Springer Theses
Recognizing Outstanding Ph.D. Research

Francesco Pandolfi

Search for the Standard Model Higgs Boson in the $H \to ZZ \to \ell^+ \ell^- q\bar{q}$ Decay Channel at CMS

Springer

Springer Theses

Recognizing Outstanding Ph.D. Research

For further volumes:
http://www.springer.com/series/8790

Aims and Scope

The series "Springer Theses" brings together a selection of the very best Ph.D. theses from around the world and across the physical sciences. Nominated and endorsed by two recognized specialists, each published volume has been selected for its scientific excellence and the high impact of its contents for the pertinent field of research. For greater accessibility to non-specialists, the published versions include an extended introduction, as well as a foreword by the student's supervisor explaining the special relevance of the work for the field. As a whole, the series will provide a valuable resource both for newcomers to the research fields described, and for other scientists seeking detailed background information on special questions. Finally, it provides an accredited documentation of the valuable contributions made by today's younger generation of scientists.

Theses are accepted into the series by invited nomination only and must fulfill all of the following criteria

- They must be written in good English.
- The topic should fall within the confines of Chemistry, Physics, Earth Sciences, Engineering and related interdisciplinary fields such as Materials, Nanoscience, Chemical Engineering, Complex Systems and Biophysics.
- The work reported in the thesis must represent a significant scientific advance.
- If the thesis includes previously published material, permission to reproduce this must be gained from the respective copyright holder.
- They must have been examined and passed during the 12 months prior to nomination.
- Each thesis should include a foreword by the supervisor outlining the significance of its content.
- The theses should have a clearly defined structure including an introduction accessible to scientists not expert in that particular field.

Francesco Pandolfi

Search for the Standard Model Higgs Boson in the $H \to ZZ \to \ell^+\ell^- q\bar{q}$ Decay Channel at CMS

Doctoral Thesis accepted by Dipartimento di Fisica, Sapienza Università di Roma, Rome, Italy

Springer

Author
Dr. Francesco Pandolfi
CERN
Sapienza Università di Roma
Rome
Italy

Supervisor
Dr. Daniele del Re
Dipartimento di Fisica & INFN
Sapienza Università di Roma
Rome
Italy

ISSN 2190-5053 ISSN 2190-5061 (electronic)
ISBN 978-3-319-00902-5 ISBN 978-3-319-00903-2 (eBook)
DOI 10.1007/978-3-319-00903-2
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013942123

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

We shall not cease from exploration
And the end of all our exploring
Will be to arrive where we started
And know the place for the first time.

T. S. Eliot

For Flavio,
my nova

Supervisor's Foreword

Particle physicists have been searching the Higgs boson since it was firstly suggested in 1964. This particle was the last missing ingredient of the so-called Standard Model of Particle Physics. Any particle which interacts with the Higgs field acquires a mass. The stronger the interaction with the field is, the larger the mass of the particle. Before the operations of the Large Hadron Collider (LHC), no experimental evidence of this hypothetical particle confirmed this theory. The CMS and ATLAS experiments at the LHC have been designed to give a final answer on this, either discovering or rejecting the presence of this elusive particle.

On the 4th of July of last year everything changed. CMS and ATLAS announced the discovery of a new boson with a mass of about 125 GeV. Its properties, as the production and decay rates, the spin and the parity look compatible with the particle predicted by Peter Higgs. This crucial discovery was the result of years of analysis optimization, tuning of Monte Carlo simulations, and precise calibrations of the detectors. To reach this important goal, several preparatory studies have been performed and the experiments worked in restricting the mass range where this state could fall. For this purpose, physicists used channels which have larger production rates but are less clean. Among these modes, the decay $H \rightarrow ZZ \rightarrow 2l2jets$ resulted to be one of the most challenging and powerful channels to investigate the presence of a Higgs boson with masses larger than 200 GeV. Compared to the standard $H \rightarrow ZZ \rightarrow 4l$ channel, it is 20 times more frequent but the presence of jets in the final state makes it very complicated because of the large expected background contributions.

Francesco Pandolfi started from the idea that this channel could play a major role in CMS. He worked on the simulation to verify its feasibility, identifying the most relevant issues for this analysis. They can be divided into two big areas: the determination of the jet energy scale and resolution and the implementation of jet-based experimental techniques to improve the sensitivity of the measurement. This thesis reflects the huge amount of work needed to deal with these two issues. It summarizes the work needed to reduce the jet energy calibration uncertainties down to 1 % level and gives details on jet calibration developed by Francesco, which used samples of the gamma+jet events and was crucial at the start of LHC collisions. In addition, the thesis describes novel techniques, as the analysis tool implemented to distinguish between jets originated by quarks and gluons. After

Francesco's implementation, they have been later inherited by other analyses of the CMS experiment and they will become even more important in the future.

In summary, Francesco's thesis is an excellent piece of work, a paradigm for studies of the Higgs boson with final states with jets. The nonexpert physicists will enjoy a complete and marvelously readable description of a proton–proton collider analysis. This is not common for a thesis at the LHC, given the average level of complexity of the analyses. At the same time, the expert analyzer will learn about searches done with jets at CMS and will be stimulated to further develop the novel techniques described.

12 March 2013, Rome, Italy Daniele del Re

Preface

The standard model of elementary particles is one of the most successful scientifical theories elaborated by mankind. Throughout the past century, its predictions have been confirmed with astonishing precision by numerous experiments, in conceptually distant fields and with a variety of experimental techniques. Its theoretical foundations, though, depended on a particle which had not been discovered.

The mass of the Higgs boson is not constrained by the theory, and is allowed to vary in a wide range. Numerous experiments have unsuccessfully pursued its search, but have been able only to exclude its existence in certain mass ranges. The analyses conducted at the detectors operating at the LEP collider at CERN have set a lower bound of 114.4 GeV, whereas the searches performed at the Tevatron at Fermilab have excluded the 156–177 GeV range.

The Large Hadron Collider (LHC) is the particle accelerator which has been built with the aim of producing definitive proof regarding the Higgs boson's existence. It is a superconducting proton collider, with a center of mass energy of 7 TeV. It has the capability of spanning a wide energy range, up to the TeV scale.

The Compact Muon Solenoid (CMS) is one of the four main experiments which analyzes the collisions produced at the LHC. It is a general-purpose detector which has been designed in order to maximize its performance in Higgs boson searches.

The discovery of a Higgs boson depends on its decay products, as it is an unstable particle. If its mass is large enough, the decay to pairs of electroweak vector bosons dominates the particle's decay channels. The production of a Z boson pair, in particular, constitutes the most promising final state in search-oriented analyses. The requirement that at least one of the two Z bosons decays to a light charged lepton pair, in fact, significantly reduces the possible sources of background at a hadron collider.

On July 4, 2012, the CMS and ATLAS experiments have both reported evidence for a narrow resonance with mass close to 125 GeV, with properties compatible with those of a Higgs boson. A crucial role in this discovery was played by the fully leptonic decay channel $\mathrm{H} \to \mathrm{ZZ} \to 4\ell$.

In this thesis, we have conducted a search for a heavy Higgs boson in the $\mathrm{H} \to \mathrm{ZZ} \to \ell^+\ell^-\mathrm{q\overline{q}}$ decay channel, with 4.6 fb^{-1} of data collected by the Compact Muon Experiment. The presence of jets in the final state poses a series of

challenges to the experimenter: both from a technical point of view, as jets are complex objects and necessitate of ad hoc reconstruction techniques, and from an analytical one, as backgrounds with jets are copious at hadron colliders; therefore, analyses must obtain high degrees of background rejection in order to achieve competitive sensitivity.

The first chapter of this thesis offers a brief introduction to the theoretical foundations of the standard model and the reasons which conjure to the postulate of the existence of the Higgs boson. It further offers an overview of the constrains on the Higgs boson mass before the LHC era, both on an experimental and on a theoretical point of views. It then inspects how Higgs boson searches may be conducted at the LHC, offering an overview of the most promising high-mass analyses and describing the $H \to ZZ \to \ell^+\ell^- q\overline{q}$ channel in more detail, highlighting its main aspects.

Chapter 2 offers a detailed description of the experimental apparatus. The LHC is introduced, and the Compact Muon Experiment is described in its various subdetectors. Particular emphasis is given to the electron and muon reconstruction algorithms.

The challenges posed by jet reconstruction are faced in Chap. 3. After an overview on the general aspects of jet reconstruction, we will give insight on the jet calibration scheme employed at CMS, and show how photon+jet events can be successfully used to measure jet reconstruction performance and resolution on data. More detail is given on the full event reconstruction technique employed at CMS, known as the 'Particle Flow', which allows a significant improvement in jet reconstruction performance over more traditional, calorimeter-based approaches.

The analysis event selection is presented in Chap. 4 , which includes a detailed account of the analyzed data samples, the trigger, and preselection requirements. The analysis, as will be shown, will be split into different categories based on jet flavor tagging information, and an optimization will be performed in each category. The main tool of background discrimination is provided by an angular likelihood discriminant, capable of selecting events likely to originate from the decay of a scalar boson, and discriminate nonresonant backgrounds. The chapter also details the strategy for evaluating the background directly on the data.

Possible sources of systematic uncertainties are investigated in Chap. 5. A number of different effects are scrutinized, ranging from trigger to object reconstruction, from theoretical uncertainties to those related to the quality of the simulation modeling.

In Chap. 6, the events passing the selection requirements in 4.6 fb^{-1} of data collected by the CMS detector are examined, in the search of a possible signal compatible with the decay of a heavy Higgs boson, the modeling of which is shown in detail. We further describe the statistical tools which are adopted to perform such an analysis, and provide the results.

Contents

Chapter 1
The Hunt for the Higgs Boson

Abstract Since the discovery of the top quark, which took place in 1995 at the Tevatron collider at Fermilab, the Higgs boson may be considered the last missing piece of the Standar Model. Its search has been undertaken at the LEP and Tevatron colliders, but no evidence of its existence was found, at the energies which were probed. The mass of the Higgs boson is an unconstrained parameter of the theory, therefore a collider which is able to explore vast energy ranges is cardinal for its discovery. The Large Hadron Collider at CERN has been designed with the aim of providing conclusive scientifical results regarding the existence of the Higgs boson. This chapter provides a quick but accurate introduction to the current theoretical panorama in elementary particle physics: we will describe the Standard Model, the motivations which bring to the introduction of the Higgs mechanism, and its consequences. We will then summarize the experimental limits on the Higgs boson mass previous to the Large Hadron Collider, and introduce the $H \rightarrow ZZ \rightarrow \ell^+\ell^- q\bar{q}$ decay channel, and the role it plays in the search for this elusive particle.

1.1 The Standard Model

The Standard Model is the physical theory which is currently adopted to provide a quantitative description of three of the four interactions in nature: electromagnetism, weak interactions and the strong nuclear force. It has been elaborated at the end of the 1960's by Glashow, Weinberg and Salam [1–3]. It is a renormalizable field theory, compatible with special relativity. Its Lagrangian presents a non-Abelian gauge symmetry which refers to the symmetry group $SU(3) \times SU(2) \times U(1)$. During the past decades its predictions have been confirmed by a large number of experiments [4], with astonishing precision.

The Standard Model (SM) can be divided in two sectors: the strong sector, known as Quantum Chromodynamics (QCD), and the electroweak sector. Hence, the SM Lagrangian may be written as the sum of two parts:

F. Pandolfi, *Search for the Standard Model Higgs Boson in the H→ZZ→ $\ell^+\ell^- q\bar{q}$ Decay Channel at CMS*, Springer Theses, DOI: 10.1007/978-3-319-00903-2_1,

$$\mathcal{L}_{\mathrm{SM}} = \mathcal{L}_{\mathrm{QCD}} + \mathcal{L}_{\mathrm{EW}}$$

Quantum Chromodynamics describes the interactions of quarks and gluons, mediated by the strong force through the colour charge. Its Lagrangian satisfies the $\mathrm{SU(3)_C}$ colour symmetry, and has the form:

$$\mathcal{L}_{QCD} = -\frac{1}{4}\sum_i F^i_{\mu\nu} F^{i\,\mu\nu} + \mathrm{i}\sum_r \bar{q}_{r\alpha}\gamma^\mu D^\alpha_{\mu\beta} q^\beta_r \tag{1.1}$$

where q_r represents the quark fields of flavour r, α, β are the colour indexes and the covariant derivative $D^\alpha_{\mu\beta}$ is defined as

$$D^\alpha_{\mu\beta} = \partial_\mu \delta^\alpha_\beta + \frac{\mathrm{i}}{2} g_F \sum_i G^i_\mu \lambda^i_{\alpha\beta}$$

where λ^i are the generator matrixes of SU(3). Expression 1.1 also presents the tensors $F^i_{\mu\nu}$, which are defined by

$$F^i_{\mu\nu} = \partial_\mu G^i_\nu - g_F f_{ijk} G^j_\mu G^k_\nu$$

where G^i $(i = 1, \ldots, 8)$ are the eight gluonic fields, g_F is the strong coupling constant and f_{ijk} are the SU(3) structure constants.

The Lagrangian which governs the electroweak sector is instead invariant under gauge transformations of the symmetry group $\mathrm{SU(2)_L} \times \mathrm{U(1)_Y}$. The $\mathrm{SU(2)_L}$ group refers to the weak isospin charge (I), and $\mathrm{U(1)_Y}$ to the weak hypercharge (Y). Left-handed (L) fermions are paired in $I = 1/2$ isospin doublets, whereas right-handed (R) fermions in $I = 0$ singlets:

$$\begin{array}{cccc} \mathbf{I = 1/2}: & & & \mathbf{I = 0}: \\ \begin{pmatrix}\nu_e \\ e\end{pmatrix}_L & \begin{pmatrix}\nu_\mu \\ \mu\end{pmatrix}_L & \begin{pmatrix}\nu_\tau \\ \tau\end{pmatrix}_L & (e)_R \quad (\mu)_R \quad (\tau)_R \\ \begin{pmatrix}u \\ d\end{pmatrix}_L & \begin{pmatrix}c \\ s\end{pmatrix}_L & \begin{pmatrix}t \\ b\end{pmatrix}_L & \begin{matrix}(u)_R & (c)_R & (t)_R \\ (d)_R & (s)_R & (b)_R\end{matrix} \end{array}$$

The presence of these local gauge symmetries introduces four vector bosons: three for the SU(2) group, the W^i fields ($i = 1, 2, 3$), and one for U(1), the B field. The physical fields are obtained as linear combinations of these fields:

$$A_\mu = \sin\theta_G\, W^3_\mu + \cos\theta_G\, B_\mu$$

$$Z_\mu = \cos\theta_G\, W^3_\mu - \sin\theta_G\, B_\mu$$

$$W_\mu^\pm = \frac{W_\mu^1 \mp \mathrm{i} W_\mu^2}{\sqrt{2}}$$

The above equations represent two neutral particles (the photon, described by the A_μ field, and the Z boson) and two charged particles (the W^+ and W^- bosons). We have further introduced the angle θ_G, which is known as the weak mixing angle.

This gives rise to a quantum field theory, invariant under local gauge symmetries, whose Lagrangian is expressed as:

$$\mathcal{L}_{\mathrm{EW}} = \mathrm{i} \sum_f \bar{f} D_\mu \gamma^\mu f - \frac{1}{4} \sum_G F_G^{\mu\nu} F_{G,\mu\nu}$$

where the sums are respectively extended over all fermionic fields f, and all vectorial fields G. Fermionic fields may be either left-handed doublets (ψ_L) or right-handed singlets (ψ_R). The covariant derivative D_μ is defined by

$$D_\mu = \partial_\mu - \mathrm{i} g_G (\lambda^\alpha G_\alpha)_\mu$$

where g_G is the generic coupling constant of a fermion to the G field, and λ^α are the generators of the symmetry group to which G refers.

This theory is necessarily incomplete: all particles it describes are massless, contradicting experimental evidence. The Lagrangian's symmetries, on the other hand, seem to forbid the introduction of mass terms without spoiling its gauge invariance. Higgs' proposal [5, 6] solves this problem by spontaneously breaking the Lagrangian's symmetry.

1.2 The Higgs Mechanism

A Lagrangian is symmetrical if it is invariant under a group of transformations. Degenerate eigenstates of a symmetrical Lagrangian will in general transform in linear combinations of each other under such a transformation. If a symmetrical Lagrangian presents a degenerate ground state, there is no univocal state which describes the system's fundamental configuration: one of the degenerate states must be chosen, but any of the chosen states will not share the Lagrangian's symmetry. This procedure of obtaining an asymmetric ground state is known as *spontaneous symmetry breaking*.

The simplest way of spontaneously breaking the SU(2) × U(1) symmetry group is that of introducing a scalar field Φ which is an isospin doublet:

$$\Phi = \begin{pmatrix} \Phi^+ \\ \Phi^0 \end{pmatrix} = \begin{pmatrix} (\Phi_1 + \mathrm{i}\Phi_2)/\sqrt{2} \\ (\Phi_3 + \mathrm{i}\Phi_4)/\sqrt{2} \end{pmatrix}$$

where we have introduced four real fields Φ_i ($i = 1, 2, 3, 4$) to manifest the complexity of the Φ^+ and Φ^0 fields.

The simplest Lagrangian of an autointeracting scalar field has the form:

$$\mathcal{L}_{\mathrm{H}} = (D_\mu \Phi)^\dagger (D^\mu \Phi) - V(\Phi)$$

where

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2$$

and the covariant derivative is defined by the operatorial identity

$$D^\mu = \partial^\mu + \frac{\mathrm{i}}{2} g \sigma_j W_j^\mu + \mathrm{i} g' Y B^\mu$$

where we imply the sum over the repeated index $j = 1, 2, 3$, we have called g and g' the coupling constants of fermions respectively to the fields W_j^μ and B^μ, σ_j are the Pauli matrices, and Y is the weak hypercharge. The Φ field is known as the Higgs field.

The potential $V(\Phi)$ depends on two parameters, μ and λ. The requirement $\lambda > 0$ ensures that the energy spectrum has a lower bound, and therefore the existence of a ground state. If the μ parameter is chosen so that $\mu^2 < 0$, the symmetry of $V(\Phi)$ may be broken, as it has a minimal value in correspondence of:

$$\Phi^\dagger \Phi = -\frac{\mu^2}{2\lambda} \equiv \frac{v^2}{2} \tag{1.2}$$

This implies that the Φ field has a vacuum expectation value $\Phi_0 = v/\sqrt{2}$.

Perturbation theory requires an expansion of Φ around its ground state, but the latter must be chosen between the set of states which satisfy Eq. 1.2, but each of them breaks the rotational symmetry of the Lagrangian. It can be shown that expanding the Higgs field around one of these fundamental states assignes a mass equal to $|qv|$ to each boson connected with the broken symmetry, where q is the charge of the Higgs field vector particles in the potential mediated by the boson in question.

When breaking the electroweak symmetry group SU(2) × U(1), some care is needed in order to avoid that the photon is assigned a spurious mass. Therefore, we need to choose a ground state Φ_0 which conserves the electric charge symmetry group U(1). From the Gell-Mann–Nishijima relation [7, 8]

$$Q = I_3 + \frac{Y}{2}$$

which connects the electric charge to weak isospin and hypercharge, we see that this condition is satisfied if we choose Φ_0 with weak isospin $I = 1/2$, $I_3 = -1/2$ and hypercharge $Y = 1$:

$$\Phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}$$

Therefore the Φ field will be expressed as

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

In this way the bosonic fields $W^{\pm}$ and Z, connected to the broken symmetry group SU(2), will acquire a mass

$$m_W = \frac{v}{2} g \qquad m_Z = \frac{v}{2} \sqrt{g^2 + g'^2}$$

We have furthermore introduced a physical particle, the Higgs boson, described by the $h(x)$ field, with mass equal to

$$m_H = \sqrt{2}\mu = \sqrt{2\lambda} v \tag{1.3}$$

We can now confer a mass to fermions by introducing a Yukawa interaction term, which couples a left-handed fermionic doublet ψ^L, a right-handed singlet ψ^R and the Higgs doublet Φ, which, for the case of the down quark d has the form

$$g^d \bar{\psi}^L d^R \Phi \ + \ \text{hermitian conjugate}$$

The non-vanishing vacuum expectation value of Φ will assign a mass $m_d = g^d v/\sqrt{2}$ to the d quark. Similarly, by defining $\tilde{\Phi} = -i[\Phi^\dagger \sigma_2]^T$, where σ_2 is the second Pauli matrix, a term of the form $g^u \bar{\psi}^L u^R \tilde{\Phi}$ confers a mass $m_u = g^u v/\sqrt{2}$ to the u quark.

Extending the concept to the three quark families, the Lagrangian which describes the interaction between quarks and the Higgs field has the form

$$\mathcal{L}_{q\phi} = \sum_{ik} g^d_{ik} \bar{\psi}^L_i d^R_k \Phi + \sum_{ik} g^u_{ik} \bar{\psi}^L_i u^R_k \tilde{\Phi} \ + \ \text{h.c.}$$

The mass terms in this expression are not diagonal in the u and d fields, so in order to obtain the physical fields a diagonalization must be performed. This is done by introducing unitary matrices V which transform the fields:

$$u^L = V_{uL} u'^L \qquad u^R = V_{uR} u'^R$$

$$d^L = V_{dL} d'^L \qquad d^R = V_{dR} d'^R$$

As the u^L and d^L fields transform in different ways, the coupling to the $W^{\pm}$ bosons is not diagonal anymore:

$$\mathcal{L}_{\mathrm{qW}} = g \sum_{ij} \bar{u}_i'^L \mathbf{V}_{ij} \gamma^\mu d_j'^L W_\mu^\dagger \; + \; \text{h.c.}$$

where we have introduced the $\mathbf{V}$ matrix, known as the Cabibbo-Kobayashi-Maskawa [9, 10] matrix, defined as

$$\mathbf{V} = V_{uL}^\dagger V_{dL} \equiv \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}$$

Conversely, it is easy to show that the couplings to the photon and the Z boson remain diagonal, thanks to the unitarity of the V matrices:

$$g\bar{u}^L \gamma^\mu u^L W_\mu^3 = g\bar{u}'^L V_{uL}^\dagger \gamma^\mu V_{uL} u'^L W_\mu^3 = g\bar{u}'^L \gamma^\mu u'^L W_\mu^3$$

A noteworthy feature of the V_{uL}, V_{uR}, V_{dL}, V_{dR} matrices is that they are determined except for a global phase, the existence of which allows the violation of the CP symmetry in the Standard Model.

Once the fermion fields are diagonalized, they acquire a mass equal to

$$m_f = \frac{v}{\sqrt{2}} g^f$$

In other words, the coupling of the Higgs boson to fermions is proportional to their mass:

$$g^f = \sqrt{2}\, \frac{m_f}{v}$$

As was shown in Eq. 1.3, the mass of the Higgs boson depends on the coupling parameter λ and the vacuum expectation value v. The value of the latter is determined by the Fermi constant (G_F), as it is simple to show that the following relation holds:

$$v = \frac{2m_W}{g} = (\sqrt{2} G_F)^{-1/2}$$

The current estimate of G_F, which comes from precise muon lifetime measurements [11, 12], allows to determine $v \simeq 247$ GeV. On the other hand, the model is not predictive on the value of the λ parameter, therefore the mass of the Higgs particle is a free parameter of the theory.

Nevertheless, we can exploit the perturbative nature of the theory to impose approximate theoretical boundaries [13] on m_H. A first limit is obtained by requiring that the breaking of symmetry actually takes place:

$$V(v) < V(0)$$

which is equivalent to requiring that λ is positive at all energies. When approaching this limit, that is for $\lambda \ll 1$, and therefore for a light Higgs boson, radiative top-loop corrections and gauge couplings become non-negligible and the above relationship can be transformed into a limit on the Higgs mass:

$$m_H > \frac{3v}{32\pi^2}(16g_t^4 - g^4 - 2g^2g'^2 - 3g'^4)\log\left(\frac{\Lambda}{m_H}\right)$$

where g_t is the coupling constant between the Higgs field and the top quark. We have here introduced a 'cut-off' energy scale Λ, above which we assume that the Standard Model is not valid. This condition corresponds to the lower curve [14–17] in Fig. 1.1.

Conversely, the requirement that the λ parameter remains finite up to the scale Λ translates into an upper bound on m_H:

$$m_H^2 < \frac{8\pi^2 v^2}{3\log(\Lambda/v^2)}$$

which is the upper curve in Fig. 1.1. As can be seen, these limits imply that, if the Standard Model is a perturbative theory up to the scale of grand unification $\Lambda_{\mathrm{GUT}} \approx 10^{16}$ GeV, the mass of the Higgs boson has to fall in the 130–190 GeV range. In other words, a Higgs boson lighter than about 130 GeV or heavier than about 190 GeV would suggest the presence of physics beyond the Standard Model at a scale inferior to Λ_{GUT}.

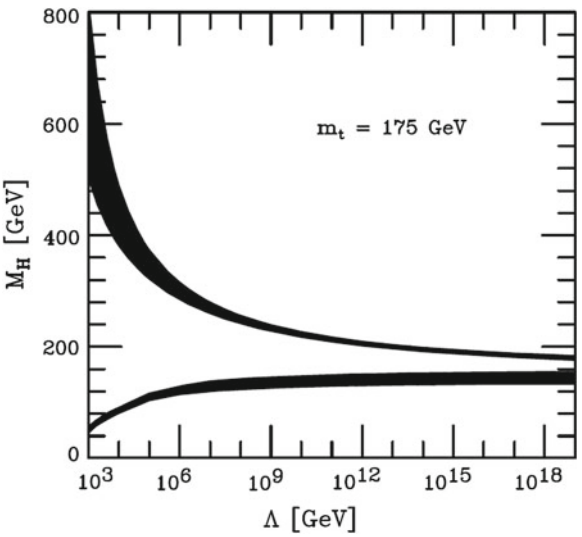

Fig. 1.1 Theoretical boundaries on the mass of the Higgs boson (M_H) as a function of the energy scale Λ at which the Standard Model is not anymore capable of describing nature. The calculation has been performed with a top quark mass value of 175 GeV

1.3 Experimental Limits on the Higgs Boson Mass

The search for the Higgs boson has conditioned most high-energy physics experiments of the past decades. Experimental constrains on its mass are of two categories: direct ones, deriving from the searches performed at colliders (such as LEP at CERN and the Tevatron at Fermilab), and indirect ones, arising mainly from precision measurements of the electroweak parameters. We will here provide a brief overview of the experimental constrains produced prior to the LHC era.

The Large Electron-Positron Collider (LEP) was an e^+e^- accelerator which was operative at CERN from 1989 to 2000. Its scientifical program was divided in two phases: the first one (LEP I), provided collisions on the Z resonance ($89 < \sqrt{s} < 93$ GeV); in the second phase (LEP II), which started in 1996, the center of mass energy was gradually increased up to a maximum of 210 GeV.

The main production mechanism of a Higgs boson at LEP II is the so-called 'Higgs-strahlung' process, where a Higgs is radiated by a virtual Z boson ($e^+e^- \to Z^* \to ZH$). The most probable outcome of this final state is events with four jets, with the Z decaying to a generic quark pair, and the Higgs to a bottom pair:

$$e^+e^- \to (H \to b\bar{b})(Z \to q\bar{q})$$

Tetra-jet events with topology compatible with Higgs-strahlung have been studied by the four detectors operative at LEP: ALEPH [18], DELPHI [19], L3 [20], OPAL [21]. A curious excess was recorded, but no conclusive evidence of a new particle, therefore a lower bound on the Higgs boson mass was placed at 114.4 GeV at 95 % confidence level [22]. This is shown in Fig. 1.2 which shows the trend of the test statistic $-2\ln Q$, defined as

$$-2\ln Q = -2\ln\frac{\Lambda_s}{\Lambda_b}$$

where Λ_b and Λ_s are respectively the likelihoods of the background only, and signal plus background hypotheses. The figure shows the expected trend of $-2\ln Q$ as a function of the hypothetical mass of the Higgs boson for the two cases: background only (black dashed) and for the presence of a signal in addition to background (brown hashed). The 68 % (green) and 95 % (yellow) probability bands are also reported for the background only hypothesis. The observed trend in the data, obtained by combining together the results of the four experiments, is marked by a solid black line. As can be seen, up to a mass of 114.4 GeV the data are compatible with the absence of signal.

The Tevatron is a proton-antiproton collider with a center of mass energy of 1.96 TeV which has taken data up to 2011. The search for a Higgs boson is focused on processes in which it is produced in association with vector bosons ($p\bar{p} \to VH$, $V \equiv W^{\pm}, Z$), and the latter are required to decay in leptonic channels. For masses larger than $m_H \sim 130$ GeV the Higgs decay to a pair of W bosons ($H \to W^+W^-$) is the most promising for its detection.

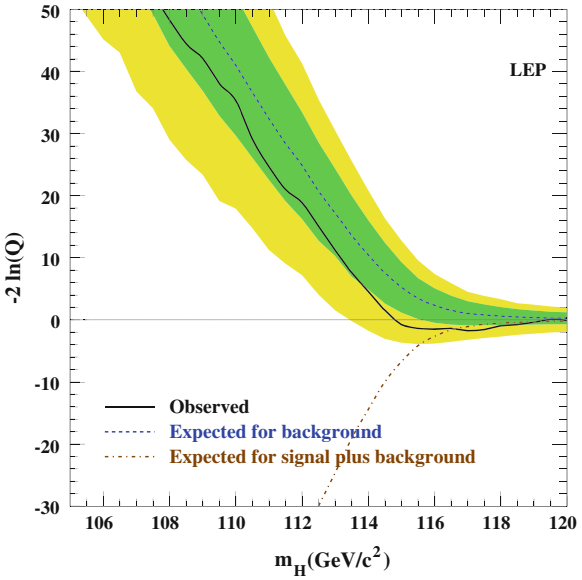

Fig. 1.2 Combined results of the four LEP detectors in the search for a Standar Model Higgs boson. The trend of the $-2\ln Q$ test statistic as a function of the Higgs mass is shown: the expected trend in absence (*black dashed*) or presence (*brown hashed*) of a Higgs signal, and the observed trend (*black solid*). The 68 % and 95 % probability intervals are shown respectively with a *green* and *yellow band*

Figure 1.3 summarizes the most recent results of the Tevatron [23], which combine the searches performed at the two detectors CDF [24] and DØ [25], by showing the trend of the 95 % bayesian confidence level upper limit on the ratio of the Higgs boson production cross section to the SM expectation (further details on this test statistic are provided in Chap. 6. The expected trend (in the absence of signal) is marked by a dashed line, and the 68 % (green) and 95 % (yellow) probability intervals are shown. As can be seen, the observed trend (solid line) reaches values inferior to unity in the 156–177 GeV range, therefore excluding the presence of a Standard Model Higgs boson in this mass interval at 95 % confidence level.

Indirect constrains on the Higgs boson mass derive from a fit performed to the precision measurements performed in the electroweak sector of the Standard Model [26]. These observables are in fact sensitive to the value of the Higgs mass as the latter modifies, through loop corrections, the vacuum polarization of Z and W bosons. It is found that these corrective terms have a logarithmic dependance on the Higgs mass. Figure 1.4 shows the variation of the χ^2 of the best fit [27] to the combined data collected at the LEP, Tevatron and SLC accelerators. As can be seen the lowest mass interval, compatible with the LEP and Tevatron direct exclusions, is favoured, but even high mass hypotheses are not completely overruled.

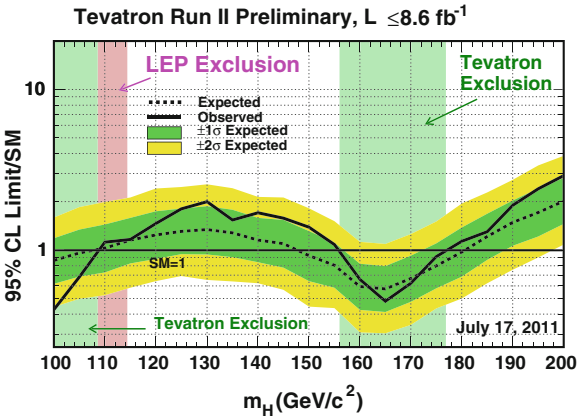

Fig. 1.3 Combined results of the CDF and DØ detectors at Tevatron in the search for a Standar Model Higgs boson, shown as the 95 % confidence level exclusion limit on the ratio of the Higgs production cross section to the Standard Model expectation: the expected trend (*dashed*) is compared to the observed trend in the data (*solid*). The 68 and 95 % probability intervals are shown respectively with a *green* and *yellow band*

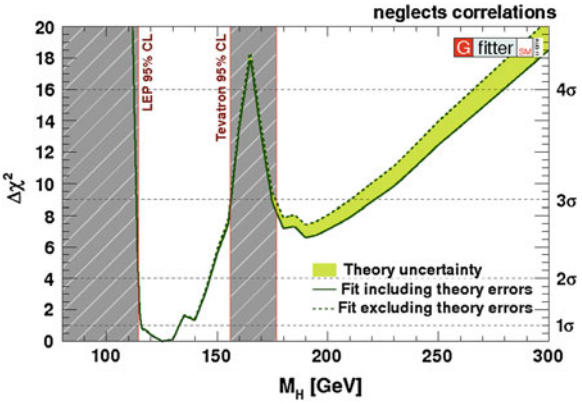

Fig. 1.4 Variation of the χ^2 of the electroweak fit as a function of the Higgs boson mass. The solid (*dashed*) lines give the results when including (ignoring) theoretical errors

1.4 The Higgs Boson at LHC

The Large Hadron Collider is a proton-proton collider with a center of mass energy of 7 TeV, operative at CERN since 2009. The main objective of its scientifical program is to shed light on Higgs sector of the Standard Model. The LHC will be described in detail in the following chapter. Here we will illustrate the expected scenario for the production and detection of a Higgs boson at such energies.

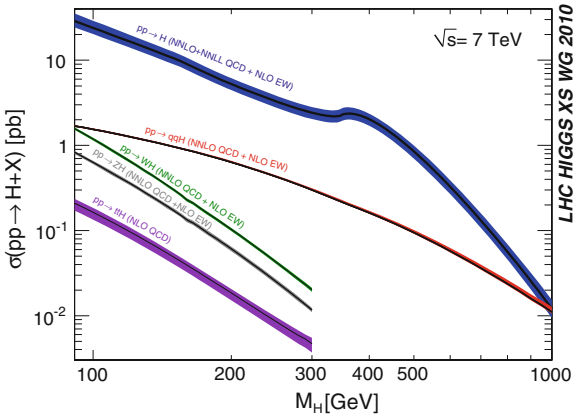

Fig. 1.5 Standard Model Higgs production cross section as a function of its mass at a center of mass energy of 7 TeV. Different production channels are shown separately, from top to bottom: gluon fusion (*blue*), VBF (*red*), associate production with a W boson (*green*), a Z boson (*grey*) and a $t\bar{t}$ pair (*violet*). All cross sections are computed to NNLO precision, except for the associate production with top quark pairs. Figure taken from [28]

The cross sections of the main production processes for a Higgs boson at a 7 TeV proton collider are shown in Fig. 1.5, as a function of the particle's mass. Across the whole mass range the gluon-fusion process ($pp \rightarrow H$, blue) is the dominant production mechanism. The sub-leading contribution, with a cross section about one order of magnitude smaller than gluon fusion, is the Vector Boson Fusion (VBF) process ($pp \rightarrow qqH$, red). Associate production processes, either in conjuction with a W boson (green), a Z boson (grey), or a $t\bar{t}$ pair (violet) are expected to play a minor contribution.

The branching ratios of the main decay channels of a Higgs boson, as a function of its mass, are shown in Fig. 1.6 (left). For low masses, the Higgs boson mainly decays

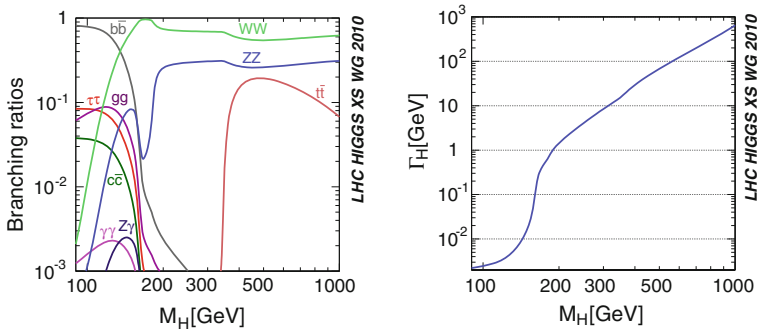

Fig. 1.6 Standard Model Higgs decay: branching ratios separated in single decay channels (*left*), and total decay width Γ_H (*right*) as a function of the Higgs boson mass. Figures taken from [28]

to a bottom quark pair. As m_H increases, decays to vector bosons pairs become energetically allowed, and constitute the dominant decay channels for masses larger than 150 GeV.

The total decay width Γ_H is shown in Fig. 1.6 (right). For masses up to about 300 GeV, the intrinsic width of the resonance is very small, therefore the detector resolutions are expected to play a dominant role in its reconstruction. For heavier masses, the contribution of the intrinsic width has to be accounted for. When approaching the very-high mass limit, towards 1 TeV, the Higgs boson cannot be considered a resonance anymore, as its width becomes comparable to its mass.

