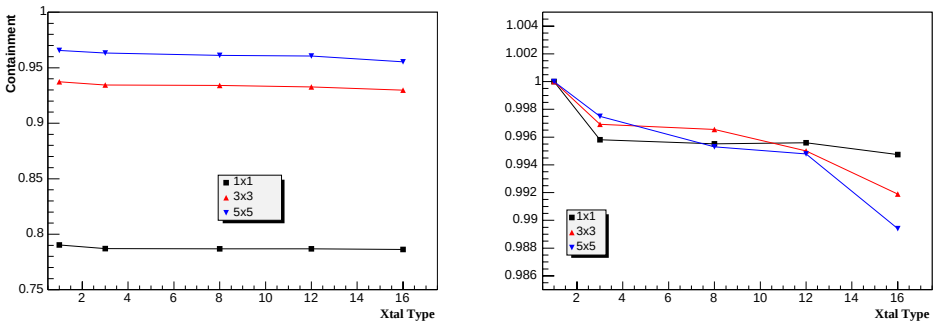


Figure 4.8: *Fraction of electron energy contained in the central crystal and in the matrix of 9 (3×3) and 25 crystals (5×5) around it versus crystal type (left), and relative variation versus crystal type (right) (data from full simulation (see 4.6) of 120 GeV electron).*

4.4.2 Conclusions

The comparison between the relative intercalibration obtained with laboratory measurements and with test beam data shows that a resolution of 4% can be achieved, although the crystals statistics is limited to reach a final conclusion. In order to deduce start-up inter-calibration of the crystals from the laboratory measurement with the maximum attainable precision, all the possible fine effects should be taken into account. A test of a full supermodule is required to better understand all these effects.

4.5 Data Validation architecture and principles

During the 2003 test beam, raw events have been automatically processed just after data acquisition. The relevant physics quantities for calibration have been made available just few minutes after the end of each run. These data have been validated online providing a prompt identification of potential problems and, in

this case, people on shift and run coordinators have been notified (via screen messages and e-mails) to take the appropriate actions. This mechanism will be of particular importance in the precalibration of a complete supermodule, since it gives confidence that the data which have been stored can be effectively used offline to determine the intercalibration factors.

The data validation architecture is described in detail in this CMS internal note [76]. The main design requirement is flexibility: the system should be able to

- automatically launch any number of data validation tasks, even if this number varies with time;
- manage data validation tasks written in any possible language;
- make data validation tasks development independent on the details of the framework;
- automatically manage alarms and notifications produced by data validation tasks.

This flexibility is achieved by completely separating the effective data validation tasks (e.g. evaluation of a first order calibration constant, analysis of the noise in electronic channels,...) from the data validation framework itself. The framework will be able to "communicate" with each indepent task looking into the output file which has been produced. The only requirements made by the framework to each data validation task are to provide an HTML formatted output and a single executable to be launched (this is not a limitation, since the single executable can be a script able to launch several other executables).

The framework itself is written in an object oriented language, Java, constructed on the so called Mediator-Observer Pattern [77]. It is able to detect *Events* from *Sensors* and dispatch them to *Solutions*; the object which is responsible for this job is the *Mediator*. *Events* can be, for example, the end of a run or an alarm issued by a specific data validation task; the *Mediator* intercepts the appropiate *Solution* for a particular event, which can be, for example, start a program, or send a mail to someone or display a message on the screen.

Each data validation task is responsible for its own ouput file, which can contain alarms hidden in HTML comment lines. A great opportunity given by the fact of having HTML formatted output files is that these files can be made directly available on a web page, which can be examined by the shifter as soon as data are made available. A relational database has been also associated with the whole

data validation process; this database is based on the table structure defined for the construction database used in the Rome Regional Center, **REDACLE** [78]. Even in this case each data validation task is responsible for the information which are filled in the database.

In the following the precalibration data validation task will be described, which is the one that has been directly developed by the author.

The other tasks used in the 2003 test beam were:

- High Voltage stability
- Pedestal stability
- Beam Profile
- Beam Alignment

4.5.1 Calibration Data Validation Task

The aim of the Calibration task in the test beam data validation is to check that the data which have been taken during a calibration run are good enough to perform a successive offline calibration. It starts its analysis from the ERF (Energy Reconstructed File) files, which contains the reconstructed amplitudes for each channel together with the beam impact point position given by the hodoscope, which are available just few minutes after the end of the run. It tries to perform an intercalibration which is quite similar to the one described in section 4.3. First it computes the two functions describing the crystal response versus x and y, $f(x)$ and $g(y)$ (defined in section 4.3), applying iterative fits with a fourth degree polynomial. From $f(x)$ and $g(y)$ the position of maximum response in both directions is also calculated; these positions are required to be within an interval of $[-8,8]$ mm both in x and y. This may indicate, for example, that the beam is in fact not aligned with the crystal or the data are corrupted or the hodoscope is not properly working. In figure 4.9 the distribution of the maximum response position for some runs identified as good is given.

Secondly, an event by event position correction is applied to the crystal response equals to:

$$\frac{f(x_{max})g(y_{max})}{f(x)g(y)}. \tag{4.7}$$

The distribution of the corrected crystal response for events in a region of 7×7 mm^2 around the maximum of response is then fitted with a gaussian function in a region of $[-1.5\sigma, 3\sigma]$. The χ^2/NDF of the fit and the efficiency of the selection

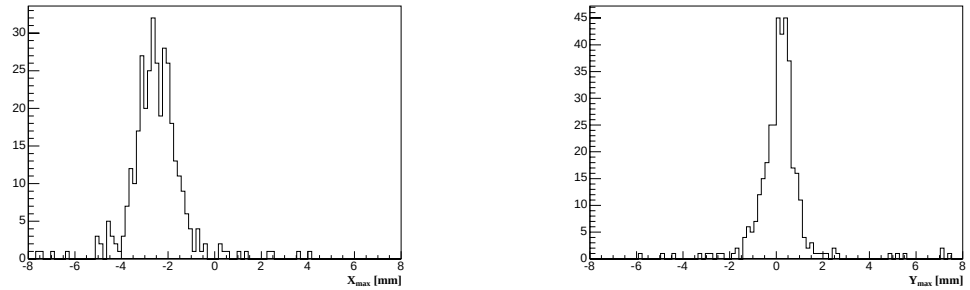


Figure 4.9: *Distribution of the crystal maximum response position for runs identified as good in the x (left) and y.*

are two other parameters used in the identification of good runs; their distribution is shown in figure 4.10. A warning is issued in case the fit χ^2/NDF is too high (greater than 6), or the selection efficiency is too small (less than 10%).

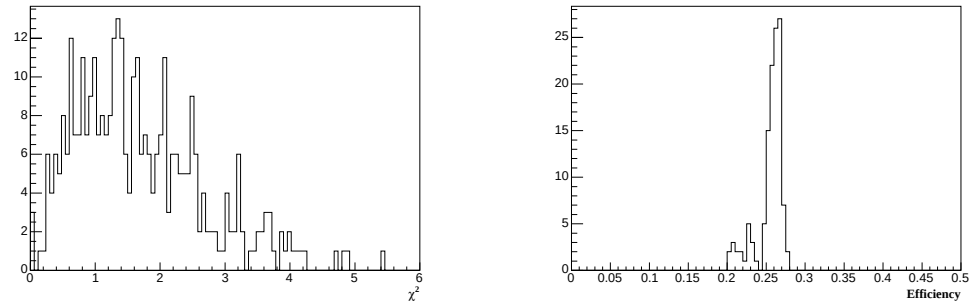


Figure 4.10: *Distributions of the χ^2/NDF (left) and of the selection efficiency (right) for good runs.*

The calibrations obtained in this way are then stored in the REDACLE database for successive retrieval; their values are found to be in agreement with those obtained with the offline analysis at a level of 0.5%.

4.6 Development of a GEANT 4 simulation of the test beam setup h4sim

`h4sim` is a GEANT 4 simulation of one full supermodule (1700 crystals) in the H4 test beam configuration. The list of physics processes activated in the simulation (*Physics List*) includes, in the current version, the electromagnetic interactions of electrons, muons, gammas. Therefore the current version is only capable of simulating electromagnetic showers and not hadronic ones.

Effort has been spent in trying to tune the simulation in order to have a good agreement with the data. This has been done trying different GEANT 4 *Production Cuts*. The general principles for tracking particles in Geant4 are that all the particles produced will be tracked up to zero range; however, each particle has a cut in range (which is converted to energy for all materials) to produce secondary particles. Playing with these cuts allows to tune the number of particles which are tracked by GEANT4, hence the simulation output. It is worth to find the set of production cuts which best reproduces real data.

Two sets of production cuts have been tested:

- 1mm for electrons, positrons and gammas the same cuts used in the full CMS detector GEANT 4 simulation OSCAR [79].
- 1mm for electrons, positrons and 100mm for gammas. This particular set of production cuts has been defined in order to have a similar cut in energy in lead tungstate (approximatively 1.15 MeV), both for the charged and neutral particles contained into the electromagnetic shower. As a backside effect of the increased cut for gammas, this particular set of cuts allows also a reduction of the CPU time needed to simulate one event of around 30% (simulating showers from electrons at 120 GeV) with respect to the standard one.

The geometry description of the calorimeter parts is shared with the official simulation, OSCAR, and is read from *.xml* files. In the actual version the simulation includes the geometry of a full supermodule along with its support aluminum frame.

In `h4sim` a parallel beam like the H4 one is simulated, giving the possibility for the primary particle to be generated from a point inside a plane perpendicular to the beam direction. A coordinate system can be defined in this plane; this coordinate system, if properly chosen, simulates the x and y coordinates reconstructed with the hodoscope system in the test beam (in case of a parallel beam). The following formula is used to relate the coordinates in the GEANT 4 reference frame (x_G, y_G, z_G) to the test beam reference frame (x, y, z) described in section 4.1:

$$\begin{pmatrix} x_G \\ y_G \\ z_G \end{pmatrix} = \begin{pmatrix} -\cos\theta\cos\phi\cos\psi + \sin\phi\sin\psi & \cos\theta\cos\phi\sin\psi + \sin\phi\cos\psi & \cos\phi\sin\theta \\ -\sin\phi\cos\theta\cos\psi - \cos\phi\sin\psi & \sin\phi\cos\theta\sin\psi - \cos\phi\cos\psi & \sin\phi\sin\theta \\ \sin\theta\cos\psi & \sin\theta\sin\psi & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \tag{4.8}$$

where θ and ϕ are the angle in the GEANT 4 reference frame of a vector parallel to the beam direction, and ψ accounts for a possible rotation of the test

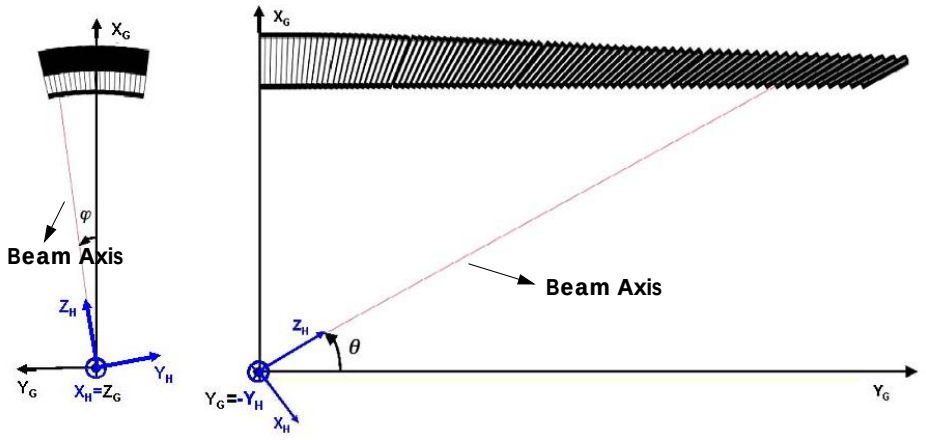


Figure 4.11: *View of the supermodule in the $x_G y_G$ plane and in the $x_G z_G$ plane. The hodoscope reference frame (x_H, y_H, z_H) is superimposed; the z_H axis is coincident with the beam line.*

beam reference frame around its axis $\vec{z}$; ψ is assumed to be zero in the simulation (see figure 4.11).

Randomly choosing the particle origin inside a window in the plane perpendicular to the beam direction allows to simulate a trigger window and a beam profile. In the current release, the possibility to have a flat or a gaussian beam profile is implemented; in both cases the beam dimension is controllable via datacards.

There is also the possibility to have different beam directions: in particular the user can decide to have a beam that points to the crystal center (read directly from the *xml* geometry) or to the "*test beam point*" which should be for each crystal the point of maximum of response to an electron of 120 GeV. Actually, the "test beam point" is fixed using an old GEANT 3 simulation, where this position has been determined qualitatively from a parabolic fit to the crystal response [80].

To simulate all the effects not present in the GEANT 4 simulation, the output of the GEANT 4 program is passed to a program that simulates the electronics response, producing an output which is readable directly from the test beam data analysis framework. The following effects have been considered:

- the photostatistics contribution to the stochastic term is simulated using the average values found for the crystal light yield which is roughly 4 photoelectrons/MeV [81], and the average estimation of the APD excess noise factor (2)

- the effect of the longitudinal non uniformity of the light yield is considered applying a simple correction to the deposited energy according to its longitudinal position inside the crystal. The values of the Front Non Uniformity

(FNUF) and of Rear Non Uniformity (RNUF) stored in the construction database for each crystal can be used.

- a constant term of 0.35%, which takes into account the average of all the effects not present in the simulation, like temperature instabilities, high voltage instabilities, crystal intercalibration,... .
- the GEANT 4 hit in each crystal is converted into ADC counts using the electronic calibration factor 17.6 MeV/ADC channel corresponding to the average absolute calibration estimated during the test beam
- generation, according to the database of real test beam pulse shapes, of the very front-end signal corresponding to the ADC converted energy value, digitized by the sampling ADC at 40 MHz
- simulation of the phase of electronics signal with respect to the ADC clock by addition of a random offset, between 0 and 1, to the TDC information, taking 2 pre-samples before the signal
- addition of noise using real Pedestal runs taken during the test beam, adding noise “samples” to each simulated ADC sample.

This procedure is performed by a modified version of the *g4Simulation.cc* program present in the *H4Analysis* package; the output is a *Raw Root File* (RRF) in the same format as the test beam data.

These RRF files can be later used as an input to produce the so called ERF files (“Energy Root Files”), which contain the values, for each crystal, of the reconstructed amplitude, obtained with the method described in section 4.2.

4.7 Comparison between h4sim and test beam data

4.7.1 Reference systems and alignment

The response of the hit crystal varies with the impact point of the electron on the crystal front face. An absolute comparison of the lateral shower development, in the real data and in the simulation, requires a transformation that connects the two different reference systems used to give the impact point position.

As a consequence, a correct position of the supermodule on the table can be revealed looking at the measured position of the “maximum containment point” in the test beam data.

On the other hand, in the simulation, the origin of the reference system is defined by center of the simulated parallel beam which, as explained in section

4.6, can point either to the center of the crystal front face or to the "maximum containment point". The latter possibility was finally chosen to reproduce the test beam set-up as much as possible.

On the test beam data, a parallel beam is also ensured applying cuts on the slope reconstructed with the hodoscope. The track slope is required to be both in the *x* and *y* direction $|X_{slope}, Y_{slope}| < 1 \cdot 10^{-4}$; a cut is made on the track quality variable too. The distribution of the beam slope in the *x* direction is shown in figure 4.12; in the *y* direction it is similar.

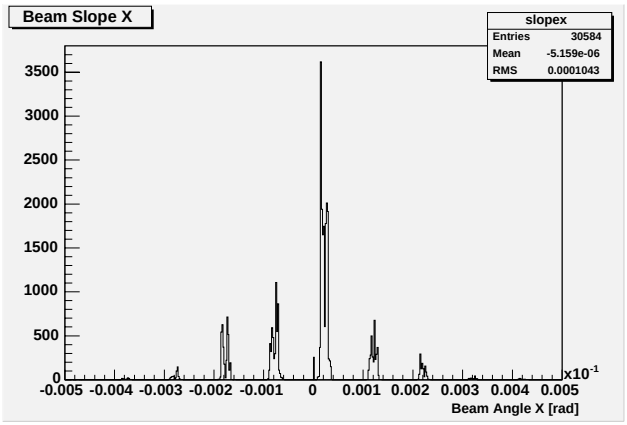


Figure 4.12: *Distribution of the variable X_{slope} variable used to select parallel tracks in test beam data. Tracks should have $|X_{slope}| < 1 \cdot 10^{-4}$.*

The crystals used in the comparison are numbered 1097 ($\eta = 78, \phi = 12$, crystal type 16) and 1522 ($\eta = 78, \phi = 17$, crystal type 16); they are respectively the center of the trigger towers 63 and 64. The beam energies used in the analysis are 20,35,50,80,120,180 GeV.

A global shift between the test beam and simulation reference system can be found comparing the position of maximum containment point at different energies. The procedure to find the maximum containment point in the test beam and simulated data is exactly the same and it is an iterative one. To find the position of the maximum in the *x* (*y*) direction the following procedure is adopted:

- select a region of $\pm$4mm around 0 in the orthogonal *y* (*x*) direction; with the events into this region it is possible to give a first estimate of the maximum position in X

- repeat the same procedure, now selecting a narrower region of $\pm$2mm around the maximum point position found in the previous iteration.

The procedure converges within the fit statistical error just after the second iteration. An example of the fit superimposed to the distribution is given in figure 4.13.

The systematic error, due also to the reproducibility of the table positioning, has been estimated to be around 100-200 μm.

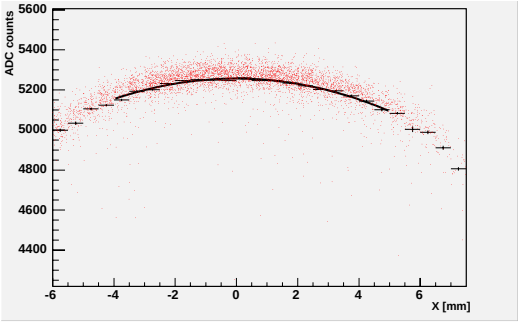


Figure 4.13: *Distribution of the response of the crystal versus the impact point position along with the iterative fit used to extract the position of the maximum of response (superimposed).*

The position in x and y for the maximum containment point at different energies in test beam and in the simulation for crystals 1097 is shown in figures 4.14. The distribution of test beam data and simulated ones can be fitted with a function

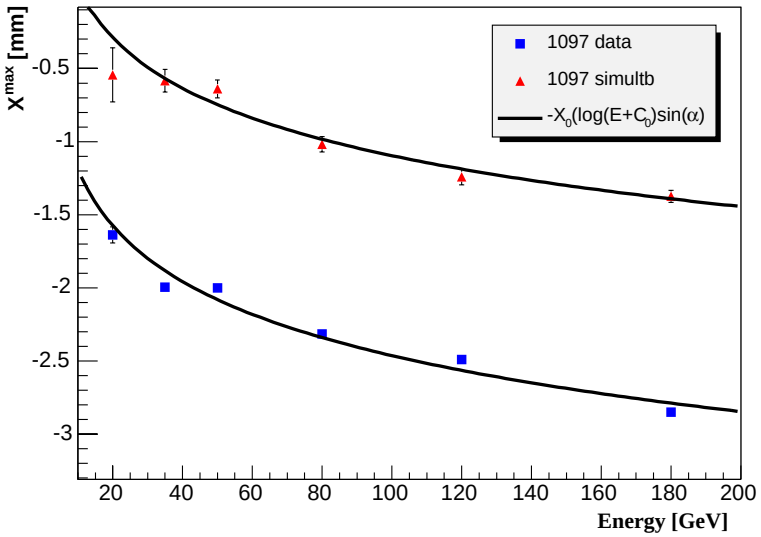
$$-X_0 \log(E + D_0) \sin\alpha, \tag{4.9}$$

where $X_0 = 8.9mm$ is the $PbWO_4$ radiation lenght, D_0 is a global shift expressed in energy, and α is the effective tilt angle. Test beam and simulation show a similar behaviour with the energy. From the measured position of the maximum containment point at 120 GeV, it is also clear that there is a shift in the position of supermodule on the table, in particular in the x direction, while in the y direction the table seems to be aligned within the experimental errors; the estimated shift in the x direction is $\Delta_X = 2.4 \pm 0.1$ mm[1].

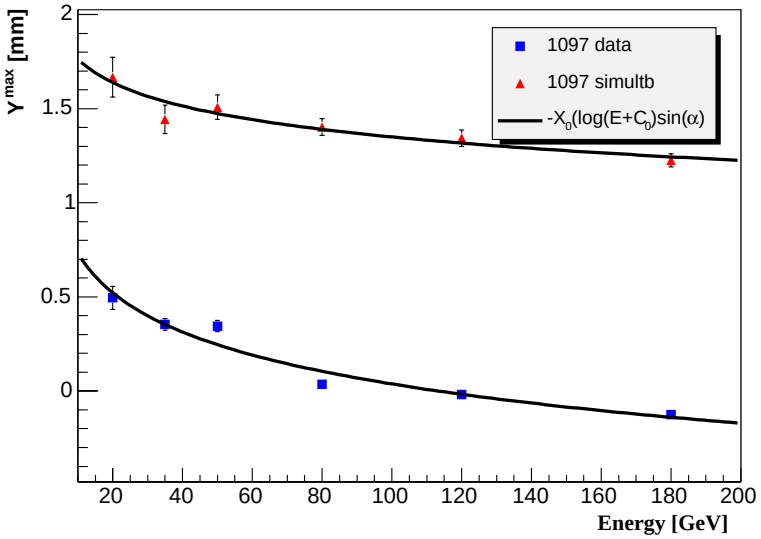
Also the angles used in the *h4sim* simulation to put the center of the beam in the maximum containment point should be recalculated for the next simulation version.

The difference between the maximum containment point positions in the simulation and in the data both in the x and y direction for crystals 1097 can be fitted with a constant function, allowing to extract a global offset between the two reference frames, which is summarized in table 4.7.1.

[1]The accuracy of supermodule positioning on the rotating table is estimated to be around 1-2



(a)



(b)

Figure 4.14: *Position in x (a) and y (b) for the maximum containment point at different energies in test beam and in the simulation for crystals 1097*

Crystal	Δ_X [mm]	Δ_Y [mm]
1097	-1.3 ± 0.1	-1.3 ± 0.1
1522	-2.3 ± 0.1	-1.1 ± 0.1

Table 4.1: *Offset in x and y direction between the impact point refence frame in the test beam and in the simulation frame for the crystals 1097 and 1522. The offset is extracted using a fit with a constant function to the difference of the measured maximum point position in the two reference frames at different energies.*

The main advantage of using the maximum containment point to find a global shift between the two reference frames is that only the energy contained within the central crystal is used, hence no intercalibration is required. However, its determination needs a large statistics, since the response of the crystal around the maximum containment position is rather flat. Another possibility, which is considered as a check of the maximum containment point method, is to use the "balance point" position, firstly introduced in [82].

The balance point is defined taking into account a matrix of 3×3 crystals; it is the point where the sum of the energies in the three left row crystals equals the sum of the right row ones, and the same happens also for the top and bottom rows. The balance point is easily determined looking at the distribution of $\frac{E_{left}-E_{right}}{E_{left}+E_{right}}$ versus x and $\frac{E_{up}-E_{down}}{E_{up}+E_{down}}$ versus y. For the balance point determination, however, since the surrounding crystals energy is used, intercalibration is needed to determine the balance point position. It is also more sensitive to the noise.

As for the "maximum containment point", a global shift can be extracted from the difference of the balance point determination in the data and in the simulation at different energies. The values obtained applying a fit with a constant function are summarized in table 4.2, both for crystal 1097 and 1522, which generally agree with those obtained using the maximum containment method.

Crystal	Δ_X [mm]	Δ_Y [mm]
1097	-1.8 ± 0.1	-1.2 ± 0.1
1522	-2.3 ± 0.1	-1.1 ± 0.1

Table 4.2: *Offset in x and y direction between the impact point refence frame in the test beam and in the simulation frame for the crystals 1097 and 1522. The offset is extracted using a fit with a constant function to the difference of the measured balance point position in the two reference frames at different energies.*

mm.

4.7.2 Lateral shower development comparisons

The lateral shower shape defines the distribution of the energy deposition in a cluster of crystals around the impact point and is therefore one the most important quantity which should be checked when comparing real data with a calorimeter simulation. To compare the lateral shower development it is possible to use the ratio of the energy contained into a single crystal (E1) over the energy contained into a 3×3 and a 5×5 crystal matrix centerd around the hit crystal (respectively E9 and E25). The channel intercalibration used to form the E9 and E25 is the one obtained from the method described in 4.3.

Data have been simulated with the two different GEANT 4 production cuts defined in section 4.6:

- 1mm for electrons, positrons and gammas, the same production cuts used in OSCAR (*OSCAR production cuts*)
- 1mm for electrons, positrons and 100 mm for gammas. (*h4sim production cuts*)

The figure 4.15 shows a comparison at different energies of the $E1/E9$, $E1/E25$ and $E9/E25$ ratios using the OSCAR production cuts; these ratios are the mean value of a gaussian fit to the distribution obtained selecting an impact point region of $4 \times 4\text{mm}^2$ around the maximum containment point.

Looking in particular at the $E1/E9$ $E1/E25$ ratios, there seems to be a disagreement of about 1% which accounts for a narrower core of the shower in the simulation. The lateral tail of the shower seems instead to be well reproduced as you can see looking at the values of the $E9/E25$.

A better agreement is found using the second set of production cuts ("h4sim production cuts"). Figure 4.16 show the data versus Monte Carlo.

Another check of the lateral shower development compatibility is the $E1/E9$ ratio versus the impact point position; as the energy contained within the central crystal is strongly dependent on the impact point position, this variable allows a detailed comparison of the lateral shower development. The comparison has been done using data at 120 GeV simulated with the "h4sim production" for the 1097 crystal. The agreement both in *x* and *y* direction is satisfactory, even if the tails are not perfectly reproduced as can be seen in figure 4.17.

4.7.3 Energy resolution comparisons

The figures 4.18, 4.19 shows the comparison for the crystal 1097 (crystal type 16) of the energy resolution at different energies for simulated events (using the "h4sim production cuts") and test beam data, when estimating the energies as the

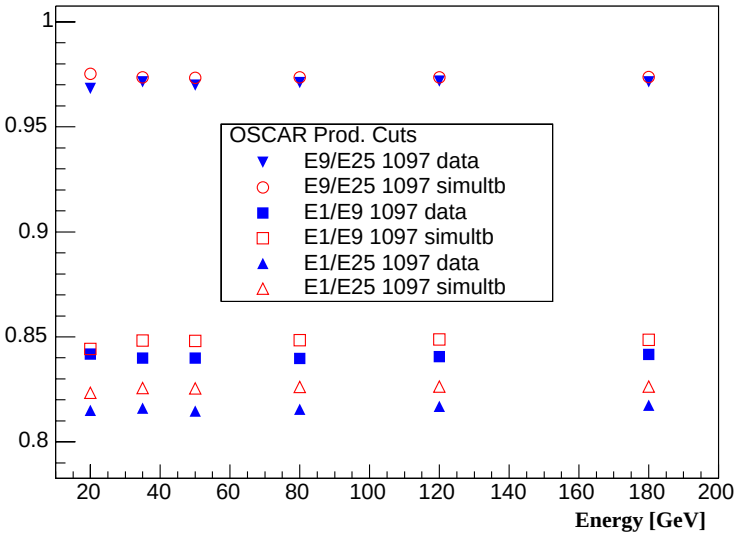


Figure 4.15: *Comparison of the ratios $E1/E9$, $E1/E25$ and $E9/E25$ between data and MC using the OSCAR production cuts. A disagreement of about 1% is present in the $E1/E9$ and $E1/E25$ almost constant with the energy.*

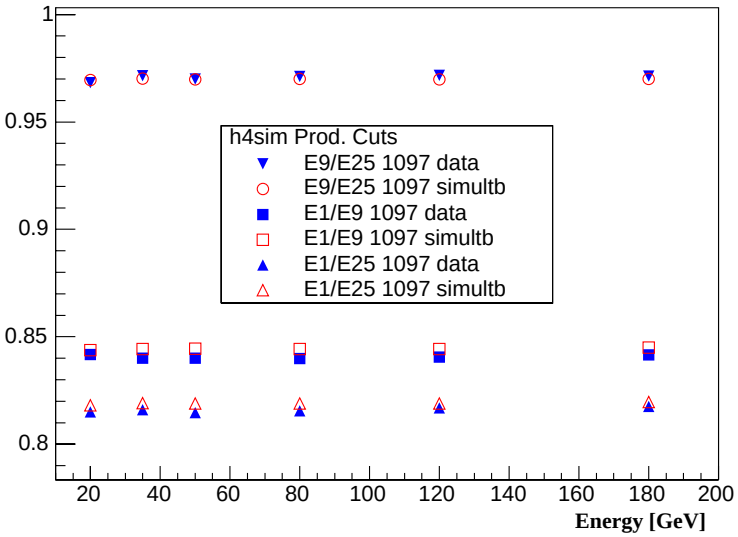
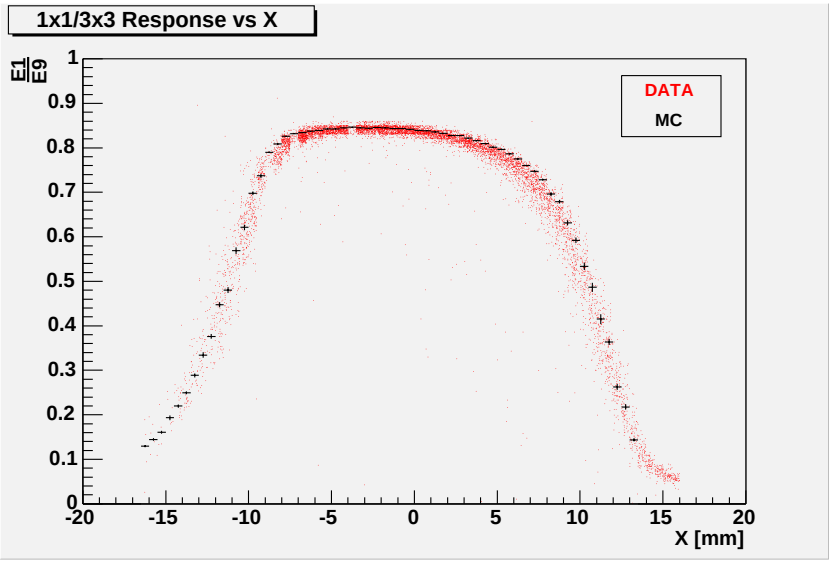
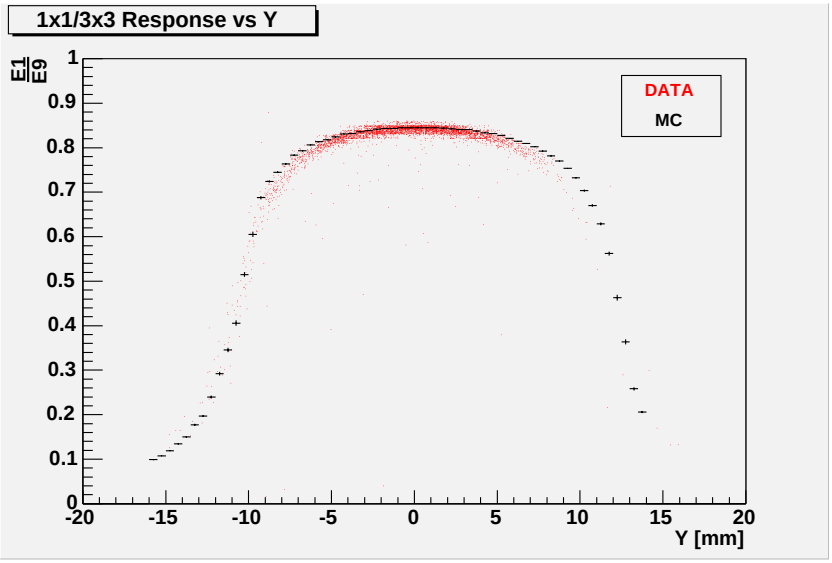


Figure 4.16: *Comparison of the ratios $E1/E9$, $E1/E25$ and $E9/E25$ between data and MC using the "h4sim production cuts".*



(a)



(b)

Figure 4.17: *$E1/E9$ ratio versus impact point position in x (a) and in y(b) of the electron (data and MC). This variable allows a detailed comparison of the lateral shower development. The agreement both in x and y direction is satisfactory, even if the tails are not perfectly reproduced. The simulation has been done using the "h4sim production cuts".*

single crystal response ($E1$), or as the sum of the energies in a 3×3 ($E9$) or 5×5 ($E25$) matrix of crystals . The energy resolution is obtained from an iterative gaussian fit performed in the region $[-1.5\sigma, 3\sigma]$ around the mean; the events were selected in a region of 4×4 mm^2 around the maximum containment point. The beam momentum spread calculated using the formula [83]

$$\frac{\sigma(P)}{P} = \frac{\sqrt{C3^2 + C8^2}}{\sqrt{3}\times 2700}, \tag{4.10}$$

where C3 and C8 are the half width of the collimators numbered 3 and 8 respectively expressed in mm, has been subtracted from the energy resolution for the test beam data. The half width of C3 and C8 for the runs considered in the analysis is 3mm; this leads to a contribution of 0.1%.

All the possible contributions to the energy resolution are identified. The lines in figure 4.18 and 4.19 are obtained from a fit with the function

$$\frac{\sigma_E}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c, \tag{4.11}$$

where *a*,*b*,*c* are respectively the stochastic, noise and constant contribution to the energy resolution, as discussed in the previous chapter. The black line (MC) is the energy resolution when considering only stochastic fluctuactions of the shower containment as simulated from the MC. The contribution to the stochastic term a coming from the shower containment is $5.0\pm0.1\%$ for the single crystal, $2.50\pm0.05\%$ for $E9$ and $2.21\pm0.04\%$ for $E25$. Then the photostatistics contribution (red line MC+PHOT) is added: a is now $5.8\pm0.1\%$ for the single crystal, $3.5\pm0.1\%$ for $E9$ and $3.2\pm0.1\%$ for $E25$. The LY longitudinal non uniformity is added next (MC+PHOT+FNUF); the latter gives almost no contribution as expected since the crystals used in the test beam, being fully qualified production crystals, have a rather flat LY uniformity profile. The simulation of the electronics (pulse shape + amplitude reconstruction) is then considered (MC+PHOT+FNUF+DIGI). It is very interesting to note that this brings a constant contribution of around 0.35% to the energy resolution. This effect has been investigated in more detail and found to be due to a phase dependent bias (see section 4.2) of the amplitude reconstruction when reconstructing simulated data. Similar effects, however, should be considered specific to the test beam setup since in the final CMS installation the phase between the signal and the ADC clock would have a constant value (as discussed in section 4.2).

The noise contribution is taken into account adding sample by sample values coming from pedestal test beam run (MC+PHOT+FNUF+DIGI+NOISE). This

procedure gives the possibility to take into account at the same time all the uncorrelated and correlated noise. The noise term is estimated to be 0.12 ± 0.02 GeV for the single crystal, 0.52 ± 0.02 GeV for $E9$, 1.04 ± 0.02 GeV for $E25$. These values, which are obtained from a fit to the data, are in good agreement with the ones which can be extracted analyzing pedestal data. The noise level in the single crystal is four times higher than what expected in the ECAL TDR. Comparing the noise levels for the single crystal and for $E9$ and $E25$, it is clear that a large fraction of the noise is coming from correlated noise. It should be stressed that these results are obtained with the old FPPA electronics; encouraging results have been obtained with the final MGPA electronics in reference [84], where the single crystal noise level is quantifed at around $\sim 44 MeV$. Other effects, which can be estimated to give a contribution of 0.35% constant term, are added to the simulation as a constant term (MC+PHOT+FNUF+DIGI+NOISE+CONST).

The final energy resolution in the data and in the MC (when all possible contributions are taken into account) can be compared:

- $E1$ DATA
$$\frac{\sigma_E}{E} = \frac{(5.7 \pm 0.1)\%}{\sqrt{E/GeV}} \oplus \frac{(0.12 \pm 0.02)\ \text{GeV}}{E} \oplus (0.44 \pm 0.02)\%, \qquad (4.12)$$
- $E1$ MC
$$\frac{\sigma_E}{E} = \frac{(5.8 \pm 0.1)\%}{\sqrt{E/GeV}} \oplus \frac{(0.12 \pm 0.02)\ \text{GeV}}{E} \oplus (0.48 \pm 0.03)\%, \qquad (4.13)$$
- $E9$ DATA
$$\frac{\sigma_E}{E} = \frac{(3.5 \pm 0.3)\%}{\sqrt{E/GeV}} \oplus \frac{(0.52 \pm 0.02)\ \text{GeV}}{E} \oplus (0.41 \pm 0.03)\%, \qquad (4.14)$$
- $E9$ MC
$$\frac{\sigma_E}{E} = \frac{(3.5 \pm 0.2)\%}{\sqrt{E/GeV}} \oplus \frac{(0.52 \pm 0.02)\ \text{GeV}}{E} \oplus (0.41 \pm 0.03)\%, \qquad (4.15)$$
- $E25$ DATA
$$\frac{\sigma_E}{E} = \frac{(3.3 \pm 0.1)\%}{\sqrt{E/GeV}} \oplus \frac{(1.02 \pm 0.02)\ \text{GeV}}{E} \oplus (0.39 \pm 0.06)\%, \qquad (4.16)$$
- $E25$ MC
$$\frac{\sigma_E}{E} = \frac{(3.2 \pm 0.2)\%}{\sqrt{E/GeV}} \oplus \frac{(1.02 \pm 0.02)\ \text{GeV}}{E} \oplus (0.39 \pm 0.05)\%. \qquad (4.17)$$

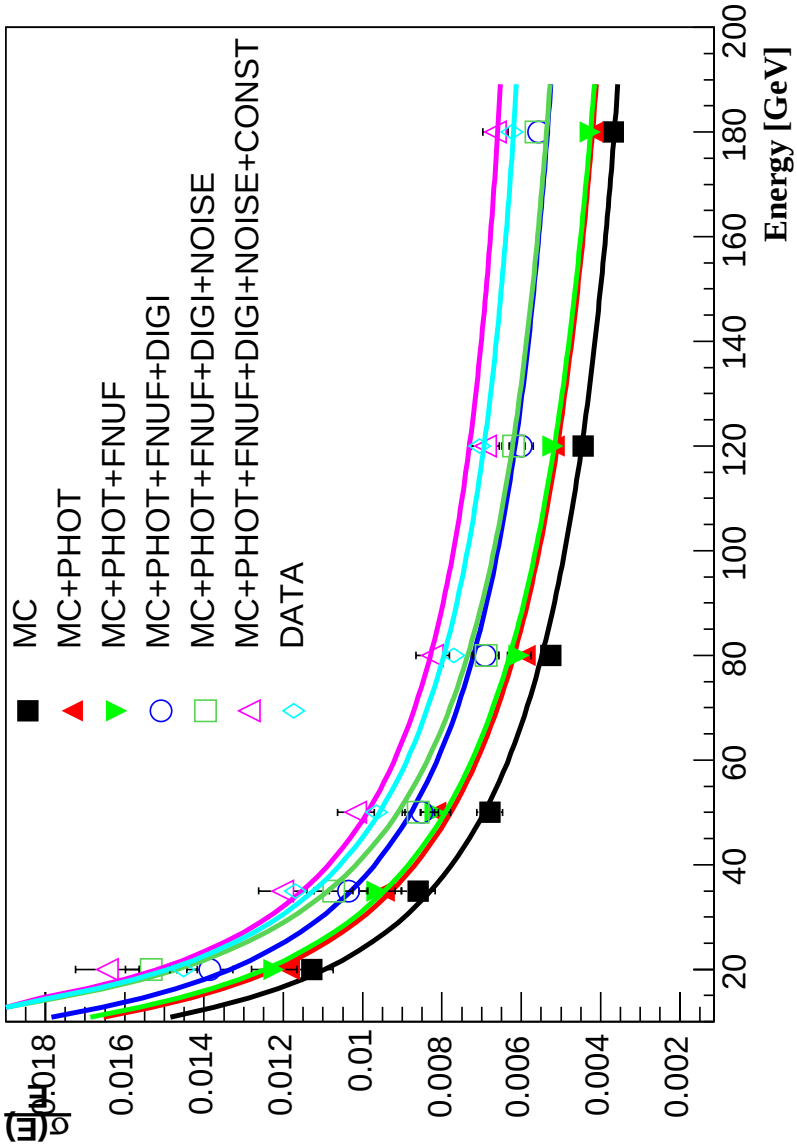


Figure 4.18: *Energy resolution versus the energy (data and MC) for crystal 1097 reconstructing the energy using the single crystal. The events are selected in a region of* 4×4 *mm*^{2} *around the maximum containment point. Different contributions to the energy resolution are identified (see the text for a detailed explanation).*

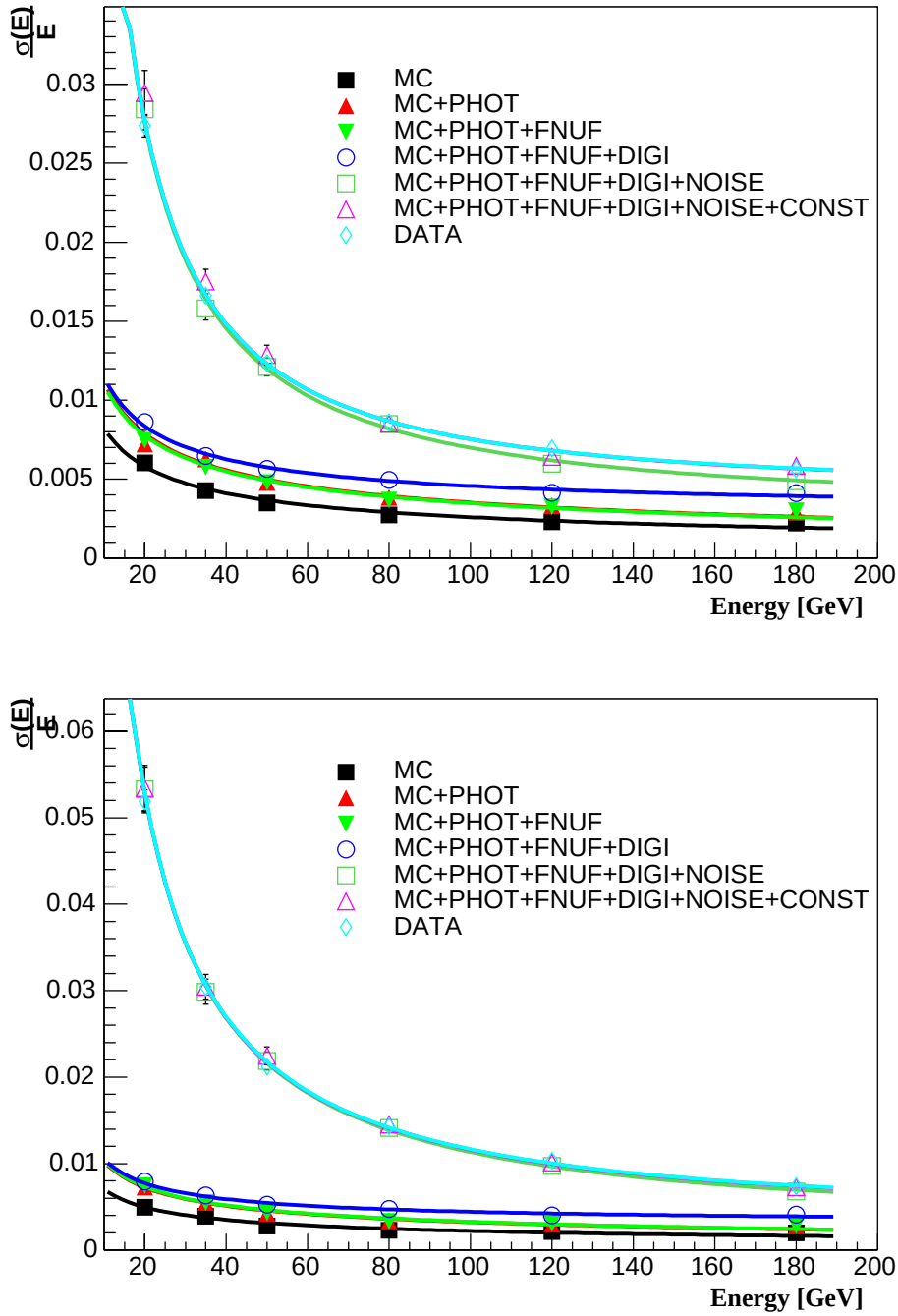


Figure 4.19: *Energy resolution versus the energy (data and MC) for crystal 1097 reconstructing the energy using the sum of a* 3×3 *(*$E9$*, top) and* 5×5 *(*$E25$*, bottom) crystals' array. The events are selected in a region of* 4×4 *mm*^{2} *around the maximum containment point. Different contributions to the energy resolution are identified (see the text for a detailed explanation).*

Only statistical errors from the fit are reported.

The good agreement found in the considered energy range indicates that the parameters used in the the simulation, namely the photostatistics value, the excess noise factor and the overall constant term, are well tuned and understood.

4.7.4 Position resolution comparisons

The distribution of the deposited energy in the 3x3 crystal matrix depends on the impact point, hence allowing to measure the impact point of the incident particle. A simple center-of-gravity method can be used to compute the impact point position, where the position of each crystal is weighted by its measured energy. In the $\tilde{X}$ coordinate the formula used to compute the center of gravity looks like:

$$\tilde{X} = \frac{\sum_i E_i \cdot x_i}{\sum_i E_i}, \tag{4.18}$$

where the i index runs over the crystals contained in a matrix of 3×3 crystals around the central one, the E_i is the reconstructed energy for the i-th crystal and x_i is the crystal x coordinate; a similar expression stands also for the $\tilde{Y}$ coordinate.

Due to the finite crystal dimension and to the exponential lateral shower profiles, a linear center-of-gravity reconstructs a position which is shifted from the true impact point position; the distribution of the difference between the true position, measured in the test beam with the hodoscope system, and the reconstructed position has a typical S-curve shape (left top plot of 4.20). A correction can be computed on the basis of the measured S-curve distribution: the corrected S-curve distribution is shown in the top right plot of figure 4.20.

The uncertainty on the reconstructed impact point position is varying with the impact point position itself; it is more precise when the impact point position is close to the crystal edges, while it gets worse when the containment in the crystal is maximal, as it can be seen in the bottom left plot of figure 4.20.

In figure 4.21 the resolution on the impact position, in the x and y direction, is shown as a function of the energy of the incident electrons for the crystal 1097 using both real data and simulated events. The quoted value for the resolution is obtained from an iterative gaussian fit in the region $[-3\sigma, 3\sigma]$ around the mean value; the events are selected in a region of $[-2, 2]$ mm in both x and y directions around the maximum containment point. For the test beam data, the resolution of the hodoscope system has been subtracted using a value of 150 μm.
Also in this case the agreement between data and Montecarlo is quite good. The different contributions, starting from the intrinsic Montecarlo resolution (as described in the previous section), have been separated. It is clear from the figure that the main contribution, apart from the intrisic shower fluctuactions, is given by

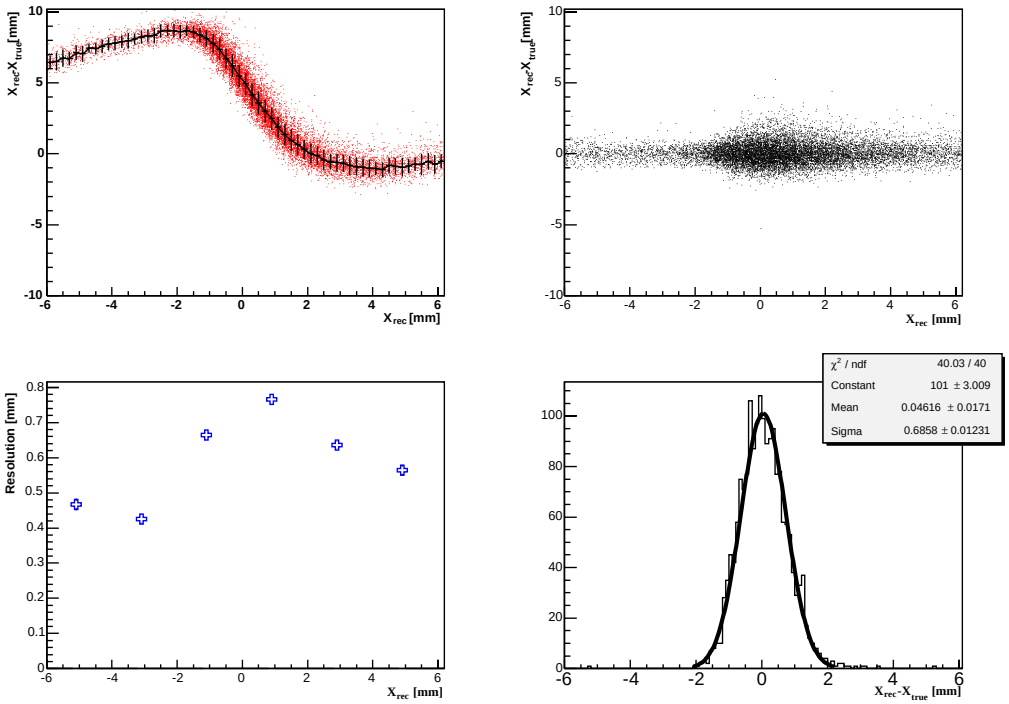


Figure 4.20: *An example of distributions involved in the position reconstruction using the center-of-gravity method. The distribution are obtained reconstructing the x coordinate of the impact point position for crystal 1097 at 120 GeV. The top left curve shows the so called S-Curve distribution, while in the top right plot there is corrected S-Curve. In the bottom left the resolution versus the reconstructed impact point position is shown (note the left-right asimmetry due to the tilt angle between the electron direction and the crystal axis) while in the bottom right there is the distribution of the difference between reconstructed and true position selecting events in an interval of* $[-2, 2]$ *mm around the maximum containment point.*

the noise, which in the case of FPPA electronics is quite high as already discussed. The final resolution found for the *x* and *y* direction is:

- X

$$\frac{\sigma_X}{E} = \frac{(5.0 \pm 0.5)\text{mm}}{\sqrt{E/GeV}} \oplus \frac{(3.6 \pm 0.2)\ \text{cm}}{E/GeV} \oplus (0.47 \pm 0.01)\text{mm}, \qquad (4.19)$$

- Y

$$\frac{\sigma_Y}{E} = \frac{(3.3 \pm 0.5)\text{mm}}{\sqrt{E/GeV}} \oplus \frac{(1.9 \pm 0.6)\ \text{cm}}{E/GeV} \oplus (0.24 \pm 0.02)\text{mm}. \qquad (4.20)$$

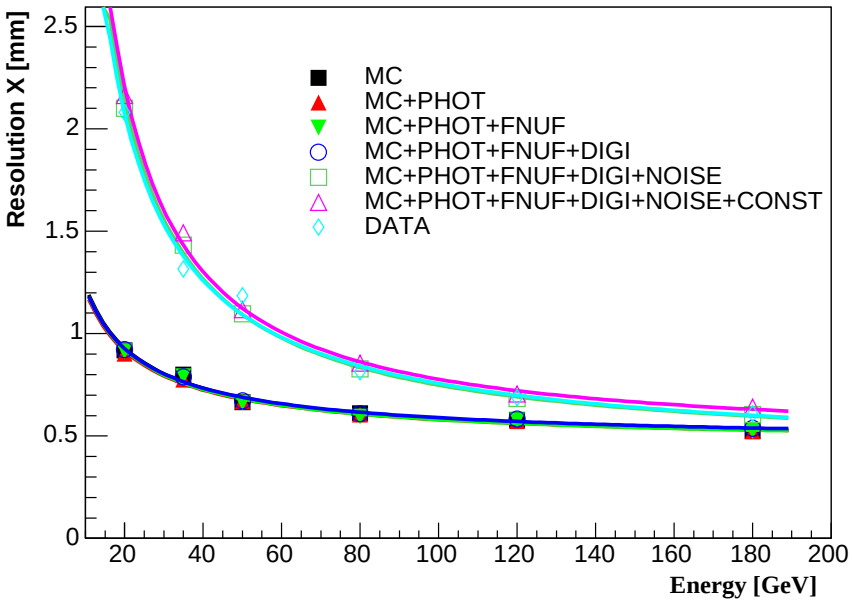
The parametrization chosen for the position resolution as a function of the energy is in some sense arbitrary, since the three terms should not be interpreted as the stochastic, noise and constant term. For example, the term which is proportional to $\frac{1}{E}$ is different from zero even when fitting the MC points without noise.

The different resolution in *x* and *y* is thought be originated from a convolution of various effects, like crystal dimensions, effective angle [2] or crystal staggering. To disentagle all these effects more studies are required.

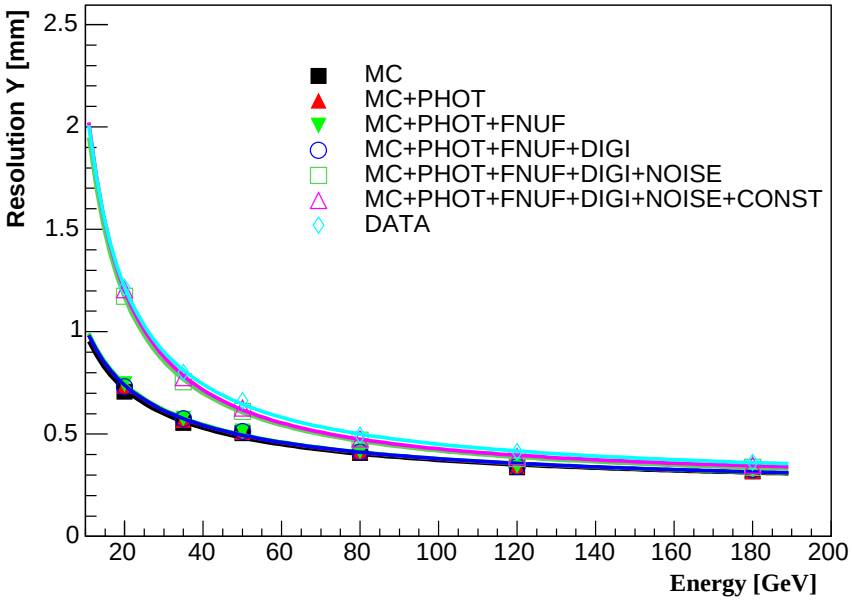
4.7.5 Conclusions

A detailed comparison between a GEANT 4 simulation and the 2003 test beam data has been made. The main interest was in comparing the lateral shower development; two different set of production cuts have been tried, the one used by the official simulation OSCAR and the other defined in such a way that the cut value, expressed in energy instead of range in material, is the same for all the particles involved in the electromagnetic shower. The comparison of the ratios of the single crystal response over the sum of 9 crystals and 25 crystals showed that in both cases lateral shower exstension is well reproduced, but the second set of production cuts gives a better agreement. Moreover the second set of production cuts allows also a reduction of the CPU time spent in the simulation of around 30-40%; the adoption of this particular set of production cust can be envisaged also for the calorimeter simulation inside the official simulation of the CMS detector

[2] The effective angle is defined as the angle between the beam direction and the crystal axis, defined as the line passing through the centers of the front and rear face. This is different from the axis that is used to define the 3^o tilt angle, which instead is the line perpendicular to the crystal front face.



(a)



(b)

Figure 4.21: *The resolution on the impact position in x (a) and in y (b) for crystal 1097 (data and MC) as a function of the electron energy. Different contributions to the energy resolution are identified (see the text for a detailed explanation).*

OSCAR. Also energy and position resolution have been compared, to check the tuning of the different parameters governing the several contributions to the resolution. The procedure which has been adopted to digitize the simulation response and to handle the noise seems to well reproduce the real data; the introduction of similar procedures in the official simulation of the detector can be foreseen.

A more refined and extended comparison should be made in 2004, when the data of a full supermodule test with the final MGPA electronics (foreseen in October 2004) will be available. It will be possible to analyze the containment for different crystal types and dimensions, the effect of the module cracks and several other effects, allowing a complete validation of the GEANT 4 detector simulation.

Chapter 5

Electron reconstruction and identification

The electron reconstruction procedure which has been developed in CMS starts from clusters reconstruction in the electromagnetic calorimeter and tracks reconstruction in the tracker. Therefore, the electron reconstruction quality relies on the performances of both the electromagnetic calorimeter and the tracker.

A specific problem in CMS is the significant amount of tracker material in front of the electromagnetic calorimeter. An electron traversing the tracker cavity radiates photons through the bremsstrahlung process which affects both the electron reconstruction efficiency and the resolution, and introduces tails in the reconstructed quantities distributions. To tackle this problem, specific algorithms have been developed in order to recover the energy lost by bremsstrahlung photons.

In the first part of this chapter, an overview of the framework used to develop the reconstruction algorithms and of the simulated samples used to evaluate their performances is given. In the second part, the clustering algorithms used in the ECAL and the tracking algorithms along with some comments on their perfomances are described. Later, the effect of bremsstrahlung on the electron reconstruction is discussed, showing some reconstructed quantities, able to to identify electrons which wich have not radiated too much. These quantities are particular important for the calorimeter calibration with electrons. A procedure to construct on optimal estimator of the electron energy, which combines both the calorimeter and tracker energy and direction measurement, is introduced. In the last part, the construction of a likelihood based variable able to distinguish between real electrons and jet reconstructed as electrons will be described. Both the optimal estimator of the electron energy and the likelihood based electron identification

will be later used in the analysis of the Higgs signal in the $ZZ^{(*)} \rightarrow 2e^{+}2e^{-}$ decay channel. The key point in the analysis of this channel are high efficiency in the electron reconstruction, good capability to identify isolated electrons and good energy and position resolution.

5.1 Simulation, Reconstruction and Analysis Framework

The CMS trigger and reconstruction software is named ORCA (Object-oriented Reconstruction for CMS Analysis) [85]. It uses the modern technology of object-oriented programming, being implemented in the `C++` language. ORCA is based on the COBRA [86] framework, which provides basic computing services (data access, user interface, flow control, etc.) and utilities (mathematical algorithms, histogrammers, etc.). The presence of a framework allows for homogeneous implementation of the basic design mechanisms of ORCA. In particular, the mechanism of on-demand reconstruction consists in the fact that a given algorithm is performed only when (and if) the corresponding piece of information is requested.

The ORCA code takes part also in the Monte-Carlo simulation of events in the CMS detector, where, however, more specific programs are used to perform specific tasks, such as for example the detector simulation. Monte-Carlo simulation of events in CMS consists of the following steps:

1. Signal and pile-up event generation
2. Detector simulation
3. Digitisation

Signal and Pile-up event generation

Event generation in CMS is done with CMKIN [87], written in `FORTRAN`, representing a common interface for different event generators. Generally CMKIN is interfaced to PYTHIA [29], a `FORTRAN` event generator.

In order to properly consider the effects of pile-up on a given physics signal event, a large number of "minimum bias" events are generated, as well as the signal sample. Minimum bias events are generated in PYTHIA with the MSEL=1 switch. This turns on all leading order inelastic processes for QCD jet production. Minimum bias pile-up interactions are generally soft but, at the LHC design luminosity of 10^{34} cm^{-2}s^{-1}, a mean of 17.5 minimum-bias interactions will occur every bunch crossing and the superposition of pile-up can significantly affect the sensitivity to signal events.

Algorithms and reconstruction software are tested with *ad hoc* simulated events: for instance couples of e^+e^- propagating in the CMS detector are used to evaluate the performances of electron reconstruction in a clean and controlled event.

Detector simulation

The event simulation step (the simulation of the passage of particles through matter) in CMS has currently two options: one, the historical one, is a `FORTRAN` program called CMSIM [88] which is based on GEANT 3 [72], the other, the most recent one, instead, is OSCAR [79], based on GEANT 4 [4] and written in `C++`. In the last year or so, extensive comparisons and tests have been made between the two different simulations by each subdetector, showing a general good agreement [59]. In both simulations the description of the CMS geometry is detailed and includes not only active subdetector volumes, but also cables and mechanical support structures. The collision point is distributed around the CMS reference frame origin according to the composition of three independent gaussian distributions: along z axis (the beam one) with σ_z=5.3 cm and along bend plane x and y axes with σ_x=σ_y=15 μm. In the future only the GEANT 4 simulation OSCAR will be used.

Monte-Carlo generated signal and pile-up events are simulated separately either with CMSIM or OSCAR, producing files which describe particle trajectories and energy depositions in the CMS detector. As the pile-up events can be reused, the signal and pile-up data are still stored separately at this stage. The interface to persistent storage is provided by COBRA.

Digitisation

Once the simulated hits have been made available, ORCA is used to digitise the event data. At this stage, pile-up and signal events are merged, and the luminosity dictates the number of pile-up events to use. Pile-up events are randomly selected from the pile-up database. Simulation of electronics and noise of all detectors is added, as well as zero suppression or selective readout algorithms. After digitisation, the event is written back to persistent storage.

5.2 Simulated data samples

Several samples have been used in the analysis of the electron reconstruction and identification performances presented in this thesis. In particular, the calorimeter and tracker reconstruction performances have been evaluated on two sam-

ples of 100000 back-to-back e^+e^- (double electron samples) simulated with OSCAR_2_4_5 and digitised with ORCA_7_6_1. One has e^+e^- generated with a flat energy distribution in the interval 5-100 GeV and in the pseudorapidity interval $|\eta|$<2.6, corresponding to the whole ECAL acceptance; the other has a flat distribution in transverse energy E_T in the interval 10-50 GeV and in the pseudorapidity interval $\eta|$<1.479 which is the ECAL barrel acceptance region. These samples have been digitised without adding pile-up effects.

The evaluation of electron identification capabilities is instead made on a sample of electrons coming from the H$\rightarrow$ZZ$^{(*)} \rightarrow 2e^+2e^-$ decay (100000 events), generated with a mass of the Higgs boson $m_H = 150$ GeV (electrons average $< E_T >$ = 40 GeV). This sample has been simulated with CMSIM 133 and digitised with ORCA_7_6_1, adding a pile-up corresponding to the low-luminosity regime 2×10^{33} cm^{-}2s^{-1}.

The background from jets has been evaluated on three samples of inclusive QCD di-jet events, simulated with OSCAR_2_4_5, digitised with ORCA_7_6_1, adding a low-luminosity pile-up. They have been preselected at generator level [89] in order to enhance the fraction of events passing the calorimetric Level-1 trigger. The three samples are divided according to the transverse momentum exchanged in the parton interaction $\hat{p}_T$, in order to cover the region $\hat{p}_T$ > 25 GeV: table 5.1 reports the cross-section (σ_{sample}) associated to each sample, the number of events which have been generated (N_{gen}), preselected (N_{sel}) and used in this analysis (N_{ana}).

$\hat{p}_T$	$\sigma_{sample}[mb]$	N_{gen}	N_{sel}	N_{ana}
25 < $\hat{p}_T$ < 50 GeV	$3.33 \cdot 10^{-1}$	38063731	500000	95639
50 < $\hat{p}_T$ < 170 GeV	$2.41 \cdot 10^{-2}$	27906220	5000000	70084
$\hat{p}_T$ > 170 GeV	$1.33 \cdot 10^{-4}$	622940	500000	68806

Table 5.1: *Parton interaction exchanged transverse momentum $\hat{p}_T$ intervals, cross-section (σ_{sample}), number of events generated (N_{gen}), preselected (N_{sel}) and analysed (N_{ana}) for the three QCD di-jet samples used in this analysis.*

5.3 Tracker Material and Bremsstrahlung

The non-negligible material budget of the tracker and its services combined with the strong magnetic field results in a significant bremsstrahlung away from the primary vertex (“external” bremsstrahlung). The expected distribution of tracker

material averaged in ϕ as a function of η is shown in figure 5.1. The material budget before the electromagnetic calorimeter, varies between 0.4 X_0 and 0.6 X_0 for $|\eta| < 0.5$, grows up to a maximum of about 1.4 X_0 around $|\eta| < 1.5$, the calorimeter barrel endcap transition region, then decreasing at 0.9 X_0 at $|\eta| = 2.5$, corresponding to the end of tracker acceptance region.

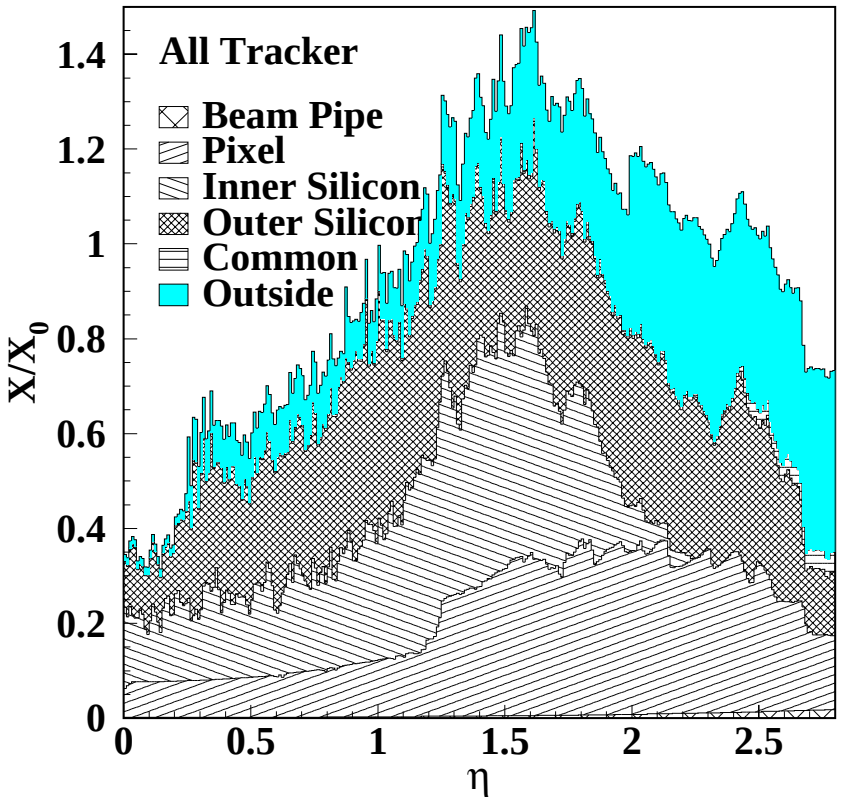


Figure 5.1: *Tracker material budget in terms of radiation lenght X_0.*

The effect of the material budget becomes immediately clear looking at the mean number of external bremsstrahlung photons as a function of the pseudo-rapity (figure 5.2). In this figure and in the following ones regarding bremsstrahlung, a cut at 10 MeV is made on the photon energy, thus requiring $E_\gamma > 10$ MeV.

As it is expected, the tracker material distribution is directly related to the probability of external bremsstrahlung emission: the mean number of bremsstrahlung photons emitted by an electron in the tracker is between 4 and 12, depending on η. The cumulated distribution of the number of photons emitted is shown in figure 5.2: in some cases up to 25 photons are emitted. In figure 5.3, the radius in the transverse plane of bremsstrahlung emission vertices has been plotted; the bremsstrahlung position distribution clearly follows the tracker layout. The elec-

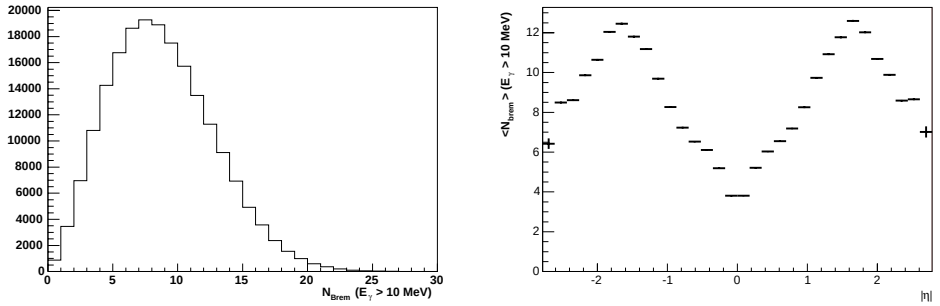


Figure 5.2: *Distribution of the number of bremsstrahlung photons (left) and mean number of external bremsstrahlung photons plotted versus the pseudorapity (right).* $E_\gamma > 10$ *MeV.*

trons which have not radiated before reaching the calorimeter, or better which have not emitted any photon with energy greater than 10 MeV are roughly 3%.

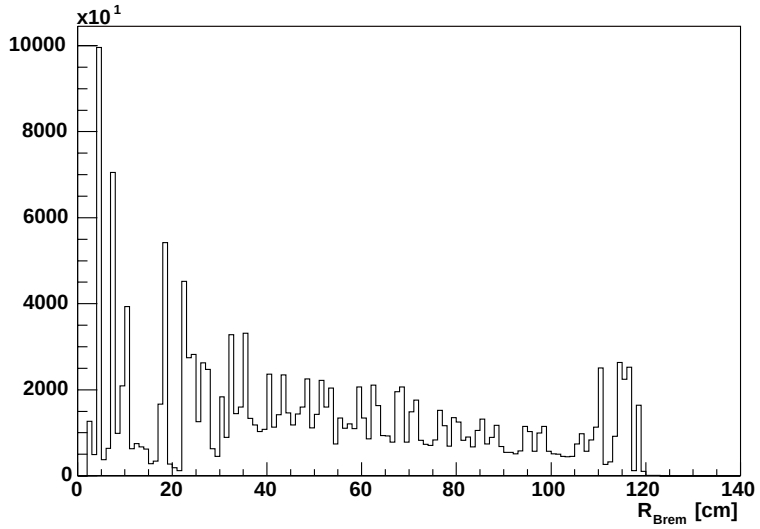


Figure 5.3: *Radius in the transverse plane of bremsstrahlung emission vertices.*

To quantify the effects of bremsstrahlung, the total fraction of bremsstrahlung energy emitted from electrons of 10-50 GeV p_T in the barrel is presented in figure 5.4. The cumulative distribution shows that 57.5% (89.1%) of electrons have lost more than 50% (10%) of their initial energies. Even more, in 19.8% of cases, electrons lose more than 90% of their initial energy. On the same figure, the energy distribution of the secondary particles is also given, together with the probability

of having a secondary particle with energy bigger than a given value: 36% of the secondary particles have energy greater that 1 GeV.

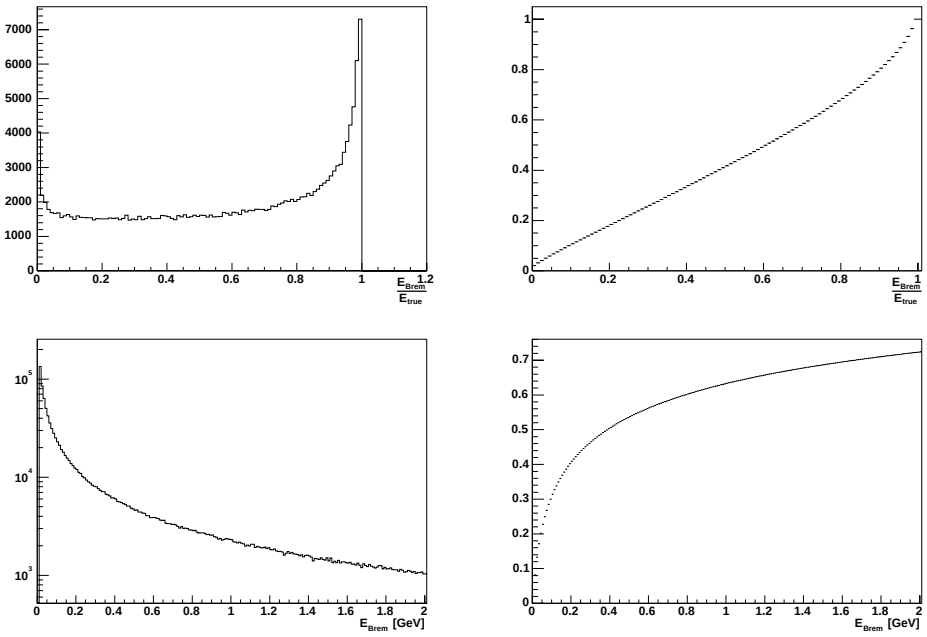


Figure 5.4: *Total fraction of bremmstrahlung energy emitted from electrons of 10-50 GeV p_T in the barrel (top left) and its cumulative distribution (top right). Energy distribution of secondary particles (bottom left) and its cumulative (top right). $E_\gamma > 10$ MeV.*

The kinematics of the bremsstrahlung process is such that the photons are radiated along the tangent to the electron trajectory. Afterwards, for these photons, if no conversion occurs, a straight line trajectory can be assumed, while the electrons are curving in the magnetic field. Thus, the bending of the electron in the 4T magnetic field results in a spray of energy reaching the ECAL. The spreading of this spray is, to good approximation, only in the ϕ-direction. As a result of this, the photon impact point in the ECAL will be more distant from the electron impact in the case of an early radiation (i.e. photon emitted at the beginning of the electron trajectory), while in case of late radiation (i.e. photon emitted near the end of the electron trajectory), the photon cluster is expected to be partially merged with the electron one. The distance between the electron and photon impact point on the calorimeter surface is also reducing when the transverse momentum of the electron is increasing, as it can be seen in figure 5.5, where this distance in the ϕ direction is given as a function of the transverse energy of the electron itself.

In the plot it is required that the photons hit more than 1.5 crystals[1] away from the electron in ϕ direction ($|\Delta\phi| > 0.0262$ rad). In case of low energy electrons, bremsstrahlung photons can be located up to $\pm\Delta\phi = 0.2$ rad far away from the electron or positron (+ or - depending by the charge), corresponding to more than 10 crystals and making the reconstruction of low energy electrons quite difficult. On the contrary, the bremsstrahlung effect for electrons with transverse momentum greater than 40-50 GeV will be much reduced in the calorimeter reconstruction, since the vast majority of the bremsstrahlung photons are contained in region of $\pm$ 1.5 crystals around the electron impact position, with a high probability to get its shower merged with the shower of the electron.

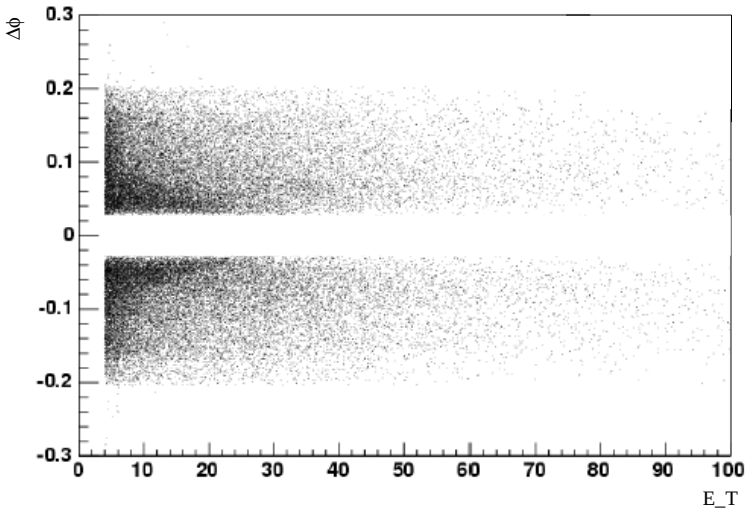


Figure 5.5: *Distance in the ϕ direction between the bremmstrahlung photons impact points and the electron impact point as a function of the transverse energy of the electron itself. In the plot it is required that the photons are hitting more than 1.5 crystals away from the electron in ϕ direction ($|\Delta\phi| > 0.0262$ rad).*

For what concerns the tracker, a late radiation will have a rather small effect on the reconstructed track parameters. On the contrary, an early radiation will lead to reconstructed track parameters moving away from the original electron momentum. All of these effects depend on the hardness of the radiated photon momentum, with increasing importance as the photon momentum increases.

[1]Crystal dimension in ϕ is 0.0175 rad.

5.4 Electron Reconstruction inside ORCA

The default ORCA algorithms for electron reconstruction were used in the data analysis. These algorithms are contained in the `ElectronPhoton` package inside ORCA; this package is the basis of the HLT (High Level trigger) event selection and offline reconstruction for electrons and photons. The ElectronPhoton subsystem includes facilities for basic calorimeter clustering right through to electron track reconstruction with track-cluster matching between the tracker and the ECAL. A detailed description of the electron selection and reconstruction algorithms is available in the DAQ TDR [90]. At the moment the offline and HLT reconstructions follow the same steps, since an optimized offline electron reconstruction has not been developed yet, waiting for detailed analysis of high statistics test beam data. It is possible to identify three main stages of reconstruction of an electron. These different stages follow the HLT steps:

- Level 2: Calorimeter clustering
- Level 2.5: Matching of calorimeter cluster to pixel hits
- Level 3: Track reconstruction starting from the pixel seeds

5.4.1 Calorimetric Reconstruction

The first stage of the reconstruction is the clustering of energy deposits in the ECAL. The details of the actual implementation of the CMS electron clustering algorithms can be found in this CMS note [91]. The CMS approach to clustering is a pattern recognition procedure, which starts from the search of the *seed* crystals, crystals having an energy above a certain threshold. Starting from the same definition of seeds, two different clustering algorithms are employed: the "Hybrid" algorithm is used in the barrel, and the "Island" algorithm is used in the endcap.

Island Clustering

The island algorithm starts by ordering the list of seed crystals, removes all the adjacent seeds, keeping the the most energetic ones. The algorithm then forms clusters of crystals around each seed, searching first in both directions in ϕ, adding to cluster crystals until a rise in the energy is found, or crystal energy is below a fixed threshold. The algorithm then moves one step in η and makes another ϕ search. The η-steps are stopped with the same conditions used in the ϕ search.

Once all the island clusters have been constructed, a bremsstrahlung recovery procedure is made, which consists in the association of multiple calorimeter

clusters with a single electron. This procedure is known as "superclustering". Superclustering starts from the most energetic cluster in a given region, and searches in ϕ for other clusters which may be bremsstrahlung photons. In this way, all the clustered energy emitted by an electron and any external bremsstrahlung photons can be fully summed. At this stage also the energy released in the preshower detector (which is about 3 X_0 thick) is summed.

Hybrid Clustering

The hybrid algorithm starts treating the seeds as the island algorithm, but now the clustering procedure is optimized for the barrel geometry. Crystals are clustered in "dominoes" (rows of 3 or 5 crystals in the η direction) tailored to the electron shower shape and the η,ϕ geometry of the barrel. In order to perform bremsstrahlung recovery, the hybrid algorithm searches dynamically in a ϕ road on either side of the primary cluster. By default the hybrid algorithm searches up to $\pm$10 crystals in ϕ. If any energy deposits above a threshold are found in this search window the algorithm clusters these as well, and associates them to the primary cluster. In effect the hybrid algorithm produces "superclusters" in a single pass. All the standard default values for hybrid clustering were used; they are shown in table 5.2. Currently their values is optimized for the noise level present in the simulation (40 MeV in the barrel).

Hybrid Parameter	**Default**
Minimum E_T for hybrid seed crystal	1 GeV
Number of crystals to search in ϕ	10
Threshold for promoting 1×3 domino $\rightarrow 1\times5$	1 GeV
Threshold for using a domino	0.1 GeV
Minimum domino to make a disconnected subcluster	0.35 GeV

Table 5.2: *Default clustering parameters for the hybrid algorithm in ORCA.*

Electron energy and position reconstruction performances

The measurement of energy in the crystals is obtained by a simple addition of the deposits measured in the crystals. Even in the areas not covered by the preshower detector, the energy containment of the clustered crystals is not complete. The reconstructed energy distribution over the true energy distribution shows a peak at a few percent less than unity, and a long tail on the low side due to unrecovered bremsstrahlung energy. The Gaussian part of the distribution corresponds, roughly, to the energy that would be reconstructed from an electron in the absence

of bremsstrahlung. At present some corrections for the energy scale are applied, designed to place the gaussian peak at 1.0, which are parametrized in terms of the number of crystals in the cluster ($f(N_{cry})$ correction). Figure 5.6 shows, as an example, E_{meas}/E_{true} as a function of the number of crystals in a reconstructed Hybrid supercluster, for electrons in the range $10 < p_T < 50$ GeV, together with a fitted polynomial function. Since the average number of crystals (N_{cry}) is dependent on E and η, this correction can help to reduce the energy scale dependence on these parameters. Figure 5.7 presents the distribution of N_{cry} for the Hybrid algorithm, and the dependence of the average number of crystals on the true energy of the electron and on the pseudorapidity (the η coordinate has been divided into 17 bins of dimension 0.0875 corresponding to the dimension of a trigger tower). The average number of crystals is increasing with the energy, since an increasing number of crystals will be over the supercluster energy threshold, and is also increasing with the crystal with η, because of the increasing tracker material (hence the increasing number of distant bremsstrahlung clusters to be recovered).

Figure 5.8 shows the distributions for the corrected supercluster energy, E_{SC}/E_{TRUE}, for electrons having energies in the intervals $5-10$ GeV, $30-35$ GeV and $80-85$ GeV. The gaussian part, corresponding roughly to the energy that would be reconstructed in the absence of bremsstrahlung, and a tail induced by the bremsstrahlung are visible. In order to characterize the gaussian part of the distributions, the parameters μ_g and σ_g of an iterative gaussian fit in the range $[-1.5\sigma, 2\sigma]$ (a range which has been chosen optimizing the absolute χ^2 of the fit and the number of bins entering in the fit) can be used; the effect of the tails can be evaluated comparing these numbers to the mean of the distributions μ and to the effective sigma σ_{eff}, defined as the half-width containing 68.3% of the distribution (if the distribution is Gaussian, then σ_{eff} is just the Gaussian sigma, while if the distribution has more significant tails, then σ_{eff} provides some measure of this).

In figure 5.9 the results for barrel electrons, reconstructed with the Hybrid supercluster are presented. The left plot of figure 5.9 shows μ and μ_g as a function of the true energy. In spite of the correction $f(N_{cry})$, there is a residual dependence of the energy scale on the energy. In the right plot of figure 5.9 the σ_g and σ_{eff} as a function of the energy are presented.

From a fit to σ_g with the resolution function

$$\frac{\sigma_g(E)}{E} = \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c, \tag{5.1}$$

the energy resolution for barrel electron in the simulation can be extracted:

$$\frac{\sigma_g(E)}{E} = \frac{3.2 \pm 0.6\ \%}{\sqrt{E/\mathrm{GeV}}} \oplus \frac{0.25 \pm 0.01\ \mathrm{GeV}}{E} \oplus 0.69 \pm 0.03\ \%. \tag{5.2}$$

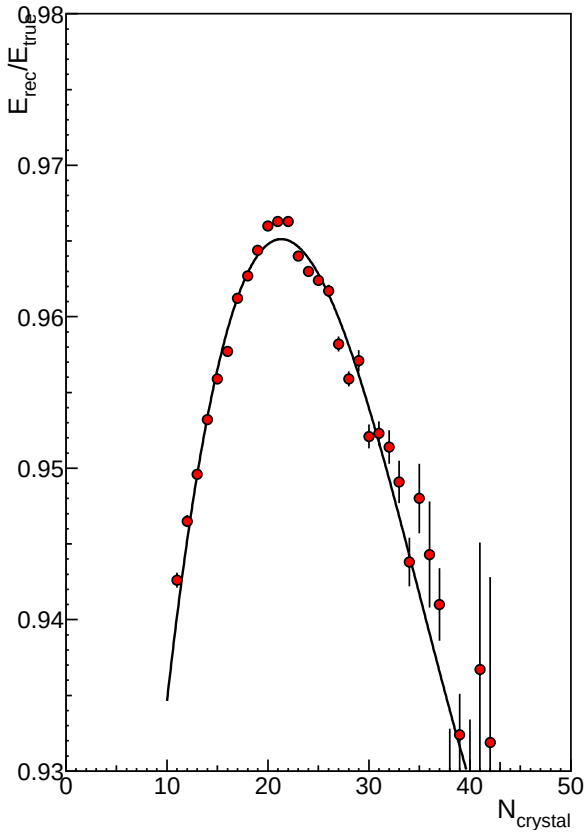


Figure 5.6: *Energy corrections parametrised in terms of the number of crystals in the cluster, $f(N_{cry})$. These corrections are designed to place the gaussian peak of the reconstructed energy distribution at the true montecarlo electron energy.*

In the simulation it is assumed a photostatistics conversion value of 4.0 p.e./MeV for the barrel and 6.0 p.e./MeV in the endcap, an uncorrelated electronic noise at a level of 40 MeV per channel in the barrel and 150 MeV in the endcap, and a constant term of 0.5%. Comparing this resolution with the one obtained in the test beam condition with a 5*x*5 algorithm (equation 4.16) it is possible to note an higher constant term, as a result of the bremsstrahlung effects; the stochastic term *a* is instead similar to what has been found in the test beam analysis. The noise level assumed in the simulation is much better than what observed in the 2003 test beam data with the old FPPA electronics, but it is comparable with what has been measured with the new MGPA electronics.