1.5 The $H \to ZZ \to \ell^+\ell^- q\bar{q}$ Channel

The discovery of a heavy ($m_H > 200$ GeV) Higgs boson would imply, as we have seen, the presence of new physical processes at energy scales inferior to the grand unification scale. At these masses, the Higgs decays predominantly to vector boson pairs, the detection of which depends on their own decay channels.

The fully leptonic decay modes

$$H \to WW \to \ell^+\nu\ \ell^-\bar{\nu}$$

$$H \to ZZ \to \ell^+\ell^-\ell^+\ell^-$$

where, from now on, we will use ($\ell = e, \mu$), are considered to be the most promising channels, as their signatures are easily reconstructable and distinguishable from background processes. The latter channel, in particular, is considered to be the 'golden' channel for its discovery, as the decay chain is fully reconstructable with high precision. This translates, at analysis level, in a narrow invariant mass peak with negligible Standard Model background contributions. However, by gaining in resolution and signal to background ratio, a price is payed: the rate of this decay chain depends on the $Z \to \ell^+\ell^-$ branching ratio, which is equal to 3.37 % [29]. Therefore, only less than 0.5 % of the total number of Higgs bosons which decay to the ZZ channel will end up in the golden, fully leptonic, final state.

About half of the Higgs bosons which decay to the ZZ channel will produce a four-jet final state, with both Z's decaying to quark pairs (BR($Z \to q\bar{q}$) = 70 %). Nevertheless, even if this channel maximizes the signal rate, it is practically not discernable from Standard Model background sources which produce topologically similar events, such as QCD processes, which are overwhelmingly more frequent at a hadron collider.

A compromise is offered by the semileptonic final state

$$H \to ZZ \to \ell^+\ell^- q\bar{q}$$

This channel still benefits from a large rate, as its branching ratio is more than 20 times larger than the fully leptonic one. On the other hand, the presence of a lepton pair originating from the Z boson decay limits the sources of background processes.

There are, of course, drawbacks:

- the resolution on jet quadrimomentum reconstruction is significantly worse than in the case of electrons or muons;
- even if the background is not submerging as in the fully hadronic case, it is still much larger than in the four-lepton search, as processes which involve jets are in general more frequent than ones which involve leptons at a hadron collider.

Any Standard Model process which presents a final state with a pair of high transverse momentum, opposite-signed electrons or muons in association with two hard jets constitutes a background for this analysis. The main contribution is expected to arise from the production of a Z boson in conjunction with a pair of QCD jets. The generation of a Z boson at LHC is a process with a cross section of about 3 nb, i.e. more than 10^4 larger than the signal we are hunting for.

Additional sources of background come from events with top quarks, and events with pairs of vector bosons. Top quark events simulate the signal signature in two cases:

- when a $t\bar{t}$ pair is produced, and it decays semileptonically to leptons of the same flavour:
$$t\bar{t} \to (W^+ \to \ell^+\nu)b \;\; (W^- \to \ell^-\bar{\nu})\bar{b}$$
- when a t quark is produced in association with a W boson, and they both produce a lepton in the final state:
$$tW^- \to (W^+ \to \ell^+\nu)b + \ell^-\bar{\nu} \quad (+1 \text{ jet})$$
$$\bar{t}W^+ \to (W^- \to \ell^-\bar{\nu})\bar{b} + \ell^+\nu \quad (+1 \text{ jet})$$

Differently from signal, these events are non-resonant in the dilepton invariant mass, therefore requiring the lepton pair to be compatible with a Z boson decay reduces significantly their contribution.

Finally, continuous production of vector boson pairs (either ZZ, WZ, or WW) is a background to this channel, if the decay chain produces two leptons of the same flavour in the final state, and two hard jets. These events could constitute an irreducible background (also because the jet invariant mass resolution is not expected to be able to distinguish $Z \to q\bar{q}$ from $W \to q\bar{q}'$), but the cross section of these processes is small enough to make it a minor source of background.

The total background cross section is substantially larger than the one of the signal. We do, however, have several handles for its discrimination: the dijet system in signal originates from the decay of a Z boson, therefore it has a resonant invariant mass, and it is made of only quark jets, democratically split into all flavours (except the energetically prohibited t), because of the couplings of the Z. Jets in the background, instead, have a substantial infiltration of gluonic jets, and heavy flavours are

suppressed. Therefore a good jet invariant mass resolution and the ability of discriminating the flavour of a jet's parton are potentially powerful means of background rejection.

Moreover, we can use an additional piece of information, which, as it turns out, will constitute our most effective means of background discrimination. In signal events, the four final state physics objects, the lepton and the jet pairs, originate respectively from two Z (spin-1) bosons, which in turn are the decay product of a spin-0 particle, the Higgs boson. The angular distribution of the decay products of any particle is determined by its spin. In the case of the $H \to ZZ \to \ell^+\ell^-q\bar{q}$ decay, the decay of a scalar particle is followed by the decay of two spin-1 particles. The final state products, therefore, exhibit the characteristics of the appropriate rotation matrices, that can be exploited to discriminate signal from non-resonant backgrounds which, conversely, gives rise to random angular distributions of the final state particles.

In this thesis we will demonstrate that we are able to successfully fight these competing processes, by using all the mentioned means of discrimination:

- optimal jet calibration and resolution in order to maximise the separation offered by the dijet invariant mass;
- jet flavour tagging, both to isolate heavy-quark jets, and to reject jets which are likely to originate from gluons;
- angular analysis of the final state, in order to select events which have a topology compatible with the decay of a spin-0 particle.

By doing so, we will show that the $H \to ZZ \to \ell^+\ell^-q\bar{q}$ channel proves to be a major player in the search for a heavy Higgs boson.

References

1. Glashow, S.: Partial symmetries of weak interactions. Nucl. Phys. **22**, 579 (1961). http://dx.doi.org/10.1016/0029-5582(61)90469--2
2. Salam, A.: Elementary Particle Physics: Relativistic Groups and Analyticity, vol. Proceedings to the Eighth Nobel Symposium, Almquvist and Wiksell (1968)
3. Weinberg, S.: A model of leptons. Phys. Rev. Lett. **19**, 1264 (1967). http://dx.doi.org/10.1103/PhysRevLett.19.1264
4. Precision electroweak measurements on the Z resonance. Phys. Rep. **427**, 257 (2006). Available from: http://dx.doi.org/10.1016/j.physrep.2005.12.006,
5. Higgs, P.W.: Broken symmetries, massless particles and gauge fields. Phys. Lett. **12**, 132 (1964). http://dx.doi.org/10.1016/0031-9163(64)91136--9
6. Higgs, P.W.: Broken symmetries and the masses of Gauge Bosons. Phys. Rev. Lett. **13**, 508 (1964). http://dx.doi.org/10.1103/PhysRevLett.13.508
7. Gell-Mann, M.: The interpretation of the new particles as displaced charge multiplets. Il Nuovo Cimento (1955–1965), **4**(848), 1956, 10.1007/BF02748000. Available from: http://dx.doi.org/10.1007/BF02748000
8. Nishijima, K.: Charge independence theory of V particles. Prog. Theor. Phys. **13** (1955), 285. Available from: http://ptp.ipap.jp/link?PTP/13/285/. http://dx.doi.org/10.1143/PTP.13.285
9. Cabibbo, N.: Unitary symmetry and leptonic decays. Phys. Rev. Lett. **10**, 531 (1963). http://dx.doi.org/10.1103/PhysRevLett.10.531

10. Kobayashi, M., Maskawa, T.: CP Violation in the Renormalizable theory of weak interaction. Prog. Theor. Phys. **49** (1973), 652. http://dx.doi.org/10.1143/PTP.49.652
11. Marciano, W., Sirlin, A.: Electroweak radiative corrections to Tau decay. Phys. Rev. Lett. **61** (1988), 1815. http://dx.doi.org/10.1103/PhysRevLett.61.1815
12. van Ritbergen, T., Stuart, R. G.: On the precise determination of the Fermi coupling constant from the muon lifetime. Nucl. Phys. **B564**, 343 (2000). http://arxiv.org/abs/hep-ph/9904240,http://dx.doi.org/10.1016/S0550-3213(99),00572--6
13. Cabibbo, N., Maiani, L., Parisi, G., Petronzio, R.: Bounds on the Fermions and Higgs Boson masses in grand unified theories. Nucl. Phys. **B158**, 295 (1979).http://dx.doi.org/10.1016/0550-3213(79),90167--6
14. Altarelli, G., Isidori, G.: Lower limit on the Higgs mass in the standard model: An Update. Phys. Lett. **B337**, 141 (1994). http://dx.doi.org/10.1016/0370-2693(94),91458--3
15. Casas, J., Espinosa, J., Quiros, M.: Improved Higgs mass stability bound in the standard model and implications for supersymmetry. Phys. Lett. **B342**, 171 (1995). http://arxiv.org/abs/hep-ph/9409458,http://dx.doi.org/10.1016/0370-2693(94)01404-Z
16. Casas, J., Espinosa, J., Quiros, M.: Standard model stability bounds for new physics within LHC reach. Phys. Lett. **B382**, 374 (1996). http://arxiv.org/abs/hep-ph/9603227,http://dx.doi.org/10.1016/0370-2693(96)00682-X
17. Hambye, T., Riesselmann, K.: SM Higgs mass bounds from theory. (1997). http://arxiv.org/abs/hep-ph/9708416
18. De Palma, M., et al.: ALEPH: Technical, Report 1983. CERN-LEPC-83-2
19. Bartl, W., et al.: DELPHI: Technical Proposal. DELPHI-83-66-1
20. von Dardel, G., Walenta, A. H., Lubelsmeyer, K., Deutschmann, M., Leiste, R., et al.: L3 Technical Proposal (1983). See the BOOKS subfile under the following call number: QCD197:E951
21. The Opal Detector. Technical Proposal. LEPC-83-4
22. Search for the standard model Higgs boson at LEP. Phys. Lett. B, **565**, 61 (2003). Available from: http://www.sciencedirect.com/science/article/pii/S0370269303006142,http://dx.doi.org/10.1016/S0370-2693(03),00614--2
23. CDF and D Ø Collaborations. Combined CDF and D0 Upper Limits onStandard Model Higgs Boson Production with up to 8.6 f b^{-1} of Data (2011). Available from: http://arxiv.org/abs/1107.5518
24. Blair, R., et al.: The CDF-II detector: Technical design, report. FERMILAB-DESIGN-1996-01
25. Abazov, V., et al.: D0 Run IIB upgrade technical design, report (2002)
26. LEP electroweak working group. http://lepewwg.web.cern.ch/LEPEWWG/ Summer 2011 results
27. GFitter. http://gfitter.desy.de/Standard_Model/. August 2011 results
28. LHC Higgs cross section working group.: Dittmaier, S., Mariotti, C., Passarino, G., Tanaka, R. (eds.) Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables. CERN-2011-002, (CERN, Geneva, 2011). http://arxiv.org/abs/1101.0593
29. K. Nakamura., et al.: The review of particle physics. J. Phys. G, 37 (2010). Available from: http://pdg.lbl.gov/

Chapter 2
The Large Hadron Collider and the CMS Experiment

Abstract This chapter is dedicated to the description of the experimental apparatus which made these measurements possible. Section 2.1 describes the Large Hadron Collider, the accelerator which provided 7 TeV proton-proton collisions which were analysed in this thesis. The collisions were reconstructed with the Compact Muon Solenoid (CMS) detector, to which Sect. 2.2 is dedicated to. The two final sections of this chapter illustrate the lepton reconstruction techniques adopted in CMS.

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) at CERN is the largest and most energetic particle accelerator ever built. It occupies the 27 km long tunnel previously hosting the LEP collider, about 100 m underneath the surface across the French-Swiss national border near Geneva. It is a superconducting proton-proton collider, capable of producing collisions at a center of mass energy of up to 14 TeV, and a maximal instantaneous luminosity of 10^{34} cm^{-2} s^{-1}. It has been delivering 7 TeV proton collisions since March 2010, and it is expected to raise its center of mass energy in the coming years.

The LHC proton injection chain is schematically shown in Fig. 2.1. Protons are accelerated three times before they enter the LHC ring: the LINAC brings them to 50 MeV, the Proton Synchrotron (PS) to 1.4 GeV, and finally the Super Proton Synchrotron (SPS) injects them into the LHC at 450 GeV. The LHC then completes the acceleration by bringing them to 7 TeV with its 400 MHz radiofrequency cavities, capable of 'kicks' which result in increases of the proton energy of 0.5 MeV per turn.[1]

Since the collisions occur between particles of the same electrical charge, two separate acceleration cavities and two different magnetic field configurations are required. The LHC is equipped with 1232 superconducting 14.2 m long Niobium-Titanium dipole magnets, cooled down to 1.9 K by means of super-fluid Helium, that

[1] The LHC is also capable of accelerating and colliding beams of lead ions at 2.76 TeV in the center of mass per nucleon. As this is not relevant for this thesis, it will not be treated.

F. Pandolfi, *Search for the Standard Model Higgs Boson in the $H \to ZZ \to \ell^+\ell^- q\bar{q}$ Decay Channel at CMS*, Springer Theses, DOI: 10.1007/978-3-319-00903-2_2,

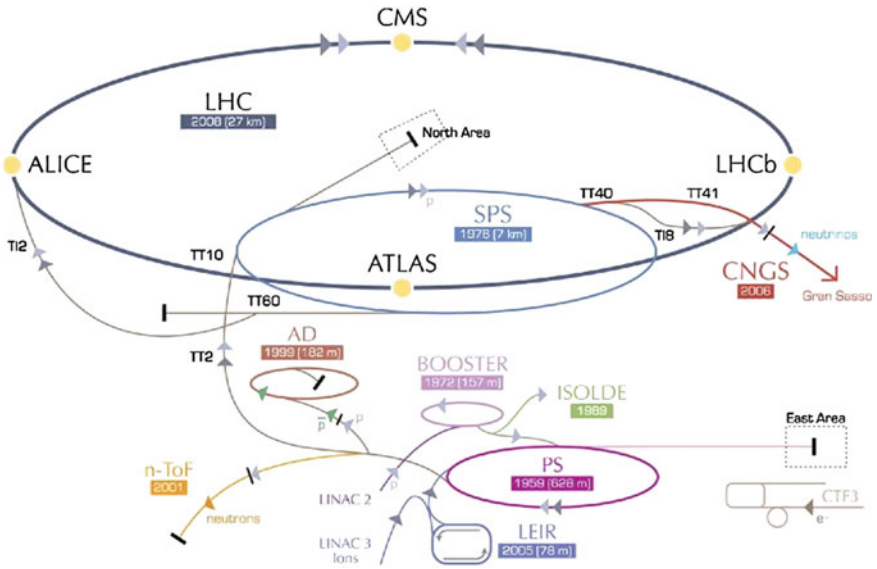

Fig. 2.1 Schematic view of the LHC injection scheme

create a bending magnetic field of about 8.3 T. The magnets are placed in the eight curved sections which connect the straight sections of the LHC ring.

The high luminosity of the LHC is obtained by a high frequency bunch crossing and a high number of protons per bunch: two beams of protons with an energy of up to 7 TeV (3.5 TeV in the initial physics runs), circulating in two different vacuum chambers, contain each up to 2,808 bunches. The bunches, with a nominal number of 10^{11} protons each, have a very small transverse spread (about 15 μm in the transverse directions) and are about 7.5 cm long in the beam direction at the interaction points. The bunches cross at the rate of up to 40 MHz, i.e. one collision each 25 ns. A summary of the principal LHC technical parameters is given in Table 2.1.

The LHC can cross its beams in four interaction points. Two of them have high luminosity and are dedicated to the general purpose experiments ATLAS [1] and CMS. The other two, at lower luminosity, serve the ALICE [2] and LHCb [3] experiments, respectively focused on heavy ion physics and CP violation measurements.

The operating conditions at the LHC are extremely challenging for the experiments. The total proton-proton cross section is estimated to be about 100 mb [4], which implies about 20 proton interactions per bunch crossing, i.e. 10^9 interactions per second. A strong online event selection is therefore required in order to reduce the event rate at $\mathcal{O}$ (100) Hz, corresponding to the maximum data storage rate sustainable by the existing device technology. The detectors must also have a fast response time (around 25 ns) and a fine granularity in order to minimize the performance degradation in simultaneous events. In addition to this, the high flux of particles coming from proton interactions implies that each component of the detector has to be radiation

Table 2.1 Summary of the principal LHC technical parameters

Circumference [km]	27
Number of magnet dipoles	1,232
Dipolar magnetic field [T]	8.33
Radiofrequency [MHz]	400
Maximal number of bunches	2,808
Magnet temperature [K]	1.9
Maximal beam energy [TeV]	7
Maximal luminosity [$cm^{-2}s^{-1}$]	10^{34}
Initial beam energy [TeV]	3.5
Protons per bunch	$1.05 \cdot 10^{11}$
Bunch spacing [m]	7.48
Minimal bunch time separation [ns]	25
Bunch length [cm]	7.5
Bunch transverse size [μm]	15
Crossing angle [rad]	$2 \cdot 10^{-4}$
Beam lifetime [h]	7
Luminosity lifetime [h]	10

resistant. Finally, to fully understand the physical processes occurring at the LHC, multi-purpose detectors are required to satisfy the following requirements:

- full hermeticity, in order to provide accurate measurements of missing transverse energy;
- excellent reconstruction of high-energy leptons and photons;
- precise determination of charged particle momenta and impact points through an efficient tracking system;
- accurate reconstruction of hadronic activity from QCD processes and heavy particle decays.

The Compact Muon Solenoid detector meets all these stringent requirements. It is described in the following.

2.2 The CMS Experiment

The Compact Muon Solenoid (CMS) experiment [5] is one of the two general-purpose detectors which take data at the LHC. One of the cardinal points of its scientifical program is the discovery of the Higgs boson. Its design philosophy has therefore been driven by such a search, and can be summarized by the following points:

- an excellent and redundant muon system. This has lead to the choice of a large, superconducting solenoidal magnet, capable of producing a 4 T field, which allows to have a compact muon spectrometer, delivering precise track and unambiguous charge measurements for muons of transverse momenta up to 1 TeV;

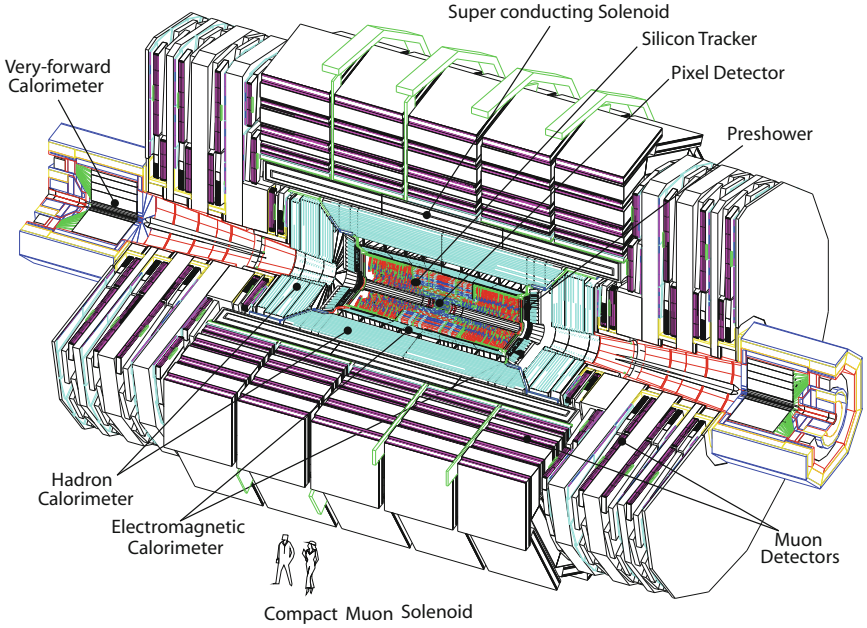

Fig. 2.2 An exploded view of the CMS detector

- the best possible electromagnetic calorimeter compatible with the magnet;
- a precise and efficient inner tracking system;
- a highly hermetic hadronic calorimeter system, capable of delivering good performance in missing transverse energy reconstruction.

The structure of the CMS detector is shown in Fig. 2.2. It has a cilindrical shape, symmetrical around the beam direction, with a radius of 7.5 m, a total length of 22 m, and weighs about 12, 500 tons. It is divided into a central section, made of several layers coaxial to the beam axis (the *barrel*), closed at its ends by two hermetic discs orthogonal to the beam (the *endcaps*).

A schematic view of a transverse section of CMS is visible in Fig. 2.3. Moving outwards starting from the beam position, it presents a silicon tracker, a crystal electromagnetic calorimeter, a hadronic calorimeter, and the superconducting solenoidal magnet, in the return yoke of which the muon drift chambers are inserted.

The coordinate system adopted in CMS is cartesian, has the origin centered in the nominal collision point at the center of the detector, and adopts the following conventions:

- the x axis points towards the center of the LHC ring;
- the z direction coincides with the CMS cylinder axis;
- the y axis points upwards, towards the surface.

The cylindrical symmetry of the apparatus drives the use of a pseudo-angular reference system, given by the triplet of variables (r,ϕ,η), where r is the radial distance

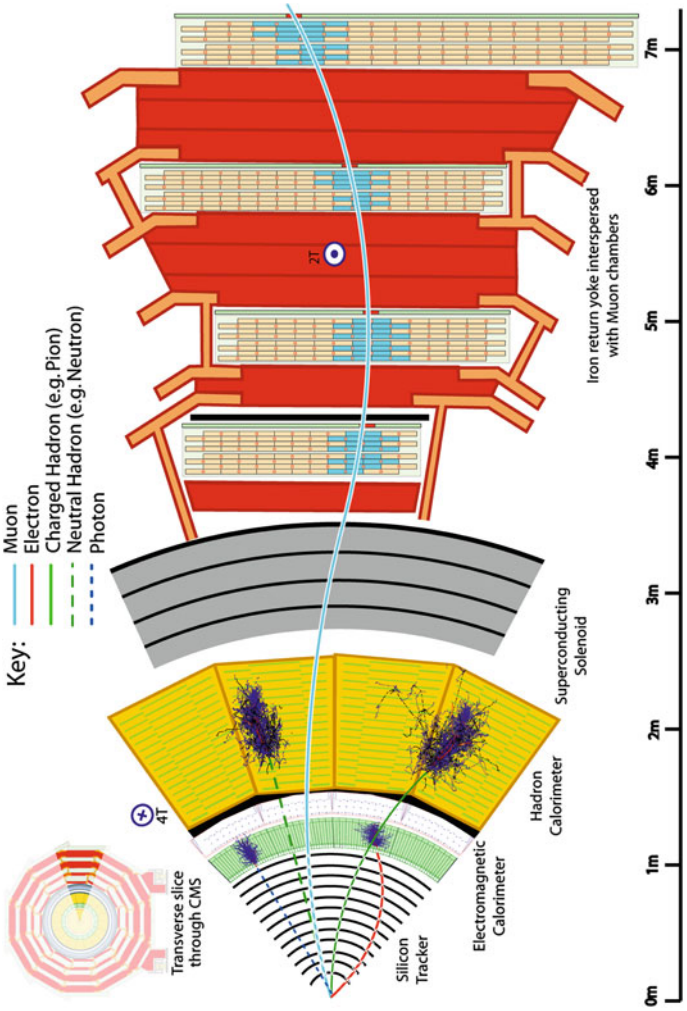

Fig. 2.3 Schematic view of a transverse slice of the central part of the CMS detector

from the beam axis, ϕ is the aximuthal angle with respect to the x axis, and η, commonly referred to as *pseudorapidity*, is defined as

$$\eta = -\ln \tan \frac{\theta}{2}$$

where θ is the polar angle with respect to the z axis. The use of the pseudorapidity is motivated by the fact that for high energies it is a good approximation of the rapidity (y) of a particle, defined as

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

where E is the particle's energy and p_{L} is the component of its momentum projected along the beam axis. The rapidity is a useful variable at a hadron collider, as it is invariant, expect for a constant additive term, under Lorentzian boosts along the axis direction. It follows from its definition that the pseudorapidity is null for $\theta = 0$ and increases in absolute value when approaching the beam pipe, asymptotically reaching infinity at $\theta = \pi/2$ (on the z axis).

We furthermore denote with p_{T} the component of a particle's momentum in the plane transverse to the beam axis, and with E_{T} its transverse energy, obtained from its energy by $E_{\mathrm{T}} = E \cdot \sin \theta$.

We will now provide a brief description of the subdetectors which constitute CMS.

2.2.1 Magnet

In order to achieve a compact and high-resolution muon detection system, a large bending power is required. This can be achieved by a relatively small solenoid, provided that an intense magnetic field is produced, as the bending starts at the collision vertex. A large enough length/radius ratio is also demanded for, in order to ensure good momentum resolution in the forward region as well.

These considerations led to the choice [6] of a 13 m long superconducting cylindrical Niobium-Titanium coil, with a diameter of 5.9 m. It provides a uniform magnetic field of 3.8 T at its center, carrying a current of 18 kA and a total stored magnetic energy of 2.4 GJ. The magnet flux is returned by a saturated iron yoke, which also works as mechanical support structure of the detector.

2.2.2 Tracker

The design goal of the inner tracking system is to reconstruct isolated, high-p_{T} electrons and muons with efficiency greater than 95 %, and tracks of particles within jets with efficiency greater than 90 %, within a pseudorapidity coverage of $|\eta| < 2.4$. At the same time it must comply to severe material budget constraints, in order not to

Table 2.2 Expected radiation dose and charged particle flux at different radii in the barrel of the CMS tracker, for the high-luminosity run of the LHC and an integrated luminosity of 500 fb^{-1}

Radius (cm)	Radiation dose (kGy)	Charged particle flux ($\text{cm}^{-2}\text{s}^{-1}$)
4	840	10^8
22	70	$6 \cdot 10^6$
115	1.8	$3 \cdot 10^5$

degrade its momentum resolution. All of this must be achieved in a high-multiplicity, highly radioactive environment such as the one created by LHC collisions.

This led to the choice of a large silicon tracker [7]. It is the first example in high-energy physics of an inner tracking system completely based on this technology alone. Referring to Table 2.2, which summarizes the expected dose of radiation and charged particle flux for the high-luminosity run of the LHC, we can identify three tracker regions:

- closest to the interaction vertex where the particle flux is highest ($\approx 10^7$/s at $r \approx 10$ cm), pixel detectors are placed. The size of a pixel is about 100×150 μm^2, giving an occupancy of about 10^{-4} per pixel per high-luminosity LHC crossing;
- the intermediate region ($20 < r < 55$ cm), where the particle flux is low enough to enable the use of silicon microstrip detectors with a minimum cell size of 10 cm $\times$ 80 μm, leading to an occupancy of $\approx$ 2–3 %/crossing;
- the outermost region ($r > 55$ cm) of the inner tracker, where the particle flux has dropped sufficiently to allow the adoption of larger-pitch silicon microstrips with a maximum cell size of 25 cm $\times$ 180 μm, whilst keeping occupancy around 1 %.

Even in heavy-ion (Pb-Pb) running, the occupancy is at the level of 1% in the pixel detectors and less than 20 % in the outer silicon strip detectors, permitting track reconstruction in such a high density environment.

The layout of the pixel detector is shown in Fig. 2.4: it features three cylindrical layers in the barrel region, placed respectively at radii of 4.7, 7.3 and 10.2 cm, and

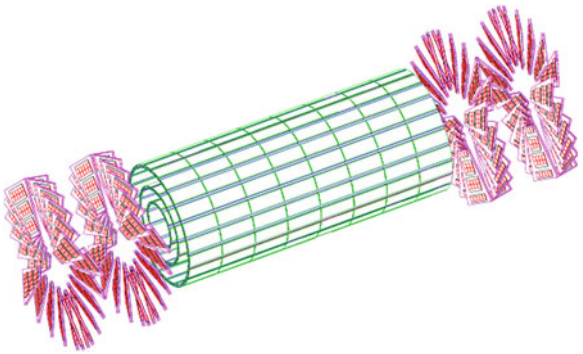

Fig. 2.4 Layout of the CMS silicon pixel detector

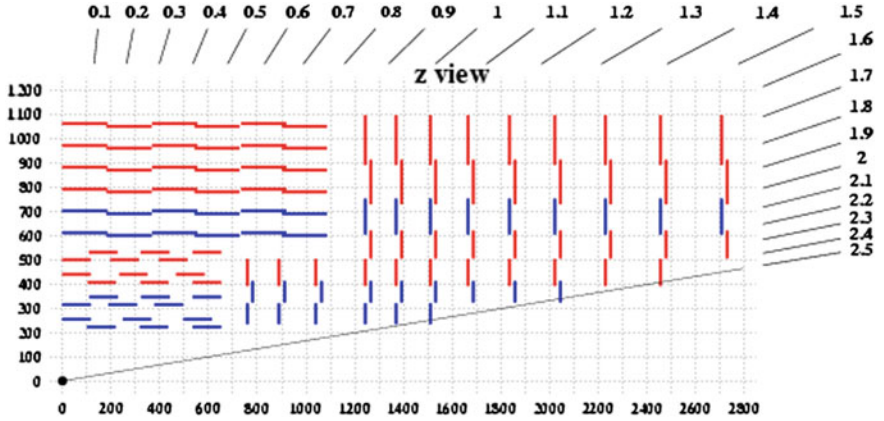

Fig. 2.5 Schematic longitudinal section of one quarter the CMS silicon microstrip detector. The nominal interaction point is in the *bottom-left corner*. Distances are marked in *millimeters on the left and bottom axes*, and pseudorapidity values are shown on the *top and right borders*

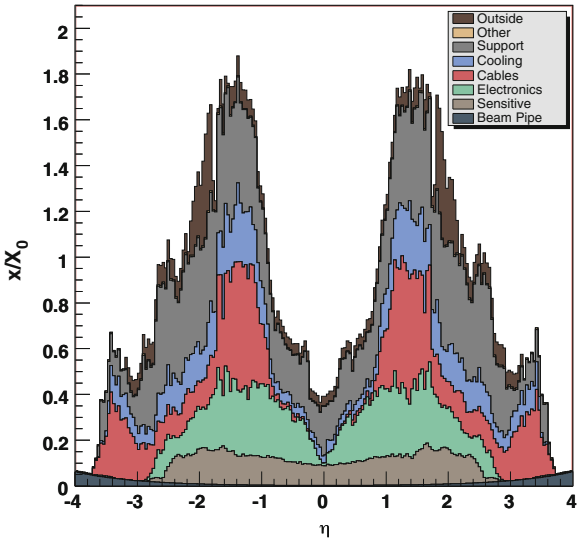

Fig. 2.6 Silicon tracker material budget as a function of pseudorapidity, expressed in units of radiation lengths (X_0). Different material categories are shown: beam pipe, silicon sensitive volumes, electronics, cables, cooling pipes and fluid, support mechanics and outer structures

two closing end-discs, extending from 6 to 15 cm in radius, and placed on each side at $|z| = 34.5$ cm and 46.5 cm. This design ensures that each charged particle produced within $|\eta| = 2.2$ releases at least two hits in the pixel detector.

The pixel layers are enveloped by a silicon microstrip detector, a section of which is schematically represented in Fig.2.5. The barrel microstrip detector is divided in

two regions: the inner and outer barrel. The inner barrel is made of four layers (the two innermost of which are double-sided), and cover the depth $20 < r < 55$ cm. The outer barrel counts six layers (the two innermost double-sided) and reach up to a radius of 110 cm. In order to avoid that particles hit the sensitive area at too small angles, the inner barrel is shorter than the outer region, and three additional disc-shaped layers have been inserted, between the inner barrel and the endcaps. The endcap detector is made of nine layers of discs, up to a maximum distance of $z = 270$ cm. The first, second, and fifth layers are double-sided.

The silicon tracker detector comprises of a total of 66 million pixel and 9.6 million strip channels. Its material budget in units of interaction lengths, as a function of pseudorapidity, is shown in Fig. 2.6. As can be seen, it adds up to less than half a radiation length in the center of the barrel, increasing to a maximum of about 1.8 X_0 in the barrel-endcap transition ($|\eta| \sim 1.5$).

2.2.3 Electromagnetic Calorimeter

A high-performance electromagnetic calorimeter is a cardinal element of a general purpose high-energy physics detector, as it allows precise measurements of photon and electron energies. The CMS collaboration has opted [8] for a hermetic, homogeneous electromagnetic calorimeter (ECAL), made of 61 200 lead tungstate ($PbWO_4$) scintillating crystals mounted in the barrel region, and closed by two endcaps, which count 7 324 crystals each.

The main characteristics of $PbWO_4$ are listed in Table 2.3. The choice of this inorganic crystal has been determined by a number of factors:

- its short radiation length (0.89 cm) allows the construction of a compact calorimeter, which can comply to the requirements imposed by the magnet radius;
- its small Molière radius (2.2 cm) ensures an efficient lateral shower containment, and therefore high granularity;
- it is characterized by a very fast light emission process (its principal and secondary scintillation components have emission times respectively of 5 and 15 ns), a crucial feature at a collider where bunch crossings are interspaced by only 25 ns;

Table 2.3 Main characteristics of lead tungstate ($PbWO_4$) The superscripts f and s respectively denote the principal (*fast*) and secondary (*slow*) scintillation emissions

Radiation length (cm)	0.89
Density (g cm^{-3})	8.3
Molière radius (cm)	2.2
Refractive index	2.29
Light Yield (γ/MeV)	30
Light emission time (ns)	5^f
	15^s
Scintillation wavelength (nm)	440^f
	480^s

- it is sufficiently radiation hard, allowing it to sustain several years of high-luminosity running with tolerable degradation of the crystals transparency, which can be corrected with a light monitoring system.

However, the relatively low light yield (30 γ/MeV) necessitates the use of photodetectors with high intrinsic gain and which can operate in a magnetic field. This led to the use of silicon avalanche photodiodes (APD) in the barrel, and vacuum phototriodes (VPT) in the endcaps.

A longitudinal section of a quarter of the ECAL is shown in Fig. 2.7. The barrel covers the pseudorapidity region $|\eta| < 1.479$ and has an inner radius of 129 cm. It is made of 18 identical supermodules, each of which covers half the barrel length. It has a granularity of 360 crystals in the azimuthal direction (ϕ), and (2×85) crystals in η. The crystals are organized in a quasi-projective geometry, so that their axes form a 3° angle with the line that connects them to the nominal interaction point. A single crystal corresponds to a 0.0174×0.0174 square in the $\eta - \phi$ plane, and its front face measures about 22×22 mm^2. They are 23 cm long, equivalent to 25.8 X_0.

The endcaps are placed at a distance of 3.144 m from the nominal interaction point and reach up to $|\eta| = 3$. They are made of identical crystals, with a front face of 28.62×28.62 mm^2 and a length of 22.0 cm (24.7 X_0). They are grouped in 5×5 mechanical units, called *supercrystals*.

The front face of the endcaps is equipped, in the $1.653 < |\eta| < 2.6$ pseudorapidity interval, with a preshower detector. It is a two-layer sampling calorimeter, which uses lead as absorber and silicon strips as active material. The thickness of the two lead absorbers is respectively 2 X_0 and 1 X_0.

The energy resolution of a homogeneous calorimeter may be parametrized with the following expression:

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

The stochastic (S), noise (N), and constant (C) terms of the ECAL have been measured at a test beam [9], and were found to have a value of:

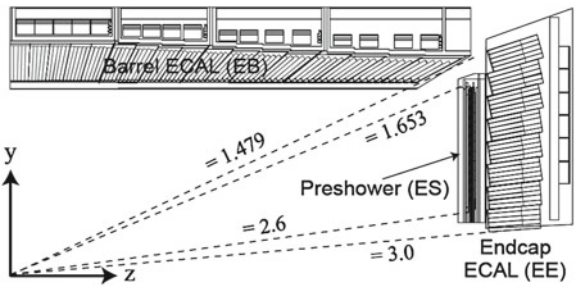

Fig. 2.7 Longitudinal section of a quarter of ECAL

$$S = 2.8\%\ \text{GeV}^{1/2} \qquad N = 124\ \text{MeV} \qquad C = 0.3\,\%$$

2.2.4 Hadronic Calorimeter

The role of the hadronic calorimeter [10] is to contain the showers of hadronic particles, and therefore measure jet quadrimomenta and the missing transverse energy of events. The two key features for these tasks are a high hermeticity and a good transverse granularity. Furthermore, a good energy resolution and a sufficient longitudinal containment are also important.

A longitudinal section of a quarter of the hadronic calorimeter is shown in Fig. 2.8. It is formed by two separate detectors: a central (HCAL) and a forward (HF) calorimeter. The HCAL covers the pseudorapidity range $|\eta| < 3$, and is in turn divided into two subdetectors: a barrel ($|\eta| < 1.3$) and two endcaps ($1.3 < |\eta| < 3$). It is a sampling calorimeter, with brass used as absorber and plastic scintillators as active material. It has a transverse granularity of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$.

The energy resolution of the HCAL is parametrized as

$$\frac{\sigma}{E} = \frac{100\,\%}{\sqrt{E\,(\text{GeV})}} \oplus 8\,\%$$

for pions of energy E. It has a total thickness of 7–10 interaction lengths (λ_i). A depth of $7\lambda_i$ is not sufficient to ensure the complete containment of a highly energetic hadronic shower, therefore an additional layer of active material was added behind the solenoid, which increases the total effective thickness by about $3\lambda_i$ and improves by 10 % the energy resolution for 300 GeV pions.