The major effect of the bremsstrahlung is however in the tails: the effective sigma is much worse than the gaussian one especially at low energies, becoming comparable only with increasing energy.

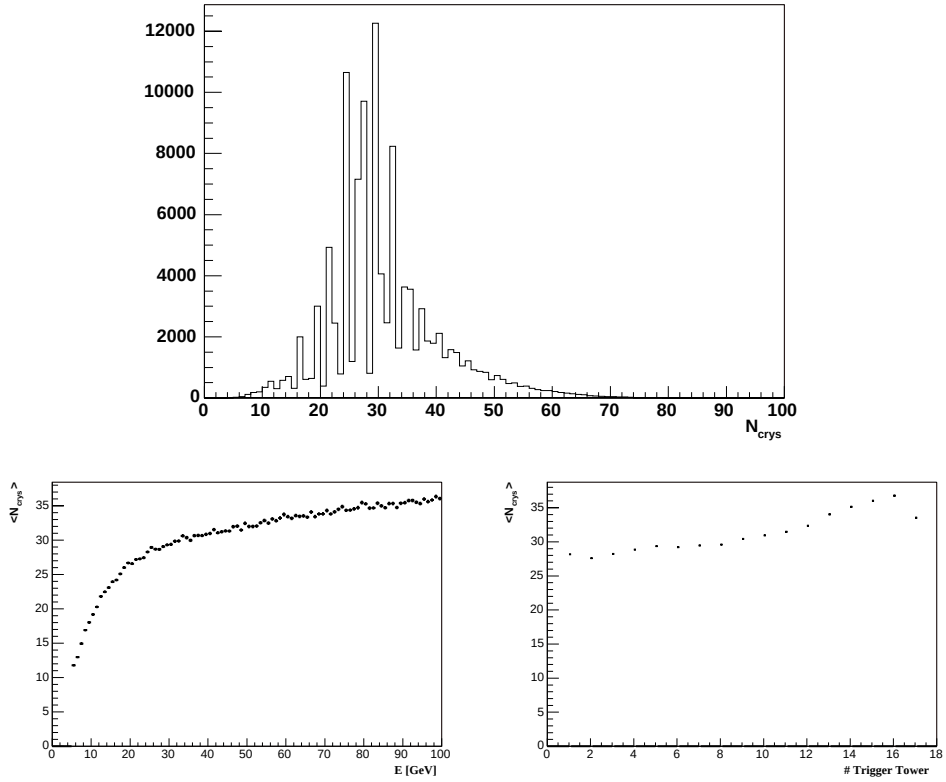


Figure 5.7: *Distribution of N_{cry} for the Hybrid algorithm (top), and dependence of the average number of crystals on the true energy of the electron (bottom left) and on the pseudorapidity (bottom right) (the* η *coordinate have been divided into 17 bins of dimension 0.0875 corresponding to the dimension of a trigger tower). The number of crystals distribution present peaks due to the fact that dominoes of fixed dimension in the* η *direction are employed.*

Another aspect which is worth mentioning is the resolution on the reconstructed position using only calorimetric informations. A simple position measurement of the shower can be obtained by calculating the mean position of the crystals in the cluster weighted with the logarithm of the energy [91]. In the CMS detector this simple picture is complicated by the non-pointing geometry, which requires some definition of the crystal (η, ϕ), that is changing along the crystal axis due to the calorimeter non pointing geometry. In the position measurement used for both Island and Hybrid super-clusters, crystal position is calculated at the shower maximum depth along the crystal axis. The performances in reconstructing the η and ϕ direction of the electron are presented in figure 5.10.

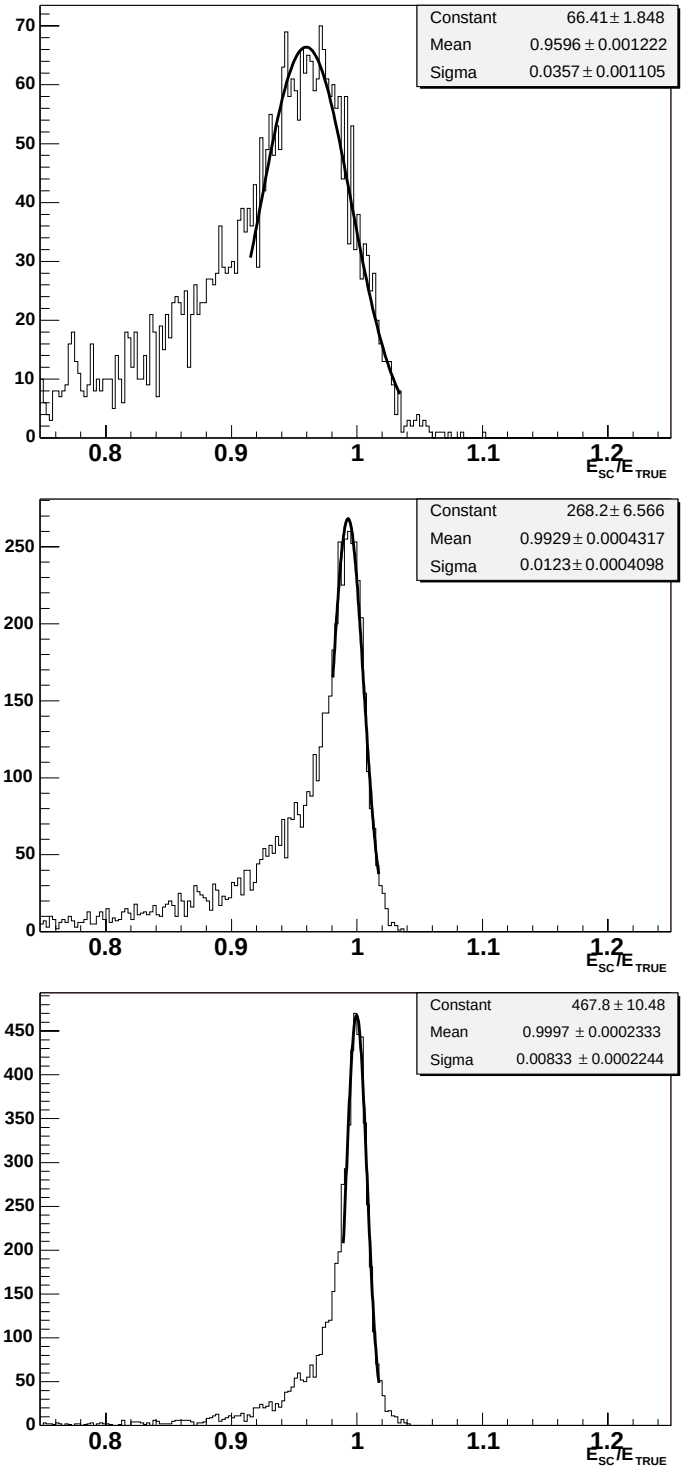


Figure 5.8: *Distributions for E_{SC}/E_{TRUE} for electrons having energies in the intervals* $5-10$ *GeV (top),* $30-35$ *GeV (center) and* $80-85$ *GeV (bottom). More details are given in the text.*

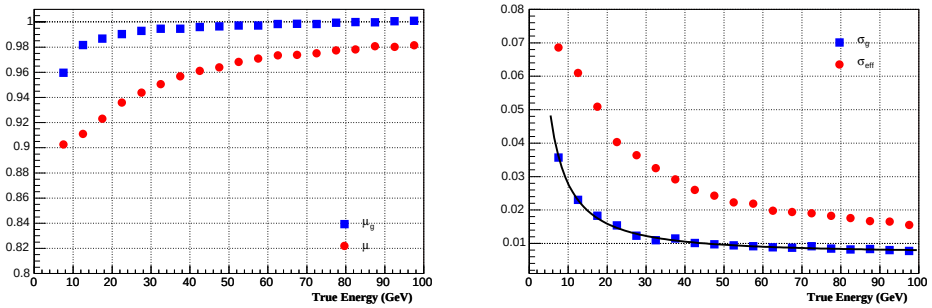


Figure 5.9: *Gaussian mean μ_g and the mean μ of the distribution of ratio of reconstructed energy over true energy as a function of the true energy (left), gaussian sigma σ_g and the effective sigma σ_{eff} as a function of the energy (right). Plots for barrel electrons reconstructed with Hybrid supercluster.*

5.4.2 Pixel Matching

Matching calorimeter clusters to hits in the pixel system allows electron-photon separation, and a first rejection of jets which may "fake" electrons. The method is described in detail in a CMS technical note [92].

First, the supercluster position and energy are used to calculate a transverse momentum for the electron. Then an electron candidate trajectory is propagated from this end point, back through the magnetic field, to the primary vertex. This defines a predicted hit position in the first layer of the pixel detector. A search area in the innermost pixel layer is defined around this predicted hit position. The search region in z is $\pm 15cm$, determined by the smearing on the primary vertex[2]. The search region in ϕ can be varied with some trade-off between efficiency and electron quality. It has been found that cuts $-0.025 < \phi_e < 0.015$ for electrons, and $-0.015 < \phi_p < 0.025$ for positrons provide a good selection, with high efficiency. If no hit is found in the innermost pixel layer the pixel matching algorithm tries again in the second layer.

If a compatible hit is found, it is used to predict the z coordinate of the primary vertex: the predicted trajectory from the cluster and pixel hit is extrapolated to the beam axis. A new track trajectory is then predicted, starting from the predicted primary vertex (0,0,z) and passing through the compatible pixel hit found in the first step. This new trajectory is propagated to the next pixel layer, where the algorithm searches for a second compatible hit. If a second hit is found, the supercluster is identified as an electron. The pixel matching algorithm performs the whole search twice, predicting trajectories once for an electron hypothesis and

[2]In CMS, $\sigma_z = 5.3$ cm

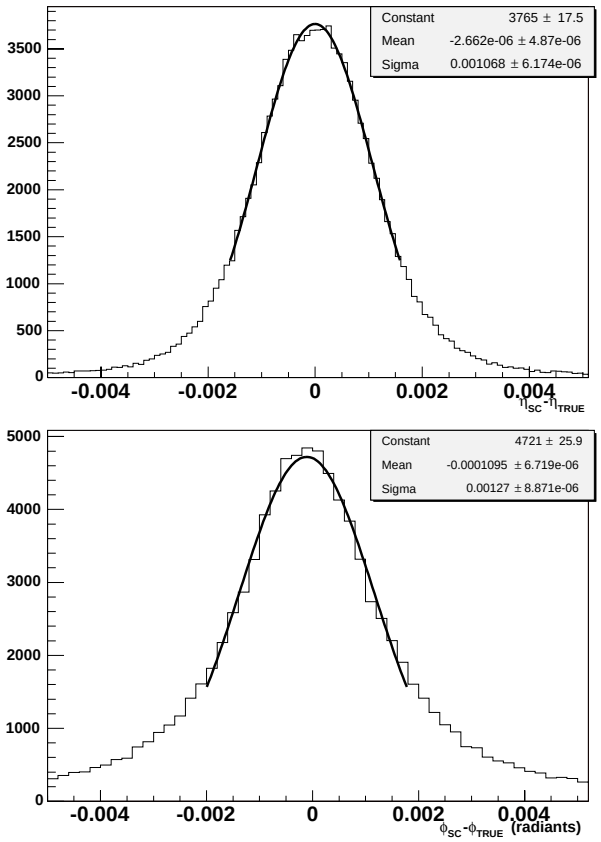


Figure 5.10: *Resolution in eta (top) and phi (bottom) using the log-weighting techinque for Hybrid Superclusters.*

once for a positron hypothesis.

5.4.3 Track Reconstruction

Once an electron track has been identified, full track reconstruction takes place starting from the pixel seed trajectory. Currently this is done using a Kalman Filter algorithm [93], although more specialised electron track reconstruction algorithms are envisaged [94]. The track fitting itself can be broken down into two basic operations, which are encapsulated into several classes in the ORCA framework:

- Propagating the track. The propagator extrapolates track parameters and their error matrix from one layer to the next. It takes into account the average energy loss and multiple scattering in each layer of the tracker. The track state vector is corrected with the expected mean value of the

bremsstrahlung energy loss. Between tracker layers, all material effects are neglected

- Updating the track parameters. Once the track parameters have been propagated onto the next tracker layer, the updator class modifies them using the reconstructed track hit in the new layer. The Kalman filter track updator class in ORCA calculates a χ^2 from the residuals of the extrapolated track parameters on the new surface and the reconstructed hit on that surface. If the χ^2 value exceeds a specified limit the hit is discarded.

If a minimum number of compatible track hits is found, which is fixed at 3, the track is then smoothed. The track smoother algorithm starts from the outermost compatible track hit and propagates the track back towards the primary vertex. At each layer the track parameters are updated using all the compatible track hits.

Reconstructed tracks in the CMS apparatus are described by six parameters:

- Impact point (x_{imp}, y_{imp}, z_{imp}) which is the point of closest approach to the beam axis in the transverse plane
- Track pointing angle at the impact point ($\cot\theta$, ϕ_0). θ is the track azimuthal angle, and ϕ_0 is the polar angle in the transverse plane, at the impact point
- Track curvature $c = \frac{B_z q}{p_T}$.

Figure 5.11 presents the distribution of the reconstructed momentum of the electron over its true energy for electrons in energy intervals $5-10$ GeV, $30-35$ GeV and $80-85$ GeV. Low energy tails are present, due to soft bremsstrahlung photons (hard bremsstrahlung effect should be reduced by the hard cut in the updator χ^2). The low momentum tail also increases when going to higher momenta, as expected from radiation losses.

The better reconstruction of low energy electrons in the tracker will be used in section 5.6 where an optimal estimator of the electron energy will be proposed, which makes an optimal combination of the measures of the electron energy and direction measurements coming separately from the calorimeter and from the tracker.

As for the calorimeter distributions, the intrinsic resolution and the effect of the tails can be quantified looking at the parameters μ_g and σ_g of an iterative gaussian fit performed in the [-1.5σ,2σ] region around the peak value compared to the mean μ and to the effective sigma σ_{eff} of the distributions. In figure 5.12 the left plot shows μ_g and μ as a function of the electron true energy; in the right one σ_g and σ_{eff} versus true energy are presented. The resolution is worsening at higher momenta, while the gaussian peak is slowly drifiting towards lower values.

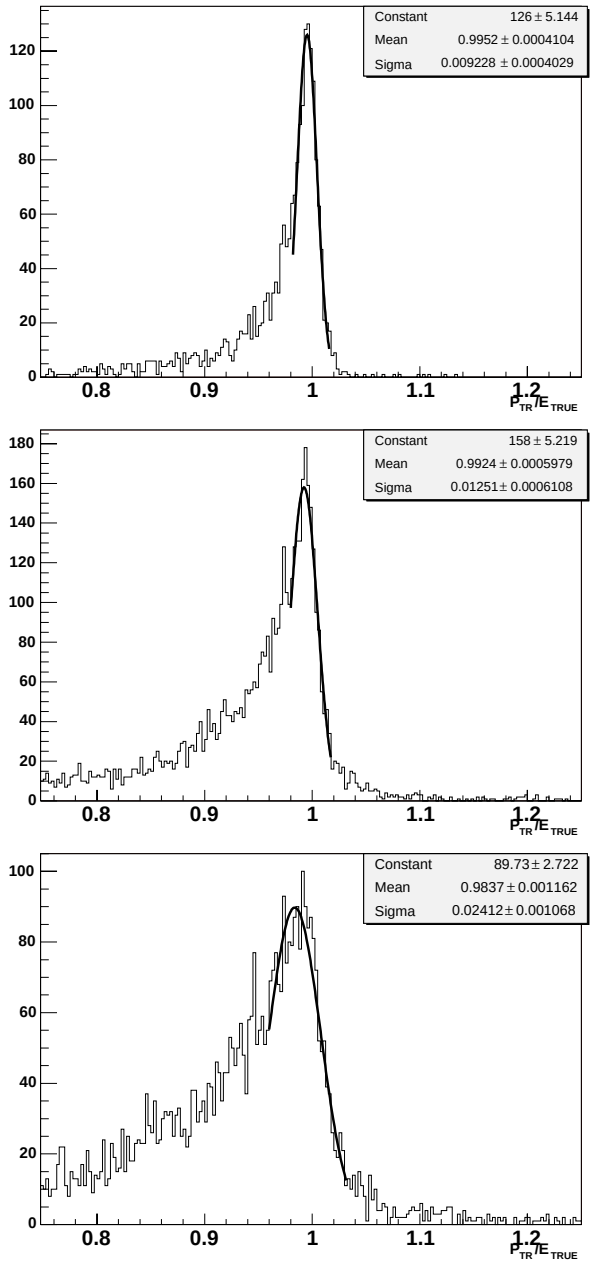


Figure 5.11: *Distributions of P_{TR}/E_{TRUE} for electrons having energies in the intervals* 5 – 10 *GeV (top),* 30 – 35 *GeV (center) and* 80 – 85 *GeV (bottom).*

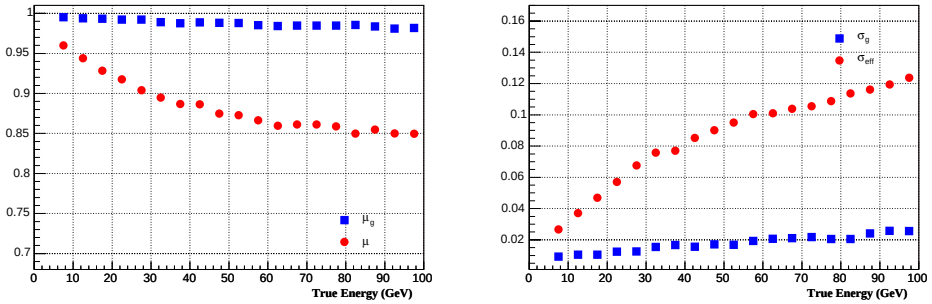


Figure 5.12: *Gaussian mean μ_g and the mean μ of the distribution of ratio of reconstructed momentum over true energy as a function of the true energy (left), gaussian sigma σ_g and the effective sigma σ_{eff} as a function of the energy (right).*

The quality of the track measurement is also dependent on the number of hits associated to the track. The number of hits of the reconstructed electrons tracks is shown in the left plot of figure 5.13, while on the right the average number of hits is given as a function of η. The behaviour of the average number hits is the combination of the two effects of the varying number of possible layers, which can be traversed by every track versus η and the overall tracker material, increasing the probability to have hard bremsstrahlung, hence to stop the trajectory building step. This means that the number of hits associated to the track can be also efficiently used to select electrons with a little radiated energy, as it will be discussed in section 5.5.

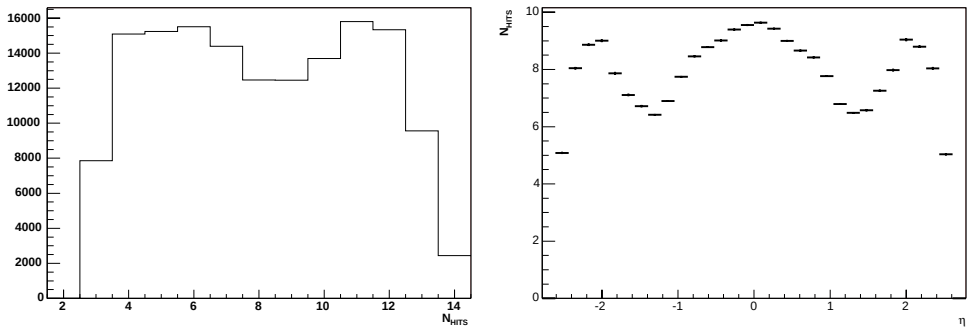


Figure 5.13: *The number of hits associated to the reconstructed electrons tracks (left); dependence of the average number of hits on η*

σ_g and σ_{eff} versus the number of hits is presented in figure 5.14; as expected, an increasing number of hits associated to the track will risult in a more precise momentum estimation.

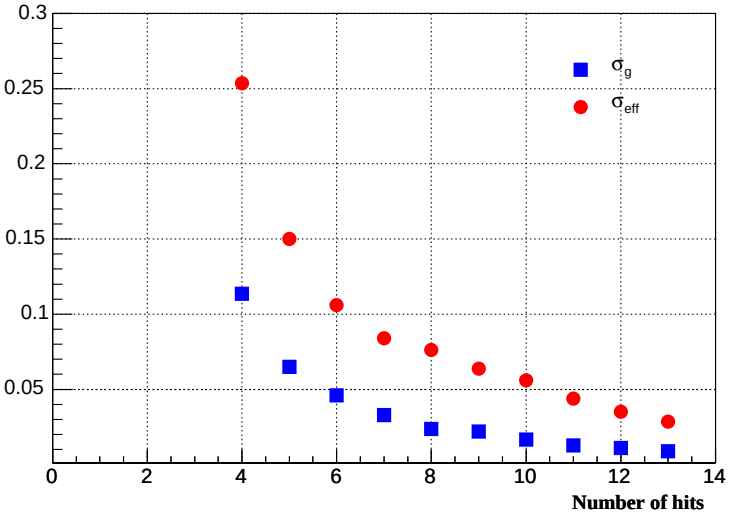


Figure 5.14: σ_g *and* σ_{eff} *of the* P_{TR}/E_{TRUE} *distribution as a function of the number of hits asscoiated to the track*

The resolution in the η and ϕ direction is given in figure 5.15, showing that a resolution better then one order of magnitude as respect to the calorimeter direction reconstruction can be achieved. In the ϕ direction tails are present due to the effect of the bremsstrahlung.

Charge reconstruction is very efficient thanks to the strong magnetic field of 4T. Charge misidentification (e^- reconstructed as e^+ or viceversa) has been evaluated to be 0.7% for electrons having energies below 150 GeV.

5.4.4 Clustering and tracking efficiencies

The electron efficiencies at the two steps of calorimetric reconstruction and tracker reconstruction are shown in figure 5.16, as a function of transverse momentum p_T and of pseudorapitidity η. The efficiency here is defined as the ratio of reconstructed electron at each step divided by the number of generated electrons. In particular the cluster reconstruction efficiency reaches a plateau of 0.99 above $p_T = 30$ GeV, presenting a sharp decrease for transverse momentum below 10 GeV. Looking at the behaviour with pseudorapidity the cluster reconstruction efficiency is almost constant in whole calorimeter, presenting a sharp drop in the transition region between barrel and endcap around $\eta = 1.5$. Including also the track reconstruction efficiency, the global efficiency achieves 0.94 for $p_T > 50$ GeV; in the p_T region between 5 and 10 GeV, the global electron reconstruction efficiency is around 0.52. In the pseudorapidity direction η the electron track re-

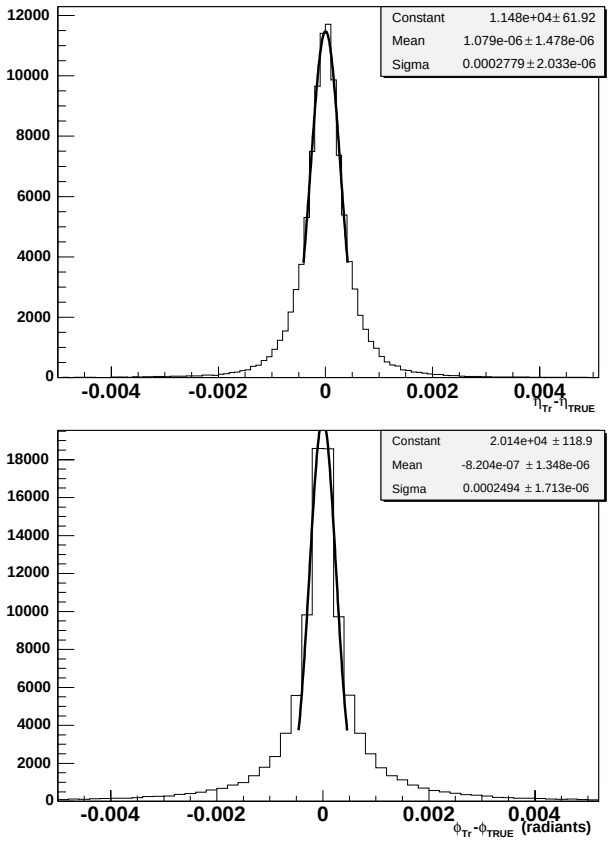


Figure 5.15: *Resolution in eta (top) and phi (bottom) of track direction measurement.*

construction efficiency follows the behaviour of the number of average track hits already discussed in section 5.4.3: in central barrel region reaches its maximum, decreasing up to the barrel-endcap transition region, then it rises up to $\eta = 1.9$, rapidly falling afterwards due to the decreasing number or tracker layers traversed by the track.

5.5 Bremsstrahlung effect on calorimeter and tracker reconstruction

In this section the effect of the bremmstrahlung is analysed in detail both for the calorimeter and for tracker electron measurement. It is important, especially for calorimeter calibration methods (see section 3.6), to deeply understand these effects, in order to disentagle it from the real intercalibration procedure. For this reason, the possibility to select only those electrons which have emitted a reduced

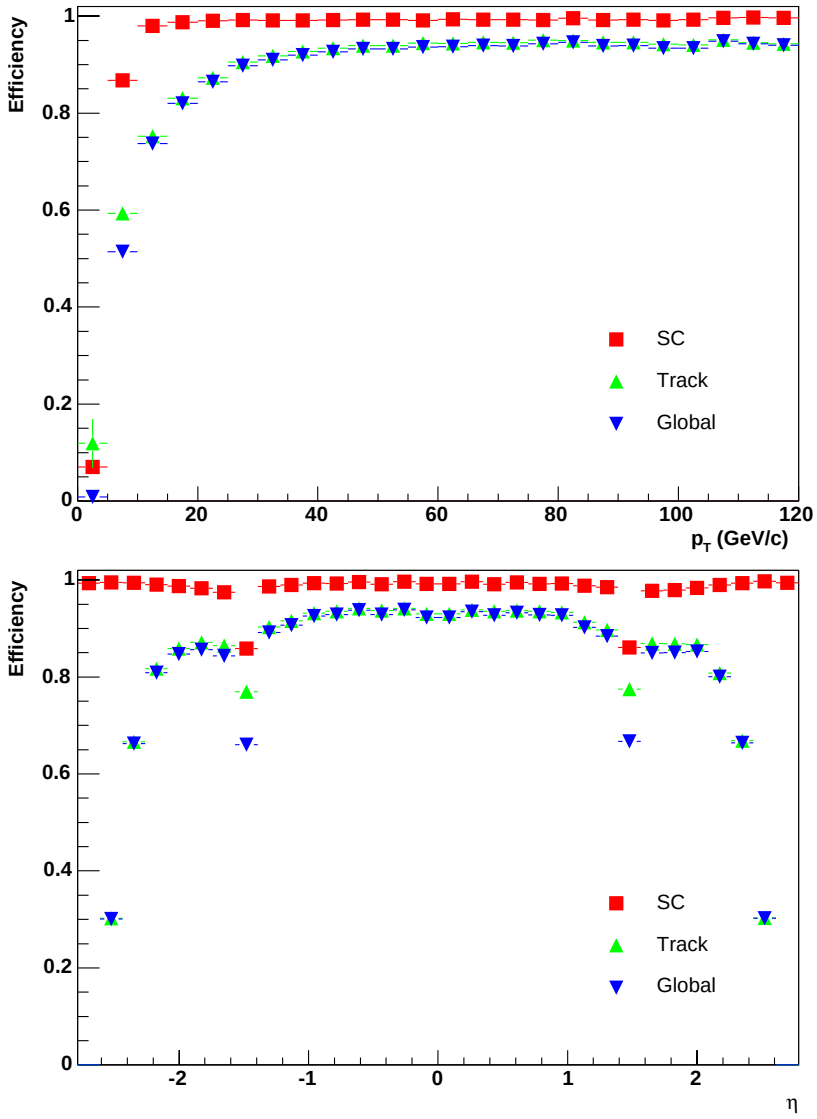


Figure 5.16: *Electron reconstruction efficiencies at the two steps of calorimetric reconstruction and tracker reconstruction as a function of transverse momentum p_T (top) and of pseudorapitidity* η *(bottom).*

fraction of their initial energy has been envisaged [95]; for this purpose some reconstructed variables have been identified and described.
This analysis has been performed on electrons in the barrel acceptance region having a transverse momentum p_T in the interval 10-50 GeV.

As a useful tool, the concept of effective radiated energy can be introduced;

this quantity, constructed from Montecarlo informations, selects only the fraction of the energy radiated by the electron which is really affecting the reconstruction. Such a variable has been identified in the sum of the energy of the bremsstrahlung photons (E_γ > 10 MeV) emitted in the transverse plane before 80 cm:

$$E_{BREM}^{EFF} = \sum_{r_{brem}<80cm} E_\gamma \tag{5.3}$$

In fact, looking either the gaussian sigma and effective sigma of the energy or momentum measurement as a function of the radius of emission of the first bremsstrahlung photon (figure 5.17), it appears that the effect on reconstruction is negligible for photons emitted after a radius of 80 cm in the transverse plane (the last point in the plots at r=129 cm, corresponding to the calorimeter barrel radius, is for electrons which have not radiated).

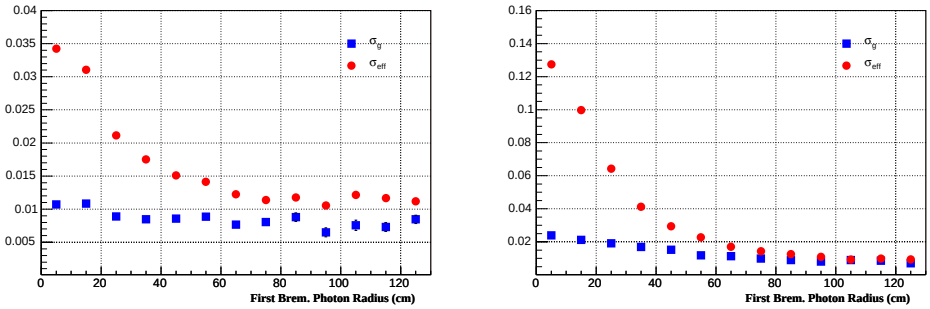


Figure 5.17: *Gaussian sigma and effective sigma of the energy (left) and momentum (right) measurement as a function of the radius of emission of the first bremsstrahlung photon.*

Thus late radiation (>80 cm) is not affecting too much the reconstruction, as it is expected from the qualitative description made in section 5.3.

Dependence of the reconstruction quality parameters with respect to the fraction E_{BREM}^{EFF}/E_{TRUE} are shown in figure 5.18.

Another aspect to be considered is that, due to the varying tracker material with pseudorapidity coordinate η, both energy scale and resolution becomes dependent on η. Gaussian mean and mean of the normalized energy and momentum measurement is given in the top plots if figure 5.19; the bottom plots show instead gaussian sigma and effective sigma.

The calorimeter reconstruction variation with the pseudorapidity η is the result of the combination of the varying material effects and of the varying crystal geometry (both dimensions and relative position). The latter effect has been estimated to give an energy scale variation of around 1.0% in the ECAL barrel acceptance

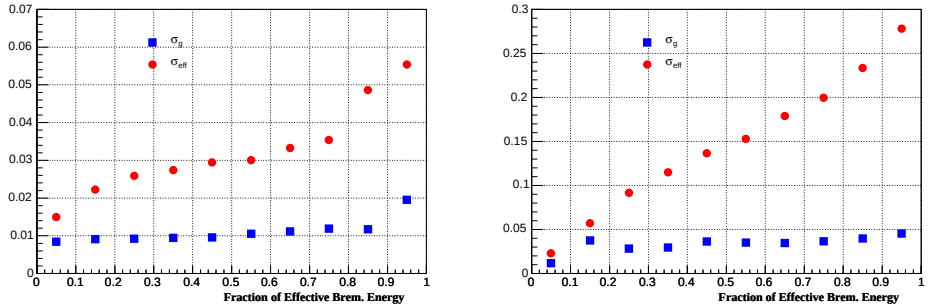


Figure 5.18: *Gaussian sigma and effective sigma of the energy (left) and momentum (right) measurement as a function of* E^{EFF}_{BREM}/E_{TRUE}

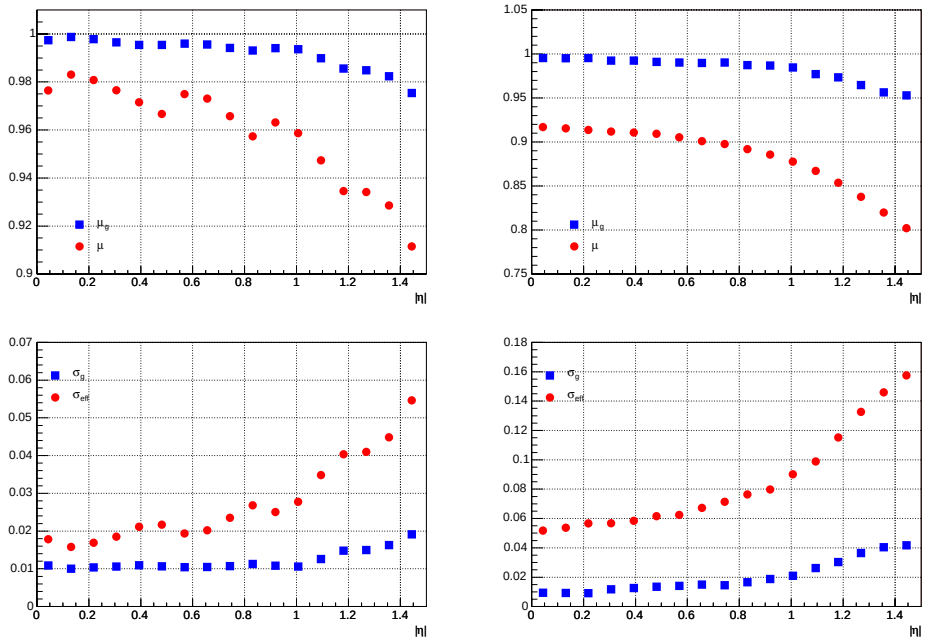


Figure 5.19: *Gaussian mean and mean of the normalized energy (top left) and momentum (top right) measurement as a function of* η*; gaussian sigma and effective sigma for energy (bottom left) and momentum (bottom right) measurement as a function of* η.

region [96] (obtained on a sample of electrons simulated removing the tracker material), mainly due to front leakage of the shower. Material effects increase this value to around 2.7%. The sharp drops around $\eta \simeq 0$, 0.45, 0.8 and 1.15 arise from the energy lost in the gaps present at the separations between the ECAL modules. No correction is applied for the latter effect.

Reconstructed variables capable of selecting electrons with a reduced fraction of radiated energy have been identified looking at the correlation between these variables and the effective radiated energy fraction E^{EFF}_{BREM}/E_{TRUE}. Three calorimeter based variables and one tracker based variable have been investigated.

Calorimeter based variable are constructed using the ratio of energy contained in an array of crystals of fixed and limited size around the seed crystal over the corrected supercluster energy. It has been considered so far:

- E_4: energy sum for highest 2×2 crystal's array energy around the seed crystal
- E_9: energy sum for the 3×3 crystal's array centred on the seed crystal
- E_{25}: energy sum for the 5×5 crystal's array centred on the seed crystal.

The correlation between these variables and E^{EFF}_{BREM}/E_{TRUE} is presented in figure 5.20.

The interpretation of the plots is similar in the three cases: there is a region for high values of the proposed cuts where the energy is mostly contained in the chosen array, giving an indication of an electron with a reduced fraction of radiated energy. Lower values of the variable are result of the fact that the energy is spread over several crystals, indicating that the electron has lost a consistent part of its energy. In the region for $E^{EFF}_{BREM}/E_{TRUE} > 0.5$, therefore in case of hard bremsstrahlung, there is instead an high probability that the cluster is constructed around the shower associated to the photon, which, being tipically separated from the electron one, will result sufficiently isolated. Therefore, the calorimeter isolation variables considered are not capable to reject electrons with a very high fraction of radiated energy, as can be seen looking at the distribution of E^{EFF}_{BREM}/E_{TRUE} applying a tight cut, for example on E_4/E_{SC}, $E_4/E_{SC} > 0.88$ (figure 5.21).

Correlation between N_{hit} and E^{EFF}_{BREM}/E_{TRUE} is given in figure 5.22. Interpretation in this case is quite simple, since with the parameters chosen for the track reconstruction, the track is stopped as soon as a hard bremmstrahlung occurs.

In figure 5.23 the effective sigma as a function of the efficiency is plotted for the four different cuts. It can be noticed that the most performing cut in the low efficiency region is N_{hit}, which is not affected by the problem of the very hard bremsstrahlung as the calorimeter isolation cuts.

The most proper set of cuts on bremsstrahlung will be the result of an inevitable trade-off between efficiency and resolution, and should be chosen and optimized for the particular analysis which is being performed. In general, since these variables are highly correlated between them, the combined cut on more than one variable (e.g. E_4/E_{SC} and N_{hit}) requires an optmization in the multi-dimensional cut phase space.

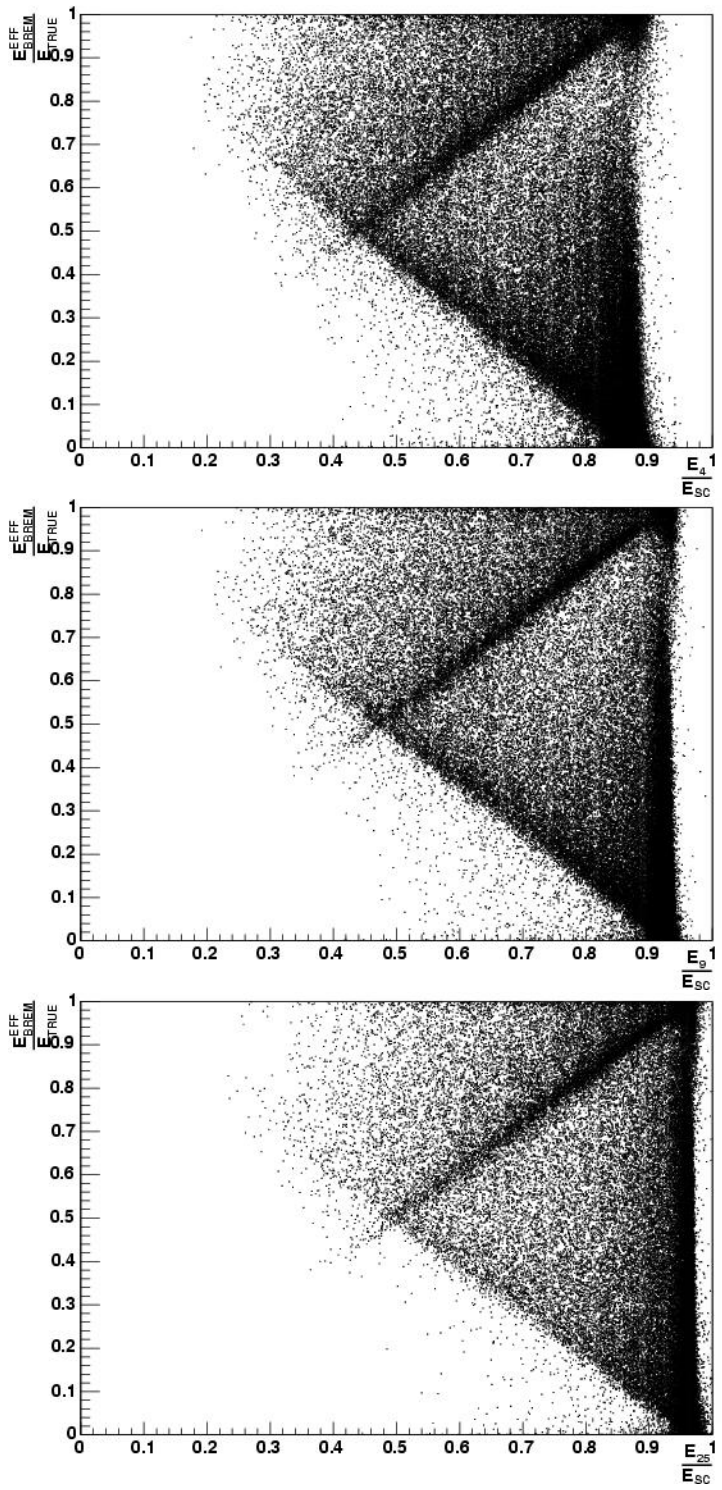


Figure 5.20: *Scatter plots of E_{BREM}^{EFF}/E_{TRUE} versus E_4/E_{SC} (top), E_9/E_{SC} (center), E_{25}/E_{SC} (bottom)*

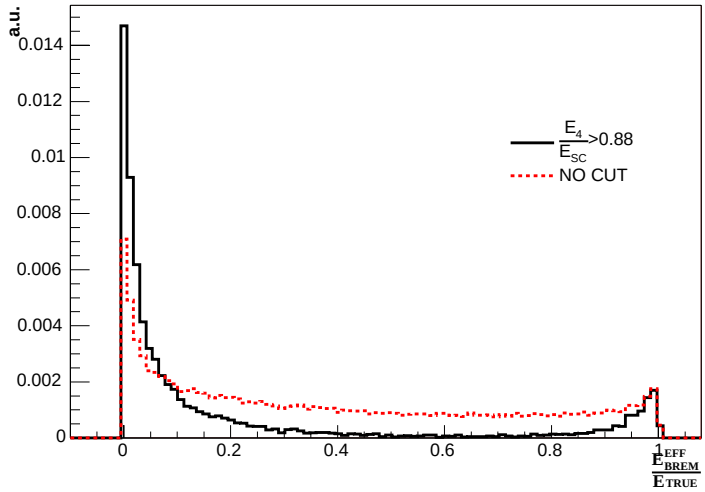


Figure 5.21: E^{EFF}_{BREM}/E_{TRUE} *distributions applying a cut* $E_4/E_{SC} > 0.88$ *(black) and without applying any cut (red). Both distributions are normalized to unity.*

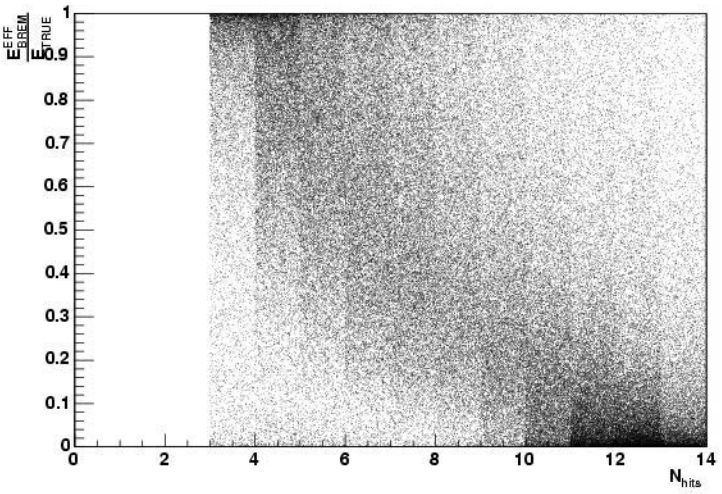


Figure 5.22: *Correlation between* E^{EFF}_{BREM}/E_{TRUE} *and* N_{hits}

5.6 Using calorimeter and tracker information: the combined energy estimator

From the analysis of the separate performances of the calorimeter and tracker, it seems that the two measures can be efficiently combined in order to get an optimal estimator of the electron quadri-impulse. It has already been stressed that the tracker seems to give a better reconstruction for low energy electrons, while the calorimeter gives better performances for higher energy electrons. In figure

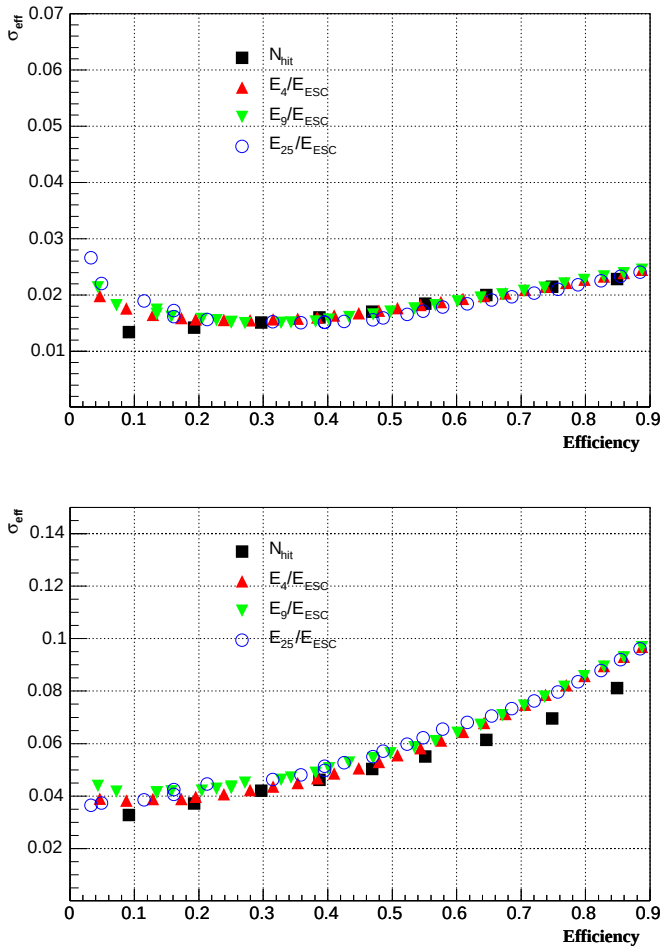


Figure 5.23: *Effective sigma of the energy (top) and momentum (bottom) measurement as a function of the efficiencies on different cuts on E_4/E_{SC} (red) , E_9/E_{SC} (green), E_{25}/E_{SC} (blue), N_{hits} (black).*

5.24 the correlation between the energy and momentum estimation normalized to the electron true energy is shown: most of the tails due to the bremsstrahlung effect of one estimator are contained in the peak region of the other. The plots shown in this section are referred to electrons with energies from 5 to 100 GeV reconstructed in the barrel acceptance region, hencewith the Hybrid algorithm. Similar plots and results are valid also for electrons in the endcap, reconstructed by the Island algorithm.

For what concerns the energy estimation, an optimal estimator can be constructed weighting the energy/momentum measurement for the inverse of their

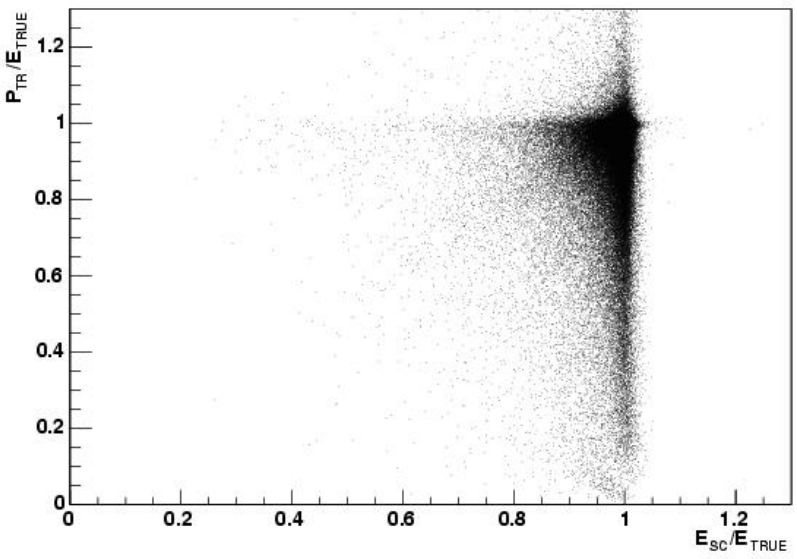


Figure 5.24: *Correlation between E_{SC}/E_{TRUE} and P_{TR}/E_{TRUE}*

variance:

$$\bar{E} = w_E \cdot E + w_P \cdot P \tag{5.4}$$

where $w_E = \frac{1}{\sigma_E^2}$ and $w_P = \frac{1}{\sigma_P^2}$.

Energy resolution σ_E can be evaluated using formula 5.2, while σ_P can be inferred from the event-based estimator given directly by the Kalman Filter algorithm, based on the residual between the track hits and the track fit.

Weighted mean, however, is an unbiased estimator only when both measurements are unbiased. To understand in which region it is possible to use it, the distribution of cluster energy (E_{SC}) and track momentum (P_{TR}) normalized to true electron energy (E_{TRUE}) is plotted as a function of the ratio of the two measurements (E_{SC}/P_{TR} called for simplicity E/P) (see figure 5.25). In order to understand better the behaviour of the E/P ratio itself, the distribution of the effective bremmed energy E_{brem}^{eff}/E_{TRUE} versus the E/P ratio is also analysed (see figure 5.26).

It is possible to identify three regions:

- E/P $\simeq$ 1. In this region both calorimeter and tracker energy measurements are unbiased. In this region it is possible to use the weighted mean
- E/P > 1. This region seems to be dominated by electrons which have had a large radiation, recovered by the calorimeter energy measurements, while affecting the momentum measurement, which is underestimated

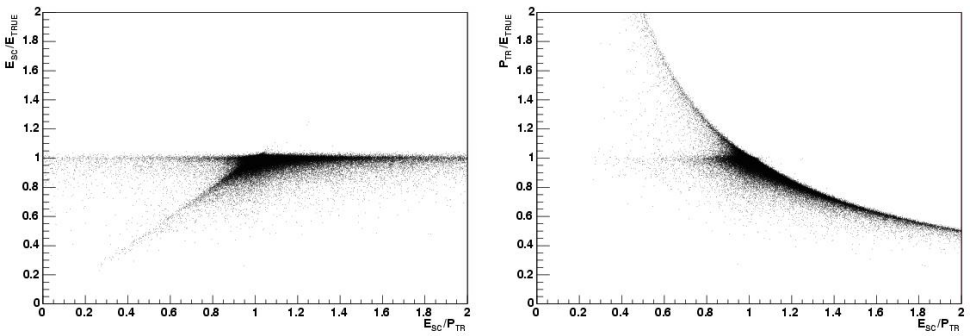


Figure 5.25: *Distribution of cluster energy (left),E_{SC}, and track momentum (right), P_{TR}, normalized to true electron energy (E_{TRUE}) as a function of the ratio of the two measurements, E_{SC}/P_{TR}.*

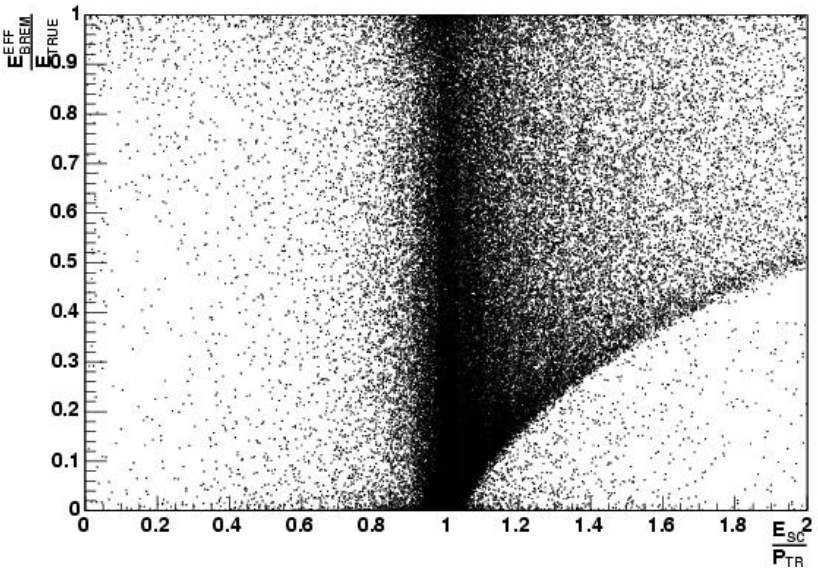


Figure 5.26: *Correlation between E^{eff}_{brem}/E_{TRUE} and E/P.*

- E/P < 1. This region is the most problematic one, since it is mainly composed by electrons which have radiated a very large fraction of their energy. There are electrons for which the calorimeter energy is underestimated (this can happen especially for low energy electrons for which the radiation recovery procedure is more difficult), while there are also events for which the momentum is overestimated, which needs to be deeply analised.

To further investigate the behaviour in the $E/P < 1$ region, the distributions of E_{SC}/E_{TRUE} and P_{TR}/E_{TRUE} versus E/P are plotted for tracks having only 3 hits associated to them and tracks with 11 hits (see figure 5.27).

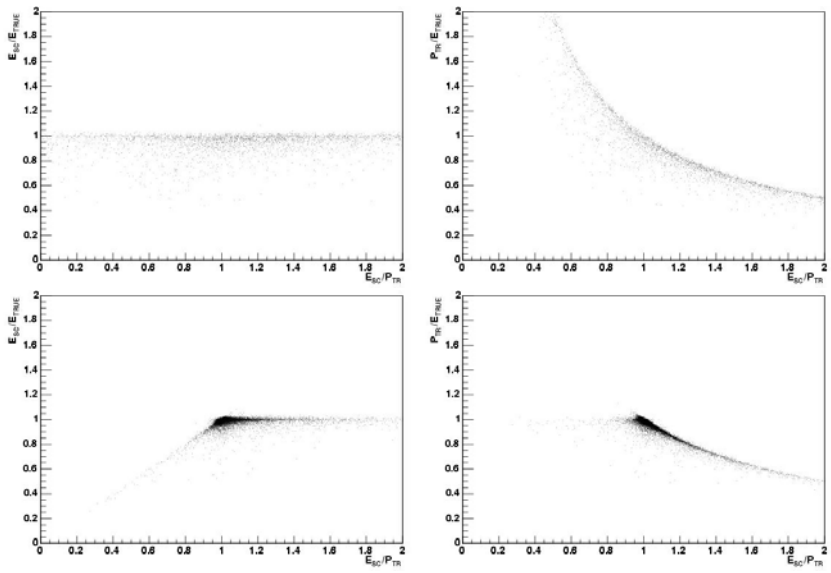


Figure 5.27: *Same distributions in figure 5.25 for tracks reconstructed using 3 hits (top) or 11 hits (bottom).*

In case of a low number of hits, the tracker measurement is not very reliable as it is expetected, while for an increasing number of hits the $E/P < 1$ region starts to be dominated by electrons having an underestimated energy calorimetric measurement. The idea is then to alternatively use in the $E/P < 1$ region the track or cluster measurement according to the number of hits belonging to the track.

The combined energy estimator is therefore defined as:

$$\bar{E} = \begin{cases} w_E \cdot E + w_P \cdot P & , \quad \text{when } \left|\frac{E}{P} - 1\right| < \alpha\sigma_{E/P} \\ E & , \quad \text{when } E/P - 1 > \alpha\sigma_{E/P} \\ p & , \quad \text{when } E/P - 1 < \alpha\sigma_{E/P} \text{ and } N_{hit} \geq 7 \\ E & , \quad \text{when } E/P - 1 < \alpha\sigma_{E/P} \text{ and } N_{hit} < 7 \end{cases} \tag{5.5}$$

An optimal choice for α is found to be equal to 2. In figure 5.28 the distribution of the optimal estimator normalized to true energy is given for electron having energies in the interval 5-10 GeV, 30-35 GeV, 80-85 GeV. Gaussian mean μ_g and mean μ of the normalized optimal energy distributions as a function of the true energy are given in the left plot of figure 5.29; σ_g and σ_{eff} are showed in the right plot.

The effect of the optimal combination is evident; resolution of less than 1% is obtained in all the energy range considered, improving in particular the reconstruction for low energy electrons. The tails of the distributions are reduced, as

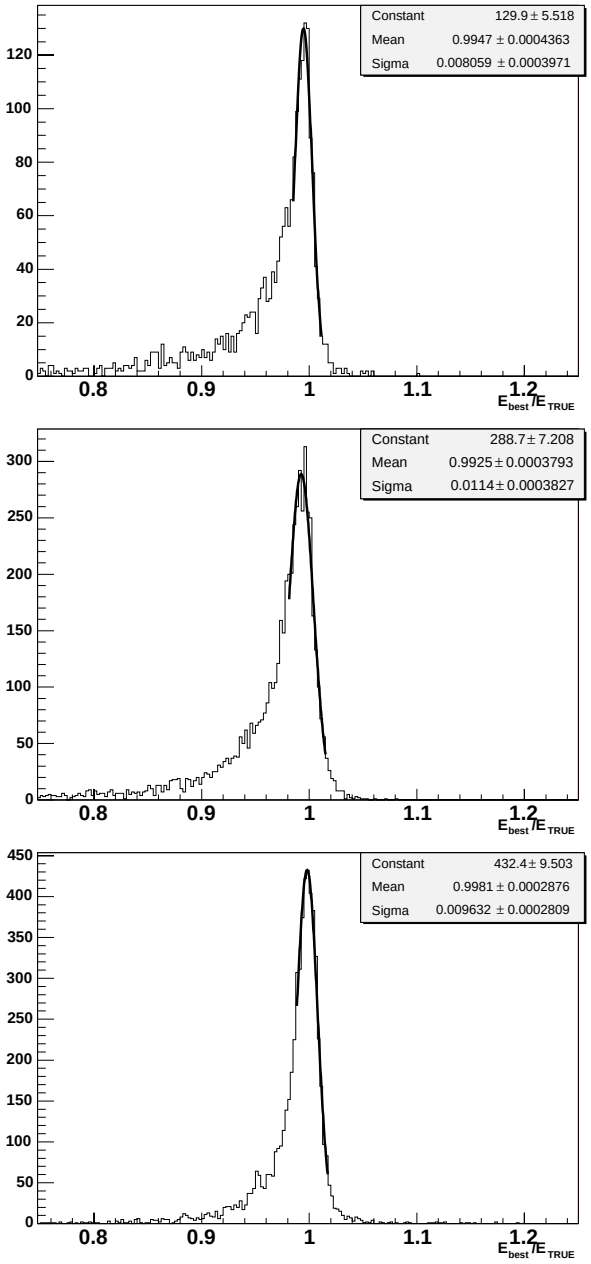


Figure 5.28: *Distributions for* $\bar{E}/E_{TRUE}$ *for electrons having energies in the intervals* 5 – 10 *GeV (left),* 30 – 35 *GeV (middle) and* 80 – 85 *GeV (right).*

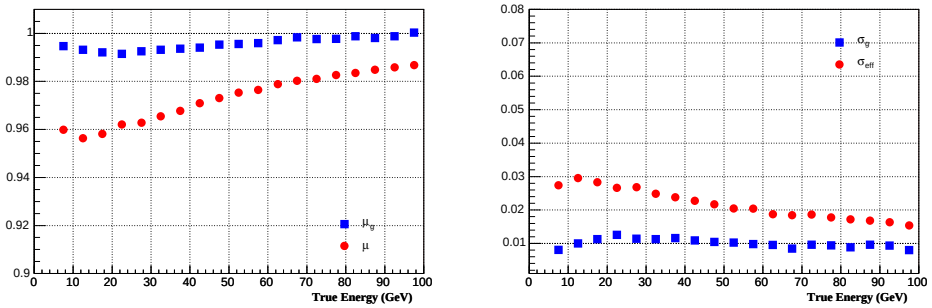


Figure 5.29: *Gaussian mean and the mean of the normalized optimal energy distributions as a function of the true energy (left), gaussian sigma and effective sigma as a function of the energy (right).*

it is clear comparing the effective sigma of the combined energy estimator with respect to those obtained from the calorimeter (see figure 5.8) and the tracker (see figure 5.12) separate reconstruction . Also the scale energy variation as a function of the energy is now reduced to less than 1% in the energy range considered.

For what concern the electron direction, the (η, ϕ) direction measured by the tracker is used everywhere.

This combined estimator of the electron four-momentum will be used in the reconstruction of a signal due to a Higgs boson decaying into 4 electrons, described in the next chapter, giving the possibility to improve the resolution on the Higgs mass, in particular in the low mass region, where the mass resolution is dominated by the experimental resolution.

5.7 Electron identification

Separation between electrons and jets reconstructed as electrons is of particular importance in a hadron collider, since it is needed to reject the overwhelming background coming from QCD jets. It is also crucial in the H$\rightarrow ZZ^{(*)} \rightarrow 2e^{+}2e^{-}$ reconstruction, since misidentification of the electron could lead to a deterioration of Higgs mass resolution and reduction of selection efficiency. In the simulated Higgs sample, about 8% of the reconstructed electrons are not coming from any of the isolated electron associated to the Higgs decay tree: these electrons are mainly coming from the Higgs underlying event, where real non isolated electron coming from b or c jets, or jets with a high electromagnetic fraction, can be reconstructed as electron (the fraction of fake electrons coming from early converted γ is negli-

gible).
A criterion to identify and separate real isolated electrons from fake electrons is thus needed, in order to select a clean sample of electrons, sufficiently pure for the analysis needs.

The study which is proposed here is based on a sample of electrons coming from the $H \rightarrow ZZ^{(*)} \rightarrow 2e^{+}2e^{-}$ decay; as background samples the three QCD di-jets events described in section 5.2 have been used.

The discriminating variable chosen is the *likelihood fraction. Likelihood fraction* is defined as:

$$Ele_{ID} = \frac{L(\vec{x};Ele)}{L(\vec{x};Ele)+L(\vec{x};Jet)}, \tag{5.6}$$

where $L(\vec{x};Ele)$ and $L(\vec{x};Jet)$ are respectively the likelihoods under the hypotheses of being a real electron or a fake electron coming from jets. The elements of the vector $\vec{x} = (x_1,\ldots,x_n)$ are n discriminating variables constructed using the available reconstructed information in the tracker and the calorimeters. The likelihood of hypotheses ψ is therefore defined as:

$$L(\vec{x};\psi) = P(\vec{x};\psi); \tag{5.7}$$

$P(\vec{x};\psi)$ is the probability of $\vec{x}$ assuming hypothesis ψ. Under the assumption (to be verified) that the discriminating variables are uncorrelated between them, it is possible to write

$$P(\vec{x};\psi) = \prod_{i=1}^{n} P_i(x_i;\psi). \tag{5.8}$$

The probability density functions (PDFs) $P_i(x_i;\psi)$ are then calculated using binned histograms of each discriminating variable both for the real electrons (signal, $\psi = Ele$) and fake electrons from jets (background, $\psi = Jet$).

The PDFs for signal are obtained from those electrons in the Higgs sample geometrically matching ($\Delta R = \sqrt{\Delta\eta^2+\Delta\phi^2} < 0.15$) with the true ones in the Higgs decay tree; background PDFs are instead obtained from the reconstructed electrons in the jet sample. Since in the latter case three different samples are used to obtain the PDFs for each x_i, each electron entering in the distribution is weighted using the sample weight w_{sample}:

$$w_{sample} = \sigma_{sample}\frac{N_{ana}}{N_{sel}}N_{gen}; \tag{5.9}$$

where $\sigma_{sample}, N_{ana}, N_{sel}, N_{gen}$ are the values reported in table 5.1. Background PDFs are later normalized to the unity.

The chosen discriminating variables are:

- E/P ratio. For real isolated electrons this ratio is expected to be around 1.

- H/E ratio. The energy deposited by an electron is almost fully contained in the electromagnetic calorimeter. On the contrary, the hadrons will tend to leave more energy in the hadronic calorimeter.

- $|\eta_{track} - \eta_{SC}|$. Clusters initiated by electrons, have their associated track pointing directly on the energy weighted cluster center.

- $E9/E25$ ratio. Clusters initiated by isolated electrons should be narrower than those iniated by hadrons or electrons contained into jets.

- $\sigma_{\eta,\eta}$. This variable measures the shower spread in η of the cluster and is defined as:
$$\sigma^2_{\eta,\eta} = \frac{\sum_i (\eta_i - \eta_{seed})^2 E_i}{\sum_i E_i}, \tag{5.10}$$
where the index i runs over the crystals in a 5×5 matrix around the seed crystal. The distribution in η is chosen since the shower spread in ϕ is affected also for real electrons by bremsstrahlung.

Electrons have been divided also into two categories: electrons in the calorimeter barrel acceptance region and electrons in the endcap, since the PDFs for the discriminating variables are different in the two cases (especially the cluster lateral shower exstension in η, $\sigma_{\eta,\eta}$, because the dimension in η of the crystals is changing between barrel and endcap).

The distributions for the 5 discriminating variables in the barrel and in the endcap for signal and background hypotheses are given in figure 5.30 and 5.31.

The uncorrelation hypotheses between the discriminating variables have been checked. In figure 5.32 scatter plots defining then correlation between E/P and the other discriminating variables are given. No important correlation is present.

5.7.1 Electron idenfication and Jet Mis-identification probability

Electron identification efficiency is measured as a function of the transverse momentum of the electron p_T (figure 5.33). Three different values of the cut are reported: Ele_{ID}>0.15, a loose cut which can be used when efficiency should be privileged to purity, Ele_{ID}>0.5, an average cut, Ele_{ID}>0.9, a tight cut when purity is more important than efficiency. The curves shows a reduction of efficiency in the low-p_T region.

The electron identification efficiency is also measured as a function of fraction of effective bremsstrahlung (figure 5.34). Tight cuts on Ele_{ID} seems also to increase the fraction in the selected electrons of low-bremming electrons.

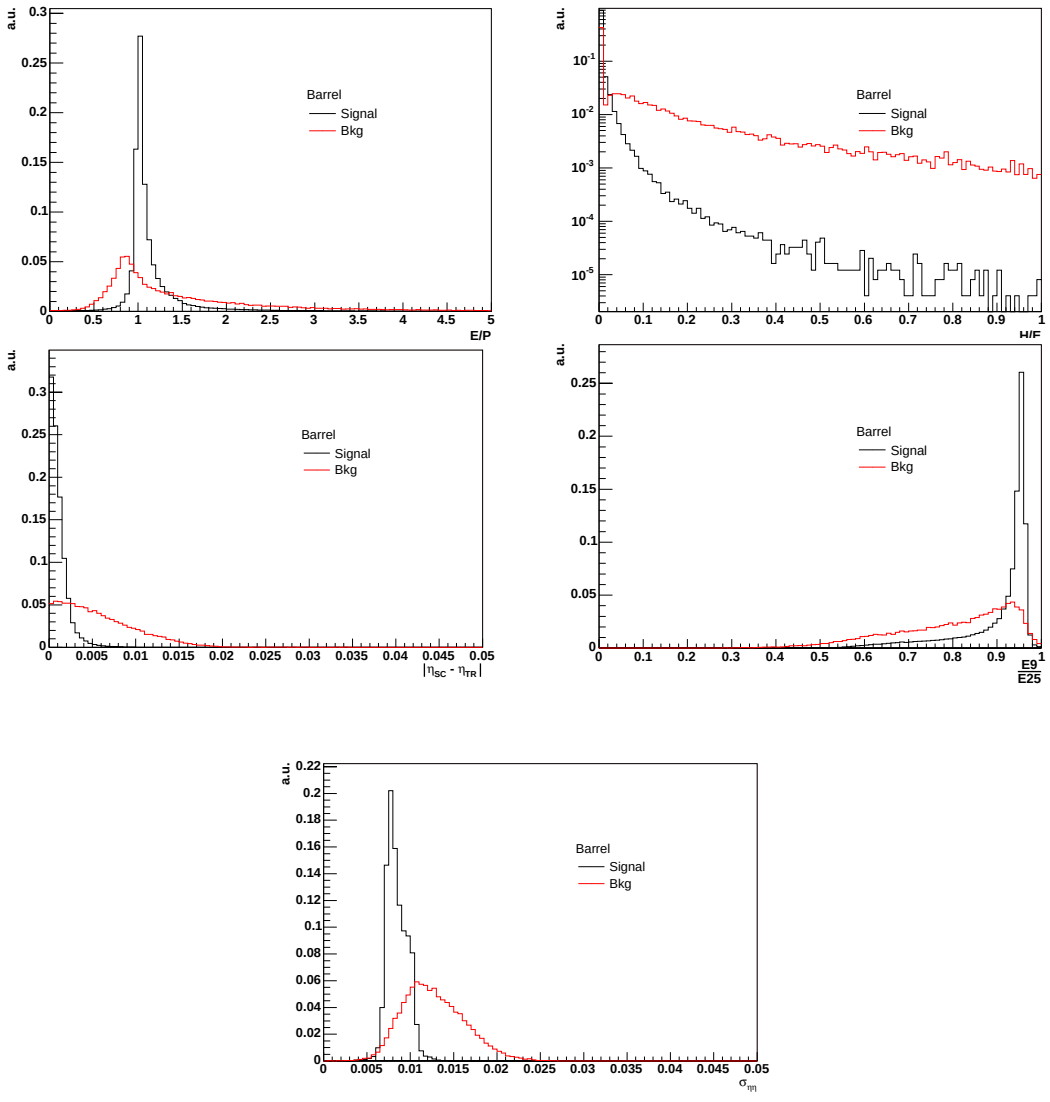


Figure 5.30: *Distributions of the 5 discriminating variables (described in the text) in the barrel for signal and background hypotheses.*

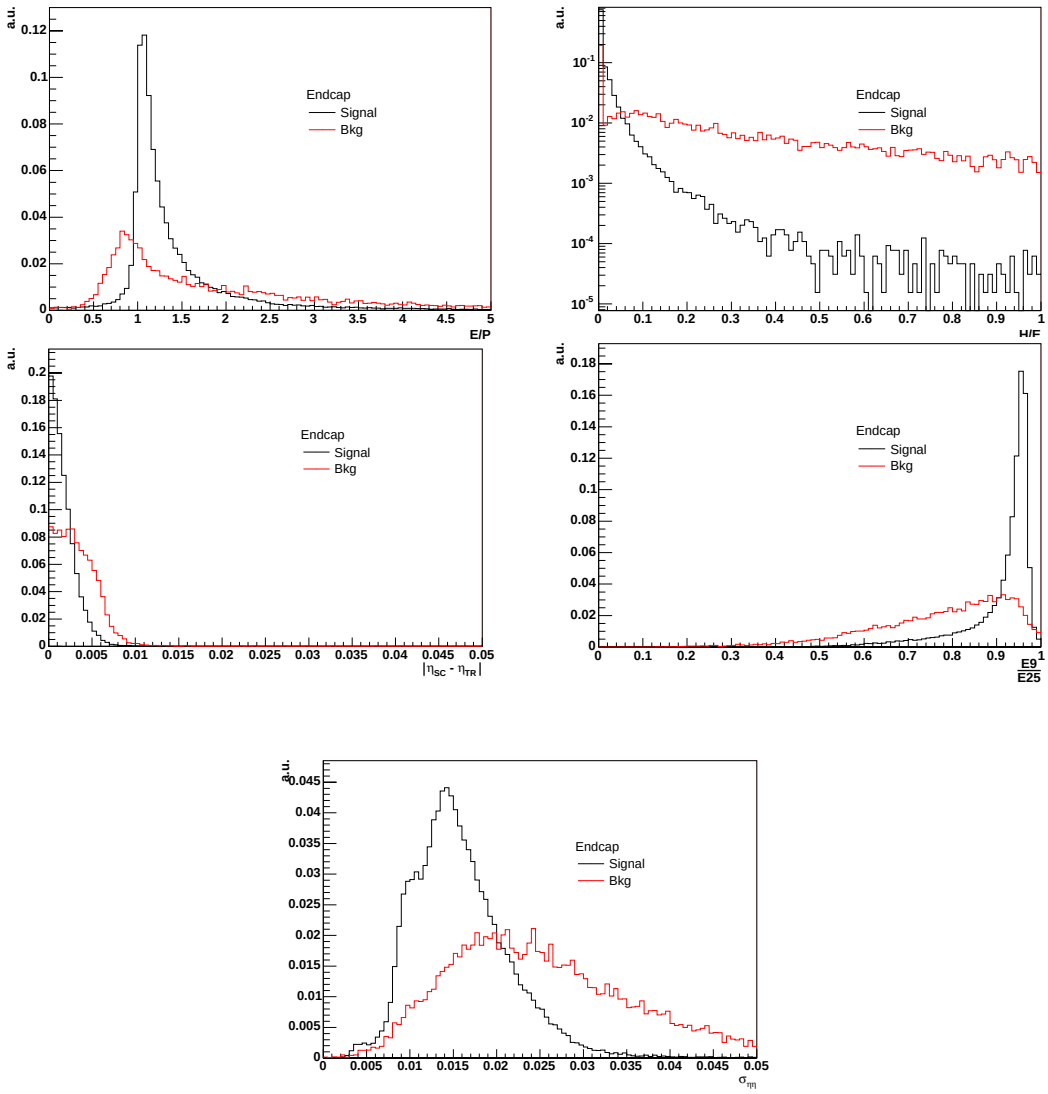


Figure 5.31: *Distributions of the 5 discriminating variables (described in the text) in the endcap for signal and background hypotheses.*

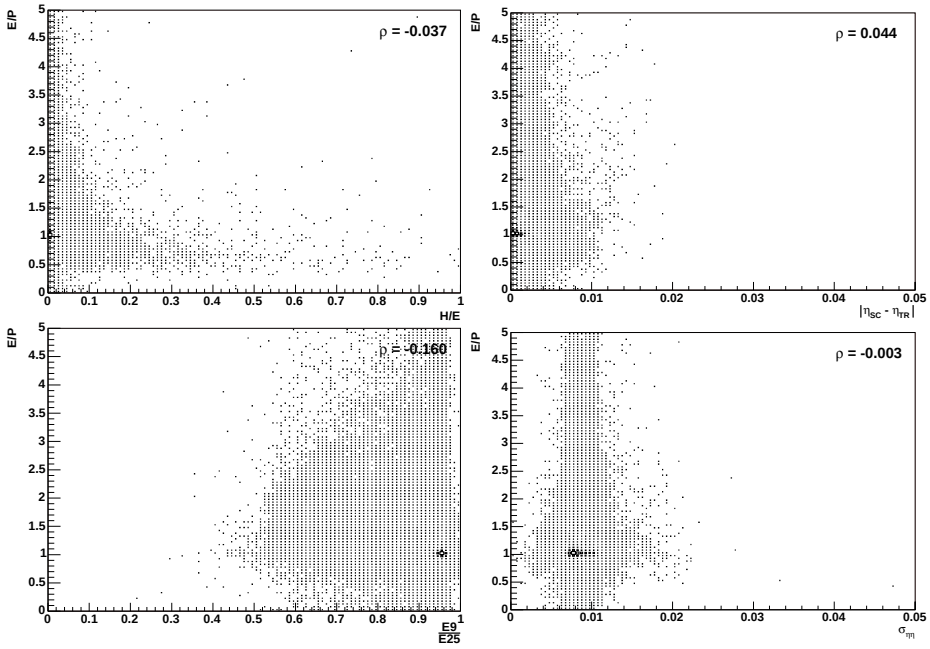


Figure 5.32: *Correlation between E/P and the other discriminating variables (barrel)*

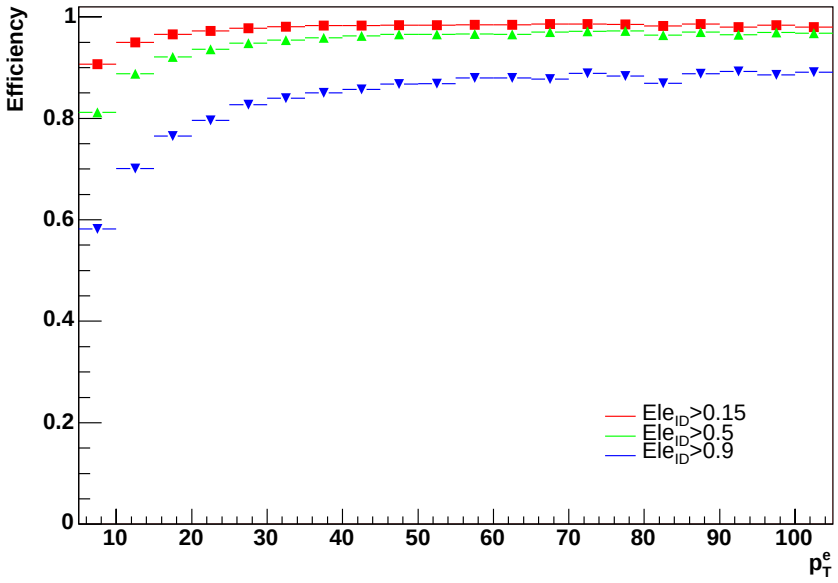


Figure 5.33: *Electron identification efficiency as a function of the transverse momentum of the electron p_T for different values of the cut: Ele_{ID}>0.15, Ele_{ID}>0.5, Ele_{ID}>0.9.*

The selection efficiency (selected electrons / reconstructed *fake* electrons) for *fake* electrons reconstructed in the jet samples, divided in the three $\hat{p}_T$ bins, is

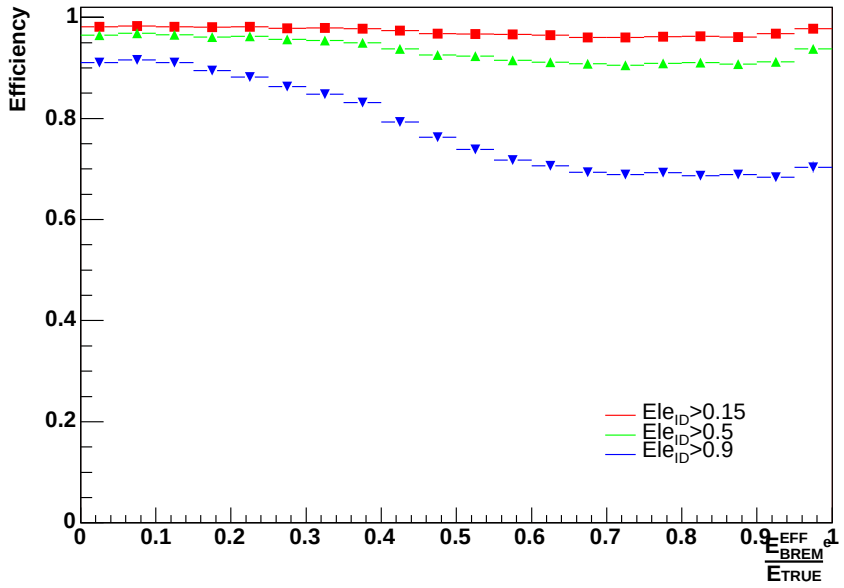


Figure 5.34: *Electron identification efficiency as a function of* E^{EFF}_{BREM}/E_{TRUE} *for different values of the cut:* Ele_{ID}*>0.15,* Ele_{ID}*>0.5,* Ele_{ID}*>0.9.*

reported in table 5.3.

$\hat{p}_T$	Ele_{ID}>0.15	Ele_{ID}>0.5	Ele_{ID}>0.9
25 < $\hat{p}_T$ < 50 GeV	30%	13%	3.8%
50 < $\hat{p}_T$ < 170 GeV	17%	6%	1.6%
$\hat{p}_T$ > 170 GeV	9%	3%	0.7%

Table 5.3: *The selection efficiency (selected electrons / reconstructed* fake *electrons) for* fake *electrons reconstructed in the three jet samples, for different values of the cut:* Ele_{ID}*>0.15,* Ele_{ID}*>0.5,* Ele_{ID}*>0.9.*

The distribution of the Ele_{ID} variable in the Higgs sample for real electrons coming from the Higgs decay tree (signal) and fake electrons coming from the underlying event (background) in the barrel acceptance region is reported in figure 5.35, together with efficiency on signal and background for different values of the cuts.

For the Higgs analysis a cut on Ele_{ID} at 0.15 is chosen, since high efficiency is required, achieving a sufficient rejection on fake electrons.

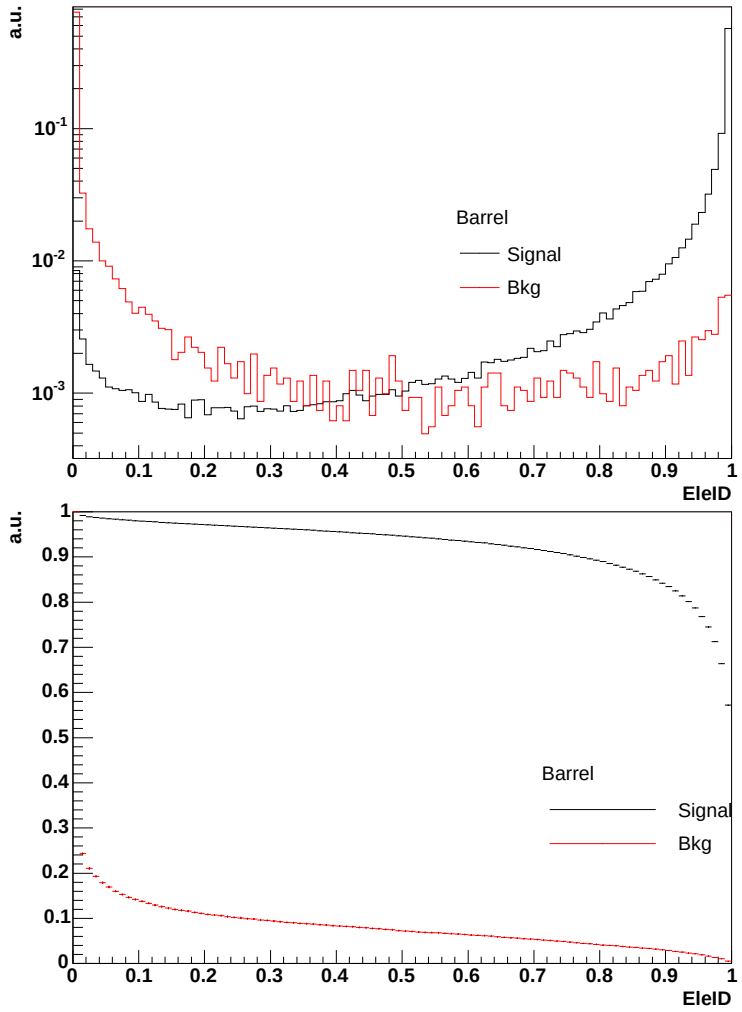


Figure 5.35: *Top: Ele_{ID} distribution in the Higgs sample (m_H=150 GeV) for real electrons coming from the Higgs decay tree (signal) and fake electrons coming from the underlying event (background). Bottom: efficiency on signal and background for different values of the cuts.*

Chapter 6

Study of the signal $H \rightarrow ZZ^{(*)} \rightarrow 2e^{+}2e^{-}$ and optimization of the discovery probability

The $H \rightarrow ZZ^{(*)} \rightarrow 4l$ channel represents an important discovery channel in the whole mass region which can be explored at LHC. The $2e^{+}2e^{-}$ final state, imposing stringent constraints on electron reconstruction and identification, will represent a benchmark of the electron identification and reconstruction algorithms discussed in the previous section.