To improve the detector's hermeticity, an additional calorimeter (HF) is placed outside the magnet yoke, 11 m away from the interaction point, on both sides. It covers the very forward pseudorapidity region $3 < |\eta| < 5$. In order to sustain the very

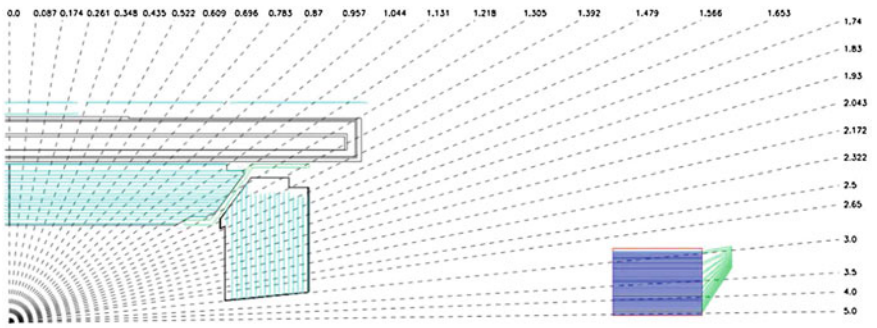

Fig. 2.8 Longitudinal section of a quarter of the CMS hadronic calorimeters: HCAL is visible on the *left*, HF far away from the interaction point on the *right*. Some values of pseudorapidity are marked

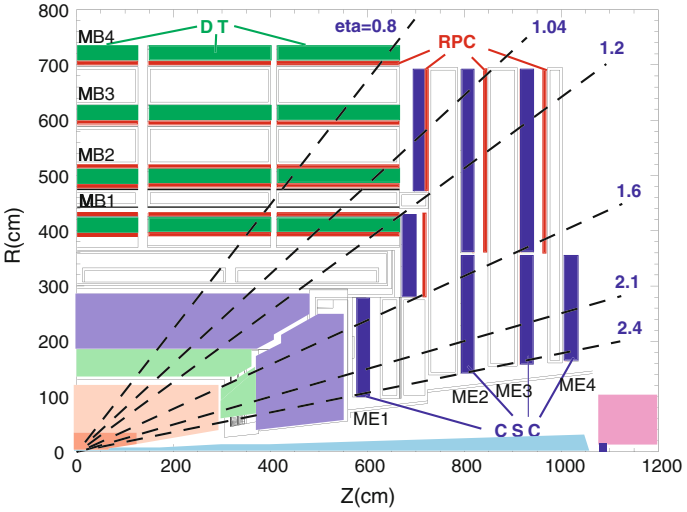

Fig. 2.9 Longitudinal section of a quarter of the CMS muon system

high doses of radiation and high particle multiplicities expected for such a region, they are sampling calorimeters made of iron and quartz fibers. The fibers are of two different lengths, the longer ones reach the front face of the calorimeter, the short ones end 22 cm before that. In this way, most of the electromagnetic component of the hadronic showers will be released in the long fibers, and can therefore be isolated by subtraction. Its granularity is $\Delta\eta \times \Delta\phi = 0.175 \times 0.175$.

2.2.5 Muon System

The muon system [11] has the aim of detecting muons, the only charged particles which are able to pass through the calorimeters without being absorbed. It is placed outside the magnet coil, and it has a pseudorapidity reach of $|\eta| < 2.4$. It is subdivided in a barrel and two endcaps, and the two regions use different technologies. Both regions are made of four layers of measuring stations, imbedded in the iron of the magnet return yoke, where the return field of the solenoid is about 1.5 T. A schematic longitudinal section of a quarter of the CMS muon system is shown in Fig. 2.9.

The barrel region ($|\eta| < 1.2$) is made of drift tube (DT) stations, each of which is made of 12 planes of tubes, for a total of 195, 000 tubes. The endcaps ($1.2 < |\eta| < 2.4$) have to cope with an intense magnetic field and a higher particle multiplicity, therefore cathode strip chamber (CSC) detectors are employed, organized in six-layer modules. CSC's are multi-wire proportional chambers in which the cathode plane has been segmented into strips.

In addition to this, both barrel and endcaps are equipped with resistive plate chamber (RPC) detectors, which are parallel-plate gas chambers with an excellent (3 ns)

time resolution. The RPC's supply a very fast trigger system, capable of identifying muons with high efficiency. They are organized in six barrel and four endcap stations, for a total of 612 chambers.

2.2.6 Trigger

At the nominal LHC luminosity the event rate is expected to reach 10^9 Hz. Given that the typical raw event size is about 1 MB, it is not possible to record all proton-proton collisions. On the other hand, the vast majority of interactions are soft collisions, which are not interesting for the search-oriented physics program CMS intends to pursue. Therefore the aim of the trigger system is that to lower the rate of acquired events to manageable levels ($\sim$100 Hz), while still retaining most of the potentially useful events.

This is achieved through a two-tier system: a Level-1 Trigger (L1) and a High-Level Trigger (HLT). The L1 system is made of a series of custom-designed, largely programmable hardware processors, whereas the HLT is a software system implemented in a computer farm made of about one thousand commercial processors. The design goal of the trigger system as a whole is that to have a reduction rate capability of 10^7.

2.2.6.1 Level-1 Trigger

The Level-1 trigger [12] reduces the rate of selected events down to 50–100 kHz. The full data are stored in pipelines of processing elements, while waiting for the trigger decision. The L1 decision whether taking or discarding data from a particular bunch crossing has to be taken in 3.2 μs; if the L1 accepts the event, the data are moved to be processed by the HLT.

To deal with the high bunch crossing rate, the L1 trigger has to take a decision in a time too short to read data from the whole detector, therefore it employs the calorimetric and muon information only, since the tracker algorithms are too slow for this purpose. The L1 trigger is organized into a Calorimeter Trigger and a Muon Trigger, whose informations are transferred to the Global Trigger which takes the final accept-reject decision.

The Calorimeter Trigger is based on trigger towers, 5×5 matrices of ECAL crystals, which match the granularity of HCAL cells. The trigger towers are grouped in 4×4 squares. The Calorimeter Trigger identifies the best four candidates of each of the following classes: electrons and photons, central jets, forward jets and τ-jets (identified from the shape of the deposited energy). The information of these objects is passed to the Global Trigger, together with the measured calorimetric missing transverse energy. The Muon Trigger is ran separately for each muon detector. The information is then merged and the best four muon candidates are transferred to the Global Trigger.

The Global Trigger takes the accept-reject decision exploiting both the characteristic of the single objects and of combinations of them.

2.2.6.2 High-Level Trigger

The High-Level Trigger [13] reduces the output rate to about 100 Hz. It is a highly-customizable software system, in which flexibility is maximized because there is complete freedom in deciding which data to access, as well as the sophistication of the adopted algorithms. The HLT software is organized in a set of algorithms (known as HLT 'paths') which are designed to select specific event topologies.

Various strategies are employed at HLT, some of which (the ones relevant for this analysis) will be shown in Sects. 2.3 and 2.4. The guiding principles are: regional reconstruction, and fast event veto. Regional reconstruction tries to avoid the complete event reconstruction, which would take time, but rather focuses on the detector regions close to where the L1 trigger has found interesting activity. Fast event veto means that uninteresting events are discarded as soon as possible, therefore freeing the processing power for the next events in line. This has led to the development of three virtual trigger levels: the first level accesses only the muon and calorimetric data, the second level adds the data of the pixel seeds, the final step reads the full event information.

2.2.7 CMS Software Components

The goals of the CMS software are to process end select events inside the HLT farm, to deliver the processed results to the experimenters within the CMS collaboration and to provide tools for them to analyze the processed information and produce physics results. The overall collection of software, now referred to as `CMSSW`, is built around a Framework, an Event Data Model, and Services needed by the simulation. The physics and utility modules are written by detector groups. The modules can be plugged into the application framework at run time, independently of the computing environment. The software should be developed keeping in mind not only performance but also modularity, flexibility, maintainability, quality assurance and documentation. CMS has adopted an object-oriented development methodology, based primarily on the `C++` programming language.

The primary goal of the CMS Framework and Event Data Model (EDM) is to facilitate the development and deployment of reconstruction and analysis software. The EDM is centered around the Event class, which holds all data that was taken during a triggered physics event as well as all data derived from the data taking (e.g. calibration and alignment constants).

The detailed CMS detector and physics simulation is currently based on the `GEANT 4` [14] simulation toolkit and the CMS object-oriented framework and event model. `GEANT 4` provides a rich set of physics processes describing electromagnetic and hadronic interactions in detail. It also provides tools for modeling the full

CMS detector and geometry and the interfaces required for retrieving information from particle tracking through these detectors and the magnetic field. The validation of GEANT 4 in the context of CMS is described in detail in [15]. The CMS GEANT 4-based simulation program uses the standard CMS software framework and utilities, as used by the reconstruction programs.

The simulation is implemented for all CMS subdetectors in both the central and forward region, including the field map of the 3.8 T solenoid. In addition, several test-beam prototypes and layouts have been simulated. The full simulation program implements the sensitive detector behavior, track selection mechanism, hit collection and digitization (i.e. detector electronic response). The detailed simulation workflow is as follows:

- a physics group configures an appropriate Monte Carlo event generator (several are used) to produce the data samples of interest;
- the production team/system runs the generator software to produce generator event data files;
- the physics group validates the generator data samples and selects a configuration for the GEANT-based simulation of CMS, with generator events as input, to produce (using the standard CMS framework) persistent hits in the detectors;
- the physics group validates these hit data which are then used as input to the subsequent digitization step, allowing for pile-up to be included. This step converts hits into digitizations which correspond to the output of the CMS electronics.

The digitization step, following the hit creation step, constitutes the simulation of the electronic readout used to acquire data by the detector. It starts from the hit positions and simulated energy losses in the detectors and produces an output that needs to be as close as possible to real data coming from CMS. Information from the generation stage (e.g. particle type and momentum) is preserved in the digitization step. The output of this step has the same format of real collision events, and therefore can be fed to the same reconstruction software chain.

Collision events are reconstructed and stored if they satisfy at least one of the High Level Trigger paths employed online. Depending on the type of HLT path which was fired, an event is stored in a given Primary Dataset, which will therefore collect events with similar topologies. Examples of Primary Datasets are Photon, DoubleMuon, DoubleElectron, SingleMuon.

2.3 Electron Reconstruction and Trigger

Electrons are reconstructed in the silicon tracker, where they form a track, and in the crystal ECAL, where they deposit their energy. The goal of electron reconstruction is therefore to successfully couple a track with an electromagnetic energy deposit, and efficiently identify these as an electron candidate, without allowing a large rate of fakes to be introduced by other charged particles, such as pions.

While traversing the tracker material, an electron not only ionizes the medium, as any charged particle, but may incur in a large energy loss via the radiation of a photon, a process commonly known as *bremsstrahlung*. As we have seen, the tracker material budget can be as large as 1.8 X_0, therefore this eventuality is not infrequent, and has to be accounted for. This is done by adapting both the track-finding algorithm and the calorimeter energy-clustering sequence.

Standard charged-track reconstruction is a pattern-recognition problem, in which the ensemble of hits released in the tracker have to be linked together in order to identify the sets of hits which arose from the passage of single charged particles. In CMS this is solved by the use of a Kalman Filter [16] algorithm, seeded in the pixel detector and in which the posterior on the particle's momentum is updated at each tracker layer. As this algorithm is optimized for particles which lose energy only via the ionization process, an indipendent track-finder is used for electrons, which has a similar functioning as the Kalman Filter, but also contemplates the possibility of a major, abrupt radiative energy loss: this algorithm is known as the Gaussian Sum Filter [17].

Bremsstrahlung not only affects the electron's trajectory, but also the shape of its ECAL energy deposit. Radiated photons will in fact give rise to indipendent satellite clusters, which must be collected in order to achieve a precise energy measurement. Because of the axial magnetic field, photons will mainly be radiated in the azimuthal direction. Therefore in the ECAL barrel the energy clustering algorithm proceeds as follows:

- search for single crystals which have collected an amount of energy above a certain threshold (1 GeV), and sort them in descreasing energy. These are the algorithm *seeds*;
- around each seed, open in the ϕ-direction a 5-crystal wide strip, reaching up to ± 17 crystals;
- add to the row which contains the seed crystal all 5-crystal rows which have a total energy larger than 0.1 GeV.

In the endcaps the algorithm is slightly different because of the geometry: it first organizes crystals in 5×5 matrices, and then groups those matrices which lie within an azimuthal distance of 0.3 rad.

The resulting sets of grouped crystals are called ECAL *superclusters*. Electron reconstruction is then simply a matter of linking a GSF track to an ECAL supercluster. This may be done both by seeding the algorithm in the calorimeter (optimal for energies larger than $\sim$50 GeV), or in the tracker (which recovers efficiency at low electron energies). Additional identification criteria, such as isolation, ECAL energy cluster shape, and track-cluster compatibility requirements are needed in order to minimize the infiltration of fake candidates. These requirements will be described in Sect. 4.2.2.

These sophisticated algorithms cannot be ran at trigger level, of course, for they would take too long. Electron triggers therefore proceed as follows:

- the L1 trigger unpacks the ECAL information in 5 $\times$ 5 crystal matrices (trigger towers);
- seeds are identified by trigger towers which pass a given transverse energy threshold;
- isolation requirements may be applied by looking at the energy of the neighboring ECAL trigger towers, and the HCAL cell directly behind the seed tower.

Level-1 seeds are passed to the High-Level trigger which further refines the electron identification process in three steps. The first step is to cluster the ECAL energy into a supercluster (as described above) and a transverse energy threshold is applied to the supercluster energy. The second step is to access the pixel information and seek hits compatible with the hypothesis that the supercluster belongs to an electron. If no hits are found the candidate is rejected. In the third and final step the full tracker information is exploited: tracks are used for isolation and the electron candidate track is matched in momentum and position to the ECAL supercluster. Details on the adopted trigger paths are given in Sect. 4.2.2.

2.4 Muon Reconstruction and Trigger

Muon tracks are reconstructed twice in the CMS detector: in the inner silicon tracker, and in the external muon chambers. Muon reconstruction starts in the muon spectrometer: the track segments which are formed in the drift tube and cathode strip chambers are linked together with a Kalman Filter algorithm, and a *standalone* muon track is formed. Given the large amount of material they have traversed in order to reach the spectrometer, the latter's resolution is degraded because of multiple scattering interactions which have modified the muon momentum. Therefore, for low and moderate transverse momenta the silicon tracker information, which measures the muons at their production in the presence of a strong magnetic field, is crucial in order to achieve high-resolution measurements.

Once the muon track is reconstructed in the spectometer, it is linked to its tracker track. This is done in two steps: first a subset of tracks is identified which roughly match the momentum and direction of the standalone track; then a more accurate tracker-spectrometer matching is performed, by considering a number of kinematic and angular variables. Figure 2.10 shows the expected muon resolution as a function of momentum when using the muon spectrometer only, the tracker only, or the full system. As can be seen for momenta of up to hundreds of GeV the tracker resolution is dominant.

The muon trigger has been designed to be redundant and efficient: all muon subdetectors are employed in the trigger logic. The DT and CSC Level-1 electronics process the information in each station, and identify stations in which hits are sought. A track finding algorithm then scans all hits and builds them into tracks, assigning them a transverse momentum. The four highest-p_{T} and best quality candidates from each subsystem are sent to the Global Muon Trigger (GMT).

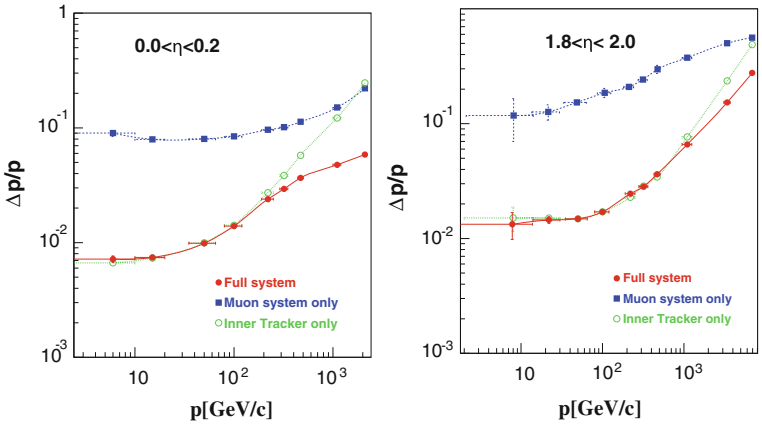

Fig. 2.10 Muon momentum resolution when using the muon spectrometers only (*squares*), the tracker only (*hollow circles*), and the full system (*solid circles*): barrel ($|\eta| < 0.2$) results are shown on the *left*, endcaps ($1.8 < |\eta| < 2.0$) on the *right*

Similarly for the RPC detectors, hits are scanned by a trigger logic and if they are aligned along a track a muon candidate is formed and assigned a transverse momentum. The candidates are ranked based on p_T and quality criteria, and the the best four in the barrel and the best four in the endcaps are kept and sent to the GMT.

The Global Muon Trigger then attempts to correlate DT and CSC candidates with the ones found in the RPC detectors. The calorimeter information is also accessed to determine the level of isolation of the muon candidate. The four best candidates are kept and passed on to the High-Level Trigger.

The muon High-Level Trigger can then impose additional selections on the candidates, such as transverse momenta and isolation requirements. This is done in two steps: the first step (Level-2) accesses the fine-grain DT and CSC information, and reconstructs the full track as seen in the muon spectrometer. The second step (Level-3) accesses the tracker information and reconstructs the complete global muon candidate.

References

1. ATLAS Collaboration.: The ATLAS experiment at the CERN large hadron collider. J. Instrum. **3**, S08003 (2008). Available from: http://stacks.iop.org/1748-0221/3/i=08/a=S08003
2. ALICE Collaboration.: The ALICE experiment at the CERN LHC. J. Instrum. **3**, S08002 (2008). Available from: http://stacks.iop.org/1748-0221/3/i=08/a=S08002
3. LHCb Collaboration.: The LHCb detector at the LHC. J. Instrum. **3**, S08005 (2008). Available from: http://stacks.iop.org/1748-0221/3/i=08/a=S08005
4. Achilli, A., Hegde, R., Godbole, R. M., Grau, A., Pancheri, G., Srivastava, Y.: Total cross-section and rapidity gap survival probability at the LHC through an eikonal with soft gluon resummation. Phys. Lett. B, **659**, 137 (2007). Comments: 15 pages, 3 figures, LaTeX

5. The CMS experiment at the CERN LHC.: J. Instrum. **3**, S08004 (2008). Available from: http://stacks.iop.org/1748-0221/3/i=08/a=S08004
6. CMS Collaboration.: The magnet project: Technical design report. CERN/LHCC 97–10, (1997)
7. CMS Collaboration.: The tracker project: Technical design Report. CERN/LHCC 98–10, (1998)
8. CMS Collaboration.: The electromagnetic calorimeter project: Technical design report. CERN/LHCC 97–33, (1997)
9. Adzic, P. et al.: Energy resolution of the barrel of the CMS Electromagnetic Calorimeter. J. Instrum. **2**, P04004 (2007). Available from: http://stacks.iop.org/1748-0221/2/i=04/a=P04004
10. CMS Collaboration.: The Hadron calorimeter project: Technical design report. CERN/LHCC 97–31, (1997)
11. CMS Collaboration.: The Muon project: Technical design report. CERN/LHCC 97–32, (1997)
12. CMS Collaboration.: The TriDAS project technical design report, vol. 1. CERN/LHCC 2000–38, (2000)
13. CMS Collaboration.: The TriDAS project technical design report, vol. 2. CERN/LHCC 2002–26, (2002)
14. Agostinelli, S. et al.: Geant 4—A Simulation Toolkit. Nucl. Inst. Meth. **A506**, 250 (2003) http://dx.doi.org/10.1016/S0168-9002(03), 01368–8
15. CMS Collaboration.: CMS physics technical design report, vol. II: Physics performance. oai:cds.cern.ch:942733. J. Phys. G. **34**, 995 (2006). Revised version submitted on 2006–09-22 17:44:47
16. Kalman, R.E.: A new approach to linear filtering and prediction problems. Trans. ASME-J. Basic Eng. **82**, 35 (1960)
17. Frühwirth, R., Speer, T.: A Gaussian-sum filter for vertex reconstruction. Nuclear instruments and methods in physics research section A: Accelerators, spectrometers, detectors and associated equipment, vol. 534, p. 217(2004). Proceedings of the IXth International Workshop on Advanced Computing and Analysis Techniques in Physics Research. Available from: http://www.sciencedirect.com/science/article/pii/S0168900204015311, http://dx.doi.org/10.1016/j.nima.2004.07.090

Chapter 3
Jet Reconstruction and Calibration

Abstract Jets represent a long withstanding challenge for both theoretical and experimental physics. Because of their composite nature, they necessitate of a clustering algorithm to be defined, and their interaction with the detector will vary significantly on a jet-by-jet basis, depending each jet's composition. The instrinsic difficulty in experimentally determining a jet's particle composition translates in an uncertainty on the measurement of the jet's energy scale, and therefore in the necessity of additional calibration procedures, specific for jets. In the first part of this chapter we will describe the jet energy scale problematic, and explain the related challenges from an experimental point of view. We will show how sophisticated jet reconstruction techniques, such as the CMS full event reconstruction known as the Particle Flow, may significantly improve jet reconstruction performance. Section 3.3 is dedicated to illustrating the jet calibration scheme employed in CMS, and showing the results on the measurement of the jet energy scale and resolution in proton-proton collisions. Finally, Sect. 3.5 demonstrates how detailed information on jet particle composition can give insight on the nature of the parton originating the jet.

3.1 Hadronization and Jets

A coloured energetic particle, such as a quark or a gluon expelled in a high-energy proton-proton collision, due to the intensity of the strong field potential is energetically incentivated into losing its colour charge in the formation of stable, colourless configurations. This is done by multiple radiations of gluons which excite the vacuum producing quark-antiquark pairs, and the quarks eventually combine themselves in the formation of mesons and baryons. This process is known as *hadronization*. Quadrimomentum conservation laws applied to the inital parton imply that the hadronization products will have a quasi collinear configuration, in the parton's direction: these sprays of particles are visible in modern detectors, and are commonly referred to as *jets*.

F. Pandolfi, *Search for the Standard Model Higgs Boson in the $H \to ZZ \to \ell^+\ell^- q\bar{q}$ Decay Channel at CMS*, Springer Theses, DOI: 10.1007/978-3-319-00903-2_3,

Hadronization energetically favours the production of light particles, such as pions and kaons. Empirically it is found that to reasonable approximation a jet's energy is on average composed in the following way:

- about 65 % of a jet's energy is carried by charged particles, predominantly charged pions and kaons;
- 20 % is converted into high-energy photons, mainly from the electromagnetic decay of neutral mesons such as π^0's and η's;
- the remaining 15 % is stored in long-lived neutral hadrons, mainly neutral kaons, neutrons and Λ baryons.

In addition, jets can register the presence of energetic neutrinos or charged leptons, in the case of semileptonic quark decays during hadronization, common in the case of jets originated from c or b quarks.

Therefore, on average, only 20 % of a jet's energy is purely electromagnetic in nature and will be efficiently measured in the ECAL. The remaining energy is carried by particles which will undergo a hadronic shower before they can be absorbed in the calorimeters.

The formation of a hadronic shower cascade is a complex process, a complete description of which is out of the scope of this thesis. A simple model [1] can nevertheless help in understanding the main ingredients which are relevant to jet energy scale calibration. In this model a particle h_0 which initiates a hadronic shower will evenly split its energy E in the creation of three pions:

$$h_0(E) \rightarrow \pi^+(E/3) + \pi^-(E/3) + \pi^0(E/3)$$

The newly formed neutral pion will decay to a photon pair ($\pi^0 \rightarrow \gamma\gamma$), whereas the charged ones will undergo a similar process, forming three new pions each, if their energy is larger than the pion production threshold. The process then iterates, until all remaining charged pions have insufficient energy to produce new particles.

As crude as this model may be, it captures the key points in hadronic shower detection:

- the shower energy is split into numerous particles, the number of which increases with the initial particle's energy;
- these particles can be divided into two broad classes: soft hadrons and high-energy photons;
- the fraction of the shower energy carried by photons (the shower electromagnetic fraction f_{em}) also increases with the initial particle's energy.

The third statement, which is the most relevant for what follows, is based on the fact that at every iteration, while energy can be moved from the hadronic to the electromagnetic sector, the inverse is highly improbable. Therefore, the larger the number of iterations, that is the larger the initial particle's energy, more energy will be stored in the form of photons.

It is immediate to show that in the simple model described above the dependance of f_{em} from the initial energy E is in the form of a power law:

$$f_{\mathrm{em}} = 1 - \left(\frac{E}{E_0}\right)^k$$

where E_0 is a scale factor and the exponent k turns out to be related to the average multiplicity and type of particles produced at each step [2]. Therefore, as the energy of the original particle increases, the electromagnetic fraction of its shower's energy increases too, asymptotically reaching unity in the limit of infinite energy.

When releasing their energy in a calibrated calorimeter, photons and hadrons will have different single-particle responses, which we may call respectively R_γ and R_h. A calorimeter is said to be *compensating* [1] if $R_h = R_\gamma$, and *non-compensating* if $R_h < R_\gamma$. The CMS calorimeters are strongly non-compensating: the CMS HCAL has been measured [3] to have $R_h/R_\gamma \approx 0.7$, and for the crystal ECAL lower values, of the order of 0.45–0.5, are assumed.

3.2 Jet Reconstruction

3.2.1 Response and Resolution

The aim of jet reconstruction is to measure the momentum of the coloured parton which initiated the hadronization process. In order to do so, final state particles, visible in the detector, have to be grouped together, through the choice of an appropriate jet algorithm, as will be shown in the following section. The same algorithm will then be applied to reconstructed objects and to generator-level particles, giving rise respectively to reconstructed and generator jets. Each reconstructed jet is then matched to its corresponding generator jet topologically, by choosing the closest generator jet on the $\eta - \phi$ plane.

Two variables are commonly employed to measure jet reconstruction performance: the jet response and resolution. The response *variable* is defined on a jet-by-jet basis as the ratio between the transverse momentum of the reconstructed jet and that of its matched generator jet:

$$R = \frac{p_{\mathrm{T}}^{\mathrm{reco}}}{p_{\mathrm{T}}^{\mathrm{gen}}} \tag{3.1}$$

This is defined only at Monte-Carlo level, as it accesses the generator information. The average value of this variable, $\langle R \rangle$, is an estimator of the *response* of a given jet reconstruction strategy. The jet *resolution*, instead, is usually defined as the width of the R variable distribution, divided by the response.

Throughout this Chapter, response and resolution are defined by truncating the R variable distributions, in order to minimize the effects of rare outliers. The truncation is a two-step procedure: first the mode of the distribution is found through an iterative gaussian fit, extended only to ± 1.5 standard deviations about the gaussian mean;

once the bin in which the mode is included is found, bins are iteratively added, symmetrically about the mode bin, until 99 % of the histogram's integral is reached. The response estimator is then defined as the average of this truncated distribution, and the resolution as its RMS, divided by the mean.

Traditional jet reconstruction strategies rely on calorimetric information only. This is justified by the fact that, for hard jets, most of the hadronization particles will have large enough momenta such that most of them will hit the detector calorimeters in proximity of each other, even in the presence of a magnetic field. The latter will deflect far away from the jet core only the soft charged particles, with a relatively small effect on the overall jet reconstruction performance.

So, for calorimeter-based jet reconstruction, we can conclude the following. We have seen in the previous section that a jet's f_{em} follows a power law as a function of the jet energy. This implies that the response of calorimeter jets will also follow a power law, with asymptotically unitary response reached at high jet energies. Furthermore, as for every calorimeter measurement, calorimeter jet resolution will improve with increasing jet energy, with infinitely accurate resolution asympotically reached at high energies.

3.2.2 Jet Algorithms

Jets are composite objects, therefore necessitate of a clustering algorithm in order to be univocally defined. In order to allow a meaningful comparison between the measurement of a physical observable and its correponding theoretical prediction, the same algorithm must be used in the two cases. The inputs, of course, will differ: reconstructed elements in the first case, final state theoretical particles in the latter. For simplicity, these inputs will generically be called 'particles' throughout this section.

Jet algorithms should be efficient in clustering particles produced by the hadronization of a parton, so that the latter's momentum can be *a posteriori* inferred by adding up the momenta of the clustered particles. Furthermore, it must satisfy two requirements in order to provide finite theoretical predictions at all orders of perturbation theory:

- *collinear safety*: if a particle of momentum p is substituted by two collinear particles of momentum $p/2$, the result of the clustering sequence must not be affected;
- *infrared safety*: if an infinitely soft gluon is added to the list of particles which have to be clustered, the result of the clustering sequence must not change.

The *anti-*k_{T} algorithm [4] is the main jet algorithm employed at CMS. It is both infrared- and collinear-safe, and proceeds as follows:

- define the distance d_{ij} between two input particles i and j as

$$d_{ij} = \min\left(\frac{1}{p^2_{\mathrm{Ti}}}, \frac{1}{p^2_{\mathrm{Tj}}}\right)\frac{\Delta R^2_{ij}}{R^2}$$

where $p_{\mathrm{Ti,j}}$ are respectively the two particles' transverse momenta, $\Delta R_{ij} = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ is the euclidean distance between them on the $\eta - \phi$ plane, and R is the algorithm's radius parameter. Define furthermore the distance between any particle i and the beam as:

$$d_{iB} = \frac{1}{p_{\mathrm{Ti}}^2}$$

- find the minimum of all d_{ij} and d_{iB};
- if the minimum is a d_{iB}, remove particle i from the list and call it a 'jet'; if it's a d_{ij}, recombine particles i and j into a new particle by summing their quadrimomenta;
- iterate until only jets are left.

It can easily be seen that the algorithm definition does not allow jets to contain particles at distances greater than R from their central axis, therefore giving rise to cone-shaped jets. This feature is valuable from an experimentalist's point of view, because it facilitates the mapping between a jet direction and the region of the detector which is actively interested in that jet's reconstruction.

All measurements with jets presented in this thesis define jets through the *anti-k_{T}* algorithm, with radius parameter R set to 0.5. The algorithm is interfaced to the CMS software framework through the FastJet [5] package. At generator level, the list of final state particles produced in the hadronization process constitute the list of objects that will be clustered; at reconstruction level a list of particle *candidates* is passed to the algorithm, produced by the full event reconstruction technique known as the CMS Particle Flow, which we will now illustrate.

3.2.3 Particle Flow Reconstruction

The Particle Flow [6] is a full event reconstruction technique which aims to reconstruct all stable particles produced in a given proton-proton collision. To do so it exploits all CMS subdetectors to their full granularity and correlates information between them in order to optimize particle reconstruction and identification performance.

The design of the CMS detector proves to be well-suited for this type of event reconstruction: its large silicon tracker and the 3.8 T magnetic field in which it is immersed allow precise and efficient charged particle detection for transverse momenta as low as 150 MeV, and its crystal electromagnetic calorimeter allows excellent resolution in the measurement of photon and electron energies. As we have seen charged particles and photons make up on average about 85 % of a jet's energy, so only 15 % of it will be reconstructed in the hadronic calorimeter alone.

The Particle Flow algorithm first collects reconstructed hits in each subdetector independently and creates a list of basic reconstructed elements (*blocks*), namely charged tracks in the tracker, clusters of energy deposits in the calorimeters. Once blocks are formed, a link algorithm connects blocks which are topologically compat-

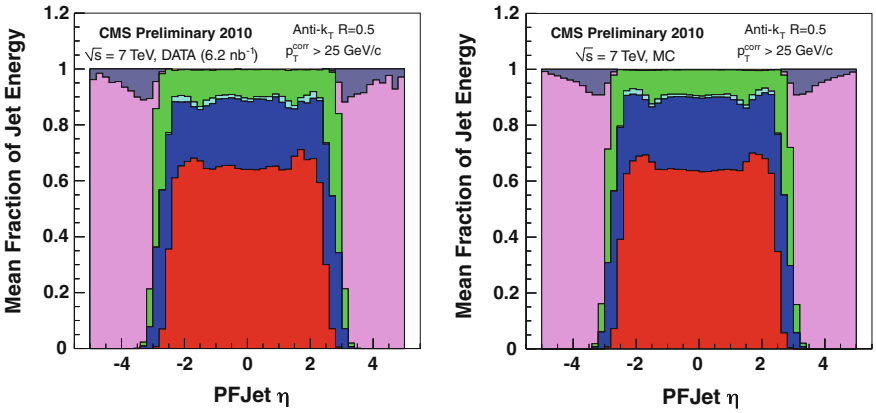

Fig. 3.1 Reconstructed jet energy fractions as a function of jet pseudorapidity. In the central region, *bottom to top*: charged hadrons (*red*), photons (*blue*), electrons (*cyan*), neutral hadrons (*green*). In the forward region: HF hadrons (*pink*), HF electromagnetic particles (*grey*)

ible, giving way to particle flow particle candidates (PFCandidates). PFCandidates may be of seven different types, depending on the type of blocks involved in their reconstruction:

- **electrons** arise from the link between a charged track and one or more ECAL clusters, provided an electron identification set of criteria is satisfied;
- charged tracks linked to any number of calorimeter (ECAL or HCAL) clusters, and which are not identified as electrons, are reconstructed as **charged hadron** candidates;
- ECAL energy deposits not compatible with charged tracks give way to **photon** candidates;
- unaccounted HCAL deposits are interpreted as **neutral hadron** candidates;
- energy deposits in the HF calorimeters are reconstructed as **HF hadronic** or **electromagnetic** particle candidates, depending on the depth at which the energy is released in the HF quartz fibres.

The formation of the PFCandidate list represents the Particle Flow interpretation of a given proton-proton collision in CMS, as it attempts to mirror the true particle composition of the event to the best of our knowledge. Particle flow jet reconstruction (PFJets) is then just a matter of choosing the jet algorithm with which the PFCandidates are to be clustered.

As PFJets are composed of PFCandidates, understanding their PFCandidate composition may give additional insight on their performance. Figure 3.1 shows the reconstructed jet energy fractions, as a function of the jet pseudorapidity, on the left for 6.2 nb^{-1} of 7 TeV data, on the right for the MC simulation. As can be seen, in the central region about 65 % of the jet energy is carried by charged hadrons (red), 20 % by photons (blue), about 2 % by electrons (cyan) and the remainder by neutral hadrons (green). Charged hadrons (and electrons) are reconstructed only in the

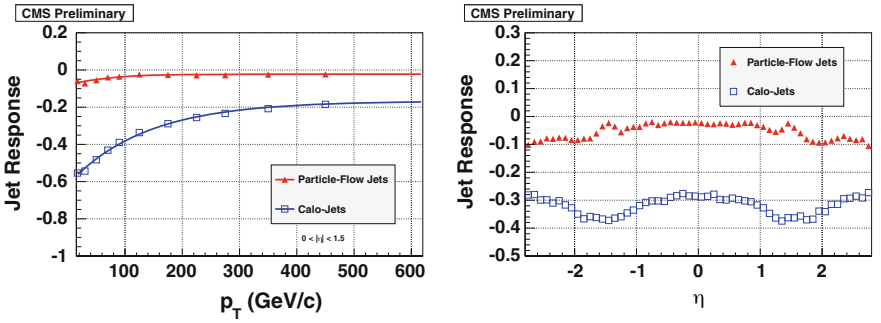

Fig. 3.2 Calorimeter (*squares*) and PFJet (*triangles*) response as a function of jet transverse momentum (*left*) and pseudorapidity (*right*). Results are based on Monte Carlo studies

tracker-covered region (up to $|\eta| = 2.5$), photons and neutral hadrons up to $|\eta| = 3$, which is the pseudorapidity coverage of the central CMS calorimeters. In the forward region, most of the energy is stored in the form of HF hadrons (pink), whereas HF electromagnetic particles (grey) contribute to less than 10 %.

Particle Flow significantly improves jet reconstruction performance at CMS, compared to traditional, calorimeter-based approaches. This is shown in Fig. 3.2, where the MC response[1] of calorimeter and PFJets are compared on the left as a function of transverse momentum, and on the right as a function of jet pseudorapidity. As can be seen, Particle Flow jets ensure a much higher response throughout the detector. Figure 3.3, instead, compares the two jet reconstruction schemes' jet resolutions as a function of jet transverse momentum, for jets reconstructed in the CMS barrel. Here calorimeter jets have been corrected with the full sequence of jet energy corrections, as described in Sect. 3.3. Even if the two approaches tend to reach similar performance in terms of resolution, the improvement offered by Particle Flow is significant, especially at moderate and low transverse momenta.

The non-uniformity of both trends shown in Fig. 3.2 makes the introduction of jet energy corrections necessary, as will be described in the following section.

3.3 Jet Calibration: The Factorized Approach

The jet energy correction scheme employed at CMS [7] is based on a factorized approach. It defines the different corrections in such a way that they address different physical aspects. They are therefore considered to be independent, so that they can be applied as a series of multiplicative factors.

The levels of corrections defined in CMS are:

[1] Actually, based on our definition of response, the Figure shows the trend of (1-response).