A complete study of this channel in the full mass range based on detailed simulation and reconstruction of both signal and backgrounds has never been realized. A previous study [97], based on an old version of the CMS detector geometry, combines results obtained with full simulation of the signal with generator level results for the backgrounds.

After a description of the signal and background samples used in this analysis, the event reconstruction and selection procedure will be described. The algorithms introduced in the previous chapter will be used for the electron identification and reconstruction.

The final values of the cuts employed in the signal selection will be optimized in order to maximize the discovery probability. Because of the small number of expected signal events, it is important to find the optimal set of cuts, giving the highest possible discovery probability in the CMS experiment. An estimator of the

discovery probability is defined, taking into account also the statistical fluctuations due to the limited statistics of the Montecarlo samples. This discovery probability estimator is used in the maximization procedure, and it is later compared to the standard significance estimators which are generally used to express the signal visibility in a future experiment.

6.1 Signal and background samples

In this study signal and backgrounds are fully simulated[1]. The low luminosity scenario has been considered; for this reason the samples are digitized superimposing the expected number of pile-up events at $2 \cdot 10^{33}$ cm^{-2}s^{-1}, 3.5 events.

Leading-order (LO) generators have been used in the event generation step for all signal and background samples. This means that higher order corrections to the kinematical distributions are taken into account via initial and final state QCD and QED radiation, instead of using explicit matrix elements at NLO. The absolute values of the LO cross section returned by the generator have been rescaled to match the value given by NLO calculations (in the terminology introduced in section 1.6 using a global K-factor), in the assumption that the corresponding kinematical distributions are in reasonable agreement.

In some cases the differences between some kimematical distributions have been evaluated; for example, in reference [97], the Higgs p_T distribution in the process $gg \rightarrow H$ obtained from analytical calculations has been compared to the one resulting from the PYTHIA Montecarlo [29]. Gluon-gluon fusion is the most important Higgs production process at LHC (see section 1.6) and the most sensitive to higher order corrections. It has been concluded that the PYTHIA distributions are in agreement with the analytical calculation, justifying the use of this event generator for the signal samples.

A detailed description of the signal and background samples used in this analysis is given in the following.

6.1.1 Signal samples

The Higgs decay into 4 leptons via an intermediate state $ZZ^{(*)}$ can be used as a discovery channel at LHC in the Higgs intermediate and high mass region (see section 1.6.2). In this analysis, the Higgs mass range from 115 to 550 GeV has been considered. This interval can be divided into two regions: the region for $m_H < 2m_Z$, where at least one of the two Z bosons in the intermediate state is

[1] Samples simulated according to the procedure introduced in section 5.1, including a detailed simulation of detector response, electronics digitization and pile-up

virtual, and the region $m_H \geq 2m_Z$ where the two Z bosons are on mass-shell.

In the interval $m_H < 2m_Z$, ten reference mass points have been chosen:

- 115, 120, 130 ... 190 GeV.

At the same time for $m_H > 2\ m_Z$, five mass points have identified:

- 200, 250, 350, 450, 550 GeV.

For each mass point a sample of 10000 events has been generated.

PYTHIA 6.223 has been adopted as event generator together with the CTEQ5L set of parton distribution functions [98].

The radiation of photons in the Z decays, the so called *internal* bremsstrahlung, is taken into account using the program PHOTOS [99], which implements an algorithm for single and double photon emission in these decays. The angular distance $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ between the photon and the nearest electron in the Z decay tree is shown in figure 6.1; in about 65% of the events the photon is distant less than 0.025, with an high probability to have its electromagnetic shower in the calorimeter merged with the electron one. The distribution of the radiated photon energy is instead presented in figure 6.2.

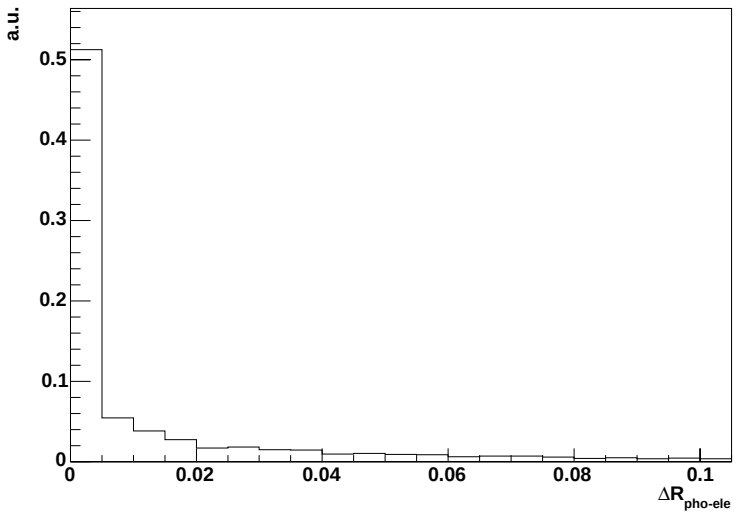


Figure 6.1: *Angular distance $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ between the internal bremsstrahlung photon and the nearest electron in the Z decay tree.*

The main characteristic of the physics channel under study is the presence of four isolated and high p_T electrons in the final state. These electrons properties

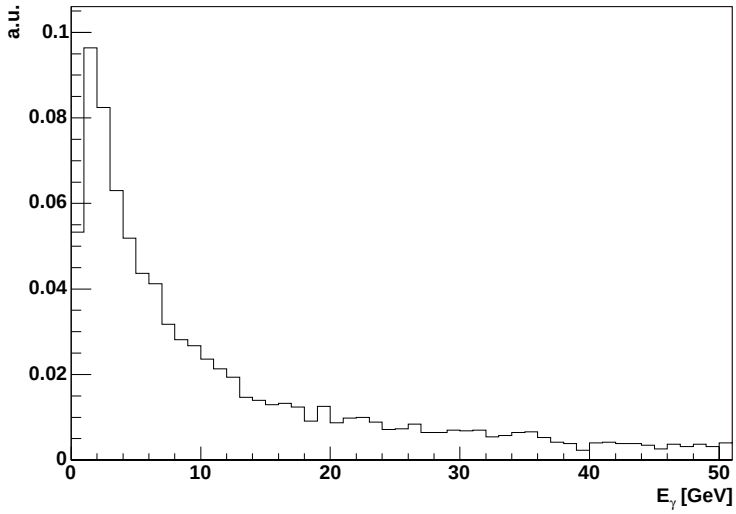


Figure 6.2: *Distribution of the radiated photon energy in the internal bremsstrahlung process.*

will be used to effectively reduce the backgrounds. A preselection at generator level is applied on these electrons in order to ensure a final sample of selected events with a reasonable high trigger and reconstruction efficiency, without biasing the analysis. The 4 electrons are required to satisfy the conditions: $p_T > 5$ GeV and $|\eta| < 2.7$.

The cross sections values for the main Higgs production modes at a center of mass energy $\sqrt{s} = 14$ TeV, as well as the branching ratio $H \to ZZ^{(*)}$ and the acceptances of the preselection cuts ε_{kin} are summarized in table 6.1 for the mass points chosen in the analysis. The cross sections have been obtained from the programs developed by M. Spira [100], while the $BR(H \to ZZ^{(*)})$ is obtained with HDECAY [7]. The gluon fusion, vector boson fusion and Higgs-strahlung processes are calculated at next-to-leading order. On the contrary, the associated production with $q\bar{q}$ pairs is calculated at the leading order only, since the NLO corrections have not yet been implemented. In the last column of table 6.1, the accepted cross-section after preselection is shown (a value of 0.03366 has been assumed for $BR(H \to ZZ^{(*)}, BR(Z \to e^+e^-)$).

6.1.2 Background samples

The main background processes for this channel are those presenting four high p_T electrons in the final state. According to their topological and kinematical characteristics they can be divided into two general classes: irreducible and reducible background.

Signal Samples	$\sigma(gg \to H)$ [pb]	$\sigma(qq \to Hqq)$ [pb]	$\sigma(q\bar{q} \to VH)$ [pb]	$\sigma(q\bar{q},gg \to Hq\bar{q})$ [pb]	σ_{tot}	$BR(H \to ZZ^{(*)})$	ε_{kin}	$\sigma_{tot} \times BR_{tot} \times \varepsilon_{kin}[fb]$
m_H=115 GeV	40.1	4.38	1.84	0.61	46.9	0.008	0.54	0.23
m_H=120 GeV	36.9	4.21	1.61	0.55	43.3	0.015	0.56	0.42
m_H=130 GeV	31.6	3.87	1.25	0.44	37.2	0.038	0.61	1.00
m_H=140 GeV	27.4	3.62	0.98	0.35	32.4	0.067	0.65	1.62
m_H=150 GeV	24.0	3.38	0.78	0.29	28.4	0.083	0.67	1.80
m_H=160 GeV	21.1	3.13	0.63	0.24	25.2	0.043	0.69	0.86
m_H=170 GeV	18.8	2.93	0.51	0.20	22.5	0.022	0.71	0.41
m_H=180 GeV	16.8	2.73	0.42	0.16	20.2	0.057	0.73	0.96
m_H=190 GeV	15.2	2.56	0.35	0.14	18.2	0.218	0.74	3.34
m_H=200 GeV	13.8	2.41	0.29	0.12	16.6	0.261	0.74	3.66
m_H=250 GeV	9.23	1.79	0.13	0.06	11.2	0.297	0.74	2.79
m_H=350 GeV	7.07	1.07	0.03	0.02	8.20	0.311	0.81	2.34
m_H=450 GeV	4.73	0.68	0.01	0.01	5.44	0.259	0.82	1.31
m_H=550 GeV	2.08	0.46	0.006	0.008	2.55	0.263	0.85	0.65

Table 6.1: *Cross sections for different Higgs production modes at* $\sqrt{s} = 14$ *TeV obtained with M. Spira programs [100] expressed in pb:* $\sigma(gg \to H)$ *HIGLU (NLO),* $\sigma(qq \to Hqq)$ *VV2H (NLO),* $\sigma(q\bar{q} \to VH)$ *V2HV (NLO),* $\sigma(q\bar{q},gg \to Hq\bar{q})$ *HQQ (LO).*
Branching ratios for $H \to ZZ^{(*)}$ *are calculated with the program HDECAY [7]. Acceptance of the kinematical preselection cuts applied on the electrons:* $p_T > 5$ *GeV and* $|\eta| < 2.7$. *Last column reports the accepted cross-section for* $H \to ZZ^{(*)} \to 2e^+2e^-$ *events.*

The main irreducible background process is represented by $ZZ^{(*)}$ production, which shows kinematical characteristics very similar to those of the signal. The cross-section of this process is comparable with the one of the Higgs samples. The reducible background is instead composed of $Zb\bar{b} \rightarrow 4e+X$ and $t\bar{t} \rightarrow 4e+X$ events, which presents a much higher cross-section (>100 times) than the signal. The reducible background can be largely suppressed requiring common vertex constraints, isolation criteria and kinematical cuts. The goal is to suppress the latter backgrounds at levels much lower than the irreducible one.

The backgrounds considered so far represents only physical backgrounds. Other backgrounds, like $Z+jets$, where jets are reconstructed as electron, even if not simulated are expected to be largely suppressed by the identification and isolation criteria used in this analysis.

In the following the cross-section and characteristics of each background sample will be discussed in detail. The same preselection cuts of the signal are applied on the different background samples, requiring at least $2e^+$ and $2e^-$ in the acceptance region $p_T > 5$ GeV and $|\eta| < 2.7$.

$ZZ^{(*)}$ background

The leading order process for gauge-boson pair production at an hadron collider is $q\bar{q}$ annihilation. Gluon-gluon fusion, $gg \rightarrow ZZ^{(*)}$, represents a process of higher order in α_s, however it gives a non negligible contribution since gluon-gluon luminosity is much higher for low values of x than quark-antiquark luminosity. In figure 6.3 the Feynman diagrams of the two processes at tree level are illistrated.

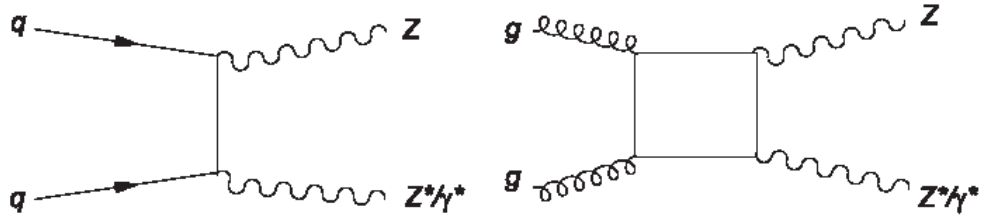


Figure 6.3: *Tree level diagrams responsible for the $ZZ^{(*)}$ production at LHC.*

In reference [101] the gluon fusion process cross section was evaluated to be at leading order around 20% of the quark-antiquark annihilation one. No NLO calculation is available for this process. NLO calculations have been performed for the process $q\bar{q} \rightarrow ZZ^{(*)}$ [102], showing a K factor of about 1.33. Taking into

account the 20% contribution of $gg \to ZZ^{(*)}$ at leading order (2.4 pb), the K factor of 1.33, the cross-section for the $ZZ^{(*)}$ production at LHC can be evaluated to be 18.2 pb.

A sample of 27250 preselected events has been generated for this background with PYTHIA 6.158. Only the quark initial state production mechanism is implemented in the PYTHIA version which has been used for the generation of the events; kinematical differences between the two production mechanisms remain as a systematic effect, which however is tought to be small.

The PHOTOS package is used to take into account single and double photon radiation from the Z bosons.

$t\bar{t}$ background

The diagrams for the processes $q\bar{q} \to t\bar{t}$ and $gg \to t\bar{t}$ at tree level are presented in figure 6.4. Full NLO calculation is available for the production cross-section; this value is is 886 ± 94 pb using CTEQ5M structure functions. The reported error is a theoretical systematic uncertainty corresponding to a 4 GeV indetermination on the top mass m_t. The central value is given for m_t=175.6 GeV. The gluon fusion process has a cross section about six times higher than the quark annihilation process.

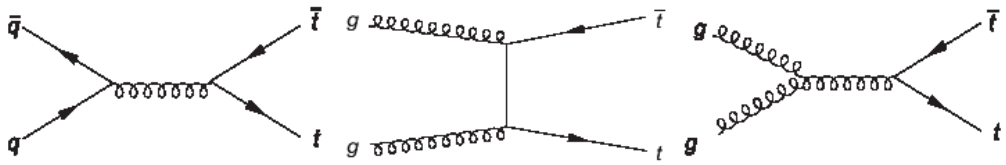


Figure 6.4: *Diagrams for the processes $q\bar{q} \to t\bar{t}$ and $gg \to t\bar{t}$ at tree level*

The top quark decays with 99.8% probability into a W boson and a b quark. As it is pictured in figure 6.5, electrons could come from many points in the top decay tree:

- from the decay of the W boson
- from the decay of a τ lepton from the W boson
- from the semi-leptonic decay in a W boson hadronic decay
- from the semi-leptonic decay of a meson produced in the b quark decay tree.

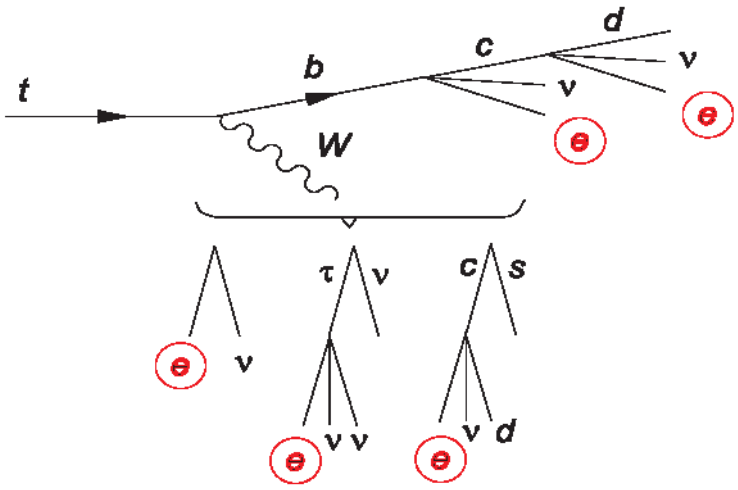


Figure 6.5: *Possible sources of electrons in a top quark decay chain.*

Simply looking at the branching ratios of the different possibilities, the most favorite situation to have $2e^+$ and $2e^-$ in the final state is when two $e^\pm$ come directly from the W bosons and two $e^\pm$ are produced in the b-quark cascade.

A sample of 75800 events generated with PYTHIA 6.158 has been produced. To speed-up the generation, the W boson has been forced to decay leptonically (the contribution of electrons coming from the hadronic decay of the W is negligible), while the b quark has been left free to decay. The preselected samples contains events with at least $2e^+2e^-$ with $p_T > 5$ GeV and $|\eta| < 2.7$.

$Zb\bar{b}$ background

Also in this case the two initial states from which a $Zb\bar{b}$ state can be produced are $q\bar{q}$ and gg, for which the tree level Feynman diagrams are reported below (6.6). PYTHIA lacks an implementation of the $gg \rightarrow Zb\bar{b}$ process, so CompHEP [103] generator has been used for generation.

Using CompHEP, a sample of 1600000 events have been generated, which are succesively passed to PYTHIA for the hadronization, final state generation and preselection (using PHOTOS package for the γ radiation from Z). At CompHEP level, in order to reduce the time needed for the generation of the events, kinemat-

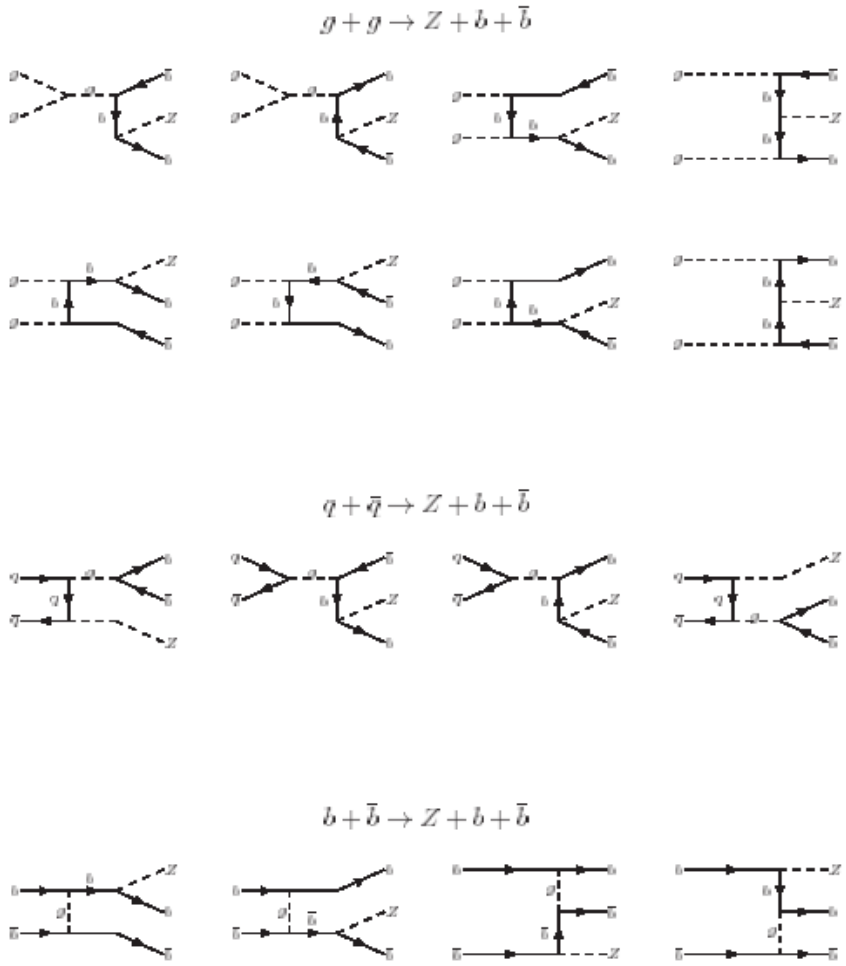


Figure 6.6: *Tree level diagrams for the processes $gg \to Zb\bar{b}$ and $q\bar{q} \to Zb\bar{b}$*

ical cuts have been imposed on the b quarks: $p_T > 1$ GeV, $|\eta| \leq 2.5$. Applying a cut at the level of b-quarks can be dangerous; the fraction of events with one electron in the acceptance region coming from a b-quark with $|\eta| > 2.5$ is less than 1%. This value has been etimated in a separate sample, generated with an acceptance region for b-quarks $|\eta| \leq 10$. Three different samples have been genenerated according to the decay properties of the $b\bar{b}$ pair. Three types of decay have been considered for the $b\bar{b}$ pair (the decay possibilities are mutually excluding each other):

- both b and $\bar{b}$ are forced to decay $b \to l + X$ (later referenced as $Zb\bar{b}$(ll))
- both b and $\bar{b}$ are forced to decay not leptonically ($Zb\bar{b}$(cc))

- b is forced to decay $b \to l + X$, $\bar{b}$ decays not leptonically ($Zb\bar{b}$(lc))

In all samples the Z boson in forced to decay e^+e^-.

The number of preselected events generated for each category are 18642 for $Zb\bar{b}$(ll), 11684 for ($Zb\bar{b}$ (cc)) and 10721 for ($Zb\bar{b}$ (lc)).

The NLO cross section has been evaluated to be 525 pb using the program MCFM [104] and CTEQ5M structure functions, applying kinematical cuts on the b quarks.

In addition to the $Zb\bar{b}$ production, any $Zq\bar{q}$ production can be studied, for example the $Zt\bar{t}$ or $Zc\bar{c}$ production. These two backgrounds are however negligible if compared to $Zb\bar{b}$. In particular, the production cross section for $Zt\bar{t}$ is 726 fb at leading order and therefore about a factor 1000 lower than the one for $Zb\bar{b}$. On the contrary, a study at generator level for the $Zc\bar{c}$ has been made [105], using events generated with CompHEP and PYTHIA. The cross-section is about 1.5 times greater than $Zb\bar{b}$, but the preselection efficiency has been found 10 times lower than $Zb\bar{b}$, because of the lower branching fraction into electrons and the softer electron momentum spectrum.

Summary of backgrounds

In table 6.2 the main sources of backgrounds are summarized reporting the estimated cross-sections for the background samples (at NLO), the preselection efficiency and the accepted cross-sections.

Background Samples	σ_{NLO} [pb]	BR	ε_{kin}	$\sigma_{NLO} \times BR \times \varepsilon_{kin}$ [fb]
$t\bar{t}$	886	$5.14 \cdot 10^{-2}$	$3.20 \cdot 10^{-3}$	146.0
$Zb\bar{b}(cc)$	525	$8.75 \cdot 10^{-4}$	$1.47 \cdot 10^{-2}$	6.77
$Zb\bar{b}(lc)$	525	$3.13 \cdot 10^{-3}$	$1.06 \cdot 10^{-2}$	17.6
$Zb\bar{b}(ll)$	525	$3.62 \cdot 10^{-4}$	$6.63 \cdot 10^{-2}$	12.6
ZZ^*	18.2	$1.13 \cdot 10^{-3}$	$2.31 \cdot 10^{-1}$	4.76

Table 6.2: *Summary of the main background processes cross-sections (NLO) and accepted cross-sections after preselection.*

The $t\bar{t}$ sample shows the largest accepted cross-section (more than 100 times the signal).

6.2 Global event reconstruction and preselection

The first step in the event reconstruction and selection is represented by the trigger. For those events which pass the Level 1 trigger and the High Level Trigger selection, the global event reconstruction is performed. The aim is to identify the electrons belonging to Higgs decay tree; this is accomplished selecting those electrons which are coming from a common vertex. Therefore, vertex reconstruction is considered in this analysis as the starting point. From the characteristics of the reconstructed vertex, a preliminary background rejection can also be obtained; in particular, both $Zb\bar{b}$ and $t\bar{t}$ events have non isolated electrons originated in the decay tree of a b quark, therefore they can be associated to tracks with large impact parameter. Common vertex reconstruction also eliminates the possibility of background due to electrons originated in pile-up events.

The two aspects of the event trigger and of common electron vertex reconstruction will be described in detail in the following.

6.2.1 Trigger

The useful triggers in the HLT trigger table, in case of events containing more than one isolated electron, are the inclusive single and double electron triggers.

The HLT electron reconstruction follows the steps described in section 5.4. Reconstruction of a high level trigger electron candidate starts from Level 1 calorimetric candidates. The standard HLT trigger thresholds for the low luminosity regime are for single electron trigger $p_T = 29$ GeV, while the double trigger threshold is $p_T = 17$ GeV for both electrons. The corresponding trigger rates for the inclusive single and double electron triggers are 33Hz and 1 Hz respectively (from table 2.6.2).

In this analysis, the global HLT response is defined as the logical OR of the inclusive single electron and double electron trigger; it is possible to assume that a future analysis on the data coming from the real pp collisions will start looking into events contained in the single and double electron streams.

The L1 and (L1+HLT) trigger efficiencies evaluated on all the Higgs samples at different Higgs masses are presented in figure 6.7; the background efficiencies are reported as well. Table 6.3 summarizes the (L1+HLT) efficiencies. The efficiency is defined as the number of triggered events divided by the number of preselected events. For Higgs masses above 180 GeV, the Level 1 Trigger is almost 100% efficient. (L1+HLT) efficiencies ranges from 80.5% for a Higgs mass of 115 GeV to above 98% for Higgs masses greater than 450 GeV. In the background samples, the $t\bar{t}$ has the lowest efficiency, 66.1% after HLT. ZZ^* and $Zb\bar{b}$ samples have similar efficiencies: 89.4% for the irreducible background, around

82% for $Zb\bar{b}$.

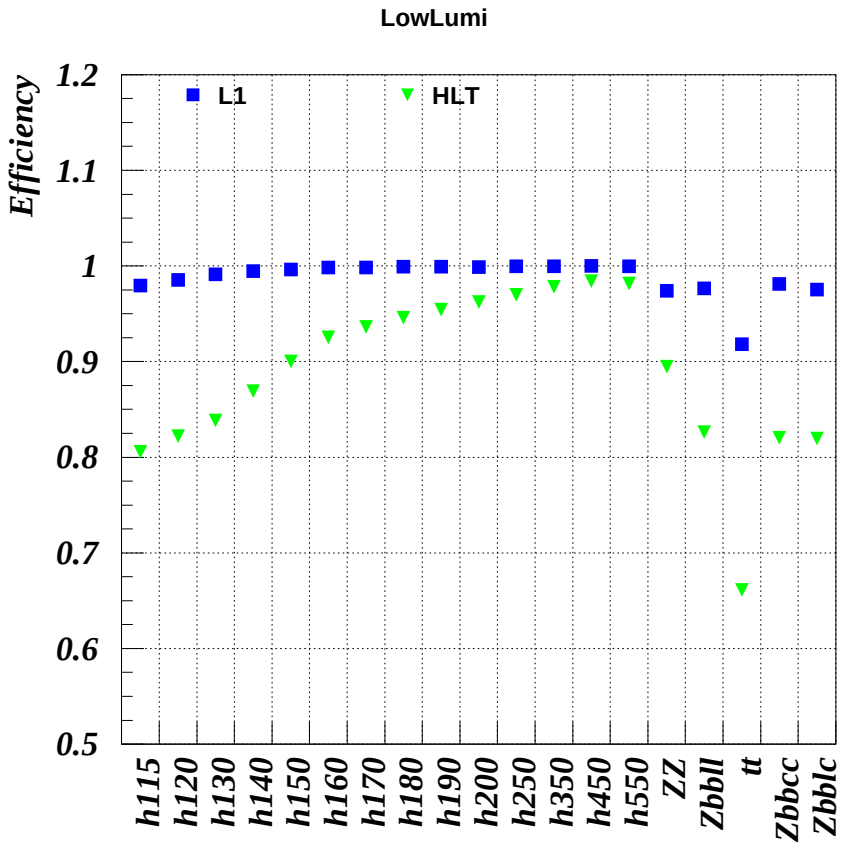


Figure 6.7: *The L1 and (L1+HLT) trigger efficiencies evaluated on all the Higgs samples at different Higgs masses and on the various backgrounds.*

In figure 6.8 the separate efficiencies of the single electron trigger alone and the double electron trigger alone are showed. Comparing the separate response to the global HLT trigger (logical or between single and double electron trigger), it can be noted that the fraction of events passing the double electron trigger only is negligible. This means that in a future analysis starting from real data, only events contained in the single electron stream can be analyzed.

The number of expected triggered events for an integrated luminosity of 20 fb^{-1} is given is figure 6.9.

6.2.2 Vertex reconstruction and selection

Vertex reconstruction is considered in this analysis as the first step in the global event reconstruction, since it allows to correctly identify the electrons contained

Signal Samples	L1+HLT efficiency
m_H=115 GeV	80.5 ± 0.04 %
m_H=120 GeV	82.2 ± 0.04 %
m_H=130 GeV	83.8 ± 0.04 %
m_H=140 GeV	86.9 ± 0.04 %
m_H=150 GeV	89.9 ± 0.03 %
m_H=160 GeV	92.5 ± 0.02 %
m_H=170 GeV	93.6 ± 0.02 %
m_H=180 GeV	94.5 ± 0.02 %
m_H=190 GeV	95.4 ± 0.02 %
m_H=200 GeV	96.2 ± 0.02 %
m_H=250 GeV	96.9 ± 0.01 %
m_H=350 GeV	97.8 ± 0.01 %
m_H=450 GeV	98.4 ± 0.01 %
m_H=550 GeV	98.1 ± 0.01 %
Background Samples	L1+HLT efficiency
$t\bar{t}$	66.1 ± 0.02 %
$Zb\bar{b}(cc)$	82.0 ± 0.04 %
$Zb\bar{b}(lc)$	81.9 ± 0.04 %
$Zb\bar{b}(ll)$	82.6 ± 0.03 %
ZZ^*	89.4 ± 0.02 %

Table 6.3: *L1+HLT efficiencies for Higgs samples at different Higgs masses and the backgrounds.*

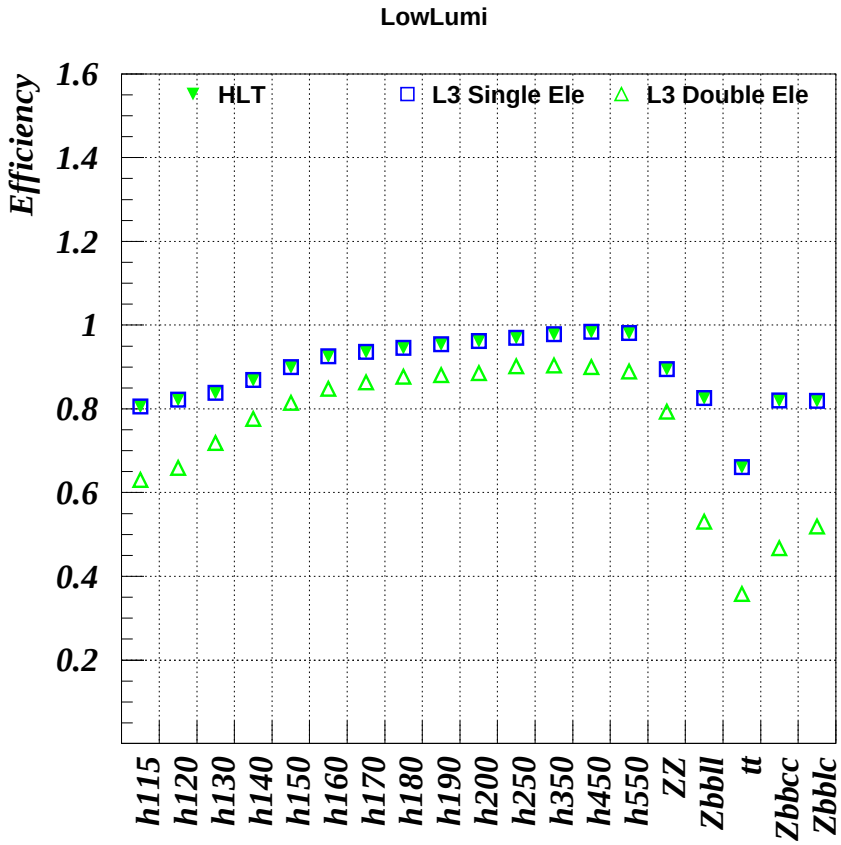


Figure 6.8: *Separate efficiencies of the single electron trigger alone and the double electron trigger alone compared to the global HLT trigger.*

in the Higgs decay tree and achieve a first rejection of reducible backgrounds.

At first electrons, reconstructed using the algorithms described in the previous chapter, are divided into e^+ and e^- categories according to their reconstructed charge. A vertex fit is tried for all the possible pairs identified by 2 e^+ and 2 e^-. All the e^+ and e^- partecipating in a vertex are required to satisfy the loose isolated electron identification criterion, introduced in section 5.7.1, which is $Ele_{ID} > 0.15$. If more than one vertex is found satisfying these criteria, the vertex with the highest algebrically summed p_T is mantained.

The Ele_{ID} requirement allows to increase from 91% to 93% the fraction of *true assigned vertices* (vertex purity), the fraction of vertices among the reconstructed ones which has correctly identified the e^+ and e^- belonging to the Higgs decay tree. Between the wrong assigned vertices (7%), around 3% are found in events where at least one of the electrons in the Higgs decay tree has not been reconstructed, while 4% are due to events where the highest summed p_T criterion

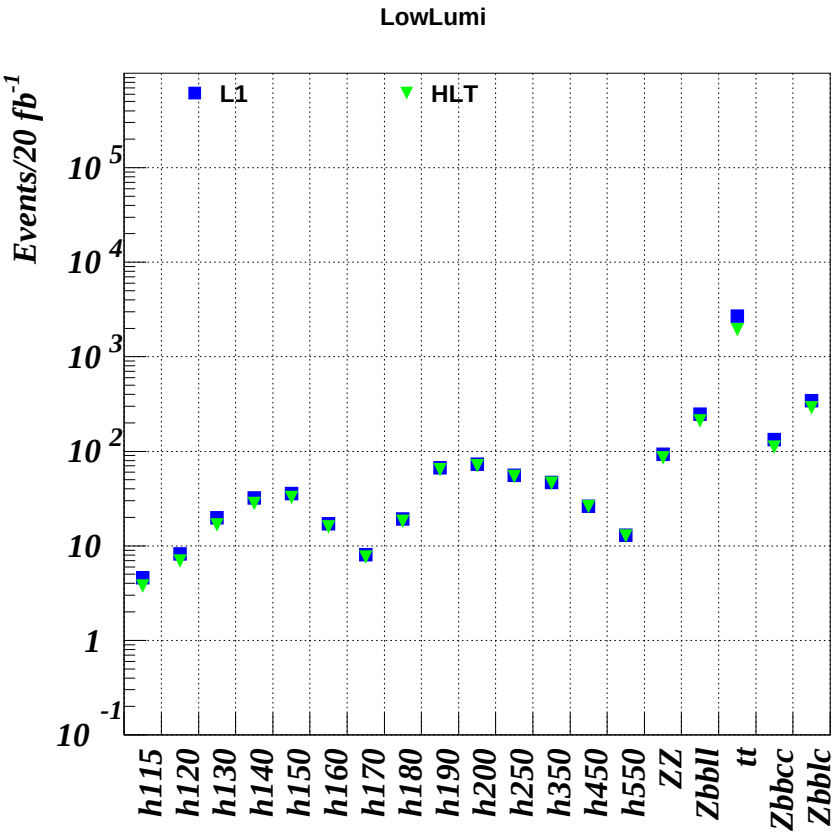


Figure 6.9: *Number of expected triggered events for an integrated luminosity of 20 fb^{-1}.*

leads to a wrong decision. Furthermore, Ele_{ID} criteria bring also an important first reduction of the reducible backgrounds, $Zb\bar{b}$ and $t\bar{t}$, proving to be powerful in rejecting non isolated real electrons contained in b jets. A rejection of around 3 is achieved.

The table 6.4 contains the fraction of events between the kinematically pre-selected events, having at least one vertex reconstructed (more than 4 electrons reconstructed with the correct charge assignment) and with at least one vertex satisfying the Ele_{ID} criteria.

The same numbers are plotted in figure 6.10.

The charge requirement made on e^{+} and e^{-} leads to a little inefficiency (charge mis-reconstruction at 0.7% level for each electron), but it is helpful in reducing the contamination coming from combinatorics. It is possible to compare the fraction of events with at least four reconstruted electrons making no charge requirement with the fraction of reconstructed vertces (applying charge requirements): in the signal sample at $m_H = 150$ GeV they are respectively 61.3% and 56.8%.

Signal Samples	Vertex Reconstruction Efficiency	Vertices with $Ele_{ID} > 0.15$ for all electrons
m_H=115 GeV	$41.8 \pm 0.6\%$	36.4 ± 0.5 %
m_H=120 GeV	$43.8 \pm 0.6\%$	38.0 ± 0.5 %
m_H=130 GeV	$48.5 \pm 0.6\%$	42.8 ± 0.5 %
m_H=140 GeV	$52.6 \pm 0.6\%$	46.2 ± 0.6 %
m_H=150 GeV	$56.8 \pm 0.6\%$	50.2 ± 0.6 %
m_H=160 GeV	$60.1 \pm 0.6\%$	53.6 ± 0.6 %
m_H=170 GeV	$62.1 \pm 0.6\%$	55.6 ± 0.6 %
m_H=180 GeV	$64.5 \pm 0.6\%$	57.0 ± 0.6 %
m_H=190 GeV	$65.6 \pm 0.6\%$	58.9 ± 0.6 %
m_H=200 GeV	$66.1 \pm 0.5\%$	59.6 ± 0.6 %
m_H=250 GeV	$68.3 \pm 0.5\%$	60.9 ± 0.6 %
m_H=350 GeV	$72.5 \pm 0.5\%$	65.4 ± 0.6 %
m_H=450 GeV	$74.5 \pm 0.5\%$	66.9 ± 0.6 %
m_H=550 GeV	$74.7 \pm 0.5\%$	67.3 ± 0.6 %
Background Samples		
$t\bar{t}$	$19.0 \pm 0.6\%$	6.7 ± 0.1 %
$Zb\bar{b}(cc)$	$18.5 \pm 0.6\%$	5.2 ± 0.3 %
$Zb\bar{b}(lc)$	$20.1 \pm 0.6\%$	11.7 ± 0.3 %
$Zbb(ll)$	$20.4 \pm 0.6\%$	12.5 ± 0.3 %
ZZ^*	$52.7 \pm 0.6\%$	47.9 ± 0.5 %

Table 6.4: *Fraction of events between the kinematically preselected events, having at least one vertex reconstructed (more than 4 electrons reconstructed with the correct charge assignment) and with at least one vertex satisfying the Ele_{ID} criteria.*

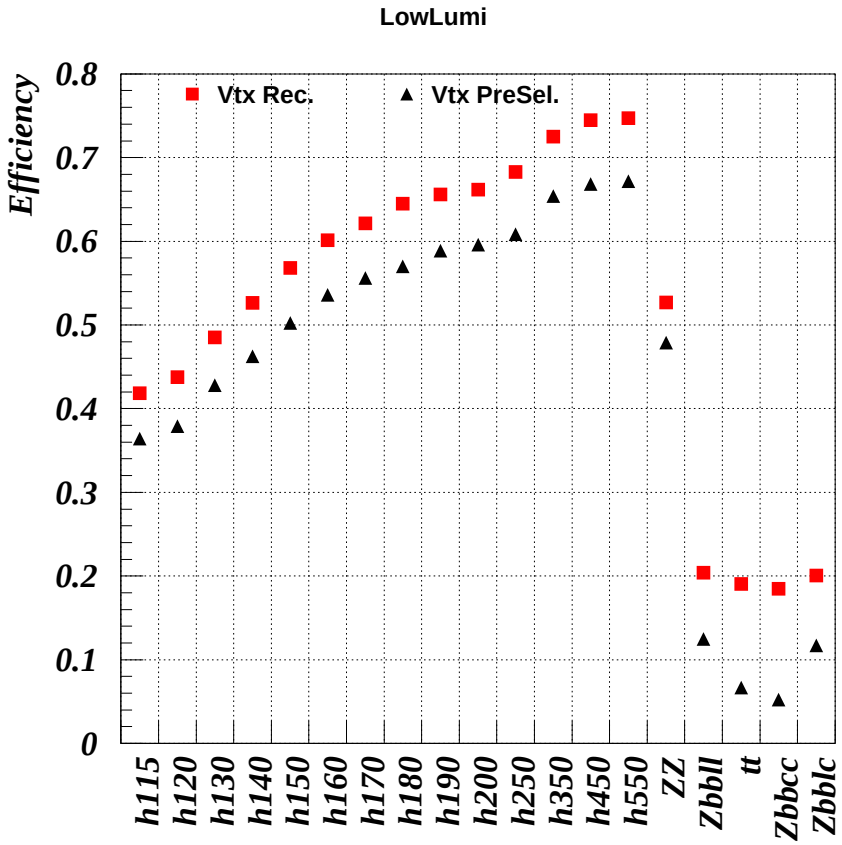


Figure 6.10: *Fraction of events between the kinematically preselected events, having at least one vertex reconstructed (more than 4 electrons reconstructed with the correct charge assignment) and with at least one vertex satisfying the Ele_{ID} criteria.*

The decreasing efficiency going towards lower Higgs masses is the effect of the inefficiency in the reconstruction of low p_T electrons. It has already been discussed that the bremsstrahlung effect makes the electron reconstruction in the low p_T region quite difficult. Therefore there is some space for improvements in the low mass region, trying to better optimize the actual electron reconstruction in the p_T interval 5-15 GeV. Another possible approach which can be tried to improve the reconstruction efficiency is to consider a global reconstruction that requires only three electron candidates plus and another isolated track with the right charge in the tracker, all coming from the same common vertex. The latter study necessitates, however, of the inclusion of additional backgrounds with three real electrons (like, for example, $Wb\bar{b} \to 3e$, $Zb\bar{b} \to 3e$, $t\bar{t} \to 3e$), which have not yet been simulated; for this reason, this approach has not been investigated so far.