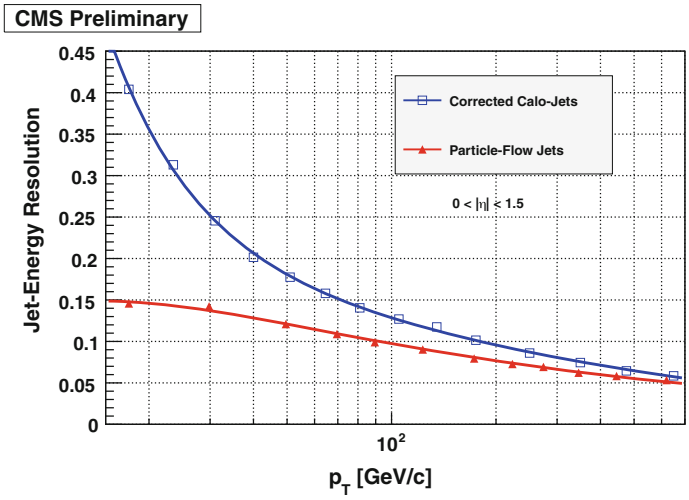

Fig. 3.3 Calorimeter (*squares*) and PFJet (*triangles*) resolution as a function of jet transverse momentum for jets reconstructed in the CMS barrel. Results are based on Monte Carlo studies

- **Level 1** or **offset** correction, which corrects jets for the effect of overlapping diproton collisions (pile-up);
- **Level 2** or **relative** correction, which minimizes the effect of non-uniformities between different CMS subdetectors;
- **Level 3** or **absolute** correction, which addresses the fundamental non-compensating nature of the CMS calorimetric system.

These three levels of corrections are considered mandatory for jets to be used as physics objects at analysis level. Additional, facultative levels of corrections (such as parton-flavour related corrections) have also been developed, but their usage is considered analysis-dependent.

The strategy adopted by CMS is to derive the values of these corrections on the full-detector software simulation and apply them on the data. The data is then used to test their effectiveness, and if a significant non-closure is found, an additional (*residual*) correction is introduced. This approach allows to minimize the effect of statistical fluctuations deriving from insufficient events in the data samples.

The aim of the offset correction is to subtract the additional energy which is irradiated inside a jet cone by secondary proton-proton collisions, on an event-by-event basis. In order to do so, a novel technique [8], based on the FastJet algorithm, is employed. This algorithm re-clusters the PFCandidates with the k_T jet algorithm [9], after having introduced a uniform flow of infinitely soft 'ghost' particles in the event. The resulting list of pseudo-jets which are created are ordered in increasing energy, and the median of the distribution is taken as an estimate of the average energy flow per unit area of the event. This energy density is then multiplied by the jet 'active area' (equal to πR^2 in the case of *anti-*k_T jets) and the result is subtracted from

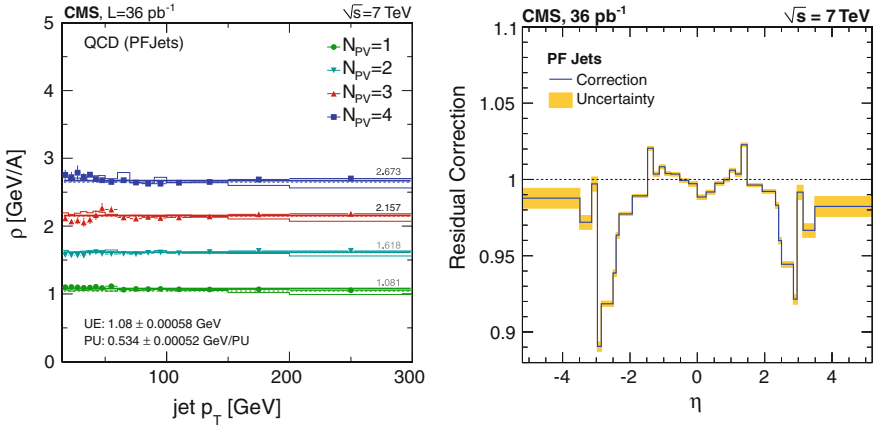

Fig. 3.4 *Left* average value of the Particle Flow energy density as a function of the transverse momentum of the leading jet in QCD multijet events. *Right* residual relative correction as a function of the probe jet pseudorapidity

the energy of the jet. It must be noted that this approach corrects the jet both from the effect of pile-up and from the contribution of the underlying event activity. The average value of this energy density, as a function of the transverse momentum of the leading jet in QCD multijet events is shown in Fig. 3.4 (left).

The relative correction uniforms the detector response to jets across its pseudorapidity range. The response at the center of the barrel is taken as reference. The correction is obtained on dijet events, in which one jet is required to be in the barrel ($|\eta| < 1.3$) and the response of the second (*probe*) jet, relative to the barrel jet, is studied as a function of the probe's pseudorapidity. The results of this measurement are shown in Fig. 3.4, where the value of the residual correction, as measured on 36 pb^{-1} of data collected by CMS, is shown as a function of the probe jet pseudorapidity, for Particle Flow jets.

Once the response to jets is uniform across the detector, the absolute jet energy correction factor, the *jet energy scale*, must be measured. This is done with photon+jet events, and will be described in the following section.

3.4 Jet Energy Scale Measurement

The absolute jet energy correction has the aim of addressing the non-compensating nature of the CMS calorimetric system, and will therefore uniform the detector response as a function of the jet transverse momentum. The extraction of this factor represents the actual measurement of a detector's jet energy scale.

The measurement of the jet energy scale at CMS is done with photon+jet events, with a technique first introduced at Tevatron experiments [10]. Their dominant

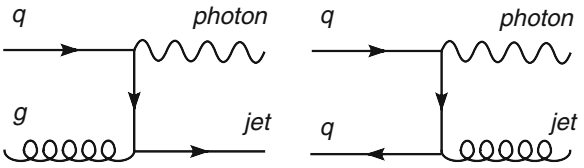

Fig. 3.5 Dominant photon+jet production diagrams at a proton-proton collider

production diagrams at a proton-proton collider are shown in Fig. 3.5. At leading order, in these events the photon and the leading jet are balanced in the transverse plane, hence the precision with which the photon is measured in the crystal ECAL can be exploited to infer the true jet transverse momentum.

The analyzed datasets are listed in Sects. 3.4.1, 3.4.2 describes the adopted photon identification criteria, which constitute the main means of event selection and background discrimination. The subsequent sections will then show the results.

3.4.1 Data Samples and Trigger

This measurement makes use of the first 1 fb^{-1} of data recorded by the CMS detector during the 2011 data taking. Signal events are stored in the `Photon` Primary Dataset after firing the single photon high level triggers. These triggers require the presence of an energy deposit in the ECAL, to which a transverse momentum requirement is applied. The names of the analyzed datasets, together with their corresponding integrated luminosities, are reported in Table 3.1.

As the accelerator's instantaneous luminosity grew, though, the level of prescales introduced in the lower transverse momentum paths increased. The presence of different prescale levels in neighboring transverse momentum ranges can create biases in the response estimation, as migrations from higher-p_{T}/less prescaled trigger paths can pollute lower-p_{T} events. In order to avoid these biases, an explicit requirement of the prescaled triggers has been introduced in the data only, as explained in Table 3.2.

Table 3.1 Analyzed data for the photon+jet analysis

Dataset	Run range	Luminosity (fb^{-1})
May 10 ReReco	160329–163869	0.2
Prompt reconstruction (v4)	165071–168437	0.8

The data are divided in two run ranges, and each is associated to its corresponding integrated luminosity

Table 3.2 Summary of HLT requirements in the photon+jet analysis selection

Photon candidate p_T range (GeV)	Required HLT path
$15 \div 22$	`HLT_Photon15_L1R`
$22 \div 32$	`HLT_Photon20_L1R`
$32 \div 53$	`HLT_Photon30_L1R`
$53 \div 80$	`HLT_Photon50_L1R`
$80 \div 150$	`HLT_Photon75_L1R`
> 150	no requirement

The number in the HLT path name indicates the transverse momentum requirement applied to the photon candidate at trigger level. These requirements are applied to the data only

Table 3.3 Photon+jet monte carlo samples

Physical Process	σ (pb)	Events	Luminosity (fb^{-1})
γ + jet, $15 < \hat{p}_T < 30$ GeV	1.717×10^5	2M	0.012
γ + jet, $30 < \hat{p}_T < 50$ GeV	1.669×10^4	2M	0.12
γ + jet, $50 < \hat{p}_T < 80$ GeV	2.722×10^3	2M	0.73
γ + jet, $80 < \hat{p}_T < 120$ GeV	4.472×10^2	2M	4.5
γ + jet, $120 < \hat{p}_T < 170$ GeV	8.417×10	2M	24
γ + jet, $170 < \hat{p}_T < 300$ GeV	2.264×10	2M	88
γ + jet, $300 < \hat{p}_T < 470$ GeV	1.493	2M	1.3×10^3
γ + jet, $470 < \hat{p}_T < 800$ GeV	1.323×10^{-1}	2M	1.5×10^4
γ + jet, $800 < \hat{p}_T < 1400$ GeV	3.481×10^{-3}	2M	5.7×10^5
γ + jet, $1400 < \hat{p}_T < 1800$ GeV	1.270×10^{-5}	2M	1.5×10^8

For each sample, its $\hat{p}_T$ interval, cross section (σ), number of analyzed events, and equivalent luminosity are given

The data is compared to Monte Carlo events generated with `PYTHIA 6` [11] and passed through a full simulation of the CMS detector, implemented in the `GEANT 4` software framework. The analyzed samples are summarized in Table 3.3. The Monte Carlo events have been reweighed in order to match the amount of pileup observed in the data (more details given in Sect. 4.1.2).

3.4.2 Photon Identification

The main background to photon+jet events is represented by QCD dijet events, in which a jet is misidentified as a photon. This could result for instance from the electromagnetic decay of an energetic π^0 or η produced during hadronization. The cross section of QCD dijet events can be 10^5 times larger than the cross section of photon+jet events, therefore to keep this background to reasonably low levels a strict set of photon identification criteria is necessary.

Two are the main handles in discriminating photons produced in jets: requirements made on the candidate's isolation and on the shape of the ECAL energy cluster.

Neutral pions which originate from hadronization are produced in association with a number of additional particles, whereas prompt photons, at tree level, are relatively well isolated. We therefore expect to register significantly more activity in the detector around the photon candidate in background events. Neutral pions (as well as η's), furthermore, decay to a *pair* of photons. Hence, unless the relativistic boost is so high that the decay products are quasi-collinear, two adjacent showers are formed in ECAL, so that the resulting energy cluster will show an elongation along the decay axis.

We therefore define the following set of photon identification requirements:

- **track isolation**: the scalar sum of the transverse momenta of all tracks reconstructed at $\Delta R < 0.35$ from the photon candidate is required to be less than 10 % of the photon transverse momentum; the total nuber of reconstructed tracks in the cone is further required to be less than three;
- **ECAL isolation**: the total reconstructed ECAL energy in a hollow cone $0.05 < \Delta R < 0.4$ around the photon direction, excluding a 5-crystal wide η-strip, is required to be less than 5 % of the photon energy, or less than 3 GeV;
- **HCAL isolation**: the total energy recorded by the hadronic calorimeter in a $\Delta R < 0.4$ cone around the photon is required to be less than 5 % of the photon energy, or less than 2.4 GeV;
- **cluster major axis**: the second moment of the energy distribution of the photon seed basic cluster in the direction of the cluster major axis is required to be less than 0.35;
- **cluster minor axis**: the second moment of the energy distribution of the photon seed basic cluster in the direction of the cluster minor axis is required to be less than 0.3.

The event selection further requires the photon to be in the ECAL barrel fiducial region ($|\eta| < 1.3$), and to have transverse momentum greater than 15 GeV.

The expected sample purity after these requirements is expected to be of the order of 90 % for photon transverse momenta greater than 100 GeV, and somewhat worse for lower transverse momenta. The bias introduced by the background contamination is expected to play a minor role: QCD events which pass the selection will present a parton which has hadronized mainly into one (or more) electromagnetic-decaying particles, so these events are very similar to true photon+jet events for practical purposes.

3.4.3 Photon-Jet Balancing

A schematic view, in the transverse plane, of a photon+jet event is shown in Fig. 3.6. In the absence of additional event radiation, transverse plane balancing is true only at *parton* level, between the parton of the leading jet and the recoiling photon. The QCD parton will then undergo hadronization, and its products will be clustered with the chosen jet algorithm. The maximum size of the jet is fixed by the algorithm,

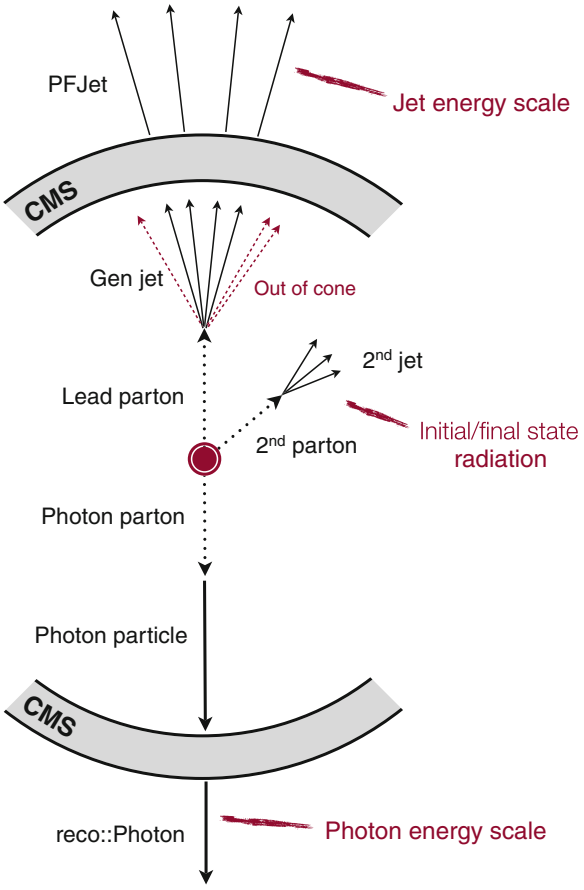

Fig. 3.6 Schematic view, in the transverse plane, of a photon+jet event. Refer to the text for details

so a number of particles will be inevitably left unclustered. This phenomenon is particularly acute at low transverse momenta, where the parton is less boosted and therefore the resulting hadrons less collimated, and for jets initiated by gluon partons, which have higher average particle multiplicities than quark jets.

This brings to the conclusion that at *reconstruction* level, due to hadronization, the photon and the jet are not exactly balanced, also in the case of no secondary event activity. Nevertheless, if we define the reconstructed balancing response estimate as the ratio between the jet and the photon transverse momenta:

$$R_{\text{balancing}} = \frac{p_{\text{T}}^{\text{recoJet}}}{p_{\text{T}}^{\gamma}}$$

it is always possibile to factorize it in the following manner:

$$R_{\text{balancing}} = \frac{p_{\text{T}}^{\text{recoJet}}}{p_{\text{T}}^{\gamma}} = \frac{p_{\text{T}}^{\text{recoJet}}}{p_{\text{T}}^{\text{genJet}}} \cdot \frac{p_{\text{T}}^{\text{genJet}}}{p_{\text{T}}^{\gamma}} \tag{3.2}$$

where we have introduced the transverse momentum of the generator jet matched to the reconstructed jet.

The new expression presents two factors. By comparing to Eq. 3.1 one can easily recognize the true response variable in the first ratio. We will define this ratio as the *intrinsic* response, and it depends on the chosen jet reconstruction scheme and on the jet transverse momentum. It is the object of the jet energy scale measurement.

The second ratio, on the other hand:

$$\frac{p_{\text{T}}^{\text{genJet}}}{p_{\text{T}}^{\gamma}}$$

is a measure of the imbalance at generator level between the photon and the leading jet. It depends on the amount of additional event activity, and on the efficiency of the chosen jet algorithm. We will call it generically *imbalance*.

Imbalance is the main source of bias in estimating the jet energy scale with photon+jet balancing. In order to reduce its effects a requirement on the transverse momentum of the subleading jet is introduced:

$$p_{\text{T}}^{\text{2ndJet}} < \max\left(0.1 \cdot p_{\text{T}}^{\gamma},\ 5\ \text{GeV}\right) \tag{3.3}$$

As will be seen, the requirement on the second jet p_{T} does not eliminate all of the bias. In order to do so, more sophisticated approaches are needed. Two methods have been devised at CMS to minimize the bias originating from imbalance: the Missing-E_{T} Projection Fraction, and the balancing extrapolation.

3.4.4 Missing-E_{T} Projection Fraction Method

The Missing-E_{T} Projection Fraction (MPF) method was first employed at the D0 detector, and, as it makes use of the event reconstruction as a whole, turns out to be particularly well-suited for Particle Flow reconstruction. It stems from the basic assumption that at generator level the vectorial sum of the transverse momenta of the photon and the full hadronic recoil will cancel each other on a per-event basis:

$$\vec{p}_{\text{T}}^{\,\gamma,\text{MC}} + \vec{p}_{\text{T}}^{\,\text{recoil}} = \vec{0}$$

When folding in the detector finite responses and resolutions, we obtain:

$$R_\gamma \vec{p}_{\mathrm{T}}^{\,\gamma,\mathrm{MC}} + R_{\mathrm{recoil}}\, \vec{p}_{\mathrm{T}}^{\,\mathrm{recoil}} = -\vec{E}_{\mathrm{T}}^{\,\mathrm{miss}}$$

where R_γ and R_{recoil} denote respectively the detector response to the photon and the recoil, and $\vec{E}_{\mathrm{T}}^{\,\mathrm{miss}}$ is the event missing transverse energy. Solving for $R_{\mathrm{recoil}}/R_\gamma$ and defining $\vec{p}_{\mathrm{T}}^{\,\gamma,\mathrm{reco}} \equiv R_\gamma \vec{p}_{\mathrm{T}}^{\,\gamma,\mathrm{MC}}$ yields:

$$R_{\mathrm{recoil}}/R_\gamma = 1 + \frac{\vec{E}_{\mathrm{T}}^{\,\mathrm{miss}} \cdot \vec{p}_{\mathrm{T}}^{\,\gamma,\mathrm{reco}}}{|\vec{p}_{\mathrm{T}}^{\,\gamma,\mathrm{reco}}|^2} \equiv R_{\mathrm{MPF}}$$

which defines the MPF response variable.

As it considers the hadronic recoil as a whole, the MPF response variable proves to be robust, showing very low sensitivity to additional event activity and pile-up. It further is an unbiased estimator of the jet response, as long as most of the recoil energy is carried by the leading jet in the event. This condition is fulfilled with a simple cut on the subleading jet transverse momentum, such as the one presented in Eq. 3.3.

Figures 3.7 show data-MC comparisons for balancing (left column) and MPF (right column) response distributions, for *anti-*k_{T} 0.5 PFJets reconstructed in the CMS barrel ($|\eta| < 1.3$), in three representative photon transverse momentum ranges. It can be noted that the MPF estimator has both a higher average and a narrower width than the balancing variable. In each photon transverse momentum bin the response estimate can be derived with the truncated mean procedure described in Sect. 3.2.1, and the result, as a function of the photon p_{T}, is shown in Fig. 3.8: the left graph shows the trend of the response estimates, for data and MC, whereas data-MC ratios are shown in the right graph. As can be seen, the simple balancing estimate presents a visible bias in the measurement of the jet response for $p_{\mathrm{T}} < 80$ GeV.

The total systematic uncertainties affecting the MPF measurement are summarized in Fig. 3.9. The total uncertainty is shown with a grey shade and is computed as the quadrature sum of all single components, namely:

- uncertainty of the **MPF method** per se (yellow band), is estimated by studying the method's sensitivity to a number of effects, including background infiltration, secondary jet activity and hadronization;
- the uncertainty on the **photon scale** is 0.9 % in the center of the barrel, according to recent calibration measurements [12];
- the uncertainty on the **extrapolation** is computed by taking the differences between the `Pythia` and `Herwig++` [13] generators; its effect is largest at high-p_{T}, where the data points are scarse;
- the **offset** uncertainty takes into account the effect of pile up, and is estimated by studying the sensitivity of the measurement to different pile up regimes;
- **residuals** are considered as an uncertainty in order to take into account imperfections in the original Monte Carlo truth jet energy calibration;
- the uncertainty related to possible mis-modelings of the **jet flavour** population in signal events is evaluated by studying the response differences between quark and gluon jets.

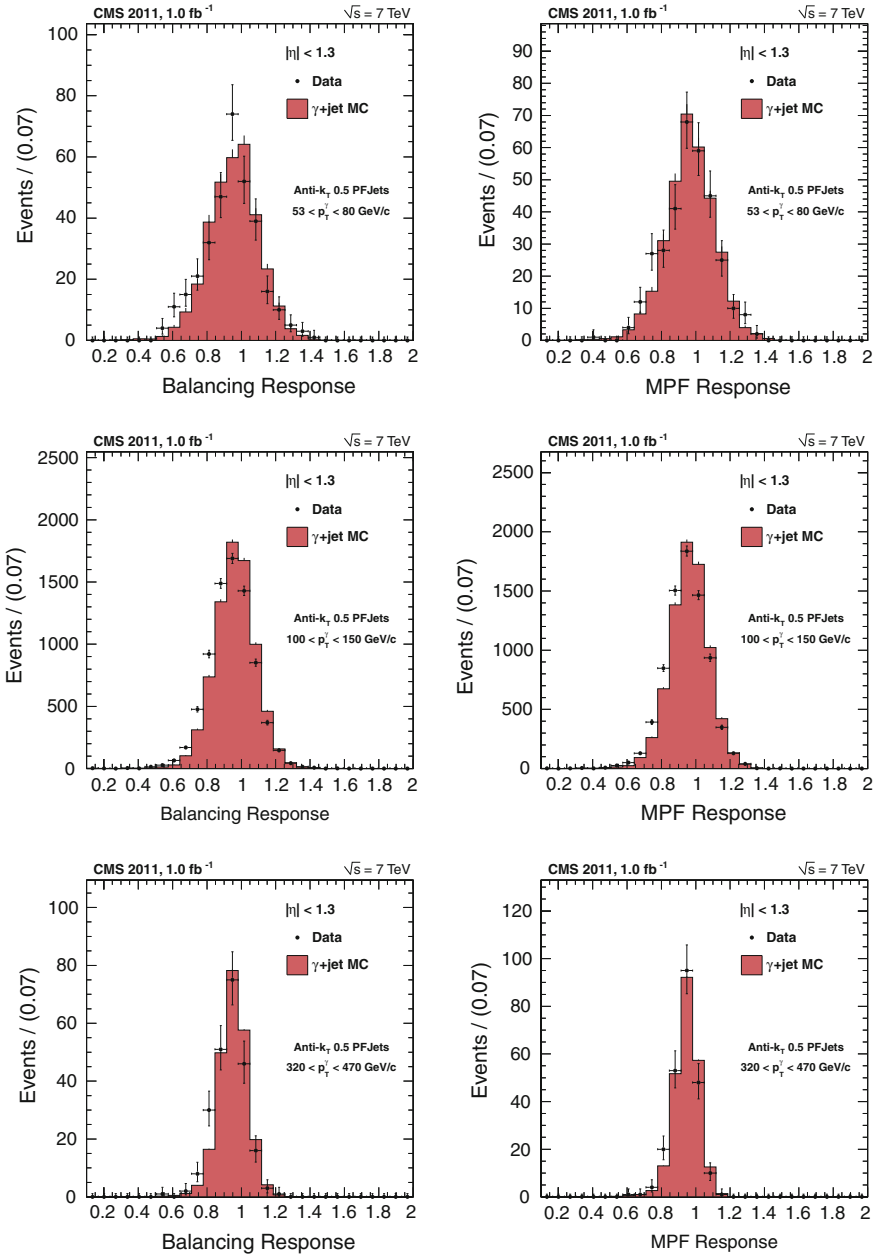

Fig. 3.7 Balancing (*left*) and MPF (*right*) response distributions in 1.0 fb^{-1} of 2011 data, in three representative transverse momentum ranges, for *anti-*k_{T} 0.5 PFJets reconstructed in the CMS barrel. The MC distributions are normalized to the integral of the data histograms

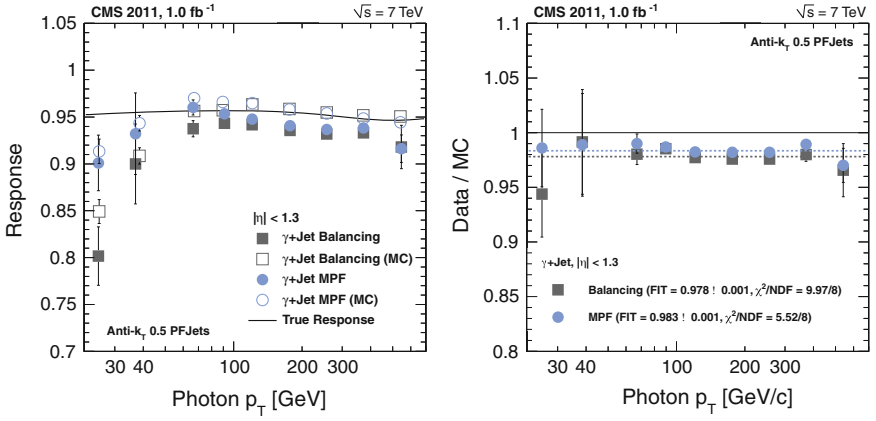

Fig. 3.8 Measurement of the response of *anti-k_T* 0.5 PFJets in the CMS barrel ($|\eta| < 1.3$). *Left*: response as a function of photon transverse momentum for the balancing (*grey squares*) and MPF (*blue circles*) methods, in 1.0 fb^{-1} of data (*solid*) and in the MC simulation (*hollow*). A comparison to the true response (*black line*) is also shown. *Right*: data/MC ratios

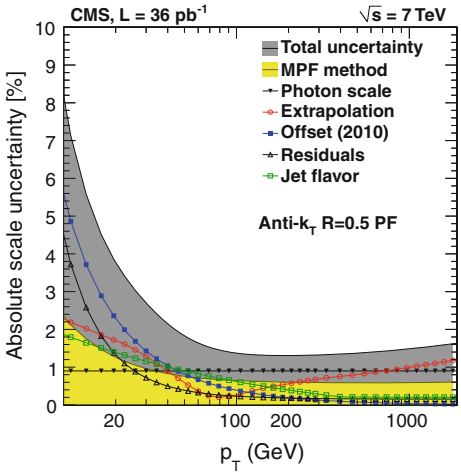

Fig. 3.9 Systematic uncertainties affecting the MPF method for *anti-k_T* 0.5 PFJets as a function of jet transverse momentum. Single contributions are marked separately, and the total, taken as the quadrature sum of all components, is shown as a *grey* shade

As can be seen in the Figure, the total uncertainty which derives from all of these contributions adds up to less than 5 % for jets with transverse momenta greater than 20 GeV, and less than 2 % for jets with $p_T > 50$ GeV. It must be specified that these uncertainty estimates were derived only on the first 36 pb^{-1} of data recorded by the CMS detector, therefore some of them are likely to drop when computed on larger datasets, once the level of comprehension of the detector has increased.

3.4.5 Balancing Extrapolation Method

The second method which minimizes the imbalance bias is the balancing extrapolation. This method is still based on a simple balancing between the leading jet and the photon, but instead of reducing the effect of additional event activity by imposing a requirement on the subleading jet, it studies the trend of the response as a function of the subleading jet's transverse momentum. The trend is then extrapolated to the ideal case of no secondary jet activity, with photon and leading jet perfectly balanced in the transverse plane.

The balancing extrapolation method is particularly suited for measuring the *corrected* jet response, i.e. the effectiveness of the MC-derived corrections on the data. This due to the fact that studying the trend of the response for very low values of the subleading jet p_{T}, the latter must be corrected in order for the measuremente to have physical sense. Furthermore, the MPF method may not be employed because it is highly non trivial to define a correction for the full hadronic recoil, especially for the fraction of it which is not clustered in jets. Therefore all results presented in this section will consider the corrected jet response, after the full chain of corrections (L1+L2+L3) derived from the simulation.

In a given photon transverse momentum range, recalling expression 3.2, which we may rewrite as $R_{\mathrm{balancing}} = R_{\mathrm{intr}} \cdot R_{\mathrm{imb}}$, we expect:

- the intrinsic response R_{intr} to be independent of the subleading jet (as long as it is 'reasonably' small), as it concerns only the leading jet;
- the imbalance R_{imb} to have a strong dependance on the subleading jet.

Our assumption is that these two effects are not correlated, so that they factorize, and therefore the response and resolution will have simple expressions:

$$\begin{aligned} \langle R_{\mathrm{balancing}} \rangle &= \langle R_{\mathrm{intr}} \rangle \cdot \langle R_{\mathrm{imb}} \rangle \\ \sigma_{\mathrm{balancing}} &= \sigma_{\mathrm{intr}} \oplus \sigma_{\mathrm{imb}} \end{aligned} \tag{3.4}$$

where we have used the symbols $\langle R \rangle$ and σ to indicate respectively response and resolution.

For what concerns the response, empirically we find that the functional dependance of R_{imb} on the subleading jet p_{T} is of quadratic form. Therefore, in a given photon p_{T} bin we will have:

$$\langle R_{\mathrm{intr}} \rangle (p_{\mathrm{T}}^{\mathrm{2ndJet}}) = c \qquad \langle R_{\mathrm{imb}} \rangle (p_{\mathrm{T}}^{\mathrm{2ndJet}}) = 1 - q - m(p_{\mathrm{T}}^{\mathrm{2ndJet}})^2 \qquad c, q, m = const$$

$$\Rightarrow \langle R_{\mathrm{balancing}} \rangle (p_{\mathrm{T}}^{\mathrm{2ndJet}}) = c \cdot \left[1 - q - m(p_{\mathrm{T}}^{\mathrm{2ndJet}})^2 \right] \tag{3.5}$$

therefore c is the object of this measurement, m describes the dependance of the imbalance on the subleading jet, and q quantifies the amount of irreducible imbalance between the photon and the leading jet. The values assumed by q in the simulation

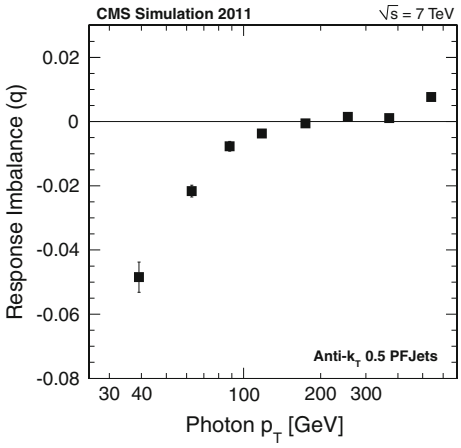

Fig. 3.10 Imbalance between the generator jet and the reconstructed photon, as a function of the latter's transverse momentum, in simulated photon+jet events

are shown in Fig. 3.10: as can be seen it is found to be negative and as large as $-5\,\%$ at low transverse momenta (dominated by jet algorithm inefficiencies), positive and of the order of $+1\,\%$ at very high transverse momenta (dominated by photon energy scale effects).

The method's operation is shown in Fig. 3.11, where the trends of the different contributions are shown as a function of the relative subleading jet transverse momentum ($p_{\mathrm{T}}^{\mathrm{2ndJet}}/p_{\mathrm{T}}^{\gamma}$), in four representative p_{T}^{γ} ranges. In each graph, the intrinsic response (blue squares) and the imbalance (black triangles) can be seen, together with their fit functions. The product of these two functions is shown with a grey line, and, if the made assumptions are correct, should constitute the predicted trend for the pseudo data points (open red markers). The observed good agreement between the two is a confirmation of the validity of the method on the simulation.

The measured trends in the data are also shown in each graph with solid red markers. The effect of the irreducible imbalance cannot be measured on data but must be accounted for, therefore the function used in the fit to the data has the functional form defined in Eq. 3.5, but with the q parameter fixed to the value obtained on the simulation.

The measured corrected response as a function of photon transverse momentum are shown in Fig. 3.12. The left plot shows the extrapolated response values, in the data and in the simulation, for simple balancing (grey) and for the extrapolation method (red), together with the expected true response (black line). The latter is visibly larger than unity at low transverse momenta: this is caused by the fact that the jet energy corrections are derived on QCD events, which are dominated by gluon jets, which have lower response than quark jets, that dominate the photon+jet events studied in this analysis. The right plot in Fig. 3.12 shows the data-MC ratios of the two methods. Consistently with what found with the MPF method on uncorrected response in the previous section, the data present a response about $1.5\,\%$ lower than what the simulation predicts.

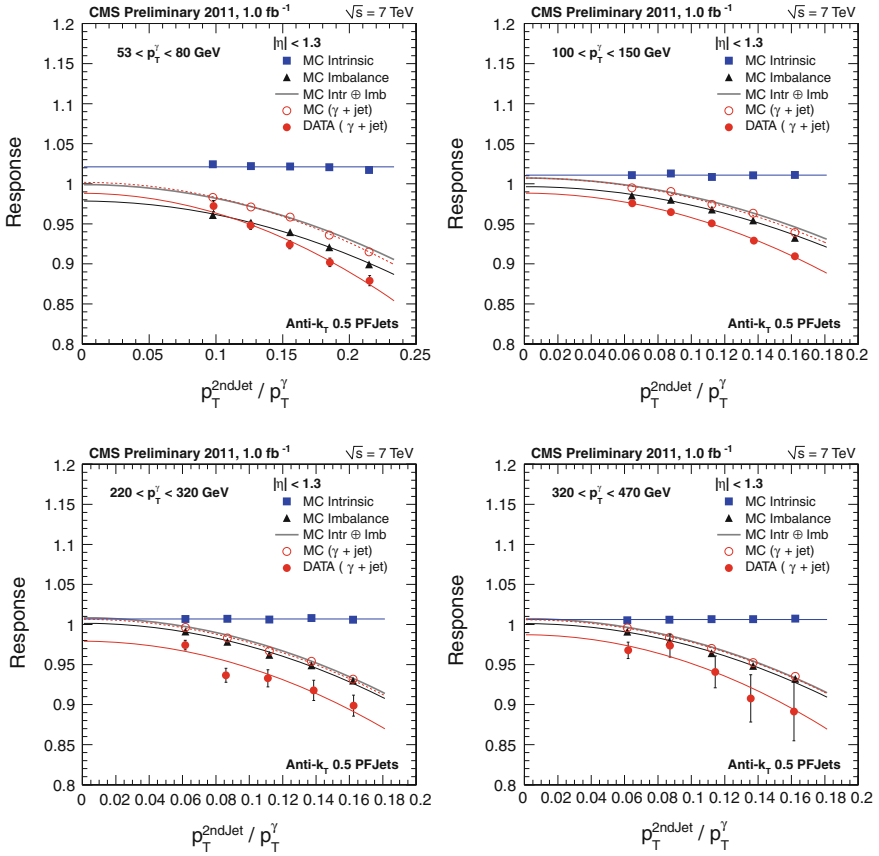

Fig. 3.11 Balancing response extrapolation in four representative transverse momentum ranges, for *anti-*k_{T} 0.5 PFJets reconstructed in the barrel

3.4.6 Jet Transverse Momentum Resolution Measurement

The balancing extrapolation method allows us to measure also the corrected jet transverse momentum resolution. Recalling Eq. 3.4, our assumptions are that, in a given p_{T}^{γ} bin, the intrinsic resolution is independent of $p_{\mathrm{T}}^{\mathrm{2ndJet}}$, whereas the imbalance effect to be linear. In formulas:

$$\sigma_{\mathrm{intr}}(p_{\mathrm{T}}^{\mathrm{2ndJet}}) = c' \qquad \sigma_{\mathrm{imb}}(p_{\mathrm{T}}^{\mathrm{2ndJet}}) = q' + m' \cdot p_{\mathrm{T}}^{\mathrm{2ndJet}} \qquad c', q', m' = const$$

$$\Longrightarrow \sigma_{\mathrm{balancing}}(p_{\mathrm{T}}^{\mathrm{2ndJet}}) = \sqrt{c'^2 + q'^2 + 2q'm' \cdot p_{\mathrm{T}}^{\mathrm{2ndJet}} + m'^2 \cdot \left(p_{\mathrm{T}}^{\mathrm{2ndJet}}\right)^2}$$

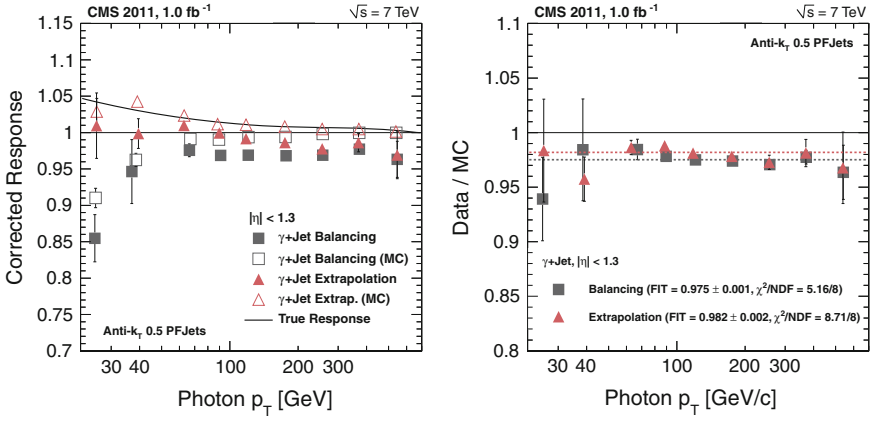

Fig. 3.12 Corrected response measurement, as a function of photon transverse momentum, for *anti-k_T* 0.5 PFJets reconstructed in the barrel. *Left* results for balancing (*grey*) and extrapolation (*red*) are shown both for data (*solid*) and the Monte Carlo simulation (*hollow*). A comparison to the expected true response (*black line*) is also shown. *Right* data-MC ratios for the two methods

The performance of the method is shown in Fig. 3.13, for the data and the simulation, in four representative p_{T}^{γ} bins. The colour coding is the same as in the response case. Again, the good agreement between the 'predicted' trend (grey line) and the reconstructed MC estimates (open red circles) proves the internal consistency of the method. The data points are fitted with the expected functional form, and, similarly as in the response case, the contribution of the irreducible imbalance (q') is fixed to the value fitted in the MC.

The results of the corrected jet p_{T} resolution as a function of transverse momentum are shown in Fig. 3.14. The left plot shows the results of the extrapolation, in data and MC, and compares them to the true response. The right plot shows the ratio of the measurements in data and MC: the resolution measured in the data is found to be about 7 % worse than the MC.

3.5 Jet Flavour Tagging: Quark-Gluon Discrimination

Detailed information on jet composition and substructure, as the one provided by the Particle Flow reconstruction, may be exploited to gain insight on the nature of the jet's underlying parton. In this section we will present a method which is able to discriminate between jets initiated by a gluon or light quark hadronization [14].

Gluons have a more intense coupling to the strong field with respect to quarks, therefore their hadronization favors the production of a larger number of stable particles. This translates, in the detector, in the observation of wider, high-multiplicity jets, when compared to those generated by final state (light) quarks. Furthermore,

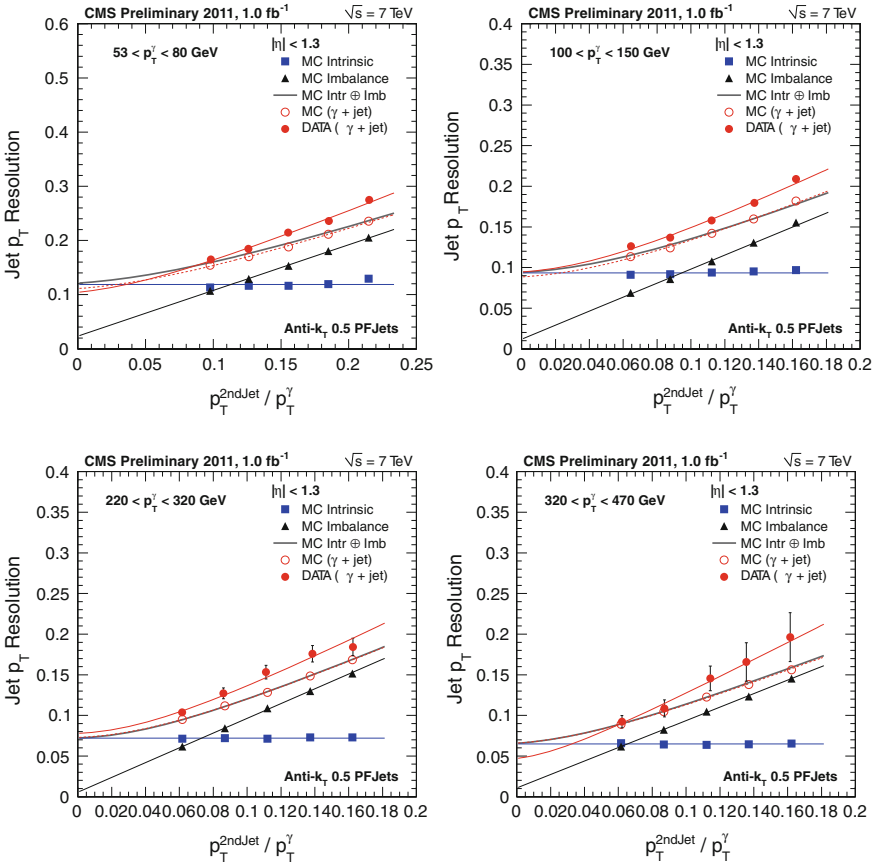

Fig. 3.13 Balancing resolution extrapolation in four representative transverse momentum ranges, for *anti-k_T* 0.5 PFJets reconstructed in the barrel

the phenomenon of 'gluon-splitting', if occurring at the beginning of hadronization, may give rise to jets made of a number of collimated quark sub-jets.

These structural differences between gluon and quark hadronization may be exploited to derive a likelihood based discriminant. In order to do so, the most precise and granular information on the jet particle composition must be accessed, such as the one provided by the CMS Particle Flow event reconstruction.

We have studied the use of three variables[2]:

- **charged hadron multiplicity**: the number of charged hadron PFCandidates clustered in the jet;

[2] We have also investigated the use of a fourth variable, i.e. the second moment of the angular distribution of the jet PFCandidates with respect to the jet axis. We have found that, as it is strongly correlated to the three variables presented here, it does not improve the discriminating power of the algorithm. The variable was therefore dropped.

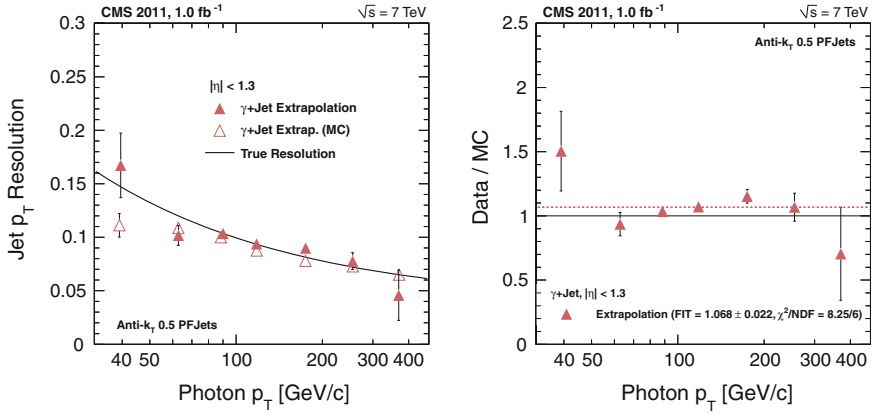

Fig. 3.14 Jet p_{T} resolution measurement, as a function of photon transverse momentum, for *anti-*k_{T} 0.5 PFJets reconstructed in the barrel. *Left*: results for data and MC; *right*: data-MC ratio

- **neutral multiplicity**: the number of PFCandidates in the jet of which are photons or neutral hadrons;
- **transverse momentum distribution** ($p_{\mathrm{T}}D$) among PFCandidates inside the jet, defined as:

$$p_{\mathrm{T}}D = \sqrt{\frac{\sum p_{\mathrm{T}}^2}{(\sum p_{\mathrm{T}})^2}}$$

  where the sums are extended to all PFCandidates inside the jet. It stems from its definition that $p_{\mathrm{T}}D \to 1$ for a jet made of one single candidate which carries the totality of its momentum, whereas $p_{\mathrm{T}}D \to 0$ for jets composed of an infinite number of particles.

The expected distributions of these variables, for quark and gluon jets, in two representative transverse momentum bins, are shown in Fig. 3.15. These shapes have been obtained on simulated QCD dijet events, by considering only the two leading jets in each event, and requiring them to be fully reconstructed in the tracker-covered pseudorapidity region ($|\eta| < 2$). The jet flavour is retrieved from the generator, by matching the jet to its closest parton in the $\eta - \phi$ plane. Reflecting our basic assumptions on quark-gluon coupling and hadronization properties, gluon jets present higher particle multiplicities and lower values of $p_{\mathrm{T}}D$, across the transverse momentum range.

These three variables are combined into a likelihood discriminant, taken as the product of the three variables' distributions, in 20 transverse momentum bins from 15 to 1000 GeV. A given PFJets identifies a vector $\vec{x}$ in the three-dimensional space of the structure variables. Probability functions for gluons (G) and quarks (Q) can then be defined as the product of each variable's probability density function (f_j^i, where $i = 1, 2, 3$ identifies the variable and $j = Q, G$ the jet parton flavour) computed at

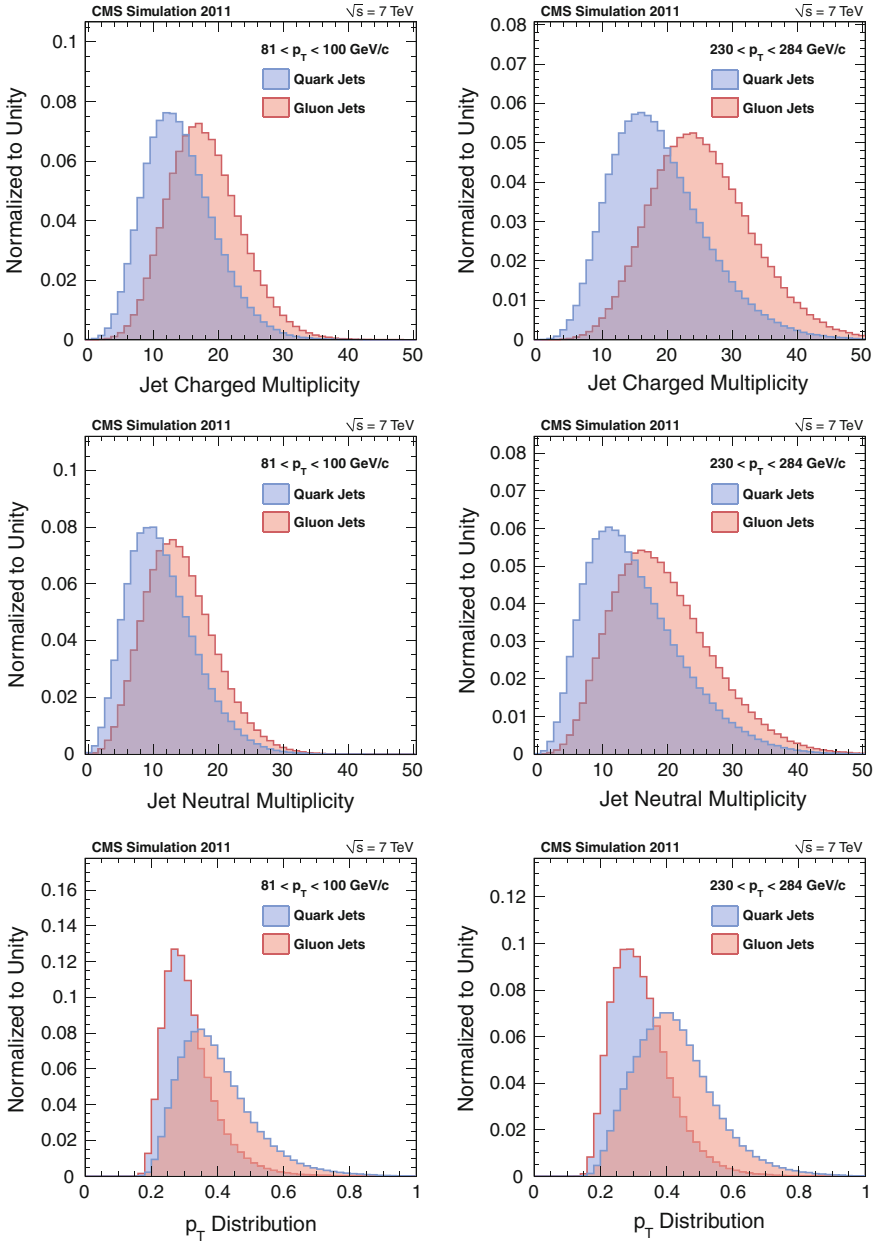

Fig. 3.15 Probability density distributions, in two representative transverse momentum ranges, for quark and gluon jets of the three considered discriminative variables: charged multiplicity (*top*), neutral multiplicity (*center*) and $p_{\mathrm{T}}D$ (*bottom*)

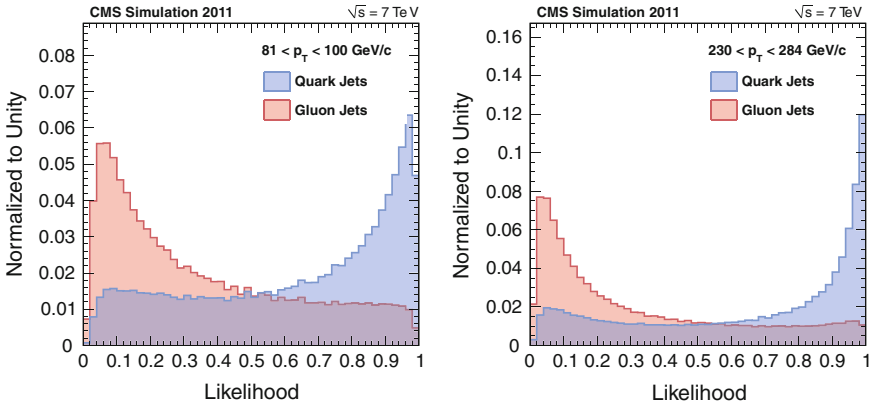

Fig. 3.16 Probability density distributions of the quark-gluon likelihood disciminant in two representative transverse momentum bins for quark and gluon jets

the given variable's value ($x[i]$):

$$G(\vec{x}) = \prod_i f_G^i\,(x[i]) \qquad Q(\vec{x}) = \prod_i f_Q^i\,(x[i])$$

A likelihood estimator can hence be defined as:

$$L(\vec{x}) = \frac{Q(\vec{x})}{Q(\vec{x}) + G(\vec{x})}$$

and interpreted as the probability of a given PFJet to be originated from a quark parton.

Figure 3.16 shows the distributions of the likelihood estimator variable for quark and gluon jets in the two representative transverse momentum bins. Figure 3.17 shows instead the quark jet efficiency - gluon jet rejection curves which are obtained by varying a simple cut on the likelihood variable distribution. This is further summarized in Fig. 3.18, where the maximum achievable gluon jet rejection is shown as a function of the jet transverse momentum for four different quark jet efficiency working points (70, 80, 90 and 95 %). As can be seen, the discriminating performance of the estimator is worst at low transverse momenta, gradually improves up to about 100 GeV, where it reaches a plateau which is maintained up to the TeV scale.

3.5.1 Treatment of Pile Up

Pile-up is expected to have a sensible effect on the distribution of jet composition variables, and therefore on the overall performance of the discriminant. Multiple proton-proton collisions within the same bunch crossing will produce a number of

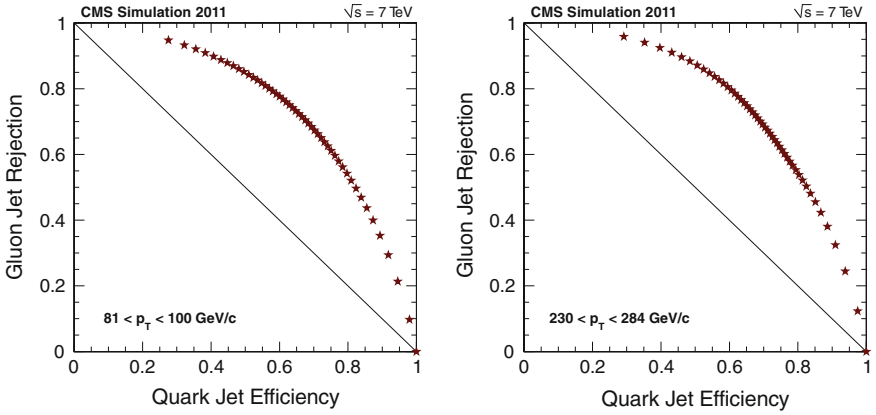

Fig. 3.17 Gluon rejection versus quark jet efficiency in two representative transverse momentum bins

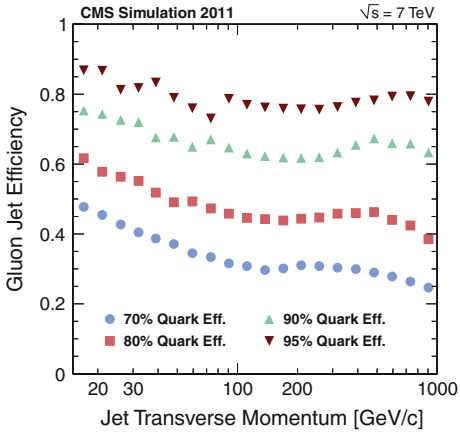

Fig. 3.18 Minimal gluon jet efficiency, as a function of transverse momentum, for fixed values of quark jet efficiency: 70 % (*circles*), 80 % (*squares*), 90 % (*upwards triangles*) and 95 % (*downwards triangles*)

diffuse soft particles, which will permeate isotropically the underlying event distribution. These additional particles will be clustered in the jets, therefore modifying their multiplicity variables, but, being generally soft, will not modify dramatically the jet transverse momentum. This has the net effect of augmenting the probability of a jet being gluon-like, as can be seen in Fig. 3.19, which shows how the charged and neutral multiplicity distributions change in the presence of pile up in a given p_{T} bin.

Whereas the shift in the multiplicity distributions is approximately constant as a function of the jet p_{T}, the effect on the $p_{\mathrm{T}}D$ variable is instead limited to the low-p_{T} region. This is due to the fact that additional particles have little weight in the

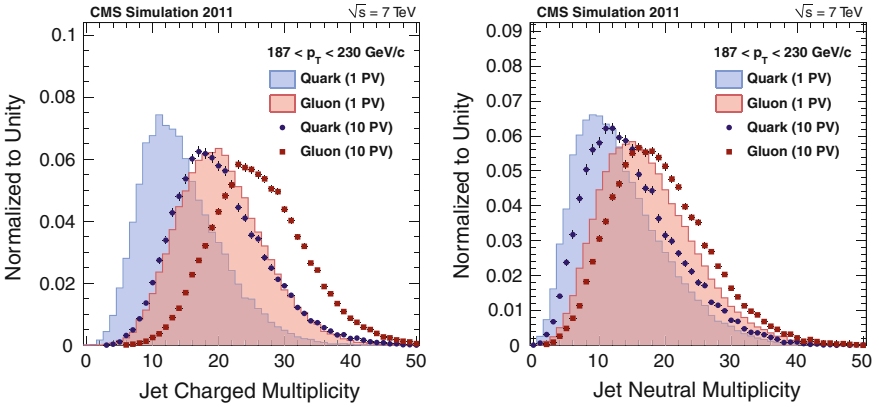

Fig. 3.19 Charged multiplicity (*left*) and neutral multiplicity (*right*) shape modifications in presence of pile-up for jets with transverse momenta between 187 and 230 GeV. *Shaded* histograms show the shapes in events with one reconstructed primary vertex, *markers* show events with ten vertexes

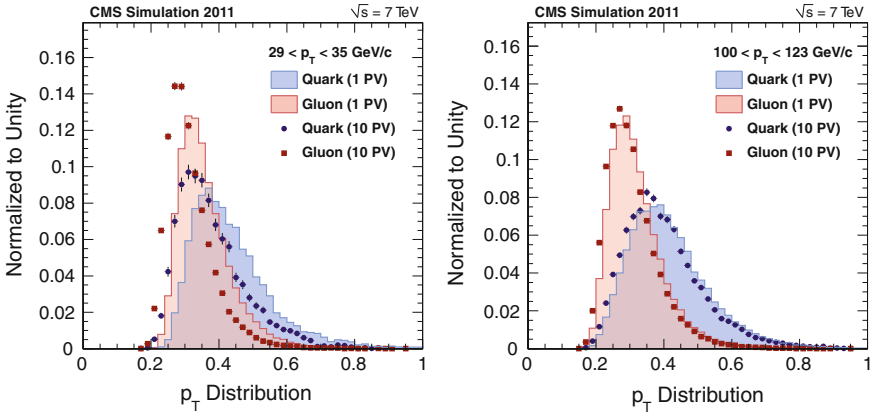

Fig. 3.20 $p_{\mathrm{T}}D$ shape modifications in presence of pile-up, in two transverse momentum intervals: $29 \div 35$ GeV (*left*) and $100 \div 123$ GeV (*right*). *Shaded* histograms show the shapes in events with one reconstructed primary vertex, *markers* show events with ten vertexes

computation of the $p_{\mathrm{T}}D$ variable when the typical jet constituents are significantly harder. Figure 3.20 shows how the $p_{\mathrm{T}}D$ variable shape changes in the presence of pile-up, in two transverse momentum ranges. As the jet transverse momentum increases, the effect of pile-up is reduced significantly, and is already almost negligible for $p_{\mathrm{T}} \gtrsim 150$ GeV.

In order to take into account the effect of pile up, the probability distribution functions for quark and gluon jets are computed in a double-differential binning: in addition to the jet transverse momentum binning, which was already showed, an additional binning in the particle flow energy density variable (ρ) has been introduced.

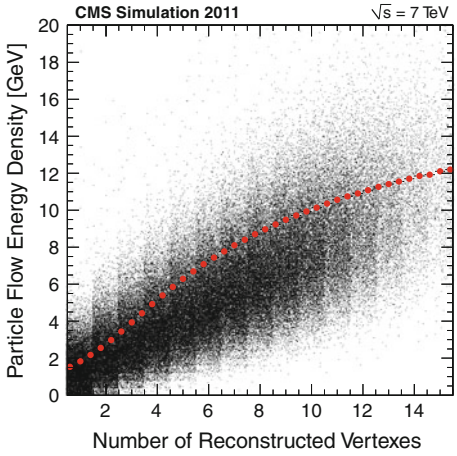

Fig. 3.21 Particle Flow energy density (ρ) as a function of the number of reconstructed primary vertexes of the event. The *markers* show the position of the average values of ρ in correspondence of given numbers of vertexes

The energy density variable ρ is the same that is used to derive the L1 Offset jet energy correction, as described in Sect. 3.3, and is a measure of the event's soft diffuse radiation (both due to pile up and underlying event activity), on an event per event basis. The ρ variable is correlated with the number of reconstructed vertexes, as can be seen in Fig. 3.21.

We have opted for 17 uniform bins in ρ, from 0 to 17 GeV. This is sufficient to account for instantaneous luminosities up to 3×10^{33} cm^{-2} s^{-1}. The closure test has been performed using quark jets from the $H \rightarrow ZZ \rightarrow 2\ell 2j$ channel (see Chap. 4 for further details). The results of the use of the double differential $p_{\mathrm{T}} \times \rho$ binning are shown in Fig. 3.22, for two representative transverse momentum bins: the red histogram shows the likelihood distribution for quark jets in signal when using probability density functions which do not take into account pile up, and as can be seen it is peaked towards 0; the yellow histogram shows the expected shape of the likelihood for quark jets in that transverse momentum range; the black markers show the signal distribution when using the double differential $p_{\mathrm{T}} \times \rho$ binning. As can be seen, the introduction of the additional binning in ρ brings the distribution in good agreement with the expected shape.

3.5.2 b-Jets

As has been previously observed by LEP experiments [15], the hadronization of a b quark yields jets which have structures similar to gluon-initiated jets, from an experimental point of view. The LEP measurement studied jets produced in the

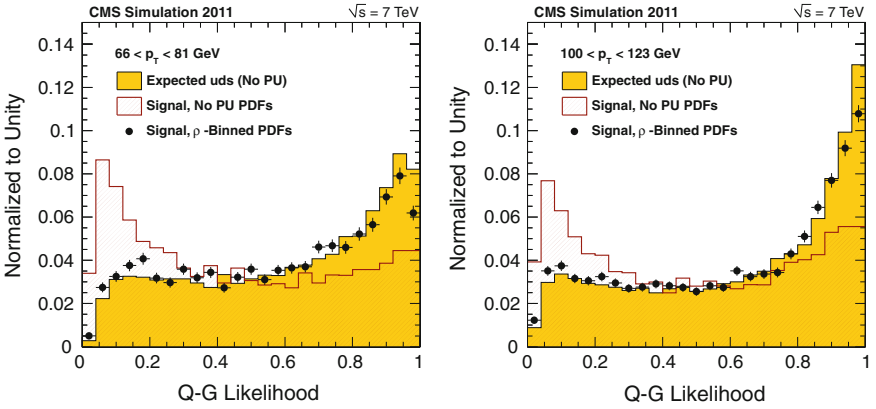

Fig. 3.22 Results of the closure test of the double differential $p_{\mathrm{T}} \times \rho$ binning in order to take into account pile up, in two representative transverse momentum bins. The *hashed* histogram shows the likelihood distribution of quark jets from the $H \to ZZ \to 2\ell 2j$ signal when using PDFs computed in the absence of pile up. The *solid* histogram shows the expected distribution for quark jets in the same transverse momentum range. The *markers* show the signal distribution when using PDFs which take into account pile up

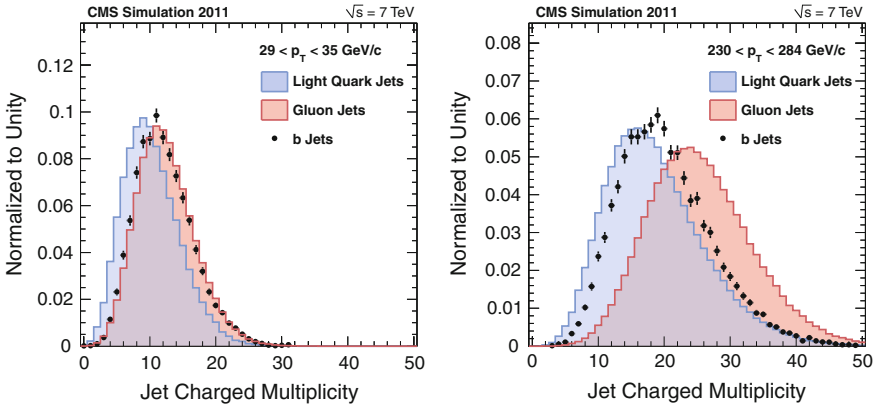

Fig. 3.23 Charged multiplicity distributions in two transverse momentum ranges for light quark jets, gluon jets and bottom quark jets

decay of an on-shell Z boson, therefore was practically limited to the energy range $E \lesssim 50$ GeV. It was concluded that for jets of these energies, while there were measurable differences in jet shapes between gluon and light quark jets, no significant ones were sought between gluon and b-jets.

Figures 3.23 and 3.24 show, respectively, the distributions of the charged multiplicity and $p_{\mathrm{T}} D$ for light quarks, gluons and b jets, in two transverse momentum bins. As can be noted, the shapes produced by b-quark hadronization are indeed very similar to the ones produced by gluons for $p_{\mathrm{T}} \lesssim 100$ GeV, which were the energies

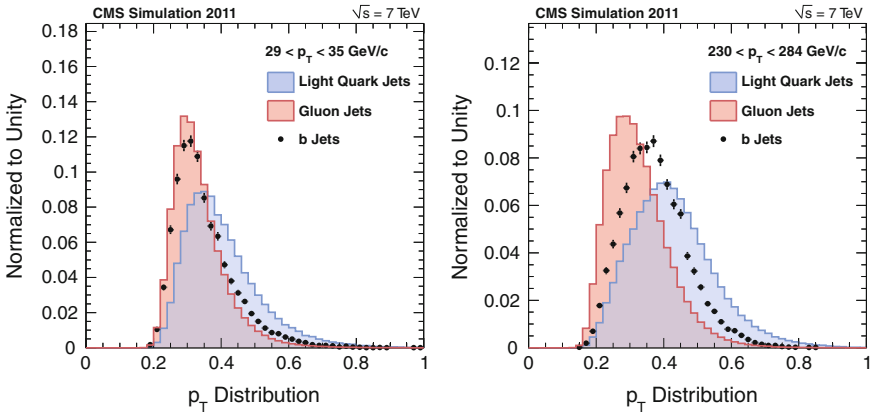

Fig. 3.24 $p_{\mathrm{T}}D$ distributions in two transverse momentum ranges for light quark jets, gluon jets and bottom quark jets

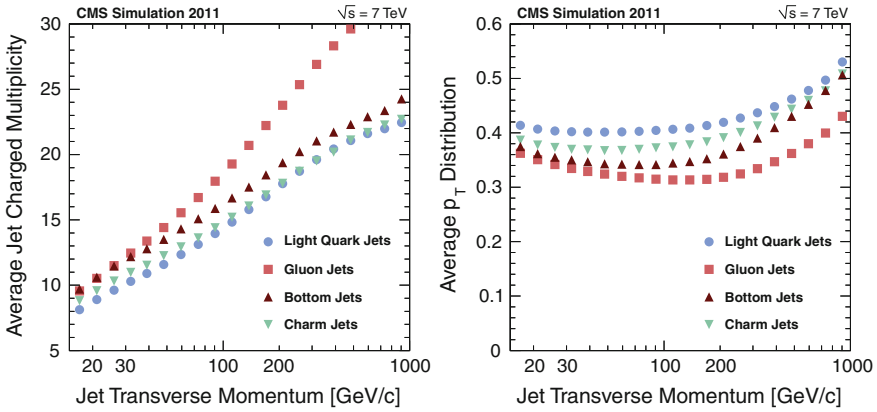

Fig. 3.25 Average jet charged multiplicity (*left*) and p_{T} distribution (*right*) as a function of jet transverse momentum for light quarks (*circles*), gluons (*squares*), bottom quarks (*upwards triangles*) and charm quarks (*upwards triangles*)

probed at LEP, but then tend to migrate towards the light quark shapes, as the transverse momentum increases. This can be more clearly seen in Fig. 3.25, where the average charged hadron multiplicity (left) and average value of $p_{\mathrm{T}}D$ (right) are shown as a function of jet transverse momentum, for light quarks (circles), gluons (squares), bottom quarks (upwards triangles) and charm quarks (upwards triangles).

We therefore conclude that this discriminant cannot be used in effectively discriminating gluon jets from b-jets. The latter, though, can be efficiently recognized by exploiting the relatively long lifetime of the B hadron, which is formed at the first step of b-quark hadronization. These techniques have been extensively used at high

energy colliders, and are commonly known as *b*-tagging. They will be described in more detail in Chap. 4.

References

1. Wigmans, R.: Calorimetry: energy measurements in particle physics. Oxford Science Publications, Oxford (2000)
2. Gabriel, T. A. et al.: Proceeding of the workshop on compensated calorimetry. Internal, Report CALTECH-68-1305
3. Abdullin, S., et al.: The CMS barrel calorimeter response to particle beams from 2 to 350 GeV. Eur. Phys. J. C: Part. Fields **60**, 359 (2009). doi:10.1140/epjc/s10052-009-0959-5
4. Cacciari, M., Salam, G. P., Soyez, G.: The anti-kt jet clustering algorithm. J. High Energy Phys. **2008**, 063 (2008). Available from: http://stacks.iop.org/1126-6708/2008/i=04/a=063
5. Cacciari, M., Salam, G. P.: Dispelling the N^3 myth for the k(t) jet-finder. Phys. Lett. **B641**, 57 (2006). doi:10.1016/j.physletb.2006.08.037
6. CMS Collaboration: Particle-flow event reconstruction in CMS and performance for jets, taus, and E_T^{miss}. CMS PAS PFT-09-001, (2009). Available from: http://cdsweb.cern.ch/record/1194487
7. CMS Collaboration. Jet calibration and resolution. CMS Physics Analysis Summary, CMS-PAS-JME-10-011 (2010). Available from: http://cdsweb.cern.ch/record/1369486
8. Cacciari, M., Salam, G. P.: Pileup subtraction using jet areas. Phys. Lett. B**659** 119–126 (2008). doi:10.1016/j.physletb.2007.09.077
9. Ellis, S. D., Soper, D. E.: Cite arxiv:hep-ph/9305266
10. Abbott, B., et al.: Determination of the absolute jet energy scale in the D0 calorimeters. Nucl. Instrum. Methods Phys. Res., Sect. A: Accelerators, Spectrometers, Detectors and Associated Equipment **424**, 352 (1999). doi:10.1016/S0168-9002(98)01368-0
11. Sjöstrand, T., Mrenna, S., Skands, P.: PYTHIA 6.4 physics and manual. JHEP **05** 026 (2007). doi:10.1088/1126-6708/2006/05/026
12. CMS Collaboration: Electromagnetic calorimeter calibration with 7 TeV data. CMS Physics Analysis Summary, CMS-PAS-EGM-10-003 (2010). Available from: http://cdsweb.cern.ch/record/1279350
13. Bähr, M., et al.: Herwig++ physics and manual. Eur. Phys. J. C: Part. Fields **58**, 639 (2008), 10.1140/epjc/s10052-008-0798-9. Available from: http://dx.doi.org/10.1140/epjc/s10052-008-0798-9
14. Marini, A.C., Pandolfi, F., del Re, D., Voutilainen, M.: Quark-Gluon jet discrimination through particle flow jet structure. CMS Analysis Note, CMS AN-2011/215 (2011)
15. Alexander, G. et al.: A comparison of b and uds quark jets to gluon jets. Zeitschrift für Physik C: Part. Fields **69**, 543 (1996), doi:10.1007/s002880050059

Chapter 4
Event Selection

Abstract The following chapter details the event selection that is employed in the analysis. Section 4.1 reports the analyzed datasets, both of data and simulated events, and Sect. 4.2 lists the preselection requirements which events are required to satisfy. Subsequent sections investigate means of background discrimination, and define the final analysis event selection procedure. Section 4.6 proceeds to optimize the selection requirements. Finally, the expected signal and background yields are reported in Sect. 4.8, and the means of background estimation employed on the data is shown in Sect. 4.9.

4.1 Datasets and Trigger

4.1.1 Data

This analysis focuses on the search of a massive Higgs boson, well above the ZZ production threshold ($m_H > 200$ GeV). The decay of one of the two electroweak bosons, which are on mass shell and will in general have an elevated relativistic boost, will produce a pair of high-p_{T} leptons. Most of the signal events, therefore, will fire the double-lepton HLT paths, and be stored in two Primary Datasets called, respectively, `DoubleElectron` and `DoubleMu`.

The results presented here make use of a total of 4.6 fb^{-1} of data collected by the CMS detector during the 2011 data taking. The data sample is divided in two running periods: Run2011A and Run2011B. The first extends up to September machine development technical stop, and can be further separated into the three secondary running periods, spaced by the intermediate technical stops which were undertaken by the LHC accelerator. We will call these run periods respectively 2011A1, 2011A2 and 2011A3. Run2011B comprises the second half of the data delivered in 2011, and extends from September to the end of the data taking, in November. Compared to Run2011A, it is characterized by significantly higher instantaneous luminosity, and

F. Pandolfi, *Search for the Standard Model Higgs Boson in the H→ZZ→ $\ell^+\ell^- q\bar{q}$ Decay Channel at CMS*, Springer Theses, DOI: 10.1007/978-3-319-00903-2_4,

Table 4.1 Subdivision of the 2011 dataset into running periods: each period is identified by its name and the corresponding run ranges and integrated luminosities

Running period	Run range	Integrated luminosity (fb^{-1})
2011A1	160329–163869	0.2
2011A2	163870–170052	0.9
2011A3	170053–172619	1.0
2011B	172620–172998	2.5
	Total	4.6

larger pile up multiplicities. The analyzed running periods and their corresponding run ranges and integrated luminosities are reported in Table 4.1.

Dimuon events are explicitly required to have fired either one of the following three HLT paths:

- `HLT_DoubleMu7`
- `HLT_Mu13_Mu8`
- `HLT_Mu17_Mu8`

The first path requires the presence of two muon candidates reconstructed at HLT level, each with transverse momentum greater than 7 GeV, and was the lowest unprescaled double muon trigger path during running period 2011A. As the LHC's instantaneous luminosity increased, `HLT_DoubleMu7` was prescaled, and a new trigger was devised (`HLT_Mu13_Mu8`), which increased the transverse momentum thresholds respectively to 13 and 8 GeV, keeping the event acceptance rates to manageable levels. Towards the end of the 2011 run, `HLT_Mu13_Mu8` was substituted with `HLT_Mu17_Mu8`, in which the transverse momentum threshold on the leading muon was raised to 17 GeV.

In order to minimize the loss in efficiency that arises from the requirement of double muon triggers, events from the `/SingleMu` Primary Dataset were added to the analyzed sample in an exclusive way. This was done by selecting events which fired the `HLT_IsoMu24` path but did not fire any of the above double muon trigger paths. This has led to an increase of about 6 % in the muon dataset dimension. In Run2011B the above mentioned single muon trigger was prescaled, therefore we relied on the path `HLT_IsoMu24_eta2p1`, which is identical in all respects except for the introduction of a pseudorapidity requirement of $|\eta| < 2.1$ which is applied to the muon candidate.

Dielectron events have been selected from the `/DoubleElectron` Primary Dataset, and are required to have fired either the `HLT_Ele17_CaloIdL_CaloIsoVL_Ele8_CaloIdL_CaloIsoVL` or the `HLT_Ele17_CaloIdT_TrkIdVL_CaloIsoVL_TrkIsoVL_Ele8_CaloIdT_TrkIdVL_CaloIsoVL_TrkIsoVL` HLT paths. Both these trigger paths require the presence of two electrons in the event, with transverse momenta respectively greater than 17 and 8 GeV. In addition to this, a number of electron identification criteria are introduced, in order to lower the rate at which QCD events could accidentally fire

these paths. No gain can be obtained by adding the /SingleElectron Primary Dataset, as the unprescaled single-electron trigger paths have very high transverse momentum thresholds.