From the point of view of the actual implementation of the vertex fitting procedure, the vertex fit is made using the `KalmanVertexFitter` class in ORCA.

The vertex resolution position has been evaluated. The distributions of the reconstructed vertex position, compared to the true vertex position in a Higgs event in the *x*, *y*, *z* directions, are showed in figure 6.11 for the Higgs sample with $m_H = 150$ GeV. Superimposed on the distributions, there are double gaussian fits. The fit is performed applying some constraints on the second gaussian parameters: the two gaussians are required to have the same mean μ, while the σ of the second gaussian is fixed to be greater than the first one. The resolutions, given by fitted σ of the first gaussian, are 16.2 ± 0.6 μm in the *x* direction, 15.1 ± 0.7 μm in *y*, 19.1 ± 0.6 μm in *z*.

A first rejection of reducible background can be achieved using several variables associated to the fitted vertex and to the tracks belonging to it. Electrons in the $Zb\bar{b}$ and $t\bar{t}$ backgrounds are expected to have larger impact parameters with respect to the fitted vertex position; at the same time, also the quality of the vertex fitting procedure in these reducible background samples should be worse.

The variables which have been considered are the following:

- χ^2 of the fit
- $D\bar{L}IP$ and $D\bar{T}IP$. $D\bar{L}IP$ and $D\bar{T}IP$ are defined respectively as the average longitudinal and transversal impact parameter of the electron tracks, associated to the vertex with respect to fitted vertex position
- $V\bar{T}CE$. $VTCE$ defines for each track belonging to the vertex the vertex-track compatibility. It is constructed as a sort of χ^2 taking into account the three-dimensional position of the vertex and of the track extrapolated to the vertex. $V\bar{T}CE$ is the average vertex-track compatibility

Background efficiency versus signal efficiency plots for all the considered variables are given in figure 6.12 for the Higgs sample with $m_H = 150$ GeV. It is possible to obtain a significant reduction of the $Zb\bar{b}$ and $t\bar{t}$ backgrounds while still preserving a very high efficiency on the signal.

The variable which better enhances the signal to background ratio is $D\bar{L}IP$. The chosen value of the cut for this variable is $DLIP < 2.25$ for all the Higgs mass hypotheses. In table 6.5 the efficiencies on signal and background samples fot this cut are reported.

6.3 Pairing

After reconstructing and selecting the $2e^+$ and $2e^-$ from the Higgs vertex, another aspect of the $H \to ZZ^{(*)} \to 2e^+2e^-$ reconstruction is the association in pair

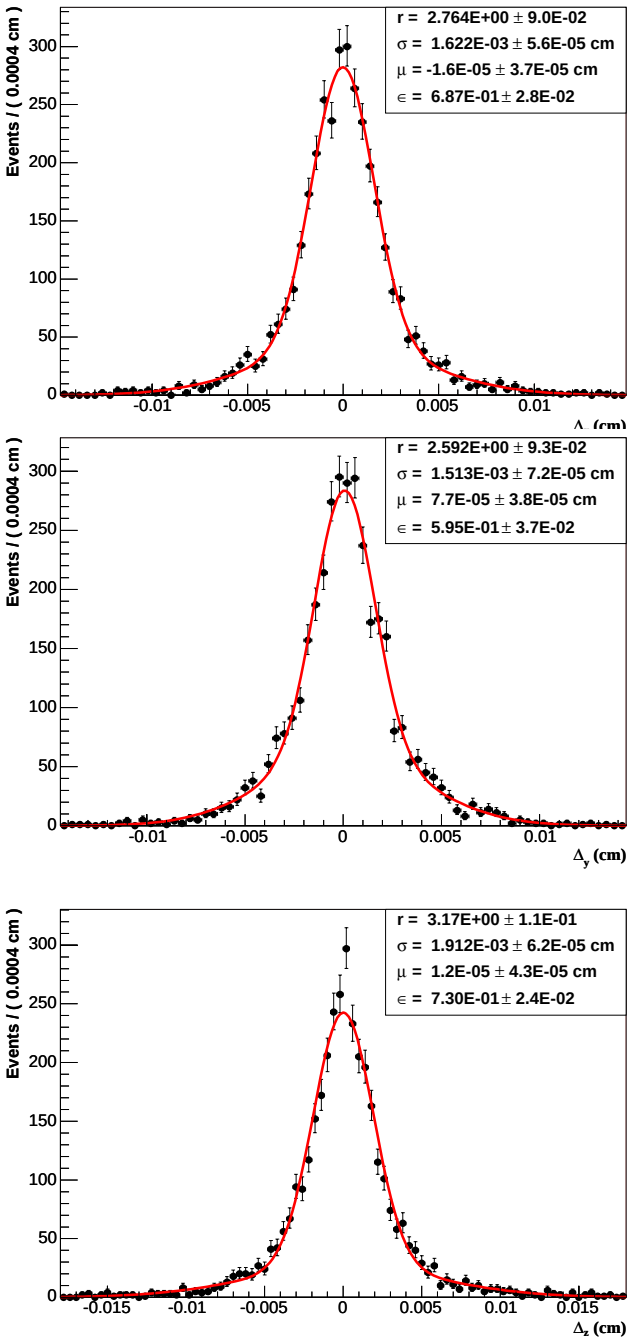


Figure 6.11: *Resolution of the reconstructed vertex position in the x, y, z directions. Superimposed on the distributions, there are double gaussian fits. The corresponding parameters of the fit are: μ, the mean of the two gaussians, σ, is the first gaussian width, r is the ratio between the σ of the two gaussians (fixed to be greater than 1), ε is the relative weight of the gaussians.*

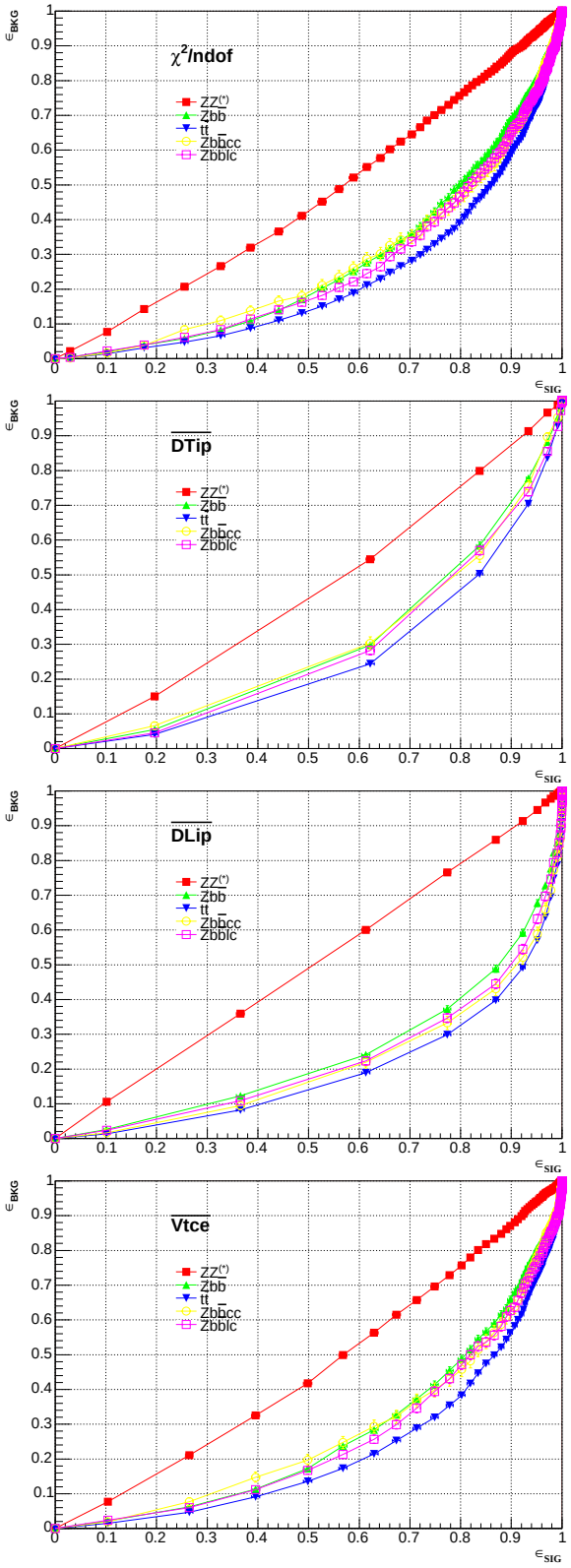


Figure 6.12: *Background efficiency versus signal efficiency plots for all the considered vertex selection variables (Higgs sample with* $m_H = 150$ *GeV).*

Signal Samples	$\varepsilon(\overline{DLIP} < 2.25)$ (Total efficiency)
m_H=115 GeV	$98.2 \pm 0.2\%$ ($35.8 \pm 0.5\%$)
m_H=120 GeV	$98.4 \pm 0.2\%$ ($37.3 \pm 0.5\%$)
m_H=130 GeV	$98.5 \pm 0.2\%$ ($42.2 \pm 0.5\%$)
m_H=140 GeV	$98.2 \pm 0.2\%$ ($45.4 \pm 0.6\%$)
m_H=150 GeV	$97.8 \pm 0.2\%$ ($49.1 \pm 0.6\%$)
m_H=160 GeV	$97.7 \pm 0.2\%$ ($52.4 \pm 0.6\%$)
m_H=170 GeV	$98.1 \pm 0.2\%$ ($54.6 \pm 0.6\%$)
m_H=180 GeV	$97.8 \pm 0.2\%$ ($55.7 \pm 0.6\%$)
m_H=190 GeV	$98.0 \pm 0.2\%$ ($57.7 \pm 0.6\%$)
m_H=200 GeV	$97.8 \pm 0.2\%$ ($58.3 \pm 0.6\%$)
m_H=250 GeV	$97.4 \pm 0.2\%$ ($59.3 \pm 0.6\%$)
m_H=350 GeV	$97.2 \pm 0.2\%$ ($63.6 \pm 0.6\%$)
m_H=450 GeV	$97.1 \pm 0.2\%$ ($65.0 \pm 0.6\%$)
m_H=550 GeV	$97.7 \pm 0.2\%$ ($65.6 \pm 0.6\%$)
Background Samples	
$t\bar{t}$	$69.5 \pm 0.7\%$ ($5.3 \pm 0.1\%$)
$Zb\bar{b}(cc)$	$71.5 \pm 1.8\%$ ($4.0 \pm 0.1\%$)
$Zb\bar{b}(lc)$	$74.6 \pm 1.2\%$ ($9.8 \pm 0.2\%$)
$Zb\bar{b}(ll)$	$77.7 \pm 0.9\%$ ($9.3 \pm 0.2\%$)
ZZ^*	$97.8 \pm 0.1\%$ ($46.8 \pm 0.5\%$)

Table 6.5: *Efficiency on signal and background samples applying the* $\overline{DLIP} < 2.25$ *cut.*

composed by one e^+ and one e^-, belonging to the decay of each Z boson. For Higgs masses below $m_H < 2m_Z$, at least one of the two Z bosons is virtual; for Higgs masses above $2m_Z$, the two Z bosons are both real. Reconstruction of the intermediate Z bosons in the Higgs decay tree is helpful since it is possible, as it will be described in section 6.4, to use the reconstructed masses of the two Z bosons to discriminate the signal from the backgrounds. The algorithm which has been used to assign the e^+e^- pair to each Z boson is the following:

- the reconstructed (e_1^+, e_2^+) (e_1^-, e_2^-) are divided into the two combinations $(e_1^+, e_1^-), (e_2^+, e_2^-)$ and $(e_1^+, e_2^-), (e_2^+, e_1^-)$
- the combination which contains the pair with a reconstructed invariant mass nearest to the nominal Z mass is chosen; this pair is associated to the reconstructed real Z boson.

This algorithm has been evaluated to assign the right combination in more than 98% of the cases in the signal sample. The same algorithm is also used in the backgrounds to construct the invariant masses of the (e^+, e^-) pairs, which in some cases are associated to a real Z boson, like in $Zb\bar{b}$, or are arbitrarily constructed like in $t\bar{t}$ events.

6.4 Selection of $H \rightarrow ZZ^{(*)} \rightarrow 2e^+2e^-$ events

In this section several cuts are proposed to further reduce both reducible and irreducible backgrounds, enhancing the signal to noise ratio. The cuts discrimate between different topological and kinematical properties of the signal and backgrounds. These cuts are applied after the first global event reconstruction, where a partial rejection of the reducible background has already been achieved. The final values for the different cut parameters will be decided according to an optimization procedure which maximize the discovery probability, and which will be discussed in section 6.6.

6.4.1 Electron Isolation

One of the main characteristic of the signal is the presence of four isolated electrons. The same is true also for the irreducible ZZ^* background, while the backgrounds $t\bar{t}$ and $Zb\bar{b}$ have two electrons coming from b-jets, which therefore are non-isolated. These latter backgrounds can be significantly suppressed by the application of isolation criteria. This isolation step is complementary to the first rejection of electrons contained in cascade mesons decay achieved from the use of the Ele_{ID} variable.

Only isolation with the tracker is studied, since calorimetric isolation for electrons is complicated by the bremsstrahlung effect, which spread in the ϕ direction the energy deposits, as it has been described in the previous chapter.

An electron can be defined as isolated by comparing the measurement of some quantity (e.g. the sum of transverse momenta of tracks) in a cone around the electron with a predefined threshold.

The algorithm starts selecting those reconstructed tracks contained in a cone of fixed $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ size around the direction of the electron candidate. The tracks are reconstructed using the standard parameters for track reconstruction, which are a minimum p_T of 0.9 GeV/c and a minimum number of valid hits 5. Since the tracks are seeded from the pixel layers, the probability to have the in the list of the isolation tracks the e^+ and e^- associated to the conversion of a bremsstrahlung photon is neglibible.

The tracks reconstructed in this way are reconstructed in a completely separate and parallel way from the track associated to the electron candidate, therefore a track associated to the electron itself can be contained within the list of the isolation tracks. To avoid this, a veto cone can be defined around the direction of each reconstructed electron candidate. The dimension of the veto cone has been chosen looking at the distance between the reconstructed offline electron track and the nearest isolation track: it has been fixed to a value of 0.015.

Starting from a list of tracks contained into the isolation cone region, there are several ways to define an isolation criterion, all giving similar results: one possibility is to define isolated those electrons which have no track contained in the isolation cone with transverse momentum greater than a certain threshold. Another one is to select the electrons for which the sum of the transverse momentum of the tracks contained in the isolation cone is below a certain threshold.

The second possibility has been preferred.

In general, the isolation algorithms are performed separately for each electron, requiring that each of the four electrons is isolated. However, this approach can lead to double counting of isolation tracks, since different cones constructed around each electron can be overlapped.

For this reason, another possibility has been considered, which is to define an *event isolation* variable, where the isolation region is defined as the sum of the isolation cones around each electron; the p_T of the isolation tracks falling into this region is summed, avoiding double counting. The two approaches of the individual isolation on all leptons and global event isolation will be respectively referred as *All Lep* and *Evt Isol*: both have been considered in the following.

In both isolation algorithms it is important to optimise the two parameters

which identify the isolation algorithm, the p_T threshold value and the size of the cone.

The optimisation procedure is divided into two steps: at first the cone size is decided looking at the cone size which gives the best signal to noise ratio in the region of signal efficiency greater than 90%. Having fixed the cone size, the p_T threshold is then optimized in the process of maximization of the Higgs signal discovery probability for different Higgs mass hypotheses.

A set of different cone sizes ΔR has been tried:

- 0.1,0.15,0.2,0.25,0.3,0.35,0.4.

In figure 6.13 and 6.14 the background efficiency for different values of cone sizes and p_T threshold is reported as a function of signal efficiency. Efficiency is defined as the number of events that pass the isolation cuts divided by the number of events that pass the vertex selection. The Higgs sample for Higgs mass of 150 GeV is used as signal sample. Since the $ZZ^{(*)}$ background contains isolated electron as the signal sample, no rejection is achieved with the isolation criteria on this sample. *Evt Isol* should be preferred to *All Lep* in the high signal efficiency region, while *All Lep* gives better results for signal efficiencies below 80%. *Evt Isol* is therefore chosen as isolation algorithm. For what concerns the cone sizes, $\Delta R = 0.2$ and $\Delta R = 0.25$ give similar results, in particular 0.2 seems to be slightly more efficient in rejecting $t\bar{t}$ background while 0.25 is better for the $Zb\bar{b}$ background. Cone size of 0.25 has been preferred, but the results are very similar to those that can be obtained using a 0.2 cone size.

The p_T thresholds, as pointed before, will be fixed together with the other selection cuts in the optimization procedure which will be described in the next section.

6.4.2 Electron p_T cuts

The distributions of the four electrons transverse momentum sorted in decreasing order are shown in figure 6.15 for some Higgs masses (m_H=130,180,250,450 GeV) and in figure 6.16 for the backgrounds. The four values of transverse momentum sorted in decreasing order will be referred as p_T^1, p_T^2, p_T^3, p_T^4.

From the distributions, it follows that a cut on p_T^1 and p_T^2 can be effective only for high values of the Higgs masses: for example, mantaining the same level of background rejection, a cut $p_T^1 > 50$ GeV retains almost all the signal events for Higgs masses above 300 GeV, but for low Higgs masses (e.g 150 GeV) it rejects around 50% of the events. On the contrary, cuts on p_T^3, p_T^4 seem to be useful in

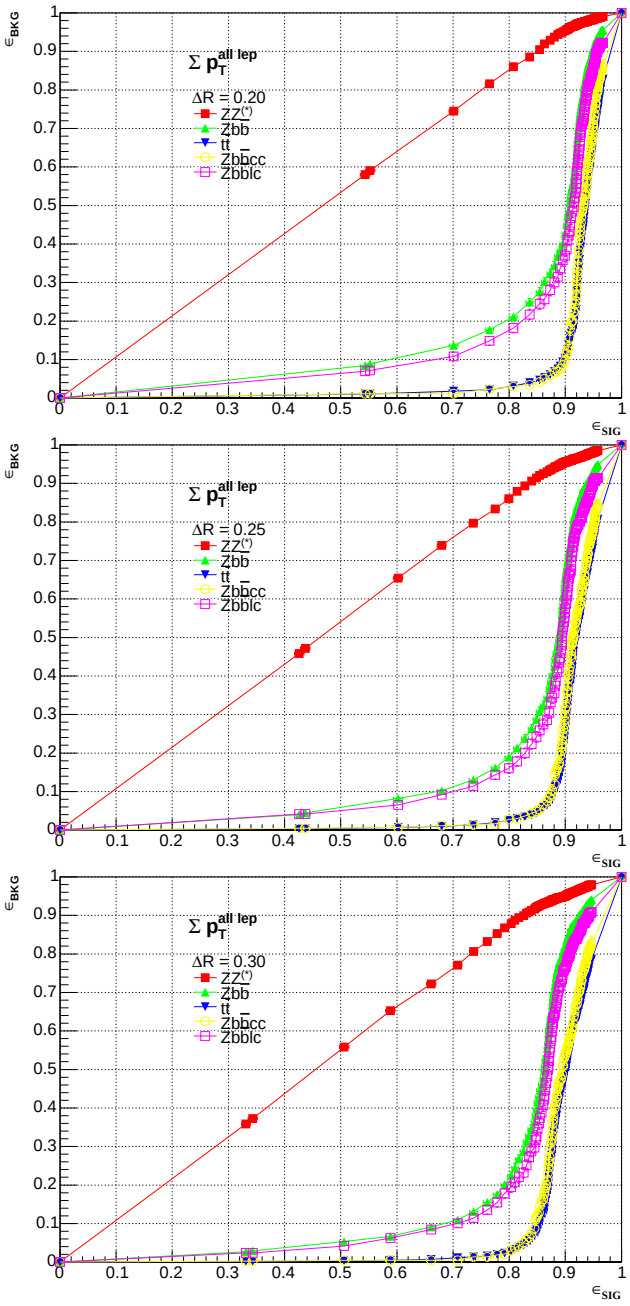


Figure 6.13: *Efficiency of the backgrounds versus signal efficiency applying isolation cuts for different coni sizes ($\Delta R = 0.2$, $\Delta R = 0.25$, $\Delta R = 0.3$). The* All Lep *approach is employed.*

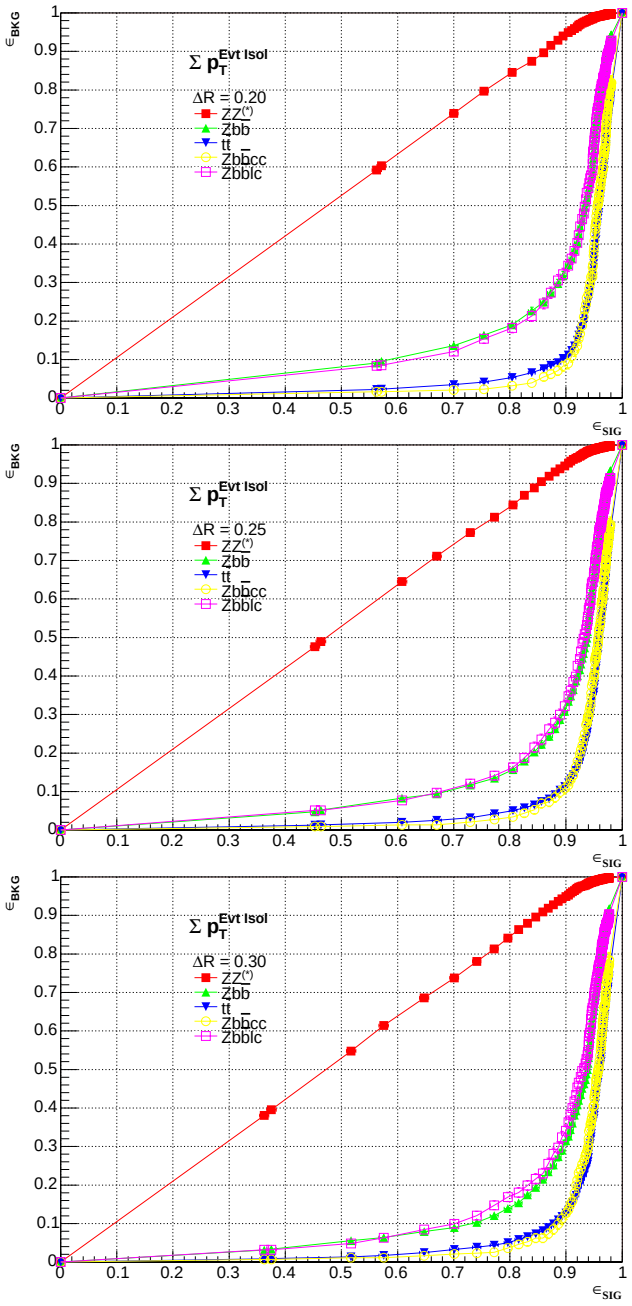


Figure 6.14: *Efficiency of the backgrounds versus signal efficiency applying isolation cuts for different coni sizes ($\Delta R = 0.2$, $\Delta R = 0.25$, $\Delta R = 0.3$). The* Evt Isol *approach is employed.*

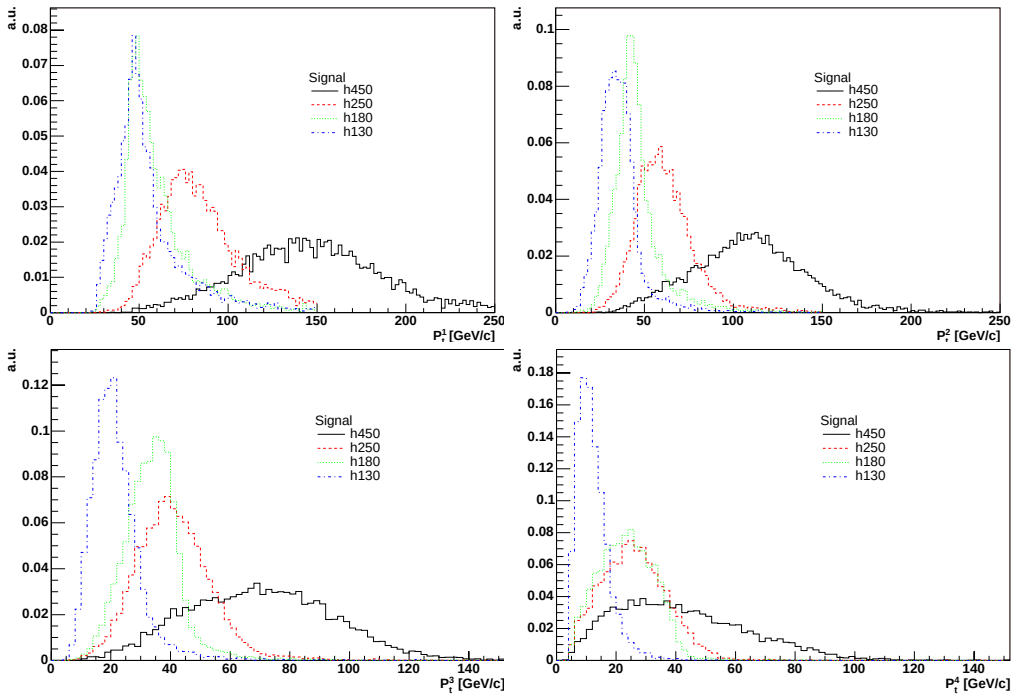


Figure 6.15: *Distributions of the four electrons transverse momentum sorted in decreasing order for some Higgs masses (m_H=130,180,250,450 GeV)*

rejecting the reducible background, $t\bar{t}$ and $Zb\bar{b}$, also in the region $m_H < 2m_Z$. The large majority of the lowest transverse momentum electrons in these two background samples, are coming from the decay tree of the b quarks, hence they present p_T distributions which are shifted towards lower values than the signal ones. However, to mantain an high efficiency on the signal, low values for the p_T^3 and p_T^4 cuts should be chosen. A cut on p_T^3 at 10 GeV is capable of rejecting nearly 50% of the $t\bar{t}$ events, with a signal selection efficiency of more than 95%. Rejection on the irreducible background $ZZ^{(*)}$ using electron p_T cuts seems to be feasible only in the Higgs high mass region, when the threshold of the p_T^1 and p_T^2 cut can be raised.

To summarize, cut on the four electrons p_T sorted in decreasing order are applied. The four values of the thresholds, $p_T^1(thr)$, $p_T^2(thr)$, $p_T^3(thr)$, $p_T^4(thr)$, will be optimized for each Higgs mass hypotheses in the discovery probability maximization step.

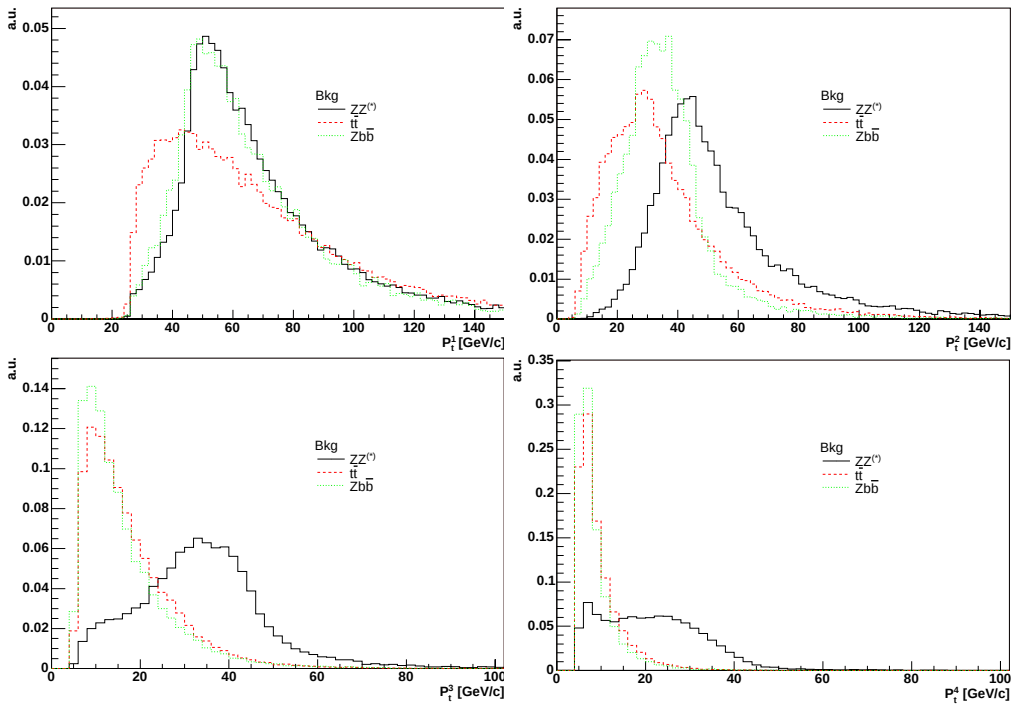


Figure 6.16: *Distributions of the four electrons transverse momentum sorted in decreasing order for the background samples.*

6.4.3 $Z(Z^{(*)})$ bosons masses cuts

The distribution of the (e^+e^-) pairs, constructed as described in section 6.3, is presented in figure 6.17 for some signal samples at different masses, and in figure 6.18 for the background samples. The distribution shows sharp edges because of the preselection cuts.

The signal events are characterized by the presence of at least a real Z boson, both for the cases when $m_H < 2m_Z$ and when $m_H > 2m_Z$. No algorithm has been applied to recover the internal bremmstrahlung; it should be noted, however, that the dominant effect, as it has been described in the previous chapter, is due to the electron radiation in tracker material. For $m_H > 2m_Z$, also the second Z boson is real, while for $m_H < 2m_Z$ it is virtual. In case of a virtual Z boson, its mass distribution presents a sharp peak at around $m_H - m_Z$.

The backgrounds events show three distinctive features:

- irreducible $ZZ^{(*)}$ background presents characteristics very similar to the signal when $m_H > 2m_Z$, which are the presence in most of the cases of two real Z boson

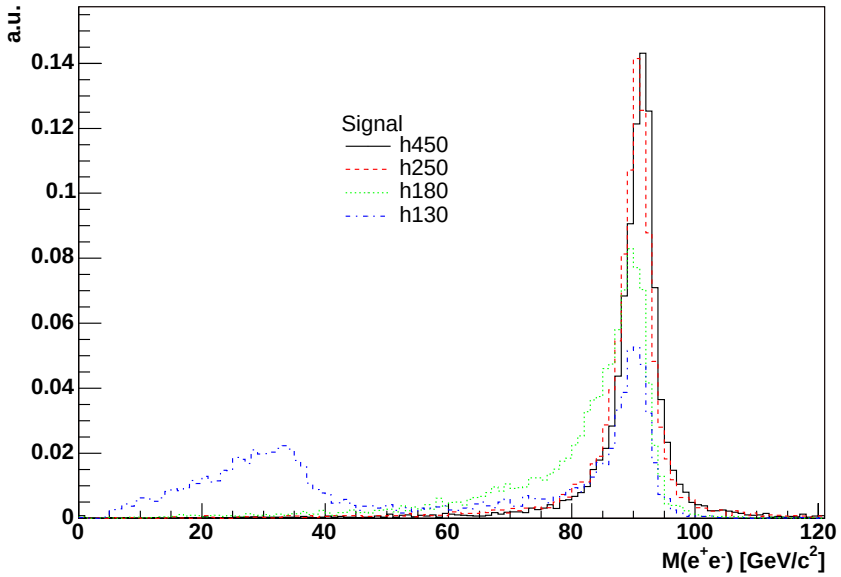


Figure 6.17: *Distribution of the (e^+e^-) pairs, constructed as described in section 6.3, for some signal samples (m_H=130,180,250,450 GeV)*

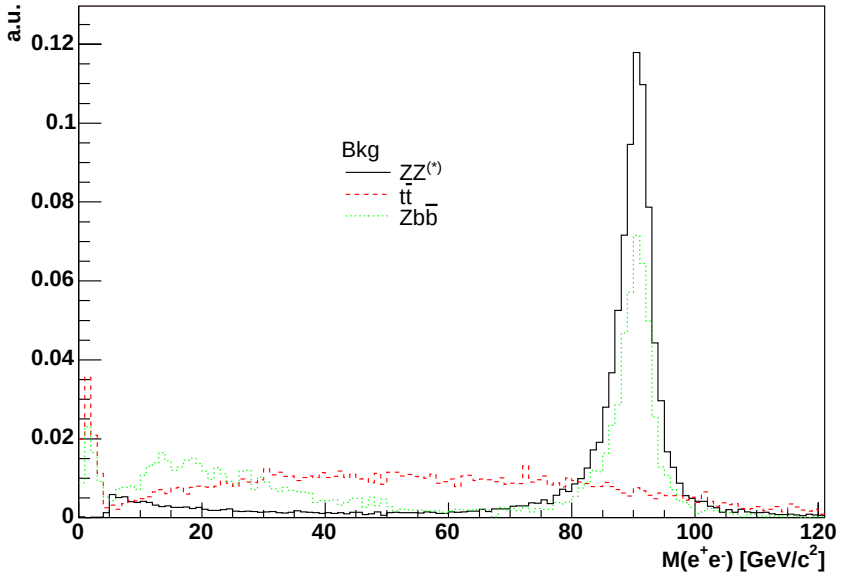


Figure 6.18: *Distribution of the (e^+e^-) pairs, constructed as described in section 6.3, for background samples.*

- $Zb\bar{b}$ events has an (e^+, e^-) pair coming from a real Z, while the other pair mass distribution presents a sharp peak at low mass due to events where the pair comes from the decay tree of the same b quark

- $t\bar{t}$ events has no real Z boson. Invariant mass of e^+e^- can lie only by chance in the mass region of a real Z boson. Also in this case a peak in the low mass region is present due to electrons originating from the same b-quark cascade.

From the analysis of the separate characteristics of signal and background, it follows that a cut on both the two e^+e^- pairs masses can efficiently reject the background. In particular, a cut on the pair associated to the real Z boson allows to reject most of the $t\bar{t}$ background, while a cut on the second pair mass allows to reject a good fraction of the irreducible background for Higgs masses $m_H < 2m_Z$ and part of the $Zb\bar{b}$ and $t\bar{t}$. For masses above $2m_Z$, the latter cut potentially could remove most of the $Zb\bar{b}$ and $t\bar{t}$ background.

Therefore two cuts are defined:

- $M_Z - \Delta^{symm}(M_Z^1) < m^1_{e^+e^-} < M_Z + \Delta^{symm}(M_Z^1)$
- $M_Z^2(min) < m^2_{e^+e^-} < M_Z^2(max)$

Three parameters are left for the cut optimization process: $\Delta^{symm}(M_Z^1)$, $M_Z^2(min)$, $M_Z^2(max)$.

6.4.4 Four electrons invariant mass cut. Higgs mass resolution

Searching the Higgs Boson in this channel, allows to reconstruct the Higgs boson invariant mass, therefore the search can be restricted in a fixed interval of the $2e^+\ 2e^-$ invariant mass defined as signal region. For masses $m_H < 2m_Z$, where the Higgs natural width is very small, the reconstructed mass is dominated by experimental resolution. An example of the Higgs mass reconstructed for $m_H =$ 120 GeV is given in figure 6.19: a tail is present due both to the internal and external bremsstrahlung.

The superimposed fit function is a gaussian with a radiative tail f_{tail}. The f_{tail} is modelled using as an additional gaussian with an exponential tail. For masses greater $m_H > 2m_Z$, the Higgs natural width, Γ_H, increasing with m_H^3, dominates the mass distribution width. The fit function used in the low mass region is also well fitting the mass distributions in this region. As an example, the distribution of the fitted invariant mass distribution for $m_H = 450$ GeV is reported in figure 6.20.

The values of the σ obtained from the gaussian fit to the mass distribution are reported in table 6.6 and are compared with the values of the Higgs natural width. This value is used to define the Higgs signal region: for masses below 250 GeV an asymmetrical cut $[m_H - 3\sigma_H, m_H + 2\sigma_H]$ is used; for masses above the 250 GeV a symmetrical window $[m_H - 3\sigma_H, m_H + 3\sigma_H]$ is exploited.

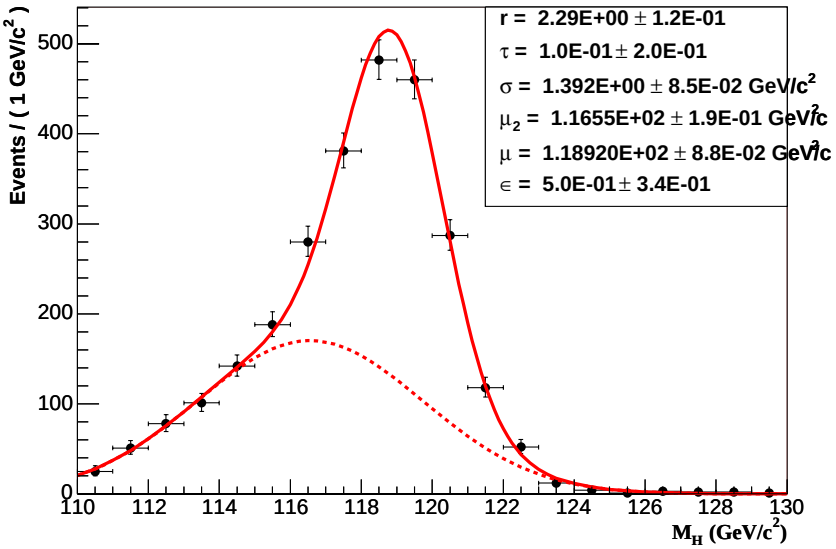


Figure 6.19: *Higgs mass reconstructed for $m_H = 120$ GeV. The fit is performed with a gaussian distribution plus a radiative tail, modelled using as an additional gaussian with an exponential tail. The parameters of the fit are: μ mean of the first gaussian, μ_2 mean of the tail sigma, σ width of the first gaussian, r is ratio between the two gaussian widths, τ is the decay parameter of the exponential tail and ε is the relative weight of the gaussian and the tail.*

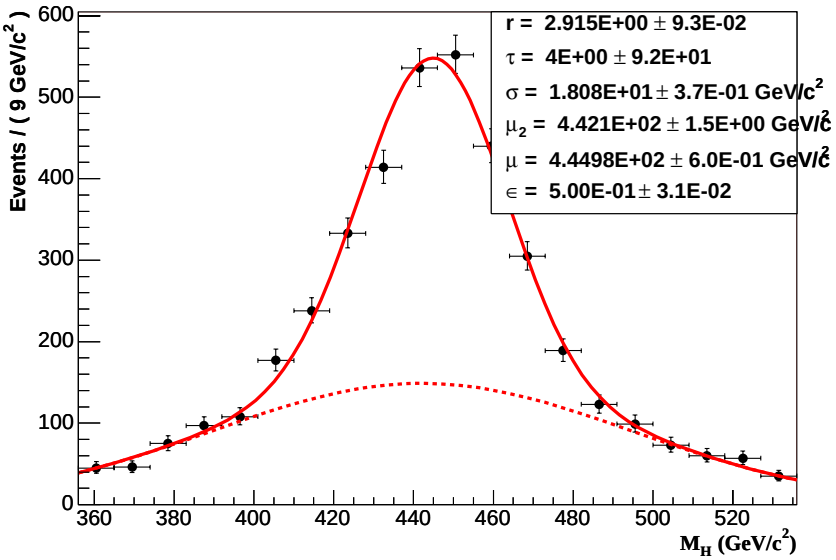


Figure 6.20: *Higgs mass reconstructed for $m_H = 450$ GeV.*

6.5 Definition of probability to observe a signal in a future experiment

The aim of this section is to define the probability to observe a particular signal with cross section σ_s over a background with cross-section σ_b, considering a

Signal Samples	Γ_H [GeV]	$\sigma(m_H)$ [GeV]
m_H=115 GeV	3.22e-03	1.33
m_H=120 GeV	3.60e-03	1.39
m_H=130 GeV	4.94e-03	1.44
m_H=140 GeV	8.06e-03	1.72
m_H=150 GeV	1.66e-02	1.77
m_H=160 GeV	7.72e-02	1.79
m_H=170 GeV	3.83e-01	1.92
m_H=180 GeV	6.28e-01	2.05
m_H=190 GeV	1.03e+00	2.20
m_H=200 GeV	1.42e+00	2.44
m_H=250 GeV	4.04e+00	4.63
m_H=350 GeV	1.54e+01	10.11
m_H=450 GeV	4.68e+01	18.08
m_H=550 GeV	9.30e+01	33.89

Table 6.6: *Higgs natuaral width compared to the reconstructed mass resolution.*

fixed level of integrated luminosity L. σ_s and σ_b denote the accepted cross sections, which take into account also the geometrical, reconstruction and selection efficiencies of the detector. Therefore, for a given luminosity L, an average number of background events $n_b = \sigma_b L$ is expected and in case of the existence of the signal an excess[2] of $n_s = \sigma_s L$ events is predicted. Therefore, in the presence of the signal a total average number of events $< n >= n_s + n_b$ is exptected.

The common practice to evaluate the possibility to observe a signal in the future experiment is to use significance estimators. Several significance estimators have been proposed and used [106], [107]:

$$S_1 = \frac{n_s}{\sqrt{n_b}}, \tag{6.1}$$

$$S_2 = \frac{n_s}{\sqrt{n_s + n_b}}, \tag{6.2}$$

$$2 \times S_{12} = 2(\sqrt{n_s + n_b} - \sqrt{n_b}). \tag{6.3}$$

$$S_L = \sqrt{2 \ln Q} \qquad Q = \left(1 + \frac{n_s}{n_b}\right)^{n_s + n_b} e^{-n_s} \tag{6.4}$$

[2]It is also possible that some particular signals proposed by new physics model can also lead to a decrease of the cross section due to destructive interference or some nonlocal formfactors. Here only the case when the signal existence leads to additional positive contribution to the background cross section is taken into account, being the consideration of the opposite case straightforward.