4.1.2 Generated Events

The same analysis which is performed on data is applied to Monte Carlo generated events, which make use of the full simulation of the CMS detector, implemented within the GEANT 4 software framework. As we have seen in Sect. 1.5, any Standard Model process which produces an opposite-signed, high-p_{T} electron or muon pair in the final state, in association with two hard jets, may constitute a background for this channel. The main processes which contribute are therefore:

- direct production of a Z boson in association with hard jets;
- continuous production of electroweak boson pairs (ZZ, WZ, ZZ);
- events with top quarks, either produced in $t\bar{t}$ pairs or in association with a W boson (tW^-, $\bar{t}W^+$).

The analyzed background Monte Carlo samples are summarized in Table 4.2: for each dataset its NLO cross section (σ), number of analyzed events and equivalent integrated luminosity is given. As shown in the following, the main analysis backgrounds are constitued by Z+jets and $t\bar{t}$: both of these processes have been generated with MADGRAPH 4.4.12 [1], which is a NLO matrix element generator. It is interfaced to PYTHIA 6 [2] for parton showering and hadronization. A similar strategy is adopted for the single top background (tW^-, $\bar{t}W^+$), where the POWHEG [3–5] generator (which contains the complete NLO calculation of the cross section of these processes) is interfaced again to PYTHIA 6. As for the diboson backgrounds (ZZ, WZ, ZZ), they have been fully generated in PYTHIA 6.4.22, therefore with LO precision, as they have a minor contribution.

The POWHEG event generator has also been utilized to generate signal events, as it contains NLO calculations and correctly describes final state angular correla-

Table 4.2 Summary of analyzed background Monte Carlo generated samples

Physical process	Generator	σ (pb)	Events	Luminosity (fb^{-1})
Z+jets	MADGRAPH 4.4.12	3048	36M	11.8
ZZ	PYTHIA 6.4.22	7.67	4M	521
WZ	PYTHIA 6.4.22	18.3	4M	218
WW	PYTHIA 6.4.22	42.9	4M	93.2
$t\bar{t}$	MADGRAPH 4.4.12	157.5	4M	25.4
tW^-	POWHEG	7.46	0.8M	107
$\bar{t}W^+$	POWHEG	7.47	0.8M	107

For each dataset, the process name, NLO cross section (σ), number of analyzed events and equivalent integrated luminosity is provided

Table 4.3 Summary of analyzed signal Monte Carlo generated samples

Higgs Mass (GeV)	$\sigma \times$BR (pb)	Events	Luminosity (fb^{-1})
190	0.134	300k	868
200	0.146	300k	1190
210	0.141	300k	1420
230	0.123	300k	1810
250	0.107	300k	2200
275	9.09×10^{-2}	300k	2680
300	7.98×10^{-2}	300k	3140
325	7.23×10^{-2}	300k	3540
350	7.25×10^{-2}	300k	3450
375	6.64×10^{-2}	300k	3210
400	5.56×10^{-2}	300k	3470
425	4.70×10^{-2}	300k	4060
450	3.63×10^{-2}	300k	4950
475	2.90×10^{-2}	300k	6170
500	2.32×10^{-2}	300k	7820
525	2.29×10^{-2}	300k	7880
550	1.51×10^{-2}	300k	1.25×10^{4}
575	1.22×10^{-2}	300k	1.58×10^{4}
600	9.85×10^{-3}	300k	2.00×10^{4}

For each Higgs boson mass point, the NLO cross section times decay branching ratio ($\sigma \times$BR), number of analyzed events and equivalent integrated luminosity is provided. All samples have been generated with the POWHEG generator

tions. Events have been generated for a total of 19 mass hypotheses between 190 and 600 GeV, and a total of 300 k events have been analyzed per mass point, as summarized in Table 4.3.

Simulated events have been generated with a default pile up configuration which has constant probability for up to ten additional interactions, and then falls off with a poissonian tail, which translates in a distribution of reconstructed primary vertices which is different from what is observed in the data, as can be seen in Fig. 4.1 (left). In order to match the amount of pile up found in the data, a reweighing procedure is undertaken: the number of pile up interactions in the Monte Carlo generated events is reweighed with the distribution measured in the data. The distribution of reconstructed primary vertices after the reweighing is shown in Fig. 4.1 (right).

No explicit trigger requirement was made on Monte Carlo events, but these were rescaled in order to take into account the measured trigger efficiencies in data. The latter were computed by measuring single-muon HLT efficiencies in the data and MC with the $Z \to \ell\ell$ 'tag-and-probe' technique [6], and taking the ratio of the two as a scaling factor to be applied to the MC. The resulting scaling factors, in different muon pseudorapidity ranges, are shown in Tables 4.4 and 4.5, respectively for double- and single-muon triggers. The event rescaling factor k_{HLT} is then obtained with the following formula:

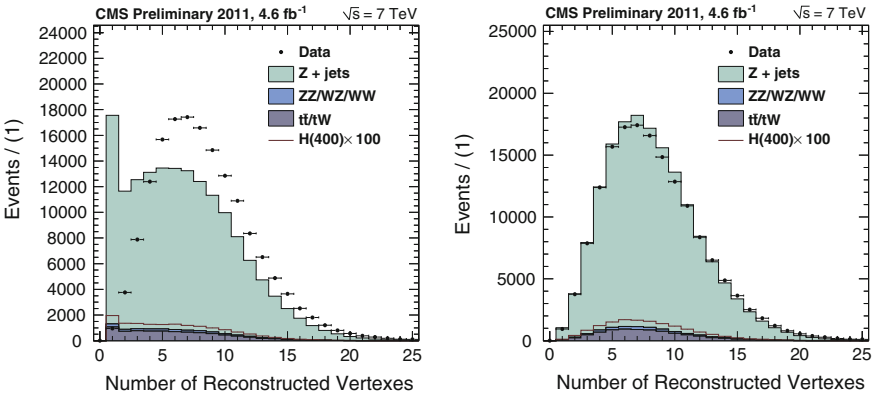

Fig. 4.1 Number of reconstructed primary vertices in data and in the simulation before (*left*) and after (*right*) the pile up reweighing procedure is applied to the simulation

Table 4.4 Single muon rescaling factors for the double muon triggers (`HLT_DoubleMu7`, `HLT_Mu13_Mu8`, `HLT_Mu17_Mu8`), for the 2011 running periods and for different muon pseudorapidity ranges

Running period	Muon pseudorapidity	MC scaling factor
2011A	$\|\eta\| < 0.8$	0.975 ± 0.001
	$0.8 < \|\eta\| < 2.1$	0.955 ± 0.001
	$2.1 < \|\eta\| < 2.4$	0.910 ± 0.001
2011B	$\|\eta\| < 0.8$	0.972 ± 0.001
	$0.8 < \|\eta\| < 2.1$	0.945 ± 0.001
	$2.1 < \|\eta\| < 2.4$	0.901 ± 0.001

Muons are required to have a transverse momentum larger than 20 GeV. Uncertainties are dominated by systematic contributions

$$k_{HLT} = \epsilon_D(\eta_1) \cdot \epsilon_D(\eta_2) + \epsilon_S(\eta_2) \cdot (1. - \epsilon_D(\eta_1)) + \epsilon_S(\eta_1) \cdot (1. - \epsilon_D(\eta_2))$$

where ϵ_D and ϵ_S indicate respictively the single-muon scaling factors for double- and single-muon triggers (which depend on the given muon's pseudorapidity), and the subscripts identify the two muons in the event. It must be noted that the formula assumes a complete correlation between the single and double muon trigger paths, i.e. if a given muon fails to be reconstructed by the double muon trigger, it automatically fails also the single muon trigger.

The adopted dielectron trigger path was found to have >99 % efficiency in $Z \rightarrow ee$ events, therefore a conservative scaling factor of 0.99 was applied to dielectron Monte Carlo events.

Table 4.5 Single muon rescaling factors for `HLT_IsoMu24`, for the 2011 running periods and for different muon pseudorapidity ranges

Running period	Muon pseudorapidity	MC scaling factor
2011A1	$\|\eta\| < 0.8$	0.986 ± 0.001
	$0.8 < \|\eta\| < 2.1$	0.807 ± 0.001
	$2.1 < \|\eta\| < 2.4$	0.608 ± 0.001
2011A2	$\|\eta\| < 0.8$	0.895 ± 0.001
	$0.8 < \|\eta\| < 2.1$	0.838 ± 0.001
	$2.1 < \|\eta\| < 2.4$	0.738 ± 0.001
2011A3	$\|\eta\| < 0.8$	0.890 ± 0.002
	$0.8 < \|\eta\| < 2.1$	0.809 ± 0.002
	$2.1 < \|\eta\| < 2.4$	0.493 ± 0.002
2011B	$\|\eta\| < 0.8$	0.870 ± 0.002
	$0.8 < \|\eta\| < 2.1$	0.790 ± 0.002
	$2.1 < \|\eta\| < 2.4$	0

Muons are required to have a transverse momentum larger than 40 GeV. Uncertainties are dominated by systematic contributions. Note that the scale factor is null for Run2011B in the pseudorapidity region $2.1 < |\eta| < 2.4$, as the adopted trigger path for that period was `HLT_IsoMu24_eta2p1`

4.2 Preselection

The signature of signal events presents two energetic Z bosons, one decaying to a pair of electrons or muons, the other to jets. Therefore the event preselection is defined as those events which pass the trigger requirements discussed in the previous section and present:

- two oppositely charged electrons or muons with transverse momenta respectively greater than 40 and 20 GeV;
- two (or more) *anti-*k_{T} 0.5 PFJets with $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.4$.

The relatively high transverse momentum requirement on the lepton pair is introduced to ensure high trigger efficiency. All physics objects (muons, electrons and jets) are required to pass a set of identification criteria, which will be discussed in the following subsections.

In the case of multiple electron or muon pairs, the oppositely-charged pair with invariant mass closest to the Z boson nominal mass is chosen. Events are then correctly identified as signal event candidates if the invariant mass of the dilepton system lies between 70 and 110 GeV. If an event is found to present both an electron and a muon pair passing this requirement, it is discarded. The dilepton invariant mass for events passing preselection requirements is shown in Fig. 4.2 (left).

The event is further required to present at least one jet pair with an invariant mass in the $75 \div 105$ GeV range. The requirement on the hadronic invariant mass is more stringent than the leptonic one, for it is a powerful handle in discriminating the main backgrounds, which do not present a real Z boson decaying to jets. It is therefore kept closest to the nominal Z boson mass, compatibly with the expected di-jet mass resolution, which is about 15 GeV for signal events. Events which pass the dilepton

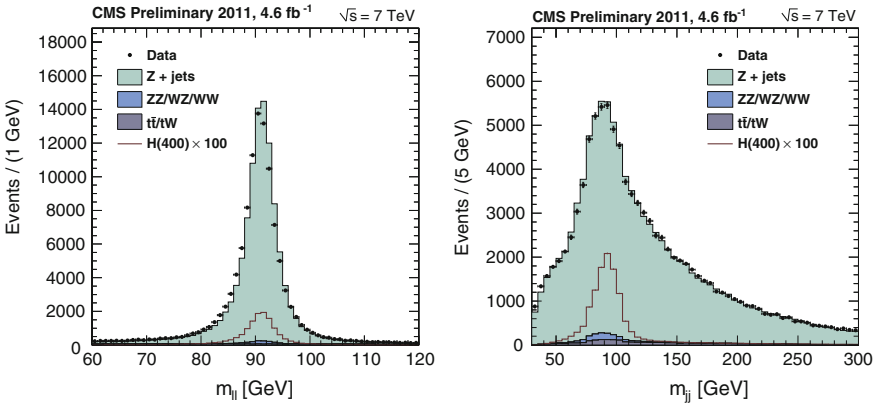

Fig. 4.2 Dilepton (*left*) and dijet (*right*) invariant mass distributions. Events passing preselection in 4.6 fb^{-1} of 2011 data are compared to the expected contribution of the dominant backgrounds in the simulation. The distribution for events coming from a 400 GeV Higgs boson decay, enhanced by a factor 100, is superimposed

mass requirement but not the dijet one are nevertheless kept, and are categorized as sideband events, as will be explained in Sect. 4.9.

In general, though, a signal event candidate will present multiple jet pairs. This is true also for true signal events, as additional jets will be created in proton fragmentation or in the process of creation of the Higgs boson. In order to minimize the effect of signal self-combinatorics, the jet pair with the invariant mass closest to the Z boson nominal mass will be selected, even if in the context of the event categorization procedure which will be described in detail in Sect. 4.5. The distribution of the invariant mass of the dijet pair with mass closest to the Z mass for events which pass preselection requirements is shown in Fig. 4.2 (right).

We will now proceed to describe the quality and identification criteria which are imposed on the used physics objects.

4.2.1 Muon Identification Criteria

Muon candidates are required to be reconstructed both in the tracker and in the muon chambers (global muons) and to satisfy the following identification criteria:

- normalized χ^2 of the global track < 10;
- the track reconstructed in the tracker must have more than ten matched hits, of which at least one in the pixel detector;
- the global track must have at least two matched segments in the muon stations;
- track transverse impact parameter $d_{xy} < 0.02$ cm and longitudinal impact parameter $d_z < 1$ cm.

The muon is furthermore required to be isolated, in order to suppress backgrounds in which muons are produced inside jets. The isolation variable I_μ is computed as the sum of all the reconstructed transverse momenta (energies) of all tracks (calorimeter deposits) found within $\Delta R < 0.3$ of the muon candidate:

$$I_\mu = \sum_{\Delta R<0.3} p_{\mathrm{T}}^{\mathrm{trk}} + \sum_{\Delta R<0.3} E_{\mathrm{T}}^{\mathrm{ECAL}} + \sum_{\Delta R<0.3} E_{\mathrm{T}}^{\mathrm{HCAL}}$$

and the value of I_μ is required to be less than 15 % of the muon candidate transverse momentum.

4.2.2 Electron Identification Criteria

A reconstructed electron candidate is the combination of a GSF track and an ECAL supercluster. Details on electron reconstruction have been given in Sect. 2.3. The set of requirements imposed on the electron candidates may be subdivided into three categories:

- electron identification criteria;
- photon conversion rejection criteria;
- isolation.

The variables used in defining the set of electron identification criteria are the following:

- $\sigma_{i\eta i\eta}$: the second moment of the ECAL cluster energy distribution along the η direction;
- $\Delta\phi$: the azimuthal difference between the GSF track position extrapolated at the calorimeter surface (with parameters computed at the vertex) and the ECAL supercluster energy baricenter;
- $\Delta\eta$: the pseudorapidity difference between the GSF track position extrapolated at the calorimeter surface (with parameters computed at the vertex) and the ECAL supercluster energy baricenter;
- H/E: the ratio between the energy deposit recorded in the HCAL tower directly behind the ECAL supercluster seed, and the ECAL supercluster energy.

The thresholds applied to these variables vary for electrons reconstructed in the barrel and in the endcaps, and the used values may be found in Table 4.6. These thresholds have been adjusted in order to ensure that the resulting requirements are more stringent than the ones introduced at HLT level.

When searching for events with prompt final-state electrons, some care must be expended in the rejection of electron candidates which arise from conversions of high-energy photons ($\gamma \to e^+e^-$) traversing the tracker material. The electron tracks are therefore required to have no more than one missing hit in the tracker layers.

Table 4.6 Electron identification requirements

Variable		Barrel	Endcaps
$\sigma_{i\eta i\eta}$	$<$	0.01	0.03
$\Delta\phi$	$<$	0.15	0.1
$\Delta\eta$	$<$	0.007	0.01
H/E	$<$	0.15	0.1

Finally, the electron candidate is required to be isolated. The electron isolation variable I_e is defined as:

$$I_e = I_e^{trk} + I_e^{ECAL} + I_e^{HCAL}$$

where I_e^{trk} is the scalar sum of the transverse momenta of all tracks reconstructed within a cone of $\Delta R = 0.3$ around the elelectron candidate direction, I_e^{HCAL} is the sum of all HCAL tower energy deposits within the same cone, and the ECAL isolation is defined as the sum of all ECAL crystal energies found in the same cone, but a pedestal energy of 1 GeV is subtracted to this value in the case of barrel electrons. The electron candidate is considered isolated if the isolation variable I_e is found to be less than 15 % (10 %) of its tranverse momentum, where the parentheses apply to the endcaps.

Electron candidates are required to be reconstructed in the tracker-covered ECAL fiducal region of $(|\eta| < 1.4442)\ ||\ (1.566 < |\eta| < 2.5)$, in this way avoiding the barrel-endcap transition.

4.2.3 Jet Identification Criteria

Particle Flow jet reconstruction makes use of all CMS subdetectors and therefore fake jet candidates, originating from calorimeter noise, can be straightforwardly removed with a simple set of requirements which make use of jet composition information. We require the reconstructed jets to be composed of at least two PFCandidates, at least one of which is a charged hadron.

4.3 Kinematic and Angular Discrimination

The dominant background to the $H \to ZZ \to \ell^+\ell^- q\bar{q}$ is constituted by the production of a real Z boson, decaying to a charged lepton pair, in association with two hard QCD jets. There are several topological differences that can help us discriminate the signal from this background: the production of the heavy Higgs resonance, on one hand, sets the energetic scale of the event, translating into harder final state kinemat-

ics. The spin of the decaying boson, on the other, defines the correlations between its decay products, and this can be exploited through an angular analysis.

In this chapter we will show the main topological differences between signal and background events. The first section is centered on an overview of the kinematical differences, whereas the following describe the angular analysis which will constitute the core ingredient of our background discrimination.

4.3.1 Kinematic Distributions

The production of a massive resonance, such as the Higgs boson, sets the energetic scale of signal events. On the contrary, backgrounds to this channel are non-resonant, and therefore favour the presence of softer final state objects.

This is shown in Fig. 4.3, where the lepton transverse momenta (top), the jet transverse momenta (center), the transverse momentum of the leptonic Z candidate (bottom left) and the ΔR spacing between jets (bottom right), are shown for three signal mass hypotheses and the dominant backgrounds. As can be seen there are significant differences between signal and backgrounds, which become more striking as the mass of the Higgs boson increases. On the other hand, it is evident that these variables have high degrees of correlations, therefore a careful treatment of them would be necessary should one decide to exploit kinematical differences between signal and backgrounds as the main means of discrimination of the latter.

Figure 4.4 shows the observed distributions of the kinematical variables in the 2011 data collected by CMS, in events which pass the analysis preselection (see Sect. 4.2 for details). The data are compared to the summed contribution of all simulated backgrounds, and an overall good agreement is observed. The expected distributions for events coming from a 400 GeV Higgs boson decay are superimposed, scaled by a factor 100 for visual puroposes.

4.3.2 Angular Distributions

The production of a heavy Higgs resonance has implications in the signal event topology which do not limit themselves to the kinematics of the final state products. This is caused by the fact that the latter are the product of a very precise decay chain: the decay of a spin-0 boson (the Higgs) to a pair of identical spin-1 bosons (the Zs), which then decay to fermions. The presence of this decay chain with a very well defined spin correlation will mirror in the angular distributions of the final state objects. For signal events, the ideal angular probability density functions can be computed analytically. In background events, on the contrary, this spin correlation is absent, as they are non resonant. Therefore we expect to observe different final state angular distributions.

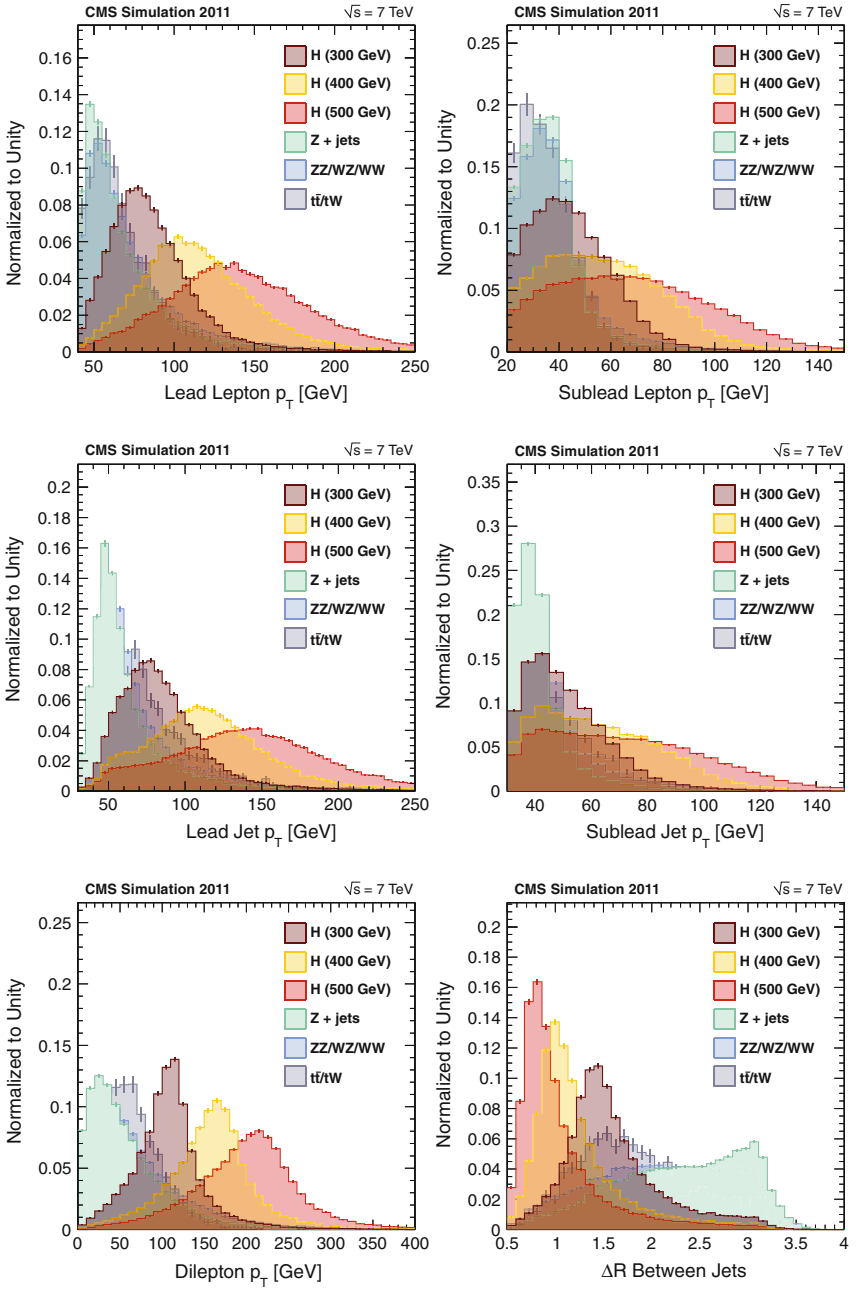

Fig. 4.3 Kinematical differences between signal and background events passing preselection: lepton transverse momenta (*top*), jet transverse momenta (*center*), transverse momentum of the leptonic Z candidate (*bottom left*) and ΔR spacing between jets (*bottom right*). The shapes originating from the main backgrounds are compared to three signal mass hypotheses (300, 400 and 500 GeV). All distributions are normalized to unit area

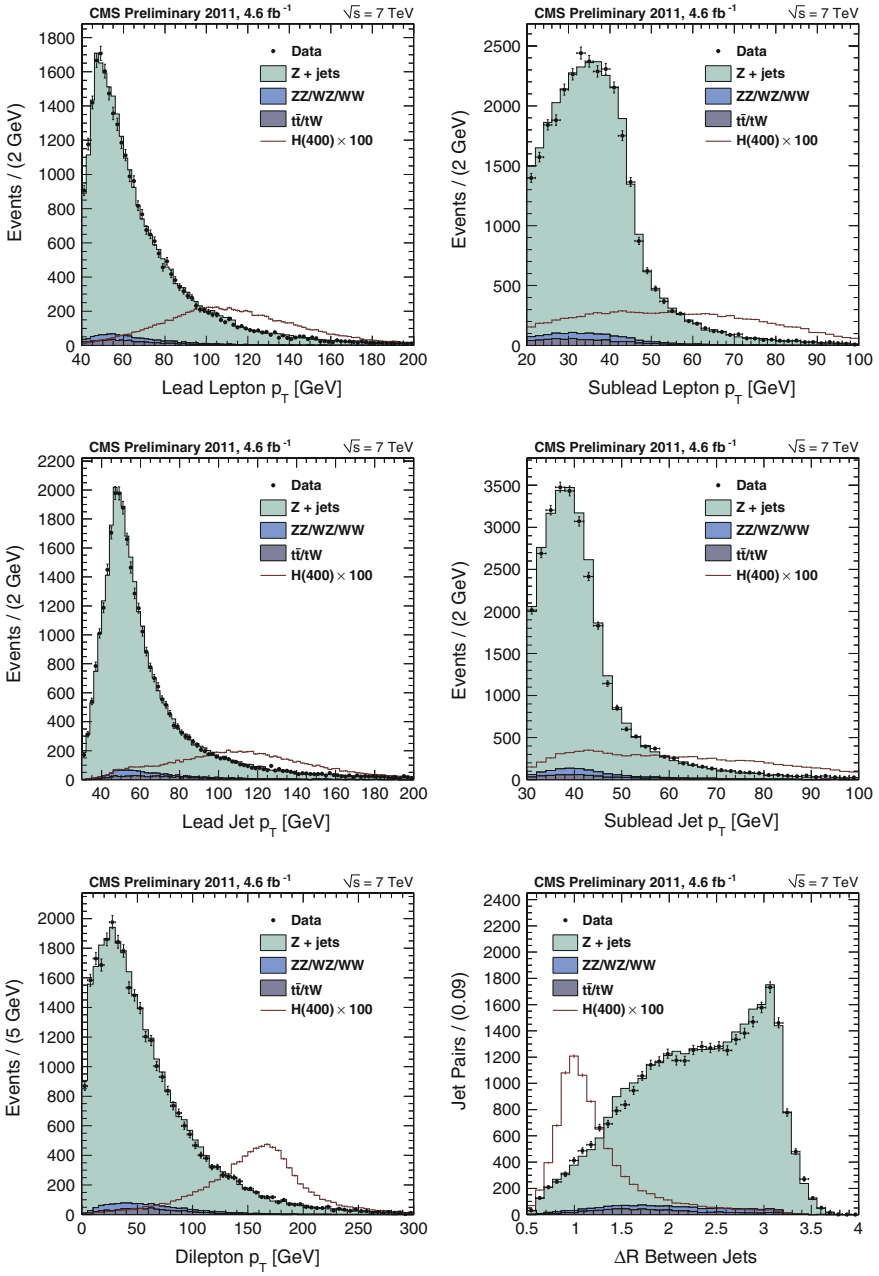

Fig. 4.4 Data-simulation comparisons for kinematical variables describing final state objects: lepton transverse momenta (*top*), jet transverse momenta (*center*), transverse momentum of the leptonic Z candidate (*bottom left*) and ΔR spacing between jets (*bottom right*). Events passing preselection in 4.6 fb^{-1} of 2011 data are compared to the expected contribution of the dominant backgrounds in the simulation. The distributions for events coming from a 400 GeV Higgs boson decay, enhanced by a factor 100 for visual purposes, are superimposed

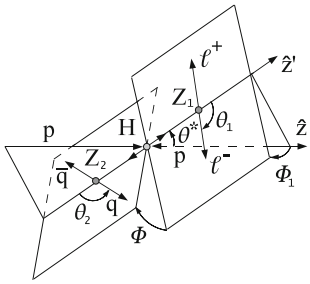

Fig. 4.5 Adopted convention in the definition of the three helicity angles (θ_1, θ_2 and Φ) and two production angles (θ^* and Φ_1) which univocally describe the $H \to ZZ \to \ell^+\ell^-q\bar{q}$ decay chain

If we do assume that the four final state objects derive from the above mentioned decay chain, the final state kinematics in the Higgs boson rest frame, once the masses of the secondary particles are fixed, are univocally determined through the definition of five angles. Following the convention used in [7], we will define them as in Fig. 4.5: they are three helicity angles (θ_1, θ_2 and Φ), respectively defined in the $Z \to \ell\ell$, $Z \to jj$ and Higgs boson rest frames, and two production angles (θ^* and Φ_1), both defined in the Higgs rest frame.

4.3.3 Angular Discriminant

The final state angular information can be exploited in order to define a likelihood discriminant (LD), able to select final state topologies which are compatible with the decay of a Higgs boson. This is a two-step procedure: first we build the probability density functions for signal and background events, which we will call respectively $\mathcal{P}_{sig}$ and $\mathcal{P}_{BG}$; then the likelihood discriminant is defined as a probability ratio:

$$LD = \frac{\mathcal{P}_{sig}}{\mathcal{P}_{sig} + \mathcal{P}_{BG}}$$

In this way events with signal-like topologies will have values of the likelihood discriminant close to unity, background-like topologies will be closer to zero. The variable can therefore be used in event selections, by requiring events to assume values above a certain threshold.

The signal probability density function is defined as a product of the ideal, fully correlated, distribution $\mathcal{P}_{\text{ideal}}$ (derived in [7]) and a set of four one-dimensional acceptance functions:

$$\begin{aligned}\mathcal{P}_{\text{sig}} = {} & \mathcal{P}_{\text{ideal}}(\theta^*, \theta_1, \theta_2, \Phi, \Phi_1; M) \\ & \cdot \mathcal{G}_{\theta^*}(\theta^*; M) \cdot \mathcal{G}_{\theta_1}(\theta_1; M) \cdot \mathcal{G}_{\theta_2}(\theta_2; M) \cdot \mathcal{G}_{\Phi_1}(\Phi_1; M)\end{aligned}$$

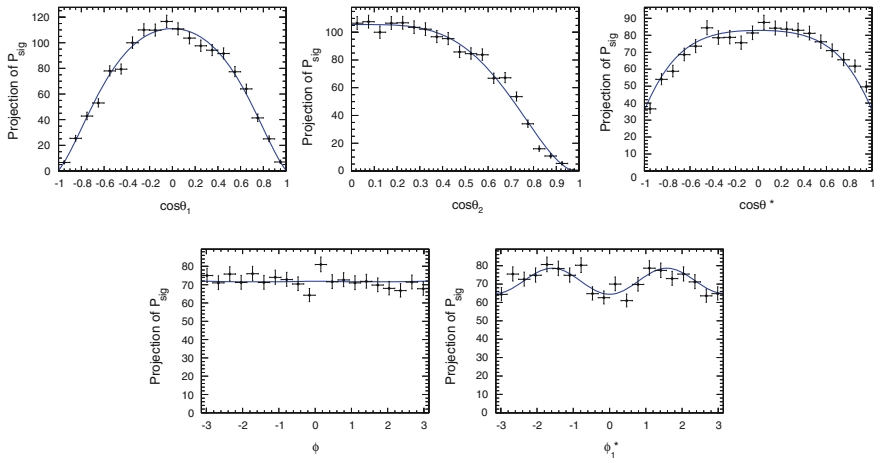

Fig. 4.6 Projections of the five-dimensional signal probability density function along the five angular axes for a 500 GeV Higgs boson: $\cos\theta_1$, $\cos\theta_2$, $\cos\theta^*$ (*top*), and Φ, Φ_1 (*bottom*). The solid line represents the function projection, the markers the values assumed by the MC simulation

where M is the mass of the reconstructed Higgs candidate. The four acceptance functions, $\mathcal{G}_{\theta^*}$, $\mathcal{G}_{\theta_2}$, $\mathcal{G}_{\theta_2}$, and $\mathcal{G}_{\Phi_1}$, have been obtained empirically from fits to the simulation. Projections of $\mathcal{P}_{\text{sig}}$ along the five angular axes can be seen in Fig. 4.6 for events produced in the decay of a 500 GeV Higgs boson. The solid line represents the function projection, and the markers the values assumed by the Monte Carlo simulation.

Whereas the ideal function, $\mathcal{P}_{\text{ideal}}$, is naturally parametrized with the Higgs candidate invariant mass M, the parameters of the four acceptance functions have all been re-parameterized in terms of M only. This was done by fitting eight different Monte Carlo samples each corresponding to a different Higgs mass and then fitting the resulting parameters with either a linear or quadratic function of M.

The probability distribution function for the background was approximated with a product of five one-dimensional functions.

$$\mathcal{P}_{\text{bkg}} = \mathcal{P}_{\theta^*}(\theta^*; M) \cdot \mathcal{P}_{\theta_1}(\theta_1; M) \cdot \mathcal{P}_{\theta_2}(\theta_2; M) \cdot \mathcal{P}_{\Phi}(\Phi; M) \cdot \mathcal{P}_{\Phi_1}(\Phi_1; M)$$

All functions were obtained empirically from fits to the simulation. Projections of $\mathcal{P}_{\text{bkg}}$, for background events with invariant mass of the reconstructed Higgs candidate in the $475 \div 550$ GeV range, can be found in Fig. 4.7. Similar to the case of $\mathcal{P}_{\text{sig}}$, the background Monte Carlo was divided into bins of M and each bin was fit with $\mathcal{P}_{\text{bkg}}$. The parameters from each fit were then fit using either linear or quadratic functions of M.

Figure 4.9 shows the observed distributions of the five angular variables in the 2011 data collected by CMS, in events which pass the analysis preselection. The data

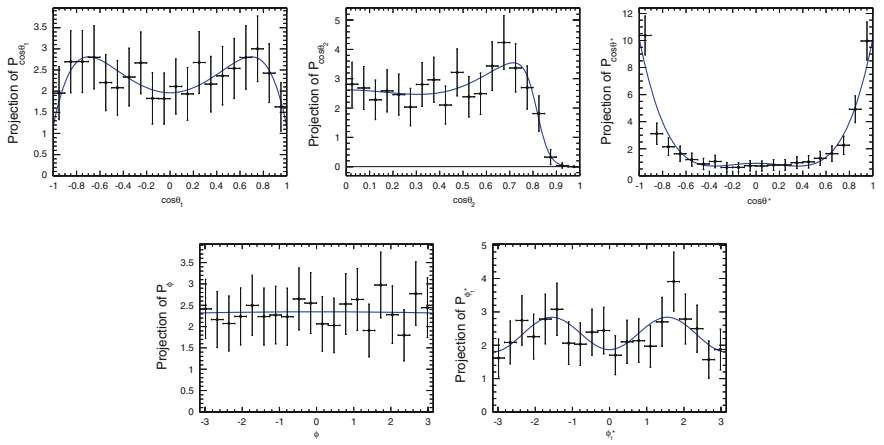

Fig. 4.7 Projections of the five-dimensional signal probability density function along the five angular axes for background events with invariant mass in the 475 ÷ 550 GeV range: $\cos\theta_1$, $\cos\theta_2$, $\cos\theta^*$ (*top*), and Φ, Φ_1 (*bottom*). The solid line represents the function projection, the markers the values assumed by the MC simulation.

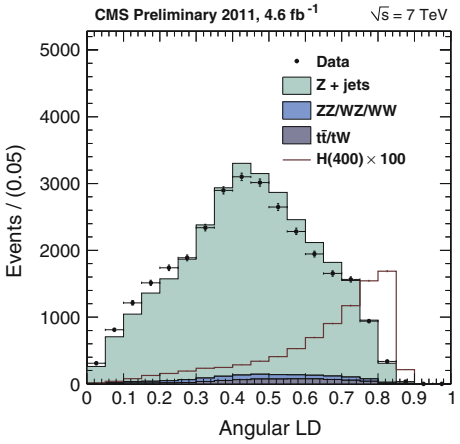

Fig. 4.8 Data-MC comparison for the angular likelihood disscriminant. Events passing preselection in 4.6 fb^{-1} of 2011 data are compared to the expected contribution of the dominant backgrounds in the simulation. The distribution for events coming from a 400 GeV Higgs boson decay, enhanced by a factor 100, is superimposed

are compared to the summed contribution of all MC backgrounds, and an overall good agreement is observed. The expected distributions for events coming from a 400 GeV Higgs boson decay are superimposed, scaled by a factor 100.

Combining $\mathcal{P}_{\text{sig}}$ and $\mathcal{P}_{\text{bkg}}$, we define the likelihood discriminant LD, which is a function of the five helicity angles and of the given event's Higgs candidate reconstructed invariant mass M. An example of the helicity likelihood discriminant is visualized in Fig. 4.8: 4.6 fb^{-1} of data is compared to the expected yield of the backgrounds, and the shape of events originating from a decay of a 400 GeV Higgs boson is overlayed, scaled by a factor 100.

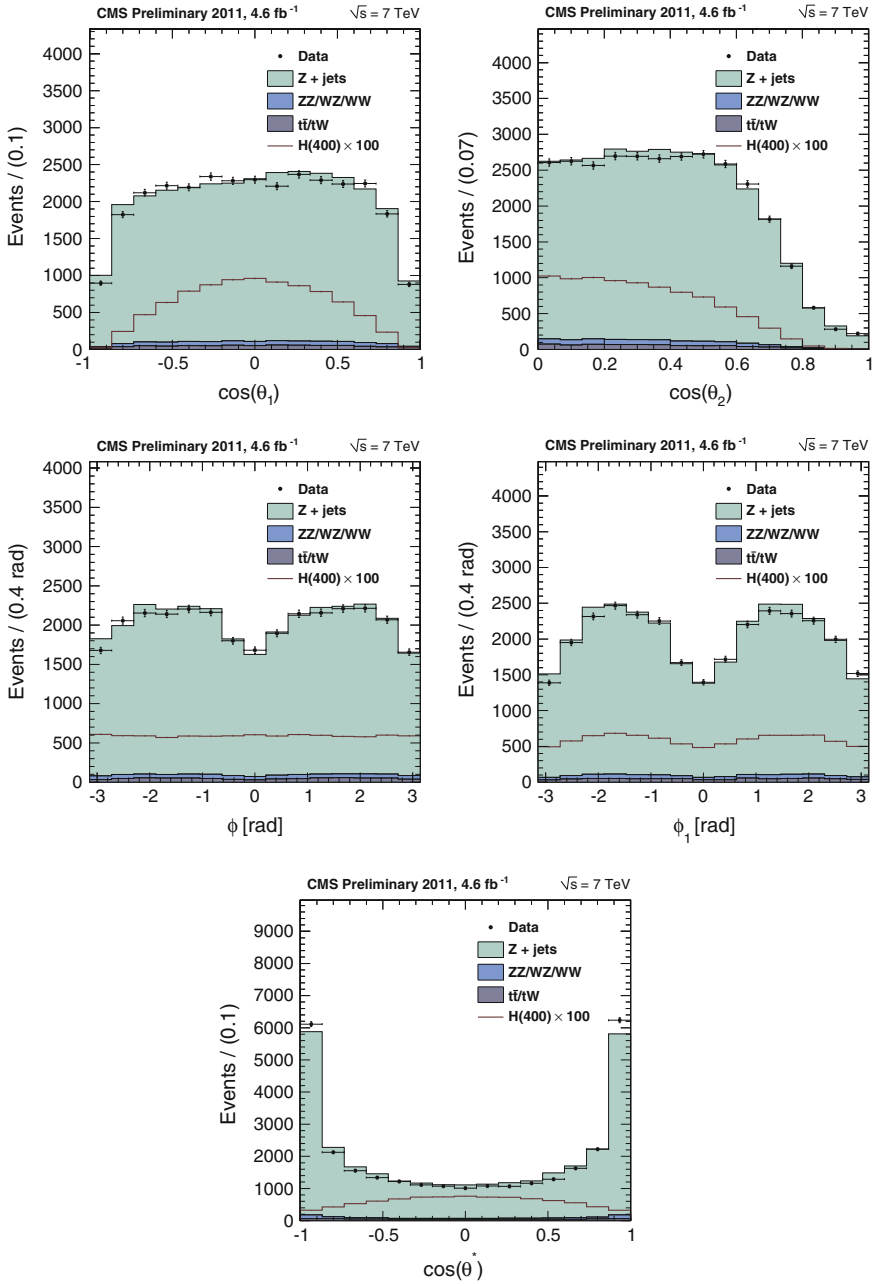

Fig. 4.9 Data-simulation comparisons for the five angular variables. Events passing preselection in 4.6 fb^{-1} of 2011 data are compared to the expected contribution of the dominant backgrounds in the simulation. The distributions for events coming from a 400 GeV Higgs boson decay, enhanced by a factor 100, are superimposed

Compared to a more traditional event selection procedure which exploits purely kinematical (scalar) information, a background discrimination strategy based on angular information presents two main advantages. Firstly, as we have seen in Fig. 4.3, the discrimination power provided by kinematical variables decreases rapidly with the mass of the Higgs boson, as signal event final states will present softer objects. This is less so in the case of angular information, which preserves its discriminating power even in different energetic regimes.[1]

Secondly, an event selection which is founded on angular variables tends to preserve the shape of the background. This can be seen for instance in Fig. 4.10: the left plot shows the resulting dilepton-dijet invariant mass after a kinematic selection, whose thresholds have been optimized for the search of a 400 GeV signal; the right plot instead shows the mass spectrum after an angular analysis. The signal and background rates in the peak region are very similar in the two cases, but the background distribution in the left plot shows a visible deformation, and peaks in a region close to the expected signal. This feature is an artifact of the requirements made on the selection variables, which are correlated to the reconstructed Higgs invariant mass. Such correlation is minimized in the case of angular variables, and as can be seen in the right plot the resulting background shape is not affected. This significantly simplifies the evaluation of the background on data, as it avoids the necessity of a detailed understanding of the correlations between kinematic variables.

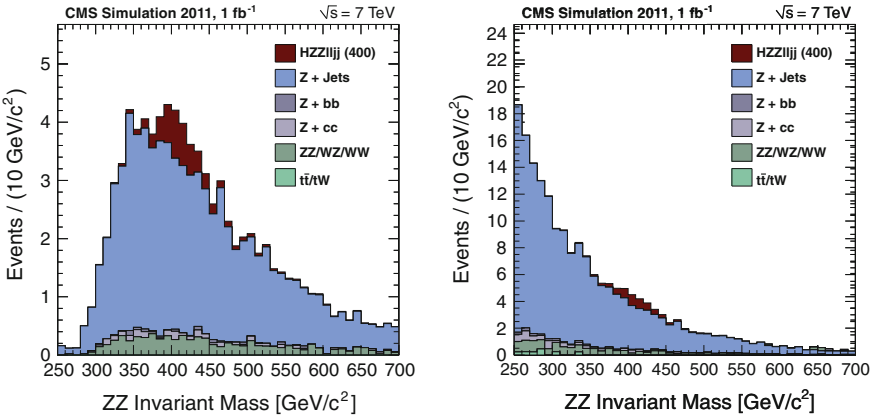

Fig. 4.10 Example of the resulting dilepton+dijet invariant mass after selections based on kinematic (*left*) and angular (*right*) variables, both optimized for the search of a 400 GeV Higgs boson. The contributions of the expected backgrounds in the simulation are shown separately. A hypothetical signal arising from the decay of a 400 GeV Higgs boson is also shown. Yields are scaled to 1 fb^{-1}

[1] In practice a smaller but still present degrading of the discrimination power is still observed, mainly originating on the worse resolution on jet position.

4.4 Kinematic Fit to the Decay Chain

The aim of the analysis is to study the invariant mass spectrum of the dilepton+dijet system, in order to search for signal-like excesses. Signal events are resonant in this variable, as the decay of a massive particle is involved. If no biases are introduced at selection level, signal events will present an invariant mass peak centered at the Higgs boson mass. The significance of the excess depends on the width of the invariant mass peak, which will have two components: an intrinsic one, which depends on the Higgs intrinsic decay width, which can be very large for massive Higgs bosons; and the effect of detector resolutions, which is dominated by the resolution on jets.

In order to contrast the effect of jet resolutions on the invariant mass peak, an additional piece of information may be exploited: jets in signal events are known to stem from the decay of a Z boson, therefore their invariant mass should be compatible with the Z boson mass (m_Z). Hence imposing to the dijet system to have an invariant mass equal to m_Z is expected to improve the final invariant mass resolution for signal events, whereas no significant effect is expected to be introduced in the main backgrounds, which are non-resonant in the dijet system.

The simplest way of imposing the m_Z mass to the dijet system is that of rescaling the dijet quadrimomentum as a whole, modifying its energy in order to obtain the needed mass. This simple procedure already significantly improves the invariant mass of signal events, as can be seen in Fig. 4.11, where the uncorrected dilepton-dijet invariant mass spectrum (black) is compared to the one obtained by applying this rescaling (blue) for three signal mass hypotheses. Though effective, this procedure is clearly suboptimal, as it treats both jets 'democratically', without exploiting the prior knowledge we have on their expected resolutions. We know for instance that jets with higher energies are expected to be reconstructed with higher precision than jets with lower energies, as well as the fact that different detector regions have different expected jet reconstruction performance.

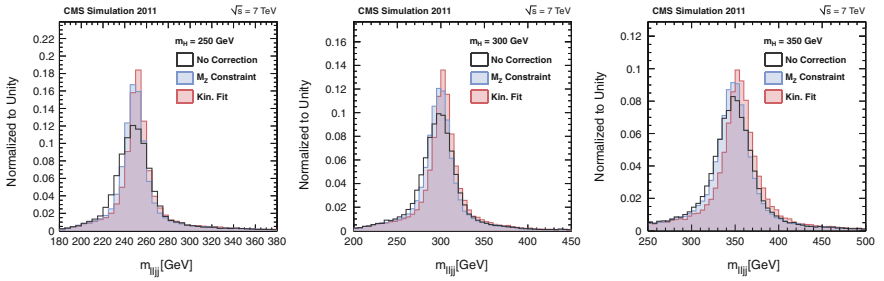

Fig. 4.11 Reconstructed Higgs invariant mass spectra in signal events for three mass hypotheses: 250 GeV (*left*), 300 GeV (*center*), 350 GeV (*right*). The black histogram respresents the uncorrected distribution, the blue histogram is obtained after imposing the Z boson mass to the dijet quadrimomentum, the red one by applying the kinematic fit. All distributions are normalized to unit area

A more powerful approach, that makes use of the information on the individual jets, is to perform a kinematic fit to the dijet system. The fit takes as input the quadrimomenta of the two jets, and makes use of the knowledge of the expected jet transverse momentum and position resolutions, as a function both of transverse momentum and pseudorapidity. It then proceeds in modifying the jet quadrimomenta, compatibly with the expected resolution, fitting the Z mass to the dijet system by minimizing a χ^2 variable of the form:

$$\chi^2 = \left(\frac{\Delta_1}{\sigma_1}\right)^2 + \left(\frac{\Delta_2}{\sigma_2}\right)^2$$

where σ_i is the expected resolution on the i jet, and Δ_i quantifies the deviation from the measured i-jet quadrimomentum.

The kinematic fit further improves the resolution on the final reconstructed Higgs invariant mass peak, as can be seen in Fig. 4.11 (red). For masses heavier than ∼400 GeV, little margin of improvement is expected, because the Higgs intrinsic width becomes the dominant factor in the determination of the invariant mass peak width.

An additional feature of the kinematic fit is that it removes the correlation between the reconstructed dijet and the diboson invariant masses. These two quantities are expected to be correlated because fluctuations in the measured jet momenta, driven by their relatively poor resolutions, will reflect with similar biases in both variables, as can be seen in Fig. 4.12 (left). Once the kinematic fit is applied, the dependance of the diboson invariant mass on jet resolutions is minimized, hence the correlation is removed (right).

As a final remark, while the mass constraint in the analysis has been done by imposing the exact Z boson mass, we have investigated the possibility of introducing

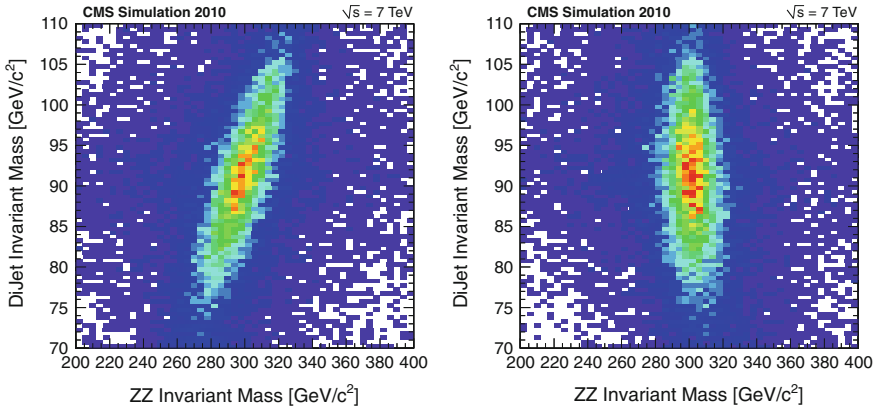

Fig. 4.12 Correlation between the reconstructed dijet invariant mass and the reconstructed diboson invariant mass in signal events with $m_H = 300$ GeV. *Left* before the kinematic fit; *right* after the kinematic fit

a finite width. This has been done by substituting in the fit the Dirac δ-function with a gaussian distribution, centered on the nominal Z boson mass. Two gaussian widths have been tested, of 2 and 5 GeV, but no significant difference in signal efficiency has been observed.

4.5 Categorization

A cardinal point of this analysis is understanding that jet flavour may provide a powerful means of background discrimination. Jets in signal events are produced in hadronic decays of a Z boson, and therefore originate from the hadronization of quark partons. The flavour of quarks in Z decays is almost equally distributed among the five types d, u, s, c, b. The dominant background, as we have seen, is represented by a leptonically-decaying Z boson produced in association with hard jets, a process in which gluon radiation is expected to play a major role. In addition to gluons, u and d quarks, valence partons of the protons, dominate the jet production associated with the Z.

The validity of above statements may be verified by looking at Fig. 4.13, which shows, for the leading jet on the left and the subleading jet on the right, the PDG identification number (PDG ID) of partons matched to jets in signal and background events, after a loose kinematic selection. The PDG ID of quarks follows the mapping scheme: $d = 1$, $u = 2$, $s = 3$, $c = 4$, $b = 5$, $t = 6$, and anti-quarks have opposite PDG ID. Gluons are assigned a value of 21.

As can be seen in the figure, the kinematic selection in signal selects only quarks jets. It is furthermore orthogonal to parton flavour, as the jets which pass the selection are equally shared between all available quark flavours (excluding the top quark

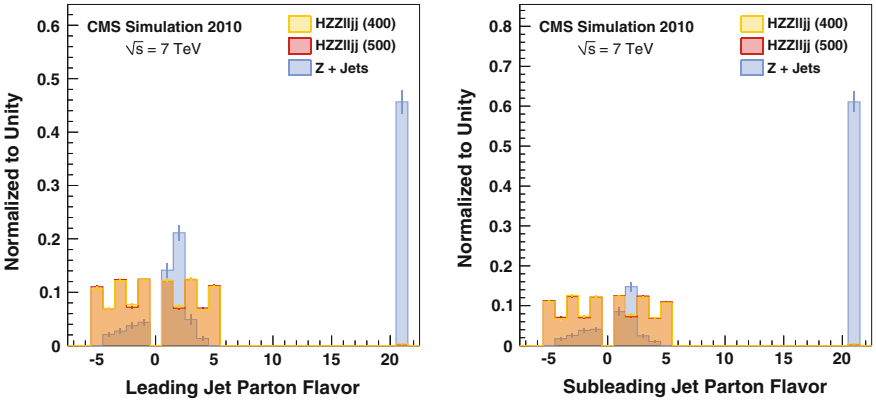

Fig. 4.13 PDG identification number for partons matched to jets in events passing loose kinematic requirements. Signal events for two Higgs masses (400 GeV in yellow and 500 GeV in red) are compared to the main background (Z+jets, blue). Distributions are normalized to unity

which is energetically forbidden). The observed enhancement of down-type quarks (d, s, b) is a direct consequence of the asymmetric coupling of the Z boson.

The jet flavour population for background Z+jets events is radically different. More than 45 % (60 %) of the selected leading (subleading) jets originate from the hadronization of a gluon. Also the quark population shows some differences: the observed u and d enhancement is mirroring the proton valence quark parton density functions (largest u contribution, and the next largest d). The contribution of heavy flavours, b in particular, is small in background, while it is about 22 % is signal.

Therefore, the main features which discriminate signal from background is the relatively large contribution of heavy flavour quarks (b and c) and absence of gluons. We take advantage of both features in the analysis by pursuing two directives: isolate heavy flavours, in order to identify an event sub-population in which only a fraction of the signal is present, but with a higher expected purity, as backgrounds are less present; limit the background gluon infiltration, trying to affect signal efficiency in a minor way.

In order to identify heavy flavour jets we will use a b-tagging discriminant which will be described in the following, whereas the gluon jet rejection is performed with the likelihood ratio introduced in Sect. 3.5. The analysis will therefore be split into four categories:

- 2 b-tag category: events in which both jets are positively identified as originating from a b quark hadronization;
- 1 b-tag category: events in which one jet is positively identified as a b-jet;
- 0 b-tag category: no jet is identified as b, and the jet pair is not incompatible with a light quark hypothesis;
- gluon-tag category: events in which jets are likely to originate from gluons.

We expect the 2 b-tag category to have the highest signal purity, but low efficiency, and the 0 b-tag category to have the highest signal efficiency, but large background yields. The gluon-tag category is dominated by background contributions.

The category of an event is defined by the values assumed on the two jets by the b-tagging algorithm known as Track Counting High Efficiency (TCHE) [8]. The TCHE variable is defined as the second-to-highest impact parameter significance S among all tracks associated with the given jet. Therefore requiring the jet to have TCHE $> x$ is equivalent to requiring the jet to be associated to at least two tracks with impact parameter significance $S > x$. A jet is considered to have a *medium* (*loose*) b-tag if it has TCHE > 3.3 (1.7). The expected mistag probabilities of these working points is 1 % for medium and 10 % for loose. Figure 4.14 shows the data-MC comparison for this tagger, for all jets passing preselection.

Recent measurements [9] have shown that the TCHE tagger has slightly worse performance than what is expected from the Monte Carlo simulation. One can define a scale factor BSF_b (BSF_{lq}) as the ratio between the data and Monte Carlo tagging efficiency for b/c (light quark) jets. In both cases the efficiency measurements have been performed in jet transverse momentum and pseudorapidity ranges, therefore the scale factors will vary accordingly. The BSF_b scale factors are on average equal to

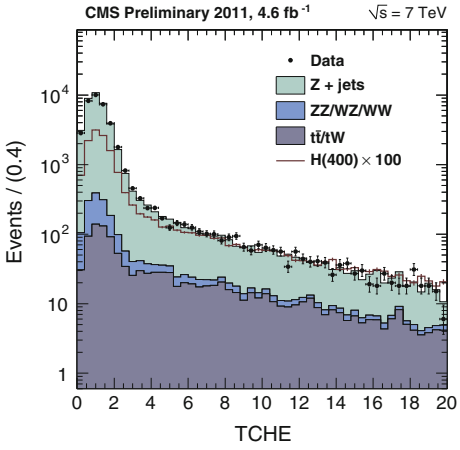

Fig. 4.14 Track Counting High Efficiency (TCHE) jet tagger for jets in events passing the analysis preselection in 4.6 fb^{-1} of 2011 data, compared to the expected contribution of the dominant backgrounds in the simulation. The distribution for events coming from a 400 GeV Higgs boson decay, enhanced by a factor 100, is superimposed

95 % (93%) for the loose (medium) working points, whereas the ratio of mistag rates BSF_{lq} are on average equal to 1.11 (1.21) for the loose (medium) working points.

An event is placed in the 2 *b*-tag category if one jet is identified with medium and the other is identified with loose requirements. Events which fail these criteria but still contain at least one jet which satisfies the loose criterion are placed in the 1 *b*-tag category. In order to simulate the effect that the lower efficiency found in data would have on Monte Carlo events, a simple simulation is used: on a jet-by-jet basis, a random number generator is used to stochastically modify the value of the given jet's *b*-tag. In this way, Monte Carlo jets which are correctly matched to a heavy flavour jet have a certain probability of being 'downgraded' to a lower *b*-tag category (medium to loose, or loose to not tagged), whereas light quark jets have a certain probability of being 'upgraded' to a higher category (not tagged to loose, or loose to medium), so that the single-tag efficiency in Monte Carlo events corresponds to the one observed in the data. By applying this mechanism on a jet-by-jet basis, no event reweighing is introduced. Rather, given the outcome of the randomization on the selected jet pair, the event may migrate to a different *b*-tag category.

Events which fail the *b*-jet identification requirements which would place them in the single or doubly-tagged categories, are then split between the 0 *b*-tag and the gluon-tag categories, by looking at the product of the two jets' quark-gluon (Q-G) likelihood discriminants, as defined in Sect. 3.5. Figure 4.15 shows the expected distributions of the leading jet Q-G discriminant (left), the subleading jet Q-G discriminant (center), and the product of the two (right), for signal events arising from the decay of a 400 GeV Higgs boson, compared to the main background of Z+jets. Events are required to have a reconstructed Higgs candidate invariant mass between 376 and 440 GeV, and all distributions are normalized to unit area. The data-MC comparison of the same variables, but after preselection requirements only, are shown in Fig. 4.16.

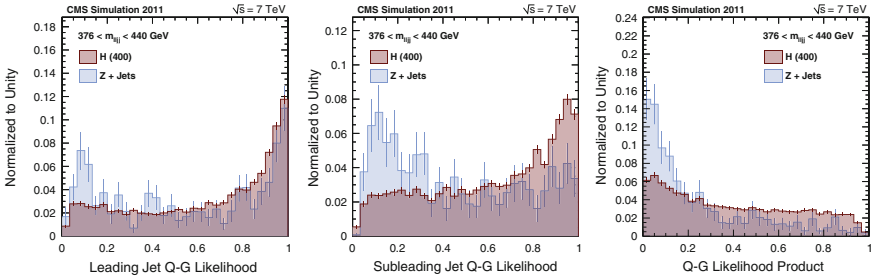

Fig. 4.15 Quark-Gluon likelihood discriminant distributions for events produced in the decay of a 400 GeV Higgs boson (*hashed*) compared to the Z+jets background (*solid*), for events passing preselection and lie in the 376–440 GeV invariant mass window

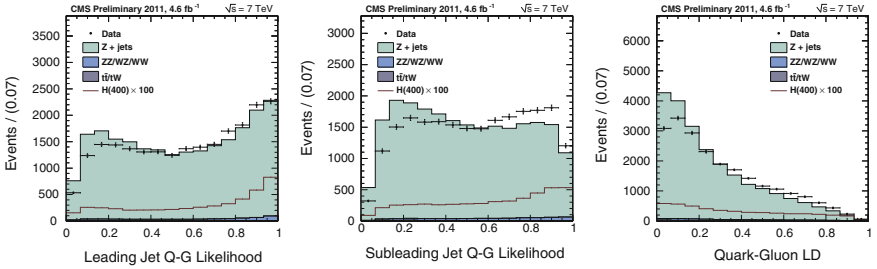

Fig. 4.16 Quark-Gluon likelihood discriminant distributions. Events passing preselection in 4.6 fb^{-1} of 2011 data are compared to the expected contribution of the dominant backgrounds in the simulation. The distribution for events coming from a 400 GeV Higgs boson decay, enhanced by a factor 100, is superimposed

Events with Q-G likelihood product less than 0.1 are rejected and placed in the gluon-tag category. This requirement has an efficiency of about 85 % on signal events, and reduces the Z+ jets background by about 34, 43, 50, and 56 % at m_{lljj} masses around 250, 300, 400, and 500 GeV.

In addition to being a means of background discrimination, the requirement on the Q-G discriminant also improves the invariant mass resolution in signal events. This is because, by selecting events in which the jet pair has composition properties which are compatible with the expectations for high-p_{T} quark jets, events with misreconstructed jets and events in which signal self-combinatorics leads to the choice of the incorrect jet pair are discarded. This may be seen in Fig. 4.17, where the dilepton-dijet invariant mass for events passing (red) and failing (blue) the requirement on the product of the two jet's Q-G likelihood discriminant, for three hypothetical Higgs boson masses: 300 (left), 400 (center), and 500 GeV (right)

In general, an event will have numerous jets, therefore multiple jet pairs. The analysis selection algorithm scans all possible jet pairs, and verifies if the given pair passes the selection requirements, which will depend on the pair's b-tag values, and on the invariant mass of the resulting reconstructed Higgs candidate, as will be described

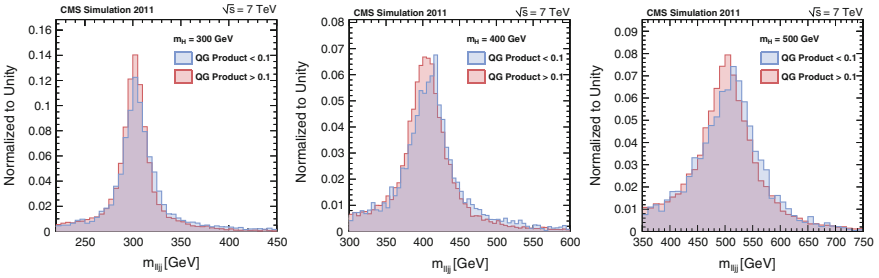

Fig. 4.17 Reconstructed Higgs candidate invariant mass distribution for events passing (*red*) and failing (*blue*) the Q-G likelihood requirement, for three hypothetical Higgs boson masses: 300 (*left*), 400 (*center*), 500 GeV (*right*). All distributions are normalized to unit area

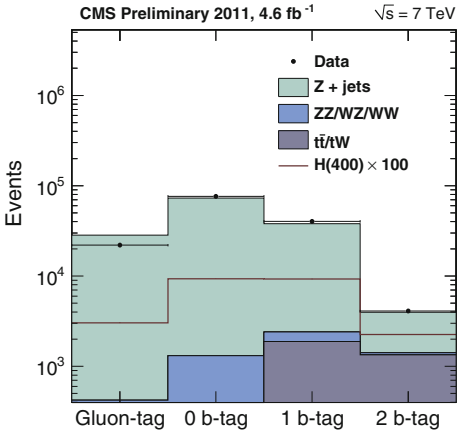

Fig. 4.18 Distribution of flavour tagging categories in events passing the analysis preselection in 4.6 fb^{-1} of 2011 data, compared to the expected contribution of the dominant backgrounds in the simulation. The distribution for events coming from a 400 GeV Higgs boson decay, enhanced by a factor 100, is superimposed

in the following section. If an event presents more than one pair which meets the requirements, the pair which belongs to the highest *b*-tag category is selected, in order to favor the highest purity samples. If the primacy is shared by more than one pair, the pair with an invariant mass closest to the nominal Z boson mass is selected. This ensures univocal classification of events, and therefore the statistical independence of the samples identified by the categories. Figure 4.18 shows the subdivision of events in the analysis categories, and as can be seen the background composition can vary significantly among them.

4.6 Selection Optimization

The main discrimination between signal and backgrounds is provided by the angular likelihood discriminant variable, on which we intend to impose a simple 'cut' requirement, by selecting events which assume values larger than a given

Table 4.7 Results of the angular likelihood discriminant threshold optimization, in the three b-tag categories

b-tag Category	Optimal AngularLD Threshold
0	$0.55 + 0.00025 \cdot m_H$ (GeV)
1	$0.302 + 0.000656 \cdot m_H$ (GeV)
2	0.5

threshold. In order to obtain the best performance, a simple optimization was conducted. This was done separately in the three b-tag categories, for the different expected backgrounds and signal purities necessarily mirror in different optimal selections.

The complexity of the selection mechanism and the correlations between the different b-tag analysis categories imposes some necessary simplifications in the optimization procedure. Each event passing preselection requirements is kept, and in order to solve the ambiguity which derives from the multiple jet pairs in the event, the pair with mass closest to the nominal Z boson mass is selected, neglecting b-tagging considerations during this step. The event is then classified according to the b-tag value of the two chosen jets, and the optimization is conducted on these events, separately in the three categories.

The optimization procedure was accomplished by analyzing the distribution of the angular likelihood discriminant variable. For a given hypothetical signal mass, and in each b-tag category, the optimal selection threshold is defined as that which minimizes the exclusion upper limit (UL) of a possible presence of signal in addition to background events, for an integrated luminosity of 1 fb^{-1}. For this purpose, the UL was computed with a Bayesian approach, with a credibility interval of 95 %, and a flat prior on the signal cross section. It must be noted that adopting this criterion does not give dramatically different results neither with respect to other exclusion recipes (such as frequentist approaches), nor to discovery-based strategies, for the analysis is expected to have large event yields.

The optimization was carried out at six pivotal hypothetical signal masses: 250, 300, 350, 400, 450, and 500 GeV. In each b-tag category the trend of the optimal angular likelihood discriminant threshold was studied as a function of the signal mass, and linear dependancies were found. They were therefore fitted with linear functions of the Higgs mass, in order to find a smooth functional dependance on the mass, and the results of this fit are summarized in Table 4.7. It must be noted that, as no significant deviation from a constant threshold was found in the 2-tag category, the requirement of 0.5 was adopted for all masses.

4.7 Missing Transverse Energy Significance

The main objective of the angular analysis is that of discriminating between resonant (signal) and non-resonant (background) events. After applying the selection, the background yield is dominated by the $Z + 2$jets contribution in the 0- and 1-tag

categories, whereas a significant role is played by top (both $t\bar{t}$ and tW) events in the 2-tag category. To contrast this source of background, an additional handle is found to provide means of discrimination: missing transverse energy.

In order to constitute a background to the $H \to ZZ \to \ell^+\ell^-q\bar{q}$ channel, top events must present two oppositely-charged, high-p_T leptons. Therefore the main contribution comes from the semileptonic decay channels:

$$t(\to \ell\nu b)\bar{t}(\to \ell\nu\bar{b})$$

$$t(\to \ell\nu b)W(\to \ell\nu) + \geqslant 1 \text{ jet}$$

Because of the presence of the final state neutrinos, both these reactions will produce in the detector significant amounts of missing transverse energy. This feature should be absent in signal events, which, at leading order, have no neutrinos in the final state, and therefore no source of true missing transverse energy.

This can be seen in Fig. 4.19 (left) where the Particle Flow missing transverse energy (PF$\not\!\!E_T$) distribution is shown for three hypothetical signal masses (300, 400, and 500 GeV), compared to what is obtained on top events. In general, one can observe that:

- top-quark events have larger values of PF$\not\!\!E_T$ than signal events;
- the average amount of PF$\not\!\!E_T$ in signal events increases with the Higgs boson mass.

The latter effect is easily explained: the main source of missing transverse energy in signal events is constitued by resolution effects, in large part originating from the reconstruction of the hadronic decay of the Z boson. As the decaying Higgs boson

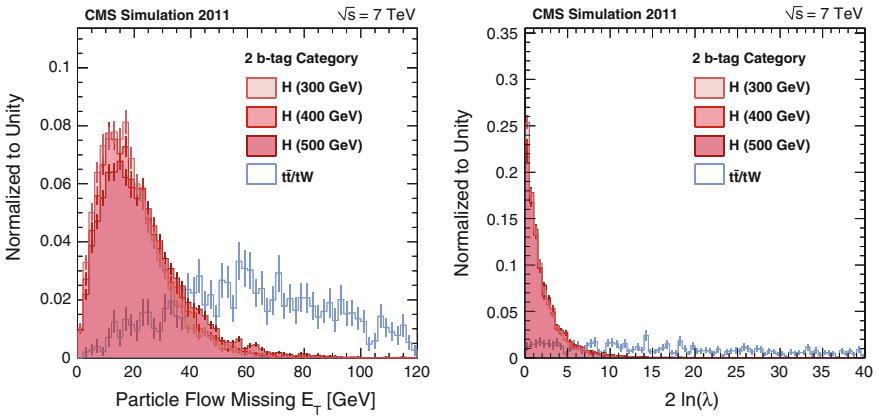

Fig. 4.19 Missing transverse energy variables after loose kinematic cuts in the 2-tag category: Particle Flow Missing E_T (*left*) and missing transverse energy significance (*right*). Figures compare the shapes obtained in signal events (*red shades*) for three hypothetical masses (300, 400, and 500 GeV) to the ones obtained on events containing top quarks (*blue shade*). All distributions are normalized to unit area. Top events have a large overflow in both cases

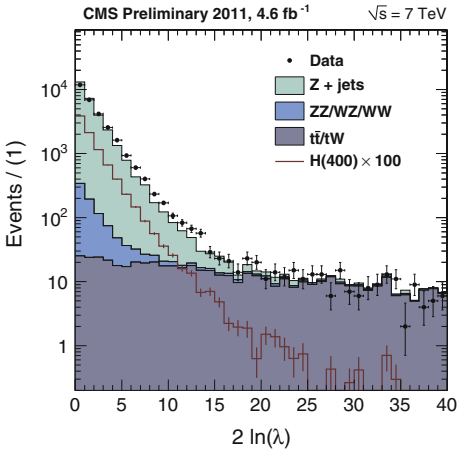

Fig. 4.20 Missing transverse energy significance ($2\ln\lambda$) distribution in events passing the analysis preselection in 4.6 fb^{-1} of 2011 data, compared to the expected contribution of the dominant backgrounds in the simulation. The distribution for events coming from a 400 GeV Higgs boson decay, enhanced by a factor 100, is superimposed

becomes heavier, the average energy of the decay products will be larger, so that the *absolute* uncertainty on the jet momentum evaluation will increase, which mirrors in higher values of PF$\not\!\!E_{\mathrm{T}}$. This implies that if one would base a selection on PF$\not\!\!E_{\mathrm{T}}$, for instance requiring events to present a level of PF$\not\!\!E_{\mathrm{T}}$ not larger than a given threshold, the loss in signal efficiency would increase with the signal mass. This is far from ideal, for non-resonant backgrounds are expected to have a larger contribution at lower invariant masses, for energetic considerations.

A more performant approach is that of considering the significance of the reconstructed PF$\not\!\!E_{\mathrm{T}}$. This is done by defining a likelihood-ratio discriminant λ, which, through the knowledge of the expected resolutions on the event's reconstructed jets, compares the hypothesis that the event presents a true missing transverse energy ($\not\!\!E_{\mathrm{T}}$) equal to the measured PF$\not\!\!E_{\mathrm{T}}$, to the null hypothesis ($\not\!\!E_{\mathrm{T}} = 0$). The distribution of $2\ln\lambda$ is shown in Fig. 4.19 (right), for signal and top events. As can be seen, the signal distribution is barely sensitive to the value of the Higgs mass.

We therefore introduce an additional requirement, in the 2-tag category only, that the event PF$\not\!\!E_{\mathrm{T}}$ satisfies the requirement $2\ln\lambda < 10$. This ensures high efficiency (>97 %) on signal events, and is expected to reject more than 50 % of the top background. The observed distribution of $2\ln\lambda$ on 4.6 fb^{-1} of data passing preselection requirements is shown in Fig. 4.20.

4.8 Summary of Selection Requirements and Yields

Table 4.8 summarizes the selection requirements for the $\mathrm{H} \to \mathrm{ZZ} \to \ell^+\ell^-\mathrm{q\bar{q}}$ analysis. The main discrimination power is provided by the angular likelihood discriminant: events are required to satisfy a threshold which depends on the reconstructed Higgs invariant mass, as shown in the table. The dependance on the Higgs mass is

different in the three categories, as found in the optimization procedure described in Sect. 4.6. Additional selections are enforced in specific categories only: namely the quark-gluon discrimination requirement in the 0-tag category, and the PF$\not{E}_{\mathrm{T}}$ significance ($2\ln\lambda$) requirement in the 2-tag category.

Table 4.9 summarizes the total expected yields per fb^{-1} of integrated luminosity, in the three b-tag categories, for background processes and for nine hypothetical Higgs boson masses. Yields are integrated over the full analyzed invariant mass range $183 \div 800$ GeV. Uncertainties derive from Monte Carlo statistics.

Tables 4.10, 4.11 and 4.12 instead show, respectively for the three b-tag categories, the expected yields and signal efficiencies per fb^{-1} of integrated luminosity in the -6 %/+10 % m_{lljj} range about the nominal Higgs boson mass, for six hypothetical signal masses. Expected event yields in the electron and muon channel are quoted separately. Backgrounds are here shown broken up in their different contributions.

Table 4.8 Summary of selection requirements in the $\mathrm{H} \to \mathrm{ZZ} \to \ell^+\ell^-\mathrm{q\bar{q}}$ analysis, split in the three analysis categories

0 b-tag	1 b-tag	2 b-tag
	Lepton HLT/ID/Isolation and $p_{\mathrm{T}} > 40/20$ GeV	
	Jet $p_{\mathrm{T}} > 30$ GeV and $\lvert\eta\rvert < 2.4$	
	$70 < m_{\ell\ell} < 110$ GeV	
	$75 < m_{jj} < 105$ GeV	
	Kinematic fit to the decay chain	
aLD $> 0.00025 \cdot m_H + 0.55$	aLD $> 0.000656 \cdot m_H + 0.302$	aLD > 0.5
QG LD > 0.1		$2\ln\lambda < 10$

The angular likelihood discriminant (aLD) requirement depends on the reconstructed Higgs boson candidate mass (m_H). The quark-gluon likelihood discriminant (QG LD) and PF$\not{E}_{\mathrm{T}}$ significance ($2\ln\lambda$) are enforced respectively only in the 0 b-tag and the 2 b-tag categories

Table 4.9 List of expected background and signal yields with 1 fb^{-1} of data after all selections and within the ZZ invariant range $183 \div 800$ GeV

	0 b-tag (events·fb)	1 b-tag (events·fb)	2 b-tag (events·fb)
$\mu^-\mu^+jj$ background	350.4 ± 5.9	345.2 ± 6.0	22.2 ± 1.6
$\mathrm{e}^-\mathrm{e}^+jj$ background	279.9 ± 5.2	286.4 ± 5.4	21.9 ± 1.6
200 GeV	2.80 ± 0.40	3.56 ± 0.51	0.78 ± 0.21
250 GeV	5.67 ± 0.80	5.06 ± 0.71	1.60 ± 0.42
300 GeV	6.18 ± 0.88	5.26 ± 0.73	1.90 ± 0.50
350 GeV	6.84 ± 1.00	5.87 ± 0.84	2.31 ± 0.60
400 GeV	5.65 ± 0.81	5.00 ± 0.69	2.10 ± 0.53
450 GeV	3.76 ± 0.56	3.47 ± 0.49	1.54 ± 0.38
500 GeV	2.36 ± 0.36	2.27 ± 0.33	1.04 ± 0.26
550 GeV	1.45 ± 0.23	1.46 ± 0.23	0.69 ± 0.17
600 GeV	0.59 ± 0.15	0.61 ± 0.15	0.29 ± 0.11

Table 4.10 Expected yields of signal (signal efficiency is shown in parentheses) and background with 1 fb^{-1} based on simulation in the 0-tag category

m_{H} (GeV)	Signal	Z+Jets	Diboson	$t\bar{t}/tW$	Total BG
250	2.2/2.5 (2.1 %/2.3 %)	81/99	2.8/3.2	0.92/0.97	85/105
300	2.4/2.5 (3.0 %/3.1 %)	40/53	1.7/2.4	0.25/0.36	42/55
350	2.5/2.5 (3.4 %/3.5 %)	21/28	1.3/1.5	0.11/0.1	23/29
400	1.8/1.8 (3.3 %/3.3 %)	11/15	0.74/0.84	0.0076/0.079	12/16
450	1.1/1.1 (2.9 %/3.0 %)	8.3/7.4	0.59/0.55	0.03/0.0067	8.9/8
500	0.61/0.67 (2.6 %/2.9 %)	3.3/3.7	0.32/0.4	0/0.0046	3.6/4.1

In each case the two numbers show $2e2j/2\mu2j$ expectations. For each considered signal mass (m_{H}), events are counted only in the $-6\,\%/+10\,\%$ window about the nominal Higgs mass

Table 4.11 Expected yields of signal (signal efficiency is shown in parentheses) and background with 1 fb^{-1} based on simulation in the 1 b-tag category

m_{H} (GeV)	Signal	Z+Jets	Diboson	$t\bar{t}/tW$	Total BG
250	1.7/1.9 (1.6 %/1.8 %)	69/81	2.4/3	7.4/8.4	79/93
300	1.7/1.9 (2.1 %/2.4 %)	39/48	1.7/1.9	2.8/3.8	44/54
350	1.9/2.1 (2.6 %/2.83 %)	23/30	0.91/1.2	1/0.93	25/32
400	1.5/1.6 (2.6 %/2.8 %)	14/19	0.71/0.75	0.34/0.23	15/20
450	0.93/0.98 (2.6 %/2.7 %)	11/11	0.43/0.5	0.18/0.026	12/11
500	0.55/0.58 (2.4 %/2.5 %)	7.6/5.6	0.36/0.49	0.065/0.051	8/6.1

In each case the two numbers show $2e2j/2\mu2j$ expectations. For each considered signal mass (m_{H}), events are counted only in the $-6\,\%/+10\,\%$ window about the nominal Higgs mass

Table 4.12 Expected yields of signal (signal efficiency is shown in parentheses) and background with 1 fb^{-1} based on simulation in the 2-tag category

m_{H} (GeV)	Signal	Z+Jets	Diboson	$t\bar{t}/tW$	Total BG
250	0.71/0.79 (0.66 %/0.74 %)	5.3/4.8	0.41/0.35	1.2/1.2	6.8/6.3
300	0.82/0.8 (1.0 %/1.0 %)	2.7/3.1	0.26/0.33	0.48/0.76	3.4/4.2
350	0.9/0.95 (1.2 %/1.3 %)	1.3/1.5	0.2/0.22	0.14/0.19	1.7/1.9
400	0.7/0.74 (1.3 %/1.3 %)	0.45/1.3	0.1/0.16	0.022/0.0084	0.58/1.5
450	0.46/0.49 (1.3 %/1.3 %)	0.63/1.3	0.097/0.16	0.0042/0.048	0.73/1.5
500	0.29/0.3 (1.3 %/1.3 %)	0.87/0.8	0.1/0.089	0/0.062	0.97/0.95

In each case the two numbers show $2e2j/2\mu2j$ expectations. For each considered signal mass (m_{H}), events are counted only in the $-6\,\%/+10\,\%$ window about the nominal Higgs mass

4.9 Background Estimation

As the adopted event selection does not depend in any way on the hypothetical Higgs boson mass, but rather on the reconstructed dilepton-dijet invariant mass m_{lljj}, after the final selection is applied to the data we have a total six m_{lljj} distributions, one per b-tag category (0,1,2) times one per lepton flavour (e, μ). We analyze these distributions for different hypothetical Higgs boson signals, as the selection is expected to

yield different efficiencies for different hypothetical Higgs masses. The distribution of background events, though, is unique in each channel.

We do not intend to fully rely on the simulation to estimate the expected background yields after applying the event selection, therefore we measure the background directly from the data. This is done by analyzing the dijet invariant mass (m_{jj}) in an extended range, and splitting events in two separate regions:

- events which pass the nominal selection ($75 < m_{jj} < 105$ GeV) are placed in the *signal region*, and are of interest for the final analysis results;
- events which fail the analysis selection because of the value of m_{jj} are kept if they lie in the broader invariant mass interval of $60 < m_{jj} < 130$ GeV, and define the *sideband region*.

The thresholds which define the sideband region are the result of a compromise: they are tight enough to ensure that the kinematics of sideband events is similar to the ones in the signal region, and wide enough so that the available amount of data is comparable in the two regions.

As Higgs events present the hadronic decay of a Z boson, the sideband region is reasonably depleted of signal. On the other hand, most of the backgrounds are not resonant in the dijet invariant mass variable (the only exception is the direct Z pair production), and are therefore expected to populate the signal and sideband region in similar fashion. Figure 4.21 shows the obtained sideband region m_{lljj} distributions for the three b-tag categories: data from 4.6 fb^{-1} of 2011 collisions are compared to the expected contribution of the dominant backgrounds.

Even if the event kinematics, and therefore the resulting m_{lljj} distributions, are similar between the signal and sideband regions, they are not identical. In order to use sideband events to estimate the background yield in the signal region, the former have to be corrected to take into account this difference. This is done by accessing the Monte Carlo simulation: for each b-tag category, the shapes of the m_{lljj} distributions in the signal and sideband regions are compared, and a bin-to-bin ratio $\alpha(m_{lljj})$ is computed. The value of $\alpha(m_{lljj})$, in the three b-tag categories, is shown in Fig. 4.22:

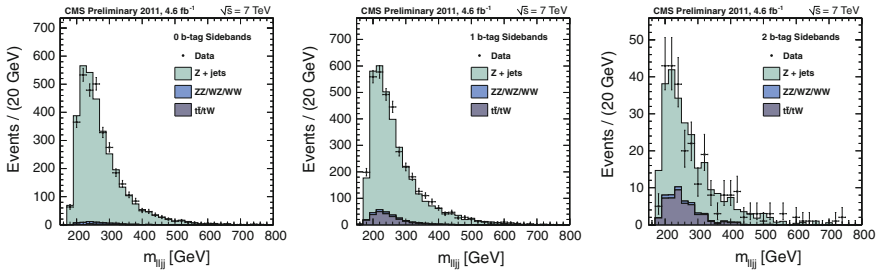

Fig. 4.21 Distributions of the dilepton-dijet invariant mass (m_{lljj}) for events in the sideband region in the three categories: 0-tag on the *left*, 1-tag in the *center*, 2-tag on the *right*. Events passing the selection in 4.6 fb^{-1} of 2011 data are compared to the expected contribution of the dominant backgrounds in the simulation

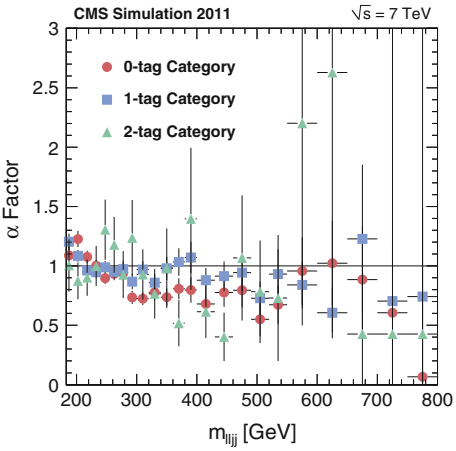

Fig. 4.22 Sideband region correction factor (α) as a function of the reconstructed dilepton-dijet invariant mass (m_{lljj}) in the three *b*-tag categories: 0-tag (*circles*), 1-tag (*squares*), 2-tag (*triangles*)

it is not very different from unity, therefore the entity of the correction (and of the possible uncertainty it implies) is small.

The comptutation of the α ratio enables us to estimate the number of background events N_{bkg} at a given m_{lljj} invariant mass. This is done by taking the number of observed events in the data sidebands (N_{sb}) and correcting it with the following formula:

$$N_{\text{bkg}}(m_{lljj}) = N_{\text{sb}}(m_{lljj}) \times \frac{N_{\text{bkg}}^{\text{MC}}(m_{lljj})}{N_{\text{sb}}^{\text{MC}}(m_{lljj})} \equiv N_{\text{sb}}(m_{lljj}) \times \alpha(m_{lljj})$$

where the corresponding Monte Carlo yields are indicated with a superscript.

The resulting α-corrected m_{lljj} sideband distribution constitutes our data-driven estimate of the signal region background yield. In order to minimize the effect of statistical fluctuations originating from the limited amount of data, the distribution is fitted with an empirical functional form, which was found to successfully describe the shape obtained on the simulation: the product of a Fermi-Dirac, for the steep low-mass turn-on, and a Crystal-Ball function, for the kinematical peak around 200 GeV and the high-mass tail.

The function has a total of six floating parameters: two from the Fermi-Dirac function (the equivalent temperature and the position of the transition), and four from the Crystal-Ball (the mean and width of the gaussian, the gaussian/power-law transition position, and the exponent of the power-law) . The function is used in an unbinned, maximum-likelihood fit to the sideband distribution in the simulation, with all parameters free to vary, taking advantage of the high number of available Monte Carlo events. It is then fitted to the the sideband distribution observed in the data, but only two Crystal-Ball parameters are kept floating in the fit procedure (the gaussian width and the power-law exponent), whereas all other parameters are fixed to the values obtained on the simulation.

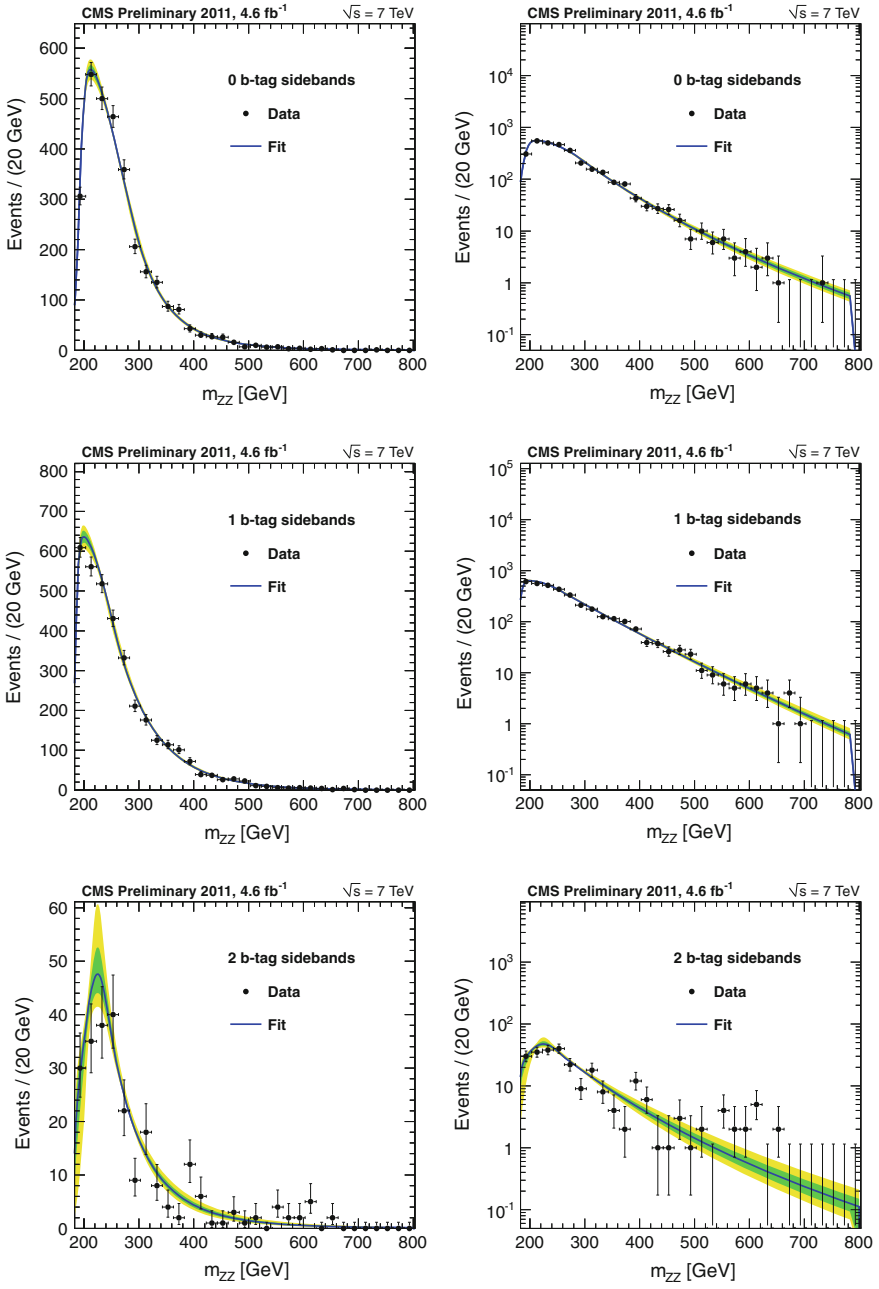

Fig. 4.23 Results of the unbinned, maximum-likelihood fits to the alpha-corrected data sidebands: 0 *b*-tag category on the top, 1 *b*-tag in the middle, and 2 *b*-tag on the bottom. Plots on the *left (right)* column are in linear (logarithmic) scale. The result of the fit is shown with a blue curve, and 68 % (95 %) fit uncertainty bands are shown with a *green (yellow) shade*

The results of the unbinned, maximum-likelihood fits to the α-corrected m_{lljj} sideband distribution are shown in Fig. 4.23: 0 b-tag category on the top, 1 b-tag in the middle, and 2 b-tag on the bottom. Plots on the left (right) column are in linear (logarithmic) scale. The result of the fit is shown with a blue curve, and 68 % (95 %) fit uncertainty bands are shown with a green (yellow) shade. These represent the background estimate for the signal region events in the three b-tag categories, as will be shown in Chap. 6.

References

1. Alwall, J., Demin, P., de Visscher, S., Frederix, R., Herquet, M., Maltoni, F., Plehn, T., Rainwater, D., Stelzer, T.: MadGraph/MadEvent v4: the new web generation. JHEP **09**, 028 (2007). http://arxiv.org/abs/0706.2334 doi:10.1088/1126-6708/2007/09/028
2. Sjöstrand, T., Mrenna, S., and Skands, P. PYTHIA 6.4 physics and manual. JHEP **05**, 026 (2007).doi:10.1088/1126-6708/2006/05/026
3. Nason, P.: A New Method for Combining NLO QCD with Shower Monte Carlo Algorithms. JHEP **11**, 040 (2004). doi:10.1088/1126-6708/2004/11/040
4. Frixione, S., Nason, P., Oleari, C.: Matching NLO QCD Computations with Parton Shower Simulations: the POWHEG method. JHEP **11**, 070 (2007). doi:10.1088/1126-6708/2007/11/070
5. Alioli, S., Nason, P., Oleari, C., Re, E.: NLO vector-boson production matched with shower in POWHEG. JHEP **07**, 06 (2008). doi:10.1088/1126-6708/2008/07/060
6. Khachatryan et al.: Measurements of inclusive W and Z cross sections in pp collisions at $\sqrt{s} = 7$ TeV. J. High Energy Phys. **2011**, 1 (2011). doi:10.1007/JHEP01(2011)080
7. Gao, Y., Gritsan, A. V., Guo, Z., Melnikov, K., Schulze, M., Tran, N. V.: Spin determination of single-produced resonances at hadron colliders. Phys. Rev.D **81**, 075022 (2010).doi:10.1103/PhysRevD.81.075022
8. CMS Collaboration.: Commissioning of b-jet identification with pp collisions at $\sqrt{s} = 7$ TeV. CMS Phys. Anal. Summary, CMS-PAS-BTV-10-001 (2010). Available from: http://cdsweb.cern.ch/record/1279144
9. CMS Collaboration.: Performance of the b-jet identification in CMS. CMS Phys. Anal. Summary CMS-PAS-BTV-11-001 (2011). Available from: http://cdsweb.cern.ch/record/1366061

Chapter 5
Systematic Uncertainties

Abstract This chapter treats the possible sources of systematic uncertainties associated with this measurement. As the background is estimated directly on the data, as shown in the previous chapter, and therefore has a separate uncertainty, here we study only effects which could affect signal efficiency.

They include uncertainties on:

- lepton reconstruction;
- jet energy scale and resolution;
- pile-up effects;
- *b*-tagging;
- quark-gluon discrimination;
- missing transverse energy requirement;
- signal cross section;
- signal production mechanism;
- Higgs boson width modeling;
- LHC luminosity.

All of these aspects are treated separately throughout the chapter. Table 5.1 summarizes the results: each source of uncertainty is shown associated to its estimated effect on signal efficiency. Uncertainties are split in the three categories, when applicable.

5.1 Lepton Reconstruction

Systematic uncertainties originating from lepton trigger, reconstruction and identification have been obtained directly on data, by measuring the relative efficiencies with a tag-and-probe method [1] applied to leptons originating from the decay of a Z boson. The efficiency of selecting a lepton object can be factorized as the product of five separate efficiencies: tracking, reconstruction, identification, isolation and trigger efficiencies. In formulas:

F. Pandolfi, *Search for the Standard Model Higgs Boson in the $H \to ZZ \to \ell^+\ell^- q\bar{q}$ Decay Channel at CMS*, Springer Theses, DOI: 10.1007/978-3-319-00903-2_5,

Table 5.1 Summary of systematic uncertainties on signal efficiency, separated by source

Source	0 *b*-tag (%)	1 *b*-tag (%)	2 *b*-tag (%)
Muon reconstruction		2.7	
Electron reconstruction		4.5	
Jet energy scale and resolution		1–5	
Pile-up		4	
b-tagging	3	1	20
Quark-gluon discrimination	4.6	-	-
Missing E_{T}	-	-	3
Higgs cross section		13–18	
Higgs production (PDF)		3	
Higgs production (HQT)	2	5	3
Higgs production (VBF)		1	
Higgs boson width		1–30	
Luminosity		4.5	

Uncertaintes common to all three *b*-tag categories are placed in the center column, whereas sources which have different effects on the three categories are reported with distinct contributions. See text for details

$$\epsilon_{lepton} = \epsilon_{tracking} \cdot \epsilon_{RECO/tracking} \cdot \epsilon_{ID/RECO} \cdot \epsilon_{ISO/ID} \cdot \epsilon_{HLT/ISO}$$

where all efficiencies are relative to the previous step, except for the tracking efficiency, which here is assumed to be 100 %. Each efficiency will be considered separately, in the following.

Reconstruction efficiency measurements are summarized in Table 5.2 for muons (top) and electrons (bottom). It must be noted that for muons this includes also identification efficiency. Isolation efficiencies are instead shown in Table 5.3 (for electrons, identification is included in this step). As for trigger efficiencies, they have been already shown in Tables 4.4 and 4.5.

Once the efficiency is measured in data and Monte Carlo events, a scale factor can be defined, as the data to Monte Carlo efficiency ratio. These scale factors, as we have seen in Sect. 4.1, are used to reweigh the Monte Carlo events, in order to reproduce the efficiency which is observed in data. For each scale factor an uncertainty σ_{SF} is provided, computed by propagating the single-efficiency errors. We evaluate the corresponding systematic uncertainty as the difference in signal yield in the $-6\,\%/+10\,\%$ invariant mass window about the nominal Higgs boson mass when varying the MC weighting factors by $\pm\sigma_{SF}$. No significant mass dependance has been observed.

Finally, the systematic uncertainty relative to the lepton momentum/energy scale was evaluated. For muons, this was done by studying the signal efficiency differences when varying the muon transverse momentum according to the function described in [2]. A similar method was applied to evaluate the uncertainty relative to the electron energy scale.

Table 5.2 Reco/ID efficiency values in 2011 data

η coverage	p_{T} range (GeV)	Efficiency (%) (Data)	Data/MC ratio
$\epsilon_{RECO} * \epsilon_{ID}$ *for muons*			
$\lvert\eta\rvert < 1.20$	20–100	96.0 ± 0.1	0.996 ± 0.001
$1.20 < \lvert\eta\rvert < 2.40$	20–100	96.0 ± 0.1	0.986 ± 0.001
ϵ_{RECO} *for electrons*			
$\lvert\eta\rvert < 0.80$	20–100	97.7 ± 0.3	0.999 ± 0.005
$0.80 < \lvert\eta\rvert < 1.44$	20–100	94.2 ± 0.3	0.964 ± 0.003
$1.44 < \lvert\eta\rvert < 1.57$	20–100	96.0 ± 0.6	0.99 ± 0.04
$1.57 < \lvert\eta\rvert < 2.0$	20–100	95.1 ± 0.5	0.992 ± 0.006
$2.0 < \lvert\eta\rvert < 2.5$	20–100	93.6 ± 0.4	1.001 ± 0.006

Table 5.3 Isolation (together with identification, for electrons) efficiency values in 2011 data

η coverage	p_{T} range (GeV)	Efficiency (%) (Data)	Data/MC ratio
ϵ_{ISO} *for muons*			
$\lvert\eta\rvert < 0.9$	20–40	94.5 ± 0.5	0.987 ± 0.006
$0.9 < \lvert\eta\rvert < 2.4$	20–40	96.5 ± 0.4	0.995 ± 0.005
$\lvert\eta\rvert < 0.9$	40–100	98.7 ± 0.2	0.994 ± 0.002
$0.9 < \lvert\eta\rvert < 2.4$	40–100	99.2 ± 0.2	0.996 ± 0.002
$\epsilon_{ID} * \epsilon_{ISO}$ *for electrons*			
$\lvert\eta\rvert < 1.5$	20–40	91.6 ± 0.5	0.988 ± 0.006
$1.5 < \lvert\eta\rvert < 2.5$	20–40	83.1 ± 0.8	0.998 ± 0.011
$\lvert\eta\rvert < 1.5$	40–100	94.8 ± 0.2	0.988 ± 0.003
$1.5 < \lvert\eta\rvert < 2.5$	40–100	94.1 ± 0.3	1.016 ± 0.064

Final results on lepton systematic uncertainties are summarized in Table 5.4, separated in the muon and electron channels. As the dielectron trigger was found to be fully efficient in data, a conservative uncertainty of 1 % has been adopted.

5.2 Jet Energy Scale and Resolution

The main uncertainty in jet reconstruction comes from the uncertainty on the jet energy scale (JES). We evaluate this source of uncertainty by systematically shifting all reconstructed jet transverse momenta by ±1 standard deviation of the measured jet energy scale uncertainty. This will affect the selection efficiency because of the jet p_{T} and dijet invariant mass requirements. Little modification is expected on the reconstructed Higgs invariant mass shape because of the role of the kinematic fit to the dijet system. The resulting changes in signal efficiency are shown in Table 5.5, for five hypothetical Higgs masses (200, 300, 400, 500, and 600 GeV). The amount by which the signal efficiency changes due to the JES shift is taken as systematic uncertainty, and is found to be of the order of 1/5 %.

Table 5.4 Systematic uncertainties relative to lepton reconstruction and trigger, for the muon and for the electron channel

	Muons(%)	Electrons(%)
Reco-ID-isolation	0.8	3.4
HLT	2.0	1.0
Momentum/energy scale	1.0	3.0

Table 5.5 Signal selection efficiency variations as a consequence of a systematic jet energy scale (JES) shift, for five hypothetical Higgs boson masses (m_H)

m_H (GeV)	Central value	JES $+1\sigma$	JES -1σ	Systematics (%): $+1\sigma, -1\sigma$
200	0.058 ± 0.001	0.061 ± 0.001	0.054 ± 0.001	$+5, -7$
300	0.195 ± 0.002	0.198 ± 0.002	0.192 ± 0.002	$+1, -2$
400	0.266 ± 0.002	0.264 ± 0.002	0.266 ± 0.002	$-1, +0$
500	0.276 ± 0.002	0.269 ± 0.002	0.280 ± 0.002	$-2, +1$
600	0.255 ± 0.002	0.248 ± 0.002	0.260 ± 0.002	$-3, +2$

The uncertainty is due to the available MC statistics. The acceptance is defined on the full reconstructed invariant mass range

A similar procedure was employed to evaluate the systematic uncertainty due to jet resolutions: an additional gaussian smearing factor was introduced to simulate a worse jet transverse momentum resolution. The entity of the smearing is taken from the recent jet energy resolution measurements [3], and are differential in jet pseudorapidity, but all within 5 %. The resulting effect on signal efficiency was found to be negligible with respect to the effects related to jet energy scale.

5.3 Pile-Up

The presence of multiple proton–proton interactions in the same bunch-crossing (pile-up) affects the analysis in multiple ways. Signal efficiency may be modified primarily by two sources:

- additional particles are clustered in jets, therefore modifying their reconstructed quadrimomentum;
- a bias will be introduced in the lepton isolation variables, lowering the efficiency of those requirements.

Both effects should be duly accounted for, as jets are corrected for pile-up on an event-by-event basis with the L1 correction, and the loss in efficiency in lepton isolation should be covered by the reweighing procedure with which Monte Carlo events are corrected.

The residual source of uncertainty derives from the simulation of the overlaid additional interactions, which might not correctly describe the topology of the

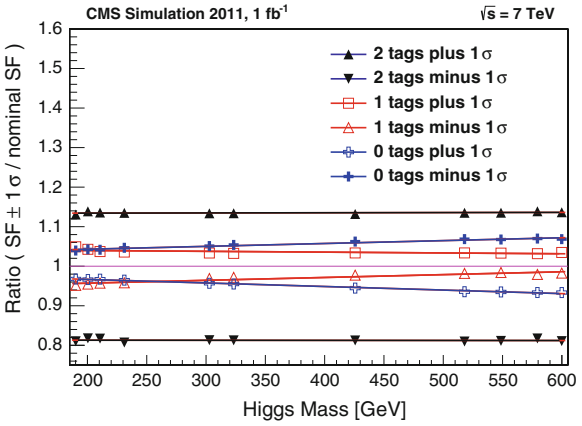

Fig. 5.1 Variation in signal efficiency, for different Higgs boson masses, when modifying the *b*-tagging scale factors by one standard deviation. Results are shown for the three categories: 0-tag (*blue*), 1-tag (*red*), 2-tag (*black*)

additional collisions in the data. In order to quantify this uncertainty, the Monte Carlo samples were divided into two subsets: those with less or more than 7 pile-up events, which is the mode of the distribution measured in the data. As an uncertainty we take the difference in signal efficiency between the full sample and the two subsets. Three mass points (200, 400 and 600 GeV) have been analysed, and the maximal difference in signal efficiency was taken as uncertainty. It was found to be equal to 4 %.

5.4 *b*-Tagging

As described in Sect. 4.5 , Monte Carlo events are corrected on a jet-by-jet basis in order to take into account the *b*-tagging efficiencies measured in the data. This is done through the use of scale factors (SF), defined as the ratio between data and Monte Carlo efficiencies. These scale factors have an uncertainty σ_{SF}, which mainly depends on the limited amount of data available for these measurements.

We have taken as a systematic uncertainty relative to this method the observed difference in signal yield when varying the *b*-tagging scale factors by $\pm\sigma_{SF}$. The results are shown in Fig. 5.1, for the three *b*-tag categories, as a function of the hypothetical Higgs boson mass. It must be noted that these variations are correlated, as, for instance, a decrease in the 2-tag category yield will imply an increase in the 1-tag yield.

5.5 Quark-Gluon Discrimination

The quark-gluon discriminant is founded on general assumptions on the structure and couplings of the QCD Lagrangian, yet does rely on the modeling of hadronization done in the generator. Mismodelings of (light) quark hadronization, which affect the chosen observables (jet multiplicities and transverse momentum distributions) would alter the predicted signal efficiency of the selection. It is therefore important to verify that the performance of the discriminant on quark jets is similar to expectations.

A control sample is identified by photon + jet events, in which the leading jet originates from light quarks in more than 90 % of the cases. In order to contrast the dominant background, constituted by QCD dijet events in which one of the two jets has fragmented mainly into a particle capable of creating a large energy deposit in ECAL (such as a neutral pion), a stringent photon identification is needed. We make use of the photon identification described in Sect. 3.4.2. In order to ensure the absence of jets originated from b-quarks, events in which the leading jet has a positive loose TCHE b-tag have been vetoed. The expected photon + jet purity of this selection is of the order of 90 % at high transverse momenta, and significantly lower at low transverse momenta (reaching $\sim$ 50 % at 20 GeV). It must be noted that a background infiltration does dilute the quark component, but not dramatically, as about 40 % of jets in QCD events are originated from quark partons.

The shape of the quark-gluon discriminant obtained on photon + jet events, in three different transverse momentum bins, is shown in Fig. 5.2, where 191 pb^{-1} of data are compared to the simulation, and the latter is normalized to the signal shape. The available amount of data decreases at lower transverse momenta because of the presence of high prescales in the photon triggers. The observed shape of the discriminant seems compatible with expectations, within the statistical precision granted by the analyzed data.

We are interested in studying a possible effect on quark efficiency, so the gluon contribution has to be subtracted. The latter is isolated in the simulation by applying a matching at MC truth level between jets and partons. Jets successfully matched to

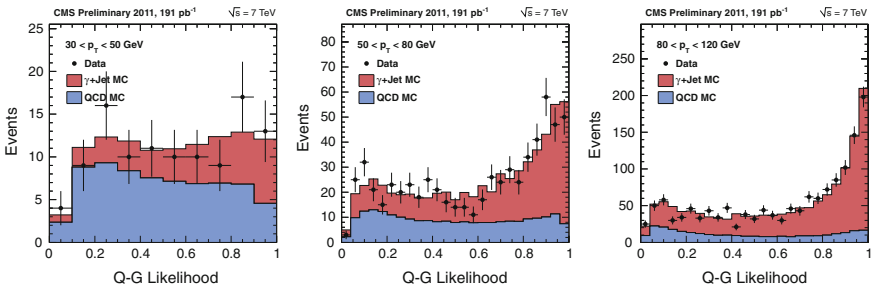

Fig. 5.2 Distributions of the quark-gluon discriminant in photon + jet events in three transverse momentum ranges. The Monte Carlo distributions are normalized to the integral of the data

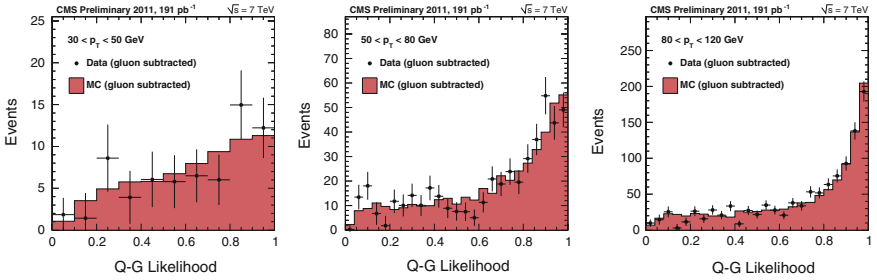

Fig. 5.3 Expected distributions of the quark-gluon discriminant for quark jets in three transverse momentum ranges. The gluon contribution has been subtracted by accessing the MC truth. The Monte Carlo distributions are normalized to the integral of the data

Table 5.6 Efficiencies of requiring the quark-gluon discriminant to be greater than 0.2 on light quark jets, in data and MC, in three transverse momentum bins

Jet p_{T} (GeV %)	MC efficiency (%)	Data efficiency (%)
$30 \div 50$	93.2	(95.1 ± 3.3)
$50 \div 80$	91.3	(91.1 ± 1.6)
$80 \div 120$	91.8	(94.0 ± 0.8)

The error on the Monte Carlo efficiency if negligible if compared to the error on data

gluons are hence subtracted both from data and MC. The gluon-subtracted distributions are shown in Fig. 5.3.

In order to evaluate the effect on signal efficiency, we have to simulate the effect of cutting on the product of two jet's likelihood, with similar kinematic properties as those expected in the case of a heavy Higgs decay. It is not possible to isolate a sample of photon + jet events with two highly quark-enriched jets with similar kinematical properties as the ones expected from the decay of a heavy Higgs boson. We will therefore have to simulate the effect of the decay kinematics, and will proceed as follows. We choose a threshold of 0.2 on the single jet Q-G likelihood distribution, as it is expected to provide an efficiency ϵ such that $\epsilon^2 \approx 85$ %, which is the expected signal efficiency of the cut applied in the analysis. The efficiency of this cut on quark jets is measured in data and MC by applying the requirement on the gluon-subtracted distributions shown in Fig. 5.3, and is reported in Table 5.6.

As no significant deviation is observed between the data and the MC prediction, the uncertainty will originate from the statistical uncertainty of the comparisons. Therefore, we expect the lowest transverse momentum bin (30–50 GeV) to play the driving role. The kinematic properties of the jets in signal will depend on the mass of the decaying Higgs boson, being on average harder as the mass increases. In order to provide a conservative estimate of this systematic uncertainty, we have considered the case of a relatively light mass Higgs boson, 250 GeV, where the weight of the lowest p_{T} bin is inflated. In the decay of a 250 GeV Higgs boson, the jets are expected to populate the three transverse momentum bins as reported in Table 5.7. We estimate the uncertainty U on the single jet as the average of the statistical error in the three

Table 5.7 Fraction of leading and subleading jets which have a transverse momentum falling the in the three considered ranges, for jets originating from the decay of a 250 GeV Higgs boson

	30–50 GeV (%)	50–80 GeV (%)	80–120 GeV (%)
Leading jet	11	64	22
Subleading jet	68	32	2.3

transverse momentum bins, weighted on the expected fractions:

$$\mathrm{U(jet)} = \sum x_i \, \Delta_i$$

where $i = 1, 2, 3$ are the three transverse momentum bins, x_i is the fraction of jets which fall in the given bin, and Δ_i is the statistical uncertainty of that bin. We find:

$$\mathrm{U(lead)} = 22\% \cdot \left(\frac{0.8\%}{94.0\%}\right) + 64\% \cdot \left(\frac{1.6\%}{91.1\%}\right) + 11\% \cdot \left(\frac{3.3\%}{95.1\%}\right) = 1.7\%$$

$$\mathrm{U(sublead)} = 2.3\% \cdot \left(\frac{0.8\%}{94.0\%}\right) + 32\% \cdot \left(\frac{1.6\%}{91.1\%}\right) + 68\% \cdot \left(\frac{3.3\%}{95.1\%}\right) = 3.0\%$$

We then take the product of the uncertainties of the two jets as an estimate of the uncertainty on the cut of the product of the two likelihoods, and therefore find a total systematic uncertainty of U(prod) = 4.6 %.

5.6 Missing Transverse Energy

Missing transverse energy affects directly only the 2 *b*-tag category. The dominant effects which could concur in generating uncertainty derive from the knowledge of the rest of the event, such as jet energy reconstruction and pile-up. Therefore, most of the related uncertainty should be covered by the studies presented in Sects. 5.2 and 5.3. The adopted requirement on missing transverse energy significance is very loose on signal events, and translates in a maximal inefficiency of 3 %. We postulate that the resulting uncertainty does not surpass this value.

5.7 Signal Production

The uncertainty on the signal may be divided in two categories: the uncertainty on its overall cross section, and an uncertainty on the selection efficiency and acceptance which stems from the uncertainty on its production mechanism. These two sources will be investigated separately.

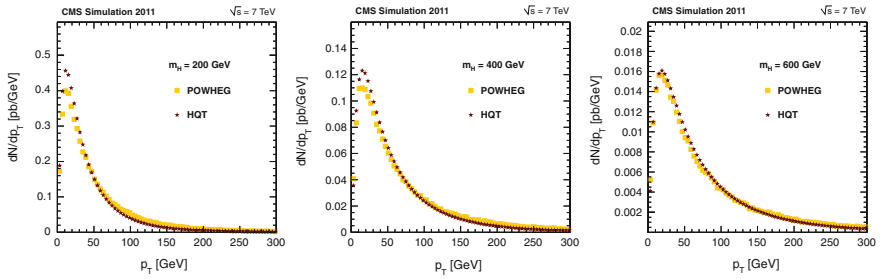

Fig. 5.4 Generated Higgs transverse momentum distributions using the nominal POWHEG (*yellow squares*) generator and the HQT program (*brown stars*), for three hypothetical Higgs boson masses: 200 (*left*), 400 (*center*), and 600 GeV (*right*)

5.7.1 Cross Section

The Higgs production cross section uncertainty depends on its production mechanism, either gluon fusion (gg) or vector boson fusion (VBF). However, since the gluon fusion mechanism dominates, it drives the total uncertainty. We use gg and VBF errors separately and for each mass point according to the Yellow Report [4] prescription. The total weighted error is in the range $(13.4 \div 18.0)\,\%$. We note that this uncertainty is relevant only for the measurement of the ratio to SM expectation R, while it does not affect the absolute cross section measurement.

5.7.2 Acceptance

We identify three different sources of uncertainty on the acceptance of signal events: limited knowledge of the proton parton distribution functions, missing higher orders of perturbation theory in the Monte Carlo simulation, and the contribution of the Vector Boson Fusion process. We