However, the values of the proposed significance estimators can differ significatively, so the question of which is the correct definition for the significance (S_1, S_2, $2 \times S_{12}$ or S_L), as well as the question of how the discovery probability can be estimated arises.

It is also imporant to stress that there is a crucial difference between what is the significance of a signal in a future experiment and significance of a signal in a real experiment, since in the latter case the total number of events n_{obs} is a given number, having already been measured. Here it is supposed to do just a simple counting experiment, where for example n_{obs} is the number of events counted (integrated) in a particular region of the global phase space, which can be defined as the *signal region*, and n_b is the average number of background events expected in that region. In the real experiment, if an excess of events $n_{obs} > n_b$ has been observed, the number of signal events is obtained as $n_s = n_{obs} - n_b$ and this number should be compared to the average number of background events n_b. Common definition of discovery of a signal [108] is when the probability of the background to reproduce the observed number of events is less than $2.85 \cdot 10^{-7}$, which corresponds to the integral of a normalized gaussian distribution $Gauss(x,0,1)$ (gaussian of mean 0 and σ 1) from 5 to infinity, $\int_5^\infty Gauss(x|0,1)dx$. Supposing that the average number of background events in the real experiment n_b is much greater than 1, the poisson distribution for background events n, $Poiss(n,n_b) = \frac{\lambda^n}{n!}e^{-n}$, is well reproduced by a gaussian distribution of mean n_b and sigma $\sqrt{n_b}$:

$$Poiss(n,n_b) = \int_{n-0.5}^{n+0.5} Gauss(x|n_b,\sqrt{n_b})dx. \tag{6.5}$$

The probability in the real experiment for the background events to be $n \geq n_{obs}$ in the case of $n_b \gg 1$ is:

$$P(n \geq n_{obs}) = \int_{n_{obs}}^{\infty} Gauss(x|n_b,\sqrt{n_b})dx = \int_{S_1}^{\infty} Gauss(y|0,1)dy, \tag{6.6}$$

where $y = \frac{x-n_b}{\sqrt{n_b}}$ and S_1 is

$$S_1 = \frac{n_{obs}-n_b}{\sqrt{n_b}} = \frac{n_s}{\sqrt{n_b}}. \tag{6.7}$$

This means that according to the common definition, a discovery can be affirmed if $S_1 \geq 5$, since in this case the probability for the background to reproduce the signal is less than $2.85 \cdot 10^{-7}$, $P(n \geq n_{obs}) \leq 2.85 \cdot 10^{-7}$, assuming no systematic uncertainty. However, this is valid only when $n_b \gg 1$. In all other situations, the

poisson distribution for n should be used, and a more general statistical significance s [106] can be defined:

$$\sum_{n=n_{obs}}^{\infty} Poiss(n, n_b) = \int_s^{\infty} Gauss(y|0,1)dy. \tag{6.8}$$

Note that s is function of n_{obs} and n_b, $s = s(n_{obs}, n_b)$, and that $s \to S_1$ when the gaussian limit can be assumed; this means that S_1 is a valid significance estimator for a real experiment only when $n_b \gg 1$.

The case of a future experiment should be treated in a different way. Here only the expected average number of signal and background events, n_s and n_b, is known, extracted from Montecarlo simulations of the signal and background experimental signatures in the detector. In this case, what should be compared is the realization of n events in the future experiment according the two predictions for the average number of events, the background only hypotheses, $< n >= n_b$, and the signal plus background hypotheses, $< n >= n_s + n_b$. In what follows some ideas expressed in this references [106] have been used and extended. This approach is based on the analysis of the uncertainty in the future hypotheses testing about the presence of the signal. The probability of realization of n events in a future experiment is described by the Poisson distribution with average $< n >$, $Poiss(n, < n >)$. The two hypotheses are represented by H_0, the signal is present ($< n >= n_s + n_b$), H_1, the signal is absent, ($< n >= n_b$), and the uncertainty is defined by the values of the probability to reject the hypotheses H_0 when is true (error of type I α) and the probability to accept H_0 when it is wrong (error of type II β). The hypotheses H_0 is accepted when the realized number of events n is greater than a crytical value n_0, $n > n_0$. Hence the two probabilities α and β are equal to:

$$\begin{aligned} 1 - \alpha &= \sum_{n_0+1}^{\infty} Poiss(n, n_s + n_b) \\ \beta &= \sum_{n_0+1}^{\infty} Poiss(n, n_b) \end{aligned} \tag{6.9}$$

The hypotheses testing and the two probabilities α and β are graphically represented in figure 6.21

The value of n_0 can be fixed in two ways using the equations presented in 6.9: for example, a value for α can be chosen (for example α=0.5), the first equation can be solved to find n_0, and β can be calculated. On the contrary, a value for β can be fixed, the second equation in 6.9 is solved to find n_0, and then the corresponding α can be extracted. In the latter approach, fixing β at $2.85 \cdot 10^{-7}$, $1 - \alpha$

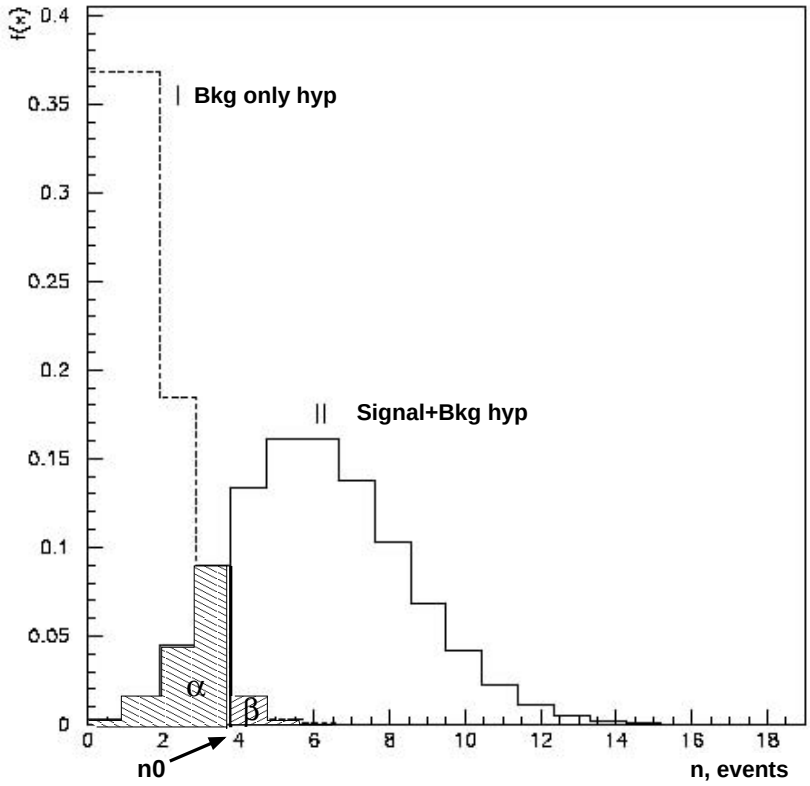


Figure 6.21: *The hypothesis testing between background only hypotheses (I) and signal plus background hypotheses (II); the two probabilities* α *and* β *defined in the text are graphically reported, together with the crytical value* n_0.

can be interpreted as the discovery probability of the future experiment, being the probability for the future experiment to observe a number of events n greater than n_0 in presence of a real signal, in such a way that the probability for the background to reproduce this number is very small. Using the other approach, which is fixing a value for α and find the corresponding β, the definition of the classical significance estimators S_1 can be obtained, supposing to be in the limiting case where the poisson distribution can be described by a gaussian one, which is when $n_b \gg 1$. In fact, if α is fixed at 0.5 and assuming valid the gaussian limit ($n_b \gg 1$), the critical value n_0 is given by $n_0 = n_b + n_s$ (the gaussian distribution for the H_0 hypotheses is symmetric around the value $n_b + n_s$) and β can be written as:

$$\beta \simeq \int_{n_s+n_b}^{\infty} Gauss(x|n_b, \sqrt{n_b})dx = \int_{S_1}^{\infty} Gauss(y|0,1)dy \qquad (6.10)$$

In this case a value of S_1 greater than 5 should be interpreted as if the future

experiment has probability of 0.5 in presence of signal to observe a number of events which can be described with probability less than $2.85{\cdot}10^{-7}$ by the background only hypotheses. Maximizing S_1 would mean to minimize the probability β that the realized number of events n in the future experiment can be described by the background only hypotheses (at a fixed level of 1-α). It should be noted that S_1 is a valid estimator only in the gaussian limit. In the general case, it is possible to define a statistical significance $s(\alpha, n_s, n_b)$ using the relation:

$$\sum_{n=n_0(\alpha,n_s,n_b)+1}^{\infty} Poiss(n, n_b) = \int_{s(\alpha,n_s,n_b)}^{\infty} Gauss(y|0,1)dy. \tag{6.11}$$

$s(\alpha, n_s, n_b)$ will tend to S_1 if α is chosen to be 0.5 and $n_b \gg 1$.

From this discussion, it follows that the correct approach to define the discovery probability for a future experiment is to evaluate $1-\alpha$ having fixed β. The value of $1-\alpha$ will be referred as $S_{poisson}$ and will have values between 0. and 1., being the probability in the future experiment to make an observation compatible with the background only hypotheses with probability less than $2.85{\cdot}10^{-7}$. Choosing different values for β, it is also possible to define estimators of the probability to have strong ($\beta = 2.85 \cdot 10^{-3}$) or weak ($\beta = 4.56 \cdot 10^{-2}$) evidence of the signal in the future experiment.

Up to now it has not been considered neither any statistical nor any systematic uncertainty in the determination of n_s and n_b; in the proposed approach they can be naturally introduced. For example, n_s and n_b would have statistical fluctuations since they are extracted from a Montecarlo simulation of limited statistics. To study this problem, it is convenient to introduce the two variables $\lambda_{s+b}(L)$ and $\lambda_b(L)$, representing the average values of the poisson distribution for the signal plus background hypotheses and for the background only hypotheses at integrated luminosity L. Also their distribution probability, $g(\lambda_{s+b}(L))$ and $g(\lambda_b(L))$, should be introduced.

Assuming the knowledge of the two distribution probabilities $g(\lambda_{s+b}(L))$ and $g(\lambda_b(L))$, the probabilities α and β can written as:

$$\begin{aligned} 1-\alpha &= 1 - \int_0^{\infty} g(\lambda_{s+b}(L)) \sum_{h=0}^{n_0} Poiss(h|\lambda_{s+b}) d\lambda_{s+b} \\ \beta &= \int_0^{\infty} g(\lambda_b(L)) \sum_{h=n_0+1}^{\infty} Poiss(h|\lambda_b) d\lambda_b \end{aligned} \tag{6.12}$$

The main difficulty is now represented by the calculation of the $g(\lambda_{s+b}(L))$ and $g(\lambda_b(L))$, and of the convolution integral.

Systematic uncertainties can be treated defining the set of parameters $\vec{\sigma}$ on which $\lambda_{s+b}(L)$ and $\lambda_b(L)$ depend (for example the signal and background theoretical cross-sections). Assuming to know the probability distribution of $\vec{\sigma}$, $g(\lambda_{s+b}(L))$ and $g(\lambda_b(L))$ can be obtained as:

$$\begin{aligned} g(\lambda_{s+b}(L)) &= \int g(\lambda_{s+b}(L)|\vec{\sigma})P(\vec{\sigma})d\vec{\sigma} \\ g(\lambda_b(L)) &= \int g(\lambda_b(L)|\vec{\sigma})P(\vec{\sigma})d\vec{\sigma}. \end{aligned} \tag{6.13}$$

In practical situations the evaluation of the integrals can be difficult, but they can be evaluated numerically.

For what concerns the statistical fluctuactions due to the limited Montecarlo statistics, a calculation of $g(\lambda_{s+b}(L))$ and $g(\lambda_b(L))$ is now proposed. This calculation exploits the use of the bayesian inference.

A first problem is represented by the fact that inference on $\lambda_{s+b}(L)$ and $\lambda_b(L)$ is in general made with several Montecarlo samples of different statistics, each of them representing the simulation of a particular process contributing either to the signal or to the background. The different samples will be treated as indipendent samples. The different Montecarlo samples may have different statistics or better corresponding Montecarlo integrated luminosity, L_{MC}, defined as

$$L_{MC} = \frac{n_{MC}}{\sigma_{MC}}, \tag{6.14}$$

where n_{MC} is the number of events generated in a particular Montecarlo sample and σ_{MC} is the corresponding cross-section for the particular process. The total number of samples is called N_{tot} , with $N_{tot} \geq 2$, of which the first $N_{bkg} \geq 1$ represents the number of background Montecarlo samples.
On each Montecarlo sample the full chain of selection is applied; this would result in a selected number of Montecarlo events n_{MC}^{SEL} integrated in the phase space region defined as the *signal region*. It is possible to define $m_i = \frac{L_i}{L}$, where the index i represents each Montecarlo sample, as the ratio of Montecarlo integrated luminosity of sample i, L_i, over the luminosity L, the luminosity at which the number of events for each sample should be evaluated. Calling $\lambda_i(L)$ the expected number of selected events for sample i at luminosity L, it is possible to infer the distribution probability of $\lambda_i(L)$, given m_i and n_i^{SEL}, $g(\lambda_i(L)|m_i, n_i^{SEL})$, which is equal to:

$$g(\lambda_i(L)|m_i, n_i^{SEL}) = \frac{m_i^{(1+n_i^{SEL})}}{n_i^{SEL}!} e^{-m_i\lambda_i(L)} \lambda_i(L)^{n_i^{SEL}}. \tag{6.15}$$

This result is based on the bayesian inference schema [109].

$g(\lambda_i(L)|m_i, n_i^{SEL})$ is represented by the so-called *Gamma distribution*[3] of parameters m_i and n_i^{SEL}, $\Gamma_{m_i,n_i^{SEL}}(\lambda_i(L))$. Treating the different $\lambda_i(L)$ as indipendent variables, it is possible to compute the probability distribution for $\lambda_{s+b}(L)$ and for $\lambda_b(L)$. Taking for example $\lambda_{s+b}(L) = \sum_i \lambda_i(L)$ where the index *i* runs over all the N_{tot} Montecarlo samples (all of them are contributing to signal and background), it is possible to write the distribution probability $g(\lambda_{s+b}(L)|\vec{m}_{s+b}, \vec{n}_{s+b})$ as:

$$g(\lambda_{s+b}(L)|\vec{m}_{s+b}, \vec{n}_{s+b}^{SEL}) = \int \delta(\lambda - \sum_i \lambda_i(L)) \prod_i g(\lambda_i(L)|m_i, n_i^{SEL}) d\lambda_1(L) \cdots d\lambda_{N_{tot}}(L), \tag{6.16}$$

where the vector $\vec{m}_{s+b}$ has components $(m_1, \cdots, m_{N_{tot}})$ and the vector $\vec{n}_{s+b}^{SEL}$ has components $(n_1^{SEL}, \cdots, n_{N_{tot}}^{SEL})$. The same expressions holds also for $g(\lambda_b(L)|\vec{m}_b, \vec{n}_b)$ writing $\lambda_b(L) = \sum_i \lambda_i(L)$, where now the index *i* runs only over the background Montecarlo samples N_{bkg}. The equation 6.12 can be rewritten as

$$\begin{aligned} \alpha &= \int_0^\infty g(\lambda_{s+b}(L)|\vec{m}_{s+b}, \vec{n}_{s+b}) \sum_{h=0}^{n_0} Poiss(h|\lambda_{s+b}) d\lambda_{s+b} \\ \beta &= 1 - \int_0^\infty g(\lambda_b(L)|\vec{m}_b, \vec{n}_b) \sum_{h=0}^{n_0} Poiss(h|\lambda_b) d\lambda_b; \end{aligned} \tag{6.17}$$

and is now possible to proceed to the evaluation of the integrals present in the two equations. They can be treated in the same way; the details of the mathematical treatment are shown in appendix. The result for α is

$$\begin{aligned} \alpha = & \left[\prod_{i=1}^{N_{TOT}} \left(\frac{m_i}{1+m_i} \right)^{n_i^{SEL}+1} \right] \sum_{h=0}^{n_0} \\ & \{ \sum_{k_1=0}^{h} \cdots \sum_{k_{N_{TOT}-1}=0}^{k_{N_{TOT}-2}} \frac{C_{h-k_1}^{n_1^{SEL}}}{(1+m_1)^{h-k_1}} \cdot \frac{C_{k_1-k_2}^{n_2^{SEL}}}{(1+m_2)^{k_1-k_2}} \cdots \\ & \cdot \frac{C_{k_{N_{TOT}-2}-k_{N_{TOT}-1}}^{n_{N_{TOT}-1}^{SEL}}}{(1+m_{N_{TOT}-1})^{k_{N_{TOT}-2}-k_{N_{TOT}-1}}} \cdot \frac{C_{k_{N_{TOT}-1}}^{n_{N_{TOT}}^{SEL}}}{(1+m_{N_{TOT}})^{k_{N_{TOT}-1}}} \}, \end{aligned} \tag{6.18}$$

[3]The Gamma distribution $\Gamma_{a,n+1}(\lambda) = \frac{a^{n+1}}{\Gamma(n+1)} e^{-\lambda} \lambda^n$, where *a* is a scale parameter, $n+1$ is a shape parameter, λ is a random variable and $\Gamma(n+1) = n!$ is the Gamma function.

while for β

$$\beta = 1 - \left[\prod_{i=1}^{N_{BKG}} \left(\frac{m_i}{1+m_i}\right)^{n_i^{SEL}+1}\right] \sum_{h=0}^{n_0} \{\sum_{k_1=0}^{h} \cdots \sum_{k_{N_{BKG}-1}=0}^{k_{N_{BKG}-2}} \frac{C_{h-k_1}^{n_1^{SEL}}}{(1+m_1)^{h-k_1}} \cdot \frac{C_{k_1-k_2}^{n_2^{SEL}}}{(1+m_2)^{k_1-k_2}} \cdots \cdot \frac{C_{k_{N_{BKG}-2}-k_{N_{BKG}-1}}^{n_{N_{BKG}-1}^{SEL}}}{(1+m_{N_{BKG}-1})^{k_{N_{BKG}-2}-k_{N_{BKG}-1}}} \cdot \frac{C_{k_{N_{BKG}-1}}^{n_{N_{BKG}}^{SEL}}}{(1+m_{N_{BKG}})^{k_{N_{BKG}-1}}}\}. \tag{6.19}$$

In the previous equations the simbol C_m^n denote the binomial combinatorial factor:

$$C_m^n = \binom{m+n}{n}. \tag{6.20}$$

These two expressions will be used for the definition of the discovery probability taking into account the limited statistics of the Montecarlo samples.

6.6 Optimization of the selection

A multivariate technique has been developed in order to find the optimal set of cuts which maximize the signal discovery probability. The *quality factor* to be maximized is hence represented by the discovery probability defined in the previous section. In particular, the expression 6.18 and 6.19 are used for the computation of the critical value n_0 (fixing β at $2.85 \cdot 10^{-7}$) and of the discovery probability $1-\alpha$. In the computation of α and β, all the background samples discussed in section 6.1.2 have been included. A code has been developed in order to have the possibility to vary in a predefined interval all the different cut parameters introduced in section 6.4. The actual implementation is represented by a `C++` set of classes, which after applying the selection cuts, returns the discovery probability for the particular set of cut parameters which have been used. The selection program has been interfaced to the minimization program MINUIT [110], which has been employed to search for a maximum of the discovery probability. In particular, the SIMPLEX algorithm has been used, since for the particular application, it has been found to be faster and more reliable than the standard minimization algorithm MIGRAD. The threshold cuts are varied together and applied in a definite order which is:

1. $Pt_{ISO} < Pt_{ISO}(thr)$

2. $p_T^1 > p_T^1(thr)$

3. $p_T^2 > p_T^2(thr)$

4. $p_T^3 > p_T^3(thr)$

5. $p_T^4 > p_T^4(thr)$

6. $|M_Z - M_Z^1| < \Delta^{symm}(M_Z^1)$

7. $M_Z^2(min) < M_Z^2 < M_Z^2(max)$

8. $M_H(min) < M_H < M_H(max)$.

The possibility to vary together all the cuts in the analysis is the main advantage of this technique, which allows to take naturally into account all the possible correlations between different cut variables. In the optimization process an integrated luminsity of $20 fb^{-1}$ has been assumed. The values of the optimized cuts are reported in table 6.7 for the different Higgs mass hypotheses. The optimization has been tried also with $10 fb^{-1}$, and very similar values for the cuts have been found.

The efficiency of each cut, defined as the number of the events which pass the considered cut over the events passing the previous one, is reported in table 6.8 and 6.9. The global efficiency is given between parentheses. The global efficiencies for the background samples can be found in table 6.10. The accepted cross sections for signal and background are instead given in table 6.11.

Looking at table 6.11, it is clear that $ZZ^{(*)}$ background remains as the most important contribution in all the mass region considered. For masses $m_H > 2m_Z$ the $t\bar{t}$ and the $Zb\bar{b}$ backgrounds can be completely neglected, being effectively reduced by the cut in the two real Z boson mass. This is also the explanation of the gradual relaxation on the isolation cut going towards higher masses. Some additional cuts can be foreseen in order to reduce the irreducible background based on other kinematical properties of the signal, compared to the background.

In figure 6.22 the expected selected events for signal and background in 20 fb^{-1} is reported.

An example of the effect of the optimized selection is visible looking, for example, at the invariant mass distribution for signal and background normalized at the expected number of events in 20 fb^{-1}. This comparison is given for three mass hypotheses, 130 GeV, 200 GeV and 550 GeV, in figures 6.23, 6.24 and 6.25.

Cut Name	m_H =115 GeV	m_H =120 GeV	m_H =130 GeV	m_H =140 GeV	m_H =150 GeV	m_H =160 GeV	m_H =170 GeV
$Pt_{ISO}(thr)$ GeV	2.1	2.8	3.3	2.9	3.6	6.2	3.4
$p_T^1(thr)$ GeV	27	20	25	22	22	29	20
$p_T^2(thr)$ GeV	17	15	16	12	16	23	15
$p_T^3(thr)$ GeV	12	13	16	15	17	15	14
$p_T^4(thr)$ GeV	4	5	5	5	7	10	7
$\Delta^{symm}(M_Z^1)$ GeV	30	30	16	30	30	30	19
$M_Z^2(min)$ GeV	7	6	9	16	14	9	34
$M_Z^2(max)$ GeV	69	58	61	69	70	83	84
$M_H(min)$ GeV	110.0	114.7	124.6	133.8	143.2	153.4	163.1
$M_H(max)$ GeV	116.7	121.7	131.8	142.4	152.0	162.4	172.7
Cut Name	m_H =180 GeV	m_H =190 GeV	m_H =200 GeV	m_H =250 GeV	m_H =350 GeV	m_H =450 GeV	m_H =550 GeV
$Pt_{ISO}(thr)$ GeV	5.1	14.3	14.4	12.1	12.1	49.0	9.3
$p_T^1(thr)$ GeV	20	27	29	25	69	76	89
$p_T^2(thr)$ GeV	15	21	24	20	20	38	24
$p_T^3(thr)$ GeV	15	15	11	3	28	44	44
$p_T^4(thr)$ GeV	6	0	0	6	0	13	17
$\Delta^{symm}(M_Z^1)$ GeV	16	17	30	12	16	18	19
$M_Z^2(min)$ GeV	47	70	69	62	47	46	40
$M_Z^2(max)$ GeV	100	130	109	140	116	140	129
$M_H(min)$ GeV	172.2	182.1	191.4	233.9	320.5	390.8	438.1
$M_H(max)$ GeV	182.5	193.2	203.6	257.1	380.0	492.2	641.4

Table 6.7: *Table with the optimized cuts for the various mass hypotheses.*

Cut Name	m_H =115 GeV	m_H =120 GeV	m_H =130 GeV	m_H =140 GeV
$Pt_{ISO} < Pt_{ISO}(thr)$	0.806 ± 0.008 (0.288 ± 0.005)	0.845 ± 0.007 (0.316 ± 0.005)	0.857 ± 0.006 (0.361 ± 0.005)	0.844 ± 0.007 (0.383 ± 0.006)
$p_T^1 > p_T^1(thr)$	0.991 ± 0.002 (0.286 ± 0.005)	1.000 ± 0.000 (0.316 ± 0.005)	1.000 ± 0.000 (0.361 ± 0.005)	1.000 ± 0.000 (0.383 ± 0.006)
$p_T^2 > p_T^2(thr)$	0.980 ± 0.003 (0.280 ± 0.005)	0.992 ± 0.002 (0.313 ± 0.005)	0.993 ± 0.002 (0.359 ± 0.005)	1.000 ± 0.000 (0.383 ± 0.006)
$p_T^3 > p_T^3(thr)$	0.850 ± 0.008 (0.238 ± 0.005)	0.859 ± 0.007 (0.269 ± 0.005)	0.829 ± 0.007 (0.297 ± 0.005)	0.915 ± 0.006 (0.351 ± 0.006)
$p_T^4 > p_T^4(thr)$	0.999 ± 0.001 (0.238 ± 0.005)	0.997 ± 0.001 (0.268 ± 0.005)	0.991 ± 0.002 (0.295 ± 0.005)	1.000 ± 0.000 (0.351 ± 0.006)
$\lvert M_Z - M_Z^1 \rvert < \Delta^{symm}(M_Z^1)$	0.780 ± 0.010 (0.186 ± 0.004)	0.887 ± 0.007 (0.238 ± 0.005)	0.834 ± 0.008 (0.246 ± 0.005)	0.986 ± 0.003 (0.346 ± 0.006)
$M_Z^2(min) < M_Z^2 < M_Z^2(max)$	0.966 ± 0.005 (0.179 ± 0.004)	0.988 ± 0.003 (0.235 ± 0.005)	0.970 ± 0.004 (0.238 ± 0.005)	0.925 ± 0.006 (0.320 ± 0.006)
$M_H(min) < M_H < M_H(max)$	0.798 ± 0.011 (0.143 ± 0.004)	0.829 ± 0.009 (0.195 ± 0.004)	0.827 ± 0.009 (0.197 ± 0.004)	0.819 ± 0.009 (0.262 ± 0.006)
Cut Name	m_H =150 GeV	m_H =160 GeV	m_H =170 GeV	
$Pt_{ISO} < Pt_{ISO}(thr)$	0.870 ± 0.006 (0.427 ± 0.006)	0.923 ± 0.004 (0.484 ± 0.006)	0.859 ± 0.006 (0.469 ± 0.006)	
$p_T^1 > p_T^1(thr)$	1.000 ± 0.000 (0.427 ± 0.006)	0.997 ± 0.001 (0.482 ± 0.006)	1.000 ± 0.000 (0.469 ± 0.006)	
$p_T^2 > p_T^2(thr)$	0.999 ± 0.001 (0.427 ± 0.006)	0.984 ± 0.002 (0.474 ± 0.006)	1.000 ± 0.000 (0.469 ± 0.006)	
$p_T^3 > p_T^3(thr)$	0.927 ± 0.005 (0.396 ± 0.006)	0.975 ± 0.003 (0.462 ± 0.006)	0.990 ± 0.002 (0.464 ± 0.006)	
$p_T^4 > p_T^4(thr)$	0.963 ± 0.004 (0.381 ± 0.006)	0.901 ± 0.005 (0.417 ± 0.006)	0.984 ± 0.002 (0.457 ± 0.006)	
$\lvert M_Z - M_Z^1 \rvert < \Delta^{symm}(M_Z^1)$	0.993 ± 0.002 (0.378 ± 0.006)	0.997 ± 0.001 (0.415 ± 0.006)	0.985 ± 0.002 (0.450 ± 0.006)	
$M_Z^2(min) < M_Z^2 < M_Z^2(max)$	0.963 ± 0.004 (0.364 ± 0.006)	0.990 ± 0.002 (0.411 ± 0.006)	0.909 ± 0.005 (0.409 ± 0.006)	
$M_H(min) < M_H < M_H(max)$	0.819 ± 0.008 (0.298 ± 0.005)	0.788 ± 0.007 (0.324 ± 0.005)	0.787 ± 0.008 (0.321 ± 0.006)	

Table 6.8: *Signal efficiency for the mass hypotheses $m_H \leq 170$ applying each optimized cut. In parentheses the global efficiency (selected events/preselected events) after each cut is reported*

Cut Name	m_H =180 GeV	m_H =190 GeV	m_H =200 GeV	m_H =250 GeV
$Pt_{ISO} < Pt_{ISO}(thr)$	0.907 ± 0.005 (0.506 ± 0.006)	0.955 ± 0.003 (0.551 ± 0.006)	0.954 ± 0.003 (0.556 ± 0.006)	0.941 ± 0.004 (0.558 ± 0.006)
$p_T^1 > p_T^1(thr)$	1.000 ± 0.000 (0.506 ± 0.006)	1.000 ± 0.000 (0.551 ± 0.006)	1.000 ± 0.000 (0.556 ± 0.006)	1.000 ± 0.000 (0.558 ± 0.006)
$p_T^2 > p_T^2(thr)$	0.999 ± 0.000 (0.505 ± 0.006)	0.998 ± 0.001 (0.550 ± 0.006)	0.995 ± 0.001 (0.553 ± 0.006)	1.000 ± 0.000 (0.558 ± 0.006)
$p_T^3 > p_T^3(thr)$	0.993 ± 0.001 (0.502 ± 0.006)	0.994 ± 0.001 (0.547 ± 0.006)	0.999 ± 0.000 (0.552 ± 0.006)	1.000 ± 0.000 (0.558 ± 0.006)
$p_T^4 > p_T^4(thr)$	0.994 ± 0.001 (0.499 ± 0.006)	1.000 ± 0.000 (0.547 ± 0.006)	1.000 ± 0.000 (0.552 ± 0.006)	0.995 ± 0.001 (0.555 ± 0.006)
$\lvert M_Z - M_Z^1 \rvert < \Delta^{symm}(M_Z^1)$	0.990 ± 0.002 (0.494 ± 0.006)	0.998 ± 0.001 (0.546 ± 0.006)	1.000 ± 0.000 (0.552 ± 0.006)	0.996 ± 0.001 (0.552 ± 0.006)
$M_Z^2(min) < M_Z^2 < M_Z^2(max)$	0.927 ± 0.004 (0.458 ± 0.006)	0.880 ± 0.005 (0.480 ± 0.006)	0.901 ± 0.005 (0.498 ± 0.006)	0.939 ± 0.004 (0.519 ± 0.006)
$M_H(min) < M_H < M_H(max)$	0.796 ± 0.007 (0.364 ± 0.006)	0.782 ± 0.007 (0.376 ± 0.006)	0.794 ± 0.007 (0.395 ± 0.006)	0.822 ± 0.006 (0.426 ± 0.006)
Cut Name	m_H =350 GeV	m_H =450 GeV	m_H =550 GeV	
$Pt_{ISO} < Pt_{ISO}(thr)$	0.928 ± 0.004 (0.590 ± 0.006)	0.966 ± 0.003 (0.628 ± 0.006)	0.924 ± 0.004 (0.607 ± 0.006)	
$p_T^1 > p_T^1(thr)$	0.963 ± 0.003 (0.568 ± 0.006)	0.971 ± 0.003 (0.609 ± 0.006)	0.967 ± 0.003 (0.587 ± 0.006)	
$p_T^2 > p_T^2(thr)$	1.000 ± 0.000 (0.568 ± 0.006)	0.999 ± 0.001 (0.608 ± 0.006)	1.000 ± 0.000 (0.587 ± 0.006)	
$p_T^3 > p_T^3(thr)$	0.956 ± 0.003 (0.543 ± 0.006)	0.901 ± 0.005 (0.548 ± 0.006)	0.945 ± 0.004 (0.554 ± 0.006)	
$p_T^4 > p_T^4(thr)$	1.000 ± 0.000 (0.543 ± 0.006)	0.965 ± 0.003 (0.529 ± 0.006)	0.956 ± 0.003 (0.530 ± 0.006)	
$\lvert M_Z - M_Z^1 \rvert < \Delta^{symm}(M_Z^1)$	0.996 ± 0.001 (0.541 ± 0.006)	0.997 ± 0.001 (0.528 ± 0.006)	0.996 ± 0.001 (0.528 ± 0.006)	
$M_Z^2(min) < M_Z^2 < M_Z^2(max)$	0.948 ± 0.004 (0.513 ± 0.006)	0.962 ± 0.003 (0.507 ± 0.006)	0.975 ± 0.003 (0.515 ± 0.006)	
$M_H(min) < M_H < M_H(max)$	0.833 ± 0.006 (0.427 ± 0.006)	0.813 ± 0.007 (0.412 ± 0.006)	0.845 ± 0.006 (0.435 ± 0.006)	

Table 6.9: *Signal efficiency for the mass hypotheses $m_H \geq 180$ GeV applying each optmized cut. In parentheses the global efficiency (selected events/preselected events) after each cut is reported*

Background Samples	m_H =115 GeV	m_H =120 GeV	m_H =130 GeV	m_H =140 GeV	m_H =150 GeV	m_H =160 GeV	m_H =170 GeV
$t\bar{t}$	0.00003 ± 0.00002	0.00008 ± 0.00003	0.00000 ± 0.00000	0.00001 ± 0.00001	0.00001 ± 0.00001	0.00000 ± 0.00000	0.00000 ± 0.00000
$Zb\bar{b}(cc)$	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000
$Zb\bar{b}(lc)$	0.00009 ± 0.00009	0.00019 ± 0.00013	0.00009 ± 0.00009	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00009 ± 0.00009	0.00000 ± 0.00000
$Zb\bar{b}(ll)$	0.00038 ± 0.00014	0.00048 ± 0.00016	0.00021 ± 0.00011	0.00032 ± 0.00013	0.00016 ± 0.00009	0.00011 ± 0.00008	0.00021 ± 0.00011
ZZ^*	0.00242 ± 0.00030	0.00341 ± 0.00035	0.00334 ± 0.00035	0.00360 ± 0.00036	0.00323 ± 0.00034	0.00503 ± 0.00043	0.00572 ± 0.00046
Background Samples	m_H =180 GeV	m_H =190 GeV	m_H =200 GeV	m_H =250 GeV	m_H =350 GeV	m_H =450 GeV	m_H =550 GeV
$t\bar{t}$	0.00001 ± 0.00001	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00003 ± 0.00002	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000
$Zb\bar{b}(cc)$	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000
$Zb\bar{b}(lc)$	0.00009 ± 0.00009	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000
$Zb\bar{b}(ll)$	0.00005 ± 0.00005	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00005 ± 0.00005	0.00000 ± 0.00000	0.00000 ± 0.00000	0.00000 ± 0.00000
ZZ^*	0.01233 ± 0.00067	0.03240 ± 0.00107	0.04143 ± 0.00121	0.04099 ± 0.00120	0.02055 ± 0.00086	0.00859 ± 0.00056	0.00712 ± 0.00051

Table 6.10: *Efficiencies (selected events/preselected events) for background samples applying the cuts for the different mass hypotheses.*

Sample	m_H =115 GeV	m_H =120 GeV	m_H =130 GeV	m_H =140 GeV	m_H =150 GeV	m_H =160 GeV	m_H =170 GeV
Signal [fb]	0.0335	0.0814	0.1965	0.4252	0.5353	0.2778	0.1302
$t\bar{t}$ [fb]	0.0039	0.0116	0.0000	0.0019	0.0019	0.0000	0.0000
$Zb\bar{b}(cc)$ [fb]	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$Zb\bar{b}(lc)$ [fb]	0.0016	0.0033	0.0016	0.0000	0.0000	0.0016	0.0000
$Zb\bar{b}(ll)$ [fb]	0.0047	0.0061	0.0027	0.0041	0.0020	0.0014	0.0027
ZZ^* [fb]	0.0115	0.0163	0.0159	0.0171	0.0154	0.0240	0.0273
Sample	m_H =180 GeV	m_H =190 GeV	m_H =200 GeV	m_H =250 GeV	m_H =350 GeV	m_H =450 GeV	m_H =550 GeV
Signal [fb]	0.3507	1.2542	1.4458	1.1883	1.0014	0.5414	0.2812
$t\bar{t}$ [fb]	0.0019	0.0000	0.0000	0.0039	0.0000	0.0000	0.0000
$Zb\bar{b}(cc)$ [fb]	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$Zb\bar{b}(lc)$ [fb]	0.0016	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$Zb\bar{b}(ll)$ [fb]	0.0007	0.0000	0.0000	0.0007	0.0000	0.0000	0.0000
ZZ^* [fb]	0.0587	0.1544	0.1974	0.1953	0.0979	0.0409	0.0339

Table 6.11: *Accepted cross section in the signal region for signal and background samples for the different mass hypotheses.*

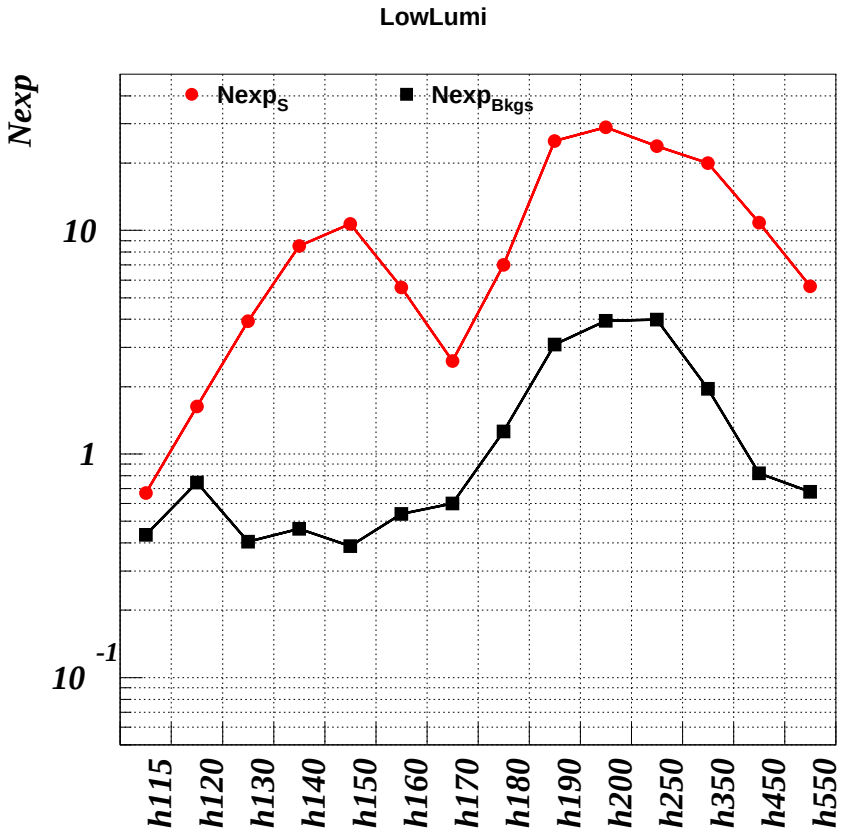


Figure 6.22: *Expected number of selected events for signal and background in 20 fb^{-1}.*

6.7 Results on the discovery probability and signal visibility

Figure 6.26 reports the discovery probability assuming an integrated luminosity of 20 fb^{-1}. The discovery probability is defined as explained in section 6.5, fixing β (the probability for the background to imitate the signal) at $2.85 \cdot 10^{-7}$, finding the critical value n_0, and integrating the poisson distribution for the signal plus background hypotheses. Two values are reported: $S_{Poisson}$ is obtained from the equation 6.9, where the limited statistics of the Montecarlo samples is not considered, while $S_{ExtendedPoisson}$ is obtained from equations 6.18 and 6.19 which takes into account the statistical fluctuactions of the expected number of signal and background samples due to the limited statistics of Montecarlo. The effect of the statistical fluctuactions is signicant only for low values of the discovery probability.

The Higgs mass intervals in which the Higgs discovery probability in the

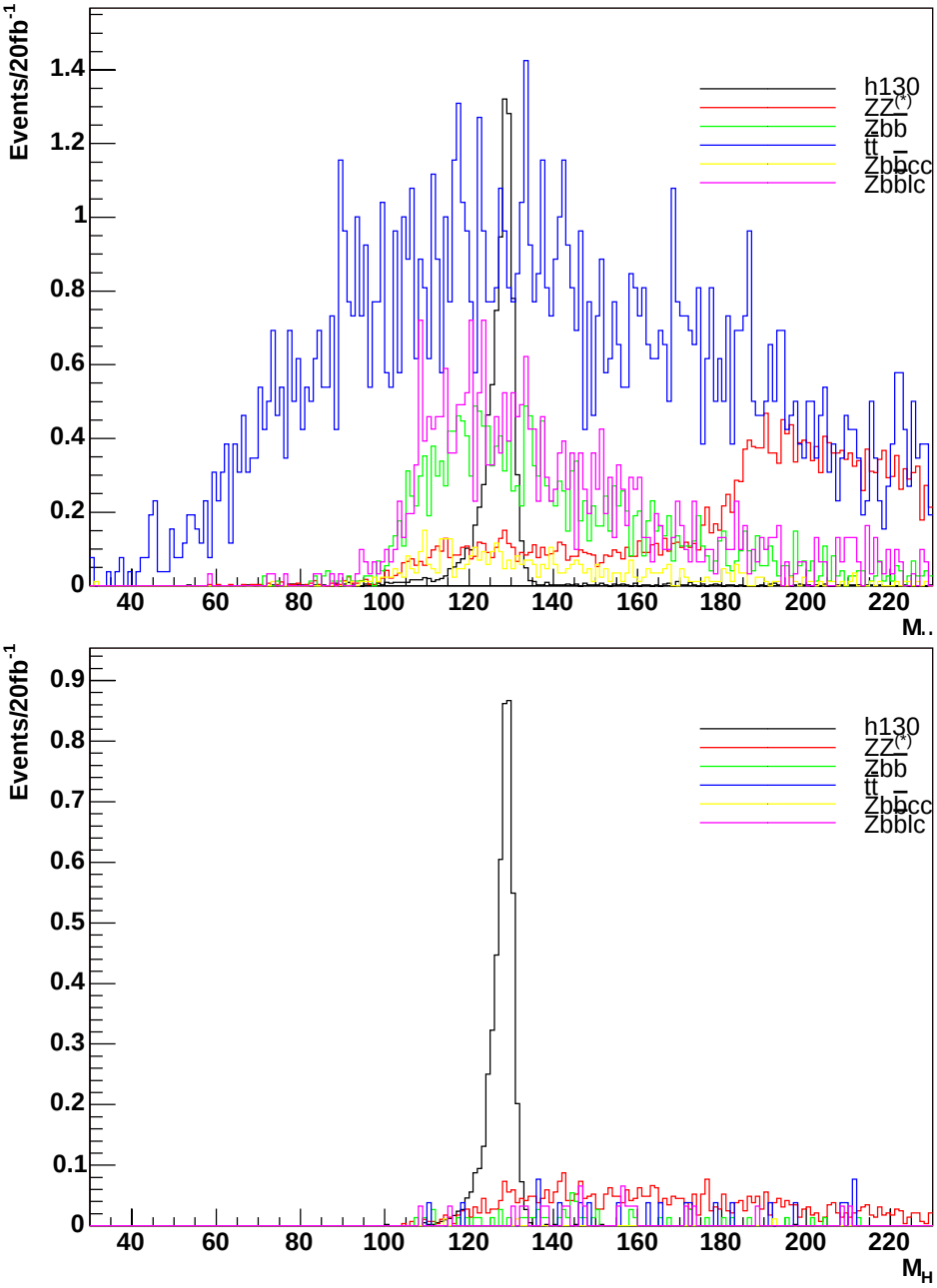


Figure 6.23: *Invariant mass distribution for signal and backgrounds normalized at the expected number of events in 20 fb^{-1} before (top) and after (bottom) the application optimized cuts ($m_H = 130$ GeV).*

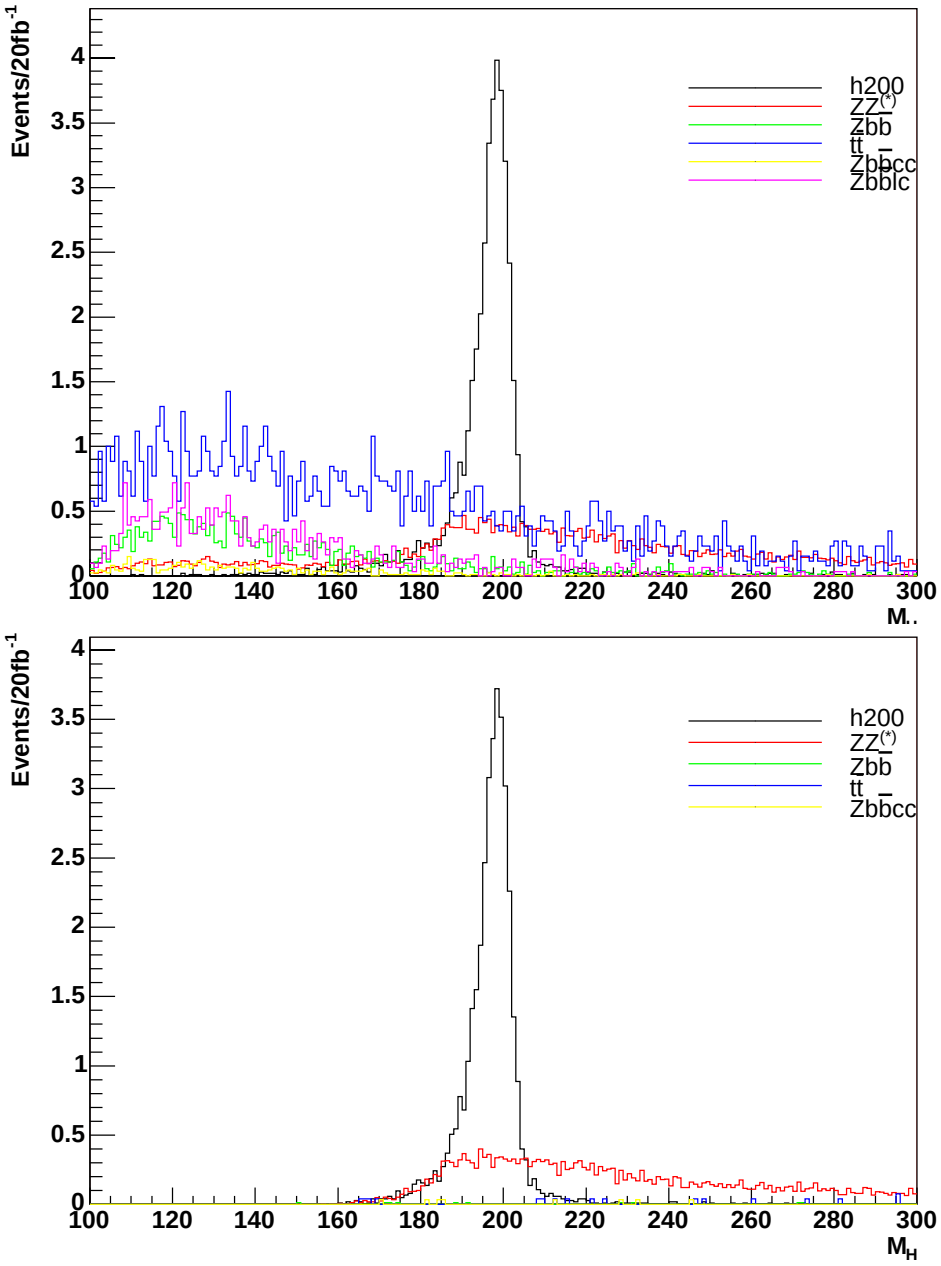


Figure 6.24: *Invariant mass distribution for signal and backgrounds normalized at the expected number of events in 20 fb^{-1} before (top)and after (bottom) the application of the optimized cuts ($m_H = 200$ GeV).*

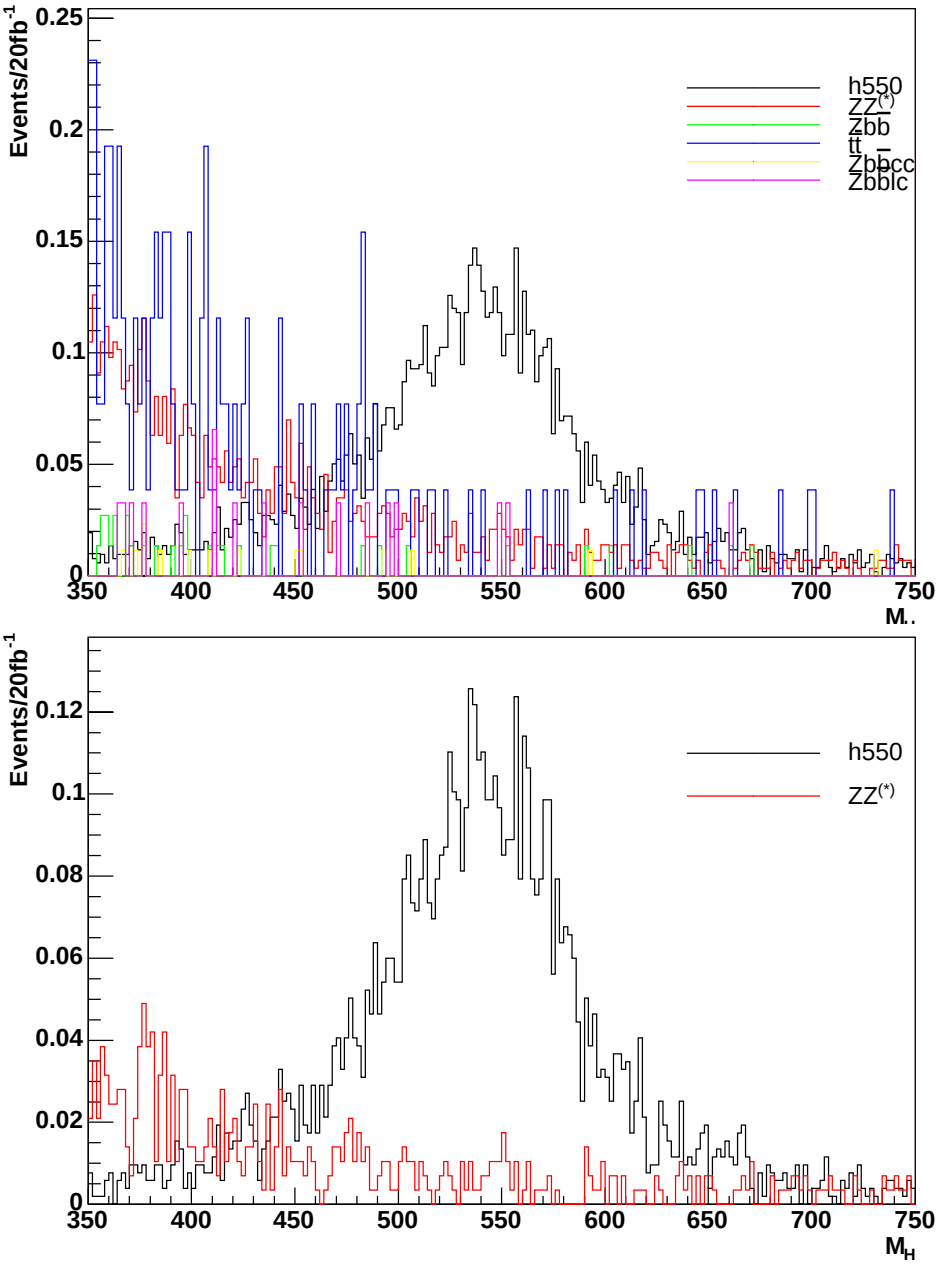


Figure 6.25: *Invariant mass distribution for signal and backgrounds normalized at the expected number of events in 20 fb^{-1} before (top) after (bottom) the application of the optimized cuts ($m_H = 550$ GeV).*

$H \rightarrow 2e^+2e^-$ channel alone is greater than 50% are $137 < m_H < 155$ GeV and $183 < m_H < 500$ GeV. This result is obtained assuming an integrated luminosity of 20 fb^{-1}.

A direct comparison with the standard definition of the significance estimators is not possible, since as it has been explained in section 6.5, the meaning of the standard significance estimators is in some sense related to the estimation of the probability β for the background to imitate the signal, at fixed discovery probability α.

In figure 6.27 the values of the estimators S_1, $2 \times S_{12}$ and S_L is reported for an integrated luminosity of 20 fb^{-1}. For example it is possible to evaluate the intervals in which each significance estimator is reporting a value greater than 5. It can be noted that looking at S_L, it is possible to find values for the interval limits which are sufficiently consistent with the one obtained using the discovery probability estimator introduced in this work; $2 \times S_{12}$ reports similar values to those of S_L, while S_1 is underestimating the background statistical fluctuactions.

However, the small number of events which characterize the search of the Higgs boson in this channel justifies the introduction of a more sophisticated discovery potential estimator, as it has been made in this work.

6.8 Conclusions

In this chapter the discovery probability of the Higgs boson in the $H \rightarrow 2e^+2e^-$ channel has been evaluated. A sophisticated statistical analysis has been made in order to define the discovery probability, since due to the low number of expected signal and background events, the standard significance estimators are not properly working. It has been showed that, assuming an integrated luminosity of 20 fb^{-1}, the discovery probability is greater than 50% in the intervals $137 < m_H < 155$ GeV and $183 < m_H < 500$ GeV. To achieve this result, it was fundamental to make an optimization of the selection cuts and the use of an isolated electron identification criterion based on the likelihood, which allowed to bring the reducible background at much lower levels than the irreducible one. Only statistical fluctuations due to the limited Montecarlo samples have been considered in this result. For the moment, systematic uncertainties have not been evaluated. It can be argued, however, that the result should not be sensitive even to consistent variation (20-30%) of the reducible background absolute cross section. Important systematics, however, can be represented by the precision of the intercalibration of the electromagnetic calorimeter, which affects both the efficiency and the resolution of the electron reconstruction, the hadronization model used in the generator, which can change both the isolation and identification dis-

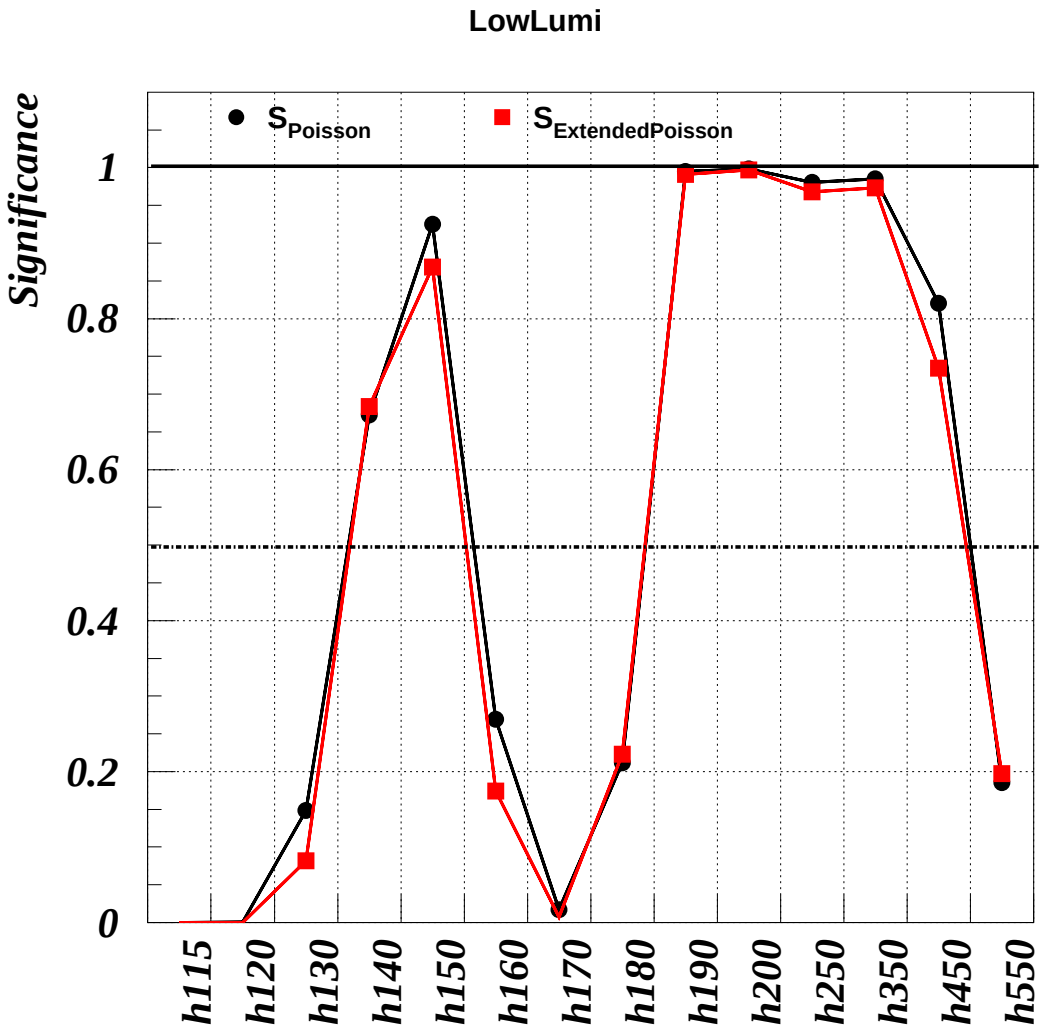


Figure 6.26: *Discovery probability in the $H \rightarrow 2e^{+}2e^{-}$ channel assuming an integrated luminosity of 20 fb^{-1} for various Higgs masses. $S_{Poisson}$ is obtained from the equation 6.9, where the limited statistics of the Montecarlo samples is not considered, while $S_{ExtendedPoisson}$ is obtained from equations 6.18 and 6.19 which takes into account the statistical fluctuactions of the expected number of signal and background samples due to the limited statistics of Montecarlo.*

tributions for the electrons, the parton distribution functions which can change the kinematical distributions for signal and backgrounds.

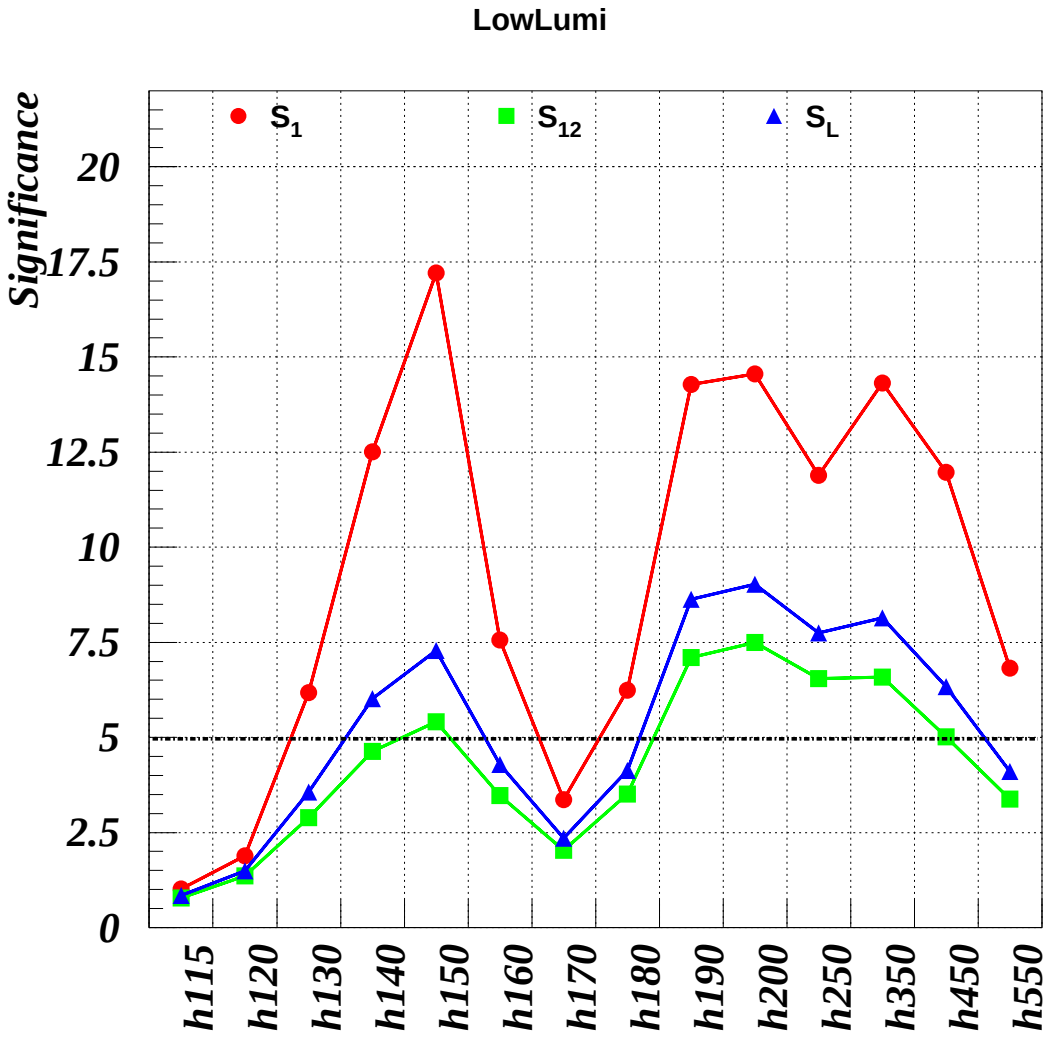


Figure 6.27: *Values of the significance estimators S_1, $2 \times S_{12}$ and S_L for an integrated luminosity of 20 fb^{-1} for various Higgs masses.*

Conclusions

This work has concerned th electron reconstruction and identification with the CMS detector. The Higgs decay with a final state represented by $2e^+$ and $2e^-$ is used as a benchmark channel, since key points in this analysis are represented by the efficiency and resolution of the electron reconstruction and isolated electron identification capabilities. The main role in the electron reconstruction process is played by the electromagnetic calorimeter. Sevaral aspects related to the the calorimeter performances have been analyzed making use of the 2003 test beam, which was performed on 100 channels equipped with the old FPPA electronics design. In particular the initial intercalibration precision from laboratory measurements have been evaluated comparing it to the beam intercalibration. A value of 4% has been obtained, even if a final conclusion requires higher statistics. In obtaining this result, it was important to combine the direct raw light yield measurement with the indirect prediction on the light yield which can be made using the transmission measurement [75].

A GEANT 4 simulation of the calorimeter in test beam condition has been also developed and compared with the test beam data; the comparison allowed to validate the simulation and to fine tune the simulations parameters. In particular, the different contributions to the energy and position resolution have been identified and characterized.

The performances for the actual electron reconstruction in the final CMS condition have been evaluated in simulated data, using a detailed detector geometry description. Tha main problem is represented by the electron bremsstrahlung, which deteriorates both efficiency and resolution in particular for low energy electrons, yielding energy scale variation as large as 4% in the energy range from 5-100 GeV. The electron momentum reconstruction performed separately by the calorimeter and by the tracker have a complementary role, and they can be combined together in an optimal estimator of the electron 4-momentum. Making use of this estimator, the energy scale variation can be mantained within 1%, and the bremsstrahlung effect can be consistenly reduced. Rejection of non isolated elec-

trons coming from jets (either real or fake) is a task of particular importance in the LHC environment higly dominated by the jet presence: an electron identification variable based on the likelihood has been constructed, able to identify real electrons with a probability of more than 90% and reject electron with a misidentification probability of less than 4%.

These algorithms are used in the analysis of the Higgs boson signal in the $H \rightarrow 2e^{+}2e^{-}$ channel, playing an important role in the signal event reconstruction and background rejection. An evaluation of the discovery potential of the Higgs boson with an integrated luminosity of 20 fb^{-1} has been performed, based on a sophisticated analysis required due to the low number of expected signal and background events. The discovery probability for the Higgs boson in the $H \rightarrow ZZ^{(*)} \rightarrow 2e^{+}2e^{-}$ has been evaluated to be greater than 50% in the intervals $137 < m_H < 155$ GeV and $183 < m_H < 500$ GeV. To achieve this result, it was fundamental to make an optimization of the selection cuts in order to maximize the discovery probability. The reducible background has been mantained to much lower levels than the irreducible one. Statistical fluctuations due to the limited Montecarlo samples have been considered, but, for the moment, no systematic uncertainties have been evaluated. It can be argued, however, that this result should not be sensitive even to consistent variation (20-30%) of the reducible background absolute cross section. Other important systematics, however, are represented by the electromagnetic energy scale and calorimeter intercalibration precision, by the the hadronization model used in the generator, which can change both the isolation and identification distributions for the electrons, by the set of parton distribution functions which can change the kinematical distributions for signal and backgrounds.

Bibliography

[1] S. Weinberg et al., Conceptual foundations of the unified theory of weak and electromagnetic interactions. *Rev. Mod. Phys.*, 52:515–523, 1980.

[2] P. W. Higgs. Broken symmetries, massless particles and gauge fields. *Phys. Lett.*, 12(2):132–133, 1964.

[3] CMS Collaboration, CMS: The Electromagnetic Calorimeter Project Technical Design Report , CERN/LHCC 97-33, CMS TDR 4, 15 December 1997.

[4] GEANT 4 Collaboration, `http://cern.ch/geant4`

[5] The LEP Collaborations ALEPH, DELPHI, L3, OPAL, the LEP Electroweak Working Group and the SLD Heavy Flavour Group, `hep-ex/0312023` Combination of Preliminary Electroweak Measurements and Constraints on the Standard Model

[6] G. t'Hooft and M. Veltman. Regularization and renormalization of gauge fields. *Nucl. Phys.*, B44:189–213, 1972.

[7] A. Djouadi, J. Kalinowski, and M. Spira. HDECAY: a program for Higgs boson decays in the Standard Model and its supersymmetric extension. *Comp. Phys.Commun.*, 108:56–74, 1998.

[8] S. Dawson, The Standard Model intermediate mass Higgs boson, Perspectives on Higgs Physics II, ed. Gordon L. Kane, World Scientific, 1997.

[9] G. Altarelli and G. Isidori Phys. Lett. B337 (1994), J. A. Casas, J. R. Espinosa and M. Quirós, Phys. Lett. B342 (1995) 171, B382 (1996) 374, T. Hambye and K. Riesselmann, Phys. Rev. D55 (1997) 7255.

[10] T. Appelquist and C.W Bernard, Phys. Rev., D22 (1980) 200, Strongly Interacting Higgs Bosons,
A. Longhitano, Nucl. Phys., B188 (1981) 118, Low-Energy Impact of a Heavy Higgs Boson Sector.

[11] M.J.G. Veltman, Acta. Phys. Polon., B8 (1977) 475, Second Threshold in Weak Interactions.

[12] ALEPH, DELPHI, L3 and OPAL Collaborations The LEP Working Group for Higgs Boson Searches hep-ex/0306033 Search for the Standard Model Higgs Boson at LEP ALEPH, DELPHI, L3 and OPAL Collaborations

[13] ALEPH Collaboration, The ALEPH Detector (Apparatus for LEp PHysics) Technical Report , CERN-LEPC-83-2, LEPC-P-1 - Geneva: CERN, 1983;
ALEPH Collaboration, Proposal for the new vertex detector for ALEPH , CERNLEPC- 93-8, LEPC-P-1-Add-1 - Geneva: CERN, 1993.

[14] DELPHI Collaboration, DELPHI: The DELPHI Detector (DEtector with Lepton Photon and Hadron Identi cation) Technical Proposal , CERN-LEPC-83-3, DELPHI-83-66, LEPC-P-2 - Geneva: CERN, 1983;
DELPHI Collaboration, Proposal for the replacement of the small angle calorimeter of DELPHI , CERN-LEPC-92-6, LEPC-P-2-Add-1 - Geneva: CERN, 1992;
DELPHI Collaboration, Proposal for the upgrade of DELPHI in the forward region , CERN-LEPC-92-13, LEPC-P-2-Add-2 - Geneva: CERN, 1992;
DELPHI Collaboration, Proposal for the DELPHI very forward tracker , CERNLEPC- 93-6, LEPC-P-2-Add-3 - Geneva: CERN, 1993.

[15] L3 Collaboration, L3 Technical Proposal , CERN-LEPC-83-5, LEPC-P-4 - Geneva: CERN, 1983;
L3 Collaboration, Some improvements to the L3 detector for LEP 200 , CERNLEPC- 94-5, LEPC-P-4-ADD-2-INS - Geneva: CERN, 1994;
L3 Collaboration, A very small angle tagger for L3 , CERN-LEPC-95-1, LEPCP- 4-ADD-3-INS - Geneva: CERN, 1995.

[16] OPAL Collaboration, OPAL: The OPAL Detector (an Omni Purpose Apparatus for Lep) Technical Proposal , CERN-LEPC-83-4, LEPC-P-3 - Geneva: CERN, 1983;
OPAL Collaboration, Proposal for a new silicon microvertex detector for OPAL , CERN-LEPC-93-13, LEPC-P-3-add-1 - Geneva: CERN, 1993;
OPAL Collaboration, Proposal for a new silicon microvertex detector for OPAL , CERN-LEPC-93-13, LEPC-P-3-add-2 - Geneva: CERN, 1993;
OPAL Collaboration, Proposal to install scintillating tiles in the OPAL end-cap , CERN-LEPC-94-09, LEPC-P-3-Add-3 - Geneva: CERN, 1994.

[17] ALEPH Collaboration, Phys. Lett., B495 (2000) 1-17, hep-ex/0011045, Observation of an Excess in the Search for the Standard Model Higgs Boson at Aleph .

[18] F. Abe et al. [CDF Collaboration], The CDF-II detector: Technical design report, FERMILAB-PUB-96-390-E(1996).

[19] F. Abe et al. [CDF Collaboration], Nucl. Instrum. Meth. A271, 387-403(1998).

[20] CDF and D0 Collaborations FERMILAB-PUB-03/320-E Results of the Tevatron Higgs Sensitivity Study.

[21] M. Carena, J. S. Conway, H. E. Haber, J. D. Hobbs, et al, hep-ph/0010338 Report of the Higgs Working Group of the Tevatron Run 2 SUSY/Higgs Workshop.

[22] L. Susskind, Phys. Rev. D20 (1979) 2619, G. 't Hooft Recent Developments in Gauge Theories, Plenum Press, 1980.

[23] S. Weinberg, Phys. Rev. D13 (1976) 974, Phys. Rev. D18 (1979) 1277.

[24] R. Casalbuoni et al., Phys. Lett. B155 (1985) 95, Phys. Lett. B282 (1987) 235, Nucl. Phys. B310 (1988) 181.

[25] J.L. Hewett and T. Rizzo, Phys. Rep. 183 (1989) 193.

[26] H.P. Nilles, Phys. Rep. 110 (1984) 1

[27] ATLAS Collaboration, Atlas Detector and Physics Performance - Technical Design Report Vol. I ATLAS TDR 14, CERN/LHCC 99-14.

[28] The LHC Study Group, The Large Hadron Collider Conceptual Design, CERN/AC 95-05 (1995).

[29] T. Sjostrand, P. Eden, C. Friberg et al., Comp. Phys. Comm. 135 (2001) 238, High-energy-physics event generation with PYTHIA 6.1.

[30] J. Pumplin, D.R. Stump, J. Huston, H.L. Lai, P. Nadolsky and W.K. Tung, JHEP, 0207, (2002) 012, hep-ph/0010017, New Generation of Parton Distributions with Uncertainties from Global QCD Analysis , D. Stump, J. Huston, J. Pumplin, W.K. Tung, H.L. Lai, S. Kuhlmann and J. Owens, hep-ph/0303013, Inclusive Jet Production, Parton Distributions, and the Search for New Physics .

[31] F. Gianotti Collider Physics:LHC, CERN-open-2000-293;

[32] H. Georgi, S. Glashow, M. Machacek and D. V. Nanopoulos, Phys. Rev. Lett. 40 (1978) 692.

[33] M. Spira, A. Djouadi, D. Graudenz and P. M. Zerwas, Nucl. Phys. B453 (1995) 17,
A. Djouadi, M. Spira and P. M. Zerwas, Phys. Lett. B264 (1991) 440
D. Graudenz, M. Spira and P. M. Zerwas, Phys. Rev. Lett. 70 (1993) 1372.

[34] M. Spira, hep-ph/9705337, QCD Effects in Higgs Physics.

[35] R. N. Cahn and S. Dawson, Phys. Lett. B136 (1984) 196,
K. Hikasa, Phys. Lett. B164 (1985) 341,
G. Altarelli, B. Mele and F. Pitolli, Nucl. Phys. B287 (1987) 205.

[36] T. Han, G. Valencia and S. Willenbrock, Phys. Rev. Lett. 69 (1992) 3274.

[37] S. L. Glashow, D. V. Nanopoulos and A. Yildiz, Phys. Rev. D18 (1978) 1724, Z. Kunzst, Z. Trocsanyi and W. J. Striling, Phys. Lett. B271 (1991) 247.

[38] T. Han and S. Willenbrock, Phys. Lett. 71 (1993) 2699.

[39] Z. Kunszt, Nucl. Phys. B247 (1984) 339, J.F. Gunion, Phys. Lett. B253 (1991) 269, W.J. Marciano and F.E. Paige, Phys. Rev. Lett. 66 (1991) 2433.

[40] Z. Kunszt, S. Morreti and W. J. Stirling, Z. Phys. C74 (1997) 479.

[41] D. Zeppenfeld, R. Kinnunen, A. Nikitenko and E. Richter-Was, Phys.Rev. D62 (2000) 013009

[42] W. Beenakker et al., Phys. Rev. Lett. 87 (2001) 201805 and Nucl. Phys. B653 (2003) 151; S.Dawson et al., Phys. Rev. Lett. 87 (2001) 201804 and Phys. Rev. D68 (2003) 034022.

[43] N. Dubinin, V.V. Ilyin and V.I. Savrin, CMS NOTE 1997/101, Light Higgs Bosons Signal at LHC in the Reactions $pp \to \gamma\gamma +$ *jet* and $pp \to \gamma\gamma$ +lepton, S. Abdullin et al. Phys. Lett. B431(1998)410.

[44] CMS Collaboration, The Compact Muon Solenoid Technical Proposal CERN/LHCC 94-38, LHCC/P1, 15 December 1994.

[45] CMS Collaboration, The Compact Muon Solenoid Letter of Intent , CERN/LHCC 92-3, LHCC/I1, 1 October 1992.

[46] CMS Collaboration, The Magnet Project: Technical Design Report, CERN/LHCC 97-10, CMS TDR 1 2 May 1997

[47] CMS Collaboration, CMS: The Tracker Project Technical Design Report , CERN/LHCC 98-06, CMS TDR 5, 15 April 1998,
CMS Collaboration, Addendum to the CMS Tracker TDR , CERN/LHCC 2000- 016, CMS TDR 5 Addendum 1, 21 February 2000.

[48] CMS Collaboration, CMS: The Hadron Calorimeter Project Technical Design Report , CERN/LHCC 97-31, CMS TDR 2, 20 June 1997.

[49] CMS Collaboration, CMS: The Muon Project Technical Design Report , CERN/LHCC 97-32, CMS TDR 3, 15 December 1997.

[50] CMS Collaboration, CMS: The Trigger and Data Acquisition Project, Volume I: The Level-1 Trigger Technical Design Report , CERN/LHCC 2000-038, CMS TDR 6.1, 15 December 2000.

[51] CMS Collaboration, CMS: The Trigger and Data Acquisition Project, Volume II: Data Acquisition and High-Level Trigger Technical Design Report , CERN/LHCC 02-26, CMS TDR 6.2, 15 December 2002.

[52] P. Lecoq et al., CMS-TN/94-308.

[53] E.Auffray et al., CMS Note 1997/054 (1997), Improvement of several properties of lead tungstate crystals with different doping ions.

[54] R. Rusack, presentation at CALOR2002, Caltech, 25-30 May 2002, Avalanche Photodiodes for the CMS Lead Tungstate Calorimeter.

[55] Yu. Blinnikov et al., Nucl. Instrum. Meth. A504(2003), 228-231

[56] C. Biino et al., The CMS ECAL Very Front end Electronics: production and tests, Proceedings of the 10th Workshop on Electronics for LHC and future Experiments, Boston USA

[57] W. Lustermann, Low Voltage Supply System for the Very Front End Readout Electronics of the CMS Electromagnetic Calorimeter, Proceedings of the 9th Workshop on Electronics for the LHC Experiments, Amsterdam Sep.2003.

[58] M.Raymond, The MGPA Electromagnetic Calorimeter Readout Chip for CMS, Proceedings of the 9th Workshop on Electronics for the LHC Experiments, Amsterdam Sep.2003.

[59] Goncalo Minderico A CMOS low power, quad channel, 12 bit, 40 MS/s pipelined ADC for application in particle physics calorimetry, Proceedings of the 9th Workshop on Electronics for the LHC Experiments, Amsterdam Sep.2003.

[60] D. J. Graham and C. Seez, CMS Note 1996/002,
C. Markou, CMS Note 1997/048.

[61] A. Bartoloni, F. Cavallari, S. Costantini, I. Dafinei, M. Diemoz, S. Guerra, E. Longo, P. Meridiani, G. Organtini, R. Paramatti, M.M. Obertino Stability of the APD HV System at the 2002 ECAL Test Beam CMS Internal Communication

[62] F. Cavallari et al., CMS RN 2004/002, Relative Light Yield comparison between laboratory and testbeam data for CMS ECAL $PbWO_4$ crystals.

[63] D. Futyan and C. Seez. Intercalibration of ECAL crystals in ϕ using symmetry of energy deposition. Technical Report CMS NOTE 2002/031, 2002.

[64] CMS-Saclay Group, CMS Internal Communication The CMS-ECAL monitoring system: data generation, acquisition and processing at LHC,

[65] P. Bonamy et al., CMS NOTE 1998/013 The ECAL Calibration: Use of the light monitoring system. Version 2.0.

[66] A. Van Lysebetten, P. Verrecchia, CMS Rapid Note 2004/001, Performance and measurements of the light monitoring system for CMS-ECAL from 2002 test beam data.

[67] E. Auffray, et al., CMS NOTE 1999/067, Performance of ACCOS, an automatic crystal quality control system for PWO.

[68] S. Baccaro et al., NIM A 459 (2001) 278-284.

[69] E. Auffray et al., CMS NOTE 2003/003. Cross-Calibration of Two Automatic Quality Control Systems for the CMS ECAL Crystals

[70] P. Sempere-Roldan, Quality control and preparation of the PWO crystals for the electromagnetic calorimeter of CMS, PHD Thesis.

[71] L. M. Barone, F. Cavallari, S. Costantini, I. Dafinei, M. Diemoz, E. Longo, P. Meridiani, G. Organtini, R. Paramatti, CMS RN 2004/003, Improvement on PbWO4 Crystal Inter-calibration Precision from Light Yield Measurements at the INFN-ENEA Regional Center.

[72] GEANT - Detector description and simulation tool, Version 3.21, CERN, Geneva, 1993.

[73] P. Paganini and I. van Vulpen, CMS NOTE-2004/025, Pulse amplitude reconstruction in the CMS ECAL using the weights method.

[74] E. Auffray-Hillemanns, F. Cavallari, P. Lecoq, E. Longo, M. Schneegans P. Sempere Roldan, M. Castellani, P. Meridiani, CMS Internal Communication Test-beam Results on the Performance of Two Matrices of PbWO4 Crystals for the CMS ECAL and Comparison with Laboratory Measurements.

[75] L. Barone, F. Cavallari, S. Costantini, I. Dafinei, M. Diemoz, E. Longo, P. Meridiani, G. Organtini, R. Paramatti, M.M. Obertino, CMS RN-2004/005 Correlation Between Light Yield and Longitudinal Transmission in PbWO4 Crystals and Impact on the Precision of the Crystal Intercalibration.

[76] G. Organtini et al., CMS Internal Communication Data Validation Software Architecture during the CMS ECAL Test Beam.

[77] E. Gamma, R. Helm, R. Johnson, J. Vlissides, Design Patterns: elements of reusable object -oriented software, Addison-Wesley

[78] Luciano M. Barone et al., CMS NOTE-2003/022, REDACLE: A Database for the Workflow Management of the CMS ECAL Construction.

[79] Object oriented Simulation for CMS Analysis and Reconstruction (OSCAR) `http://cmsdoc.cern.ch/OSCAR/`

[80] A. Givernaud, private communication.

[81] F.Cavallari et al., CMS NOTE in preparation.

[82] G. Daskalakis, I. van Vulpen, CMS Internal Communication Position Resolution in the CMS ECAL Using 2003 Testbeam Data.

[83] P. Grafstrom, SL-Note-97-81/EA, The momentum resolution of the H4 beam for secondary beams in Filter Mode optics.

[84] G. Dewhirst, R. Bruneliere, CMS RN-2004/004, Energy Resolution of the CMS ECAL Barrel Super-Module Using MGPA Electronics.

[85] `http://cmsdoc.cern.ch/orca/`

[86] V.Innocente et al., CMS NOTE-2000/047, CMS Software Architecture Software framework, Services and Persistency in High Level Trigger, Reconstruction and Analysis, `http://cobra.web.cern.ch/cobra/`

[87] `http://cmsdoc.cern.ch/cmsoo/projects/CMKIN/index.html`

[88] `http://cmsdoc.cern.ch/cmsim/cmsim.html`

[89] C. Seez and D. Barney, CMS IN-2001/023, Summary of ECAL-egamma Autumn 2000 Monte-Carlo Data Production Parameters and Specifications.

[90] CMS Collaboration, CMS: Data Acquisition & High Level Trigger TDR, CERN/LHCC/2002-26, CMS TDR 6.2, 15 December 2002

[91] E. Meschi et al., CMS NOTE-2001/034, Electron Reconstruction in the CMS Electromagnetic Calorimeter.

[92] G. Daskalakis, K. Lassila-Perini, CMS NOTE-2002/039, Jet Rejection Using the Pixel Matching for the Low and the High Luminosity.

[93] R. Fruehwirth et al. CMS Internal Communication Track Fitting Methods in ORCA.

[94] W. Adam et al., CMS RN-2003/001, Reconstruction of Electron Tracks With the Gaussian-Sum Filter in the CMS tracker at LHC.

[95] P. Meridiani, The Calibration Strategy of CMS Electromagnetic Calorimeter, Presented at XI International Conference on Calorimetry in Particle Physics.

[96] M. Takahashi, CMS IN-2004/039, Rapidity Dependence of the Shower Containment in the CMS ECAL

[97] I. Puljak, Potentiel de decouverte du boson de Higgs dans le canal $H \rightarrow ZZ^* \rightarrow 4e$. Doctorat de l'universite Paris, 2000.

[98] The Coordinated Theoretical-Experimental Project on QCD `http://www.phys.psu.edu/~cteq/`

[99] E. Barberio and Z. Was, Comput. Phys. Commun. 79 (1994) 291-308.

[100] M. Spira. `http://people.web.psi.ch/spira/proglist.html`

[101] C.Zecher et al, ZEITSCHRIFT FUR PHYSIK C, 1994, p.219

[102] Proceedings of the workshop on Standard Model (and more) at the LHC, CERN 2000-04.

[103] `http://theory.sinp.msu.ru/comphep`

[104] J. Campbell and K. Ellis, MCFM - monte carlo for femtobarn processes, `http://mcfm.fnal.gov`

[105] D. Giordano, Presentation at the CMS Higgs group meeting

[106] S.I. Bityukov and N.V. Krasnikov, Nucl. Instr. and Meth. A452 (2000) 518.

[107] V. Bartsch and G. Quast, Expected signal observability at future experiments, CMS Internal Communication.

[108] S. Eidelman et al., Phys. Lett. B592, 1 (2004)

[109] Giulio D' Agostini, Bayesian Reasoning In Data Analysis, 2003 World Scientific.

[110] MINUIT, Function Minimization and Error Analysis CERN Program Library Long Writeup D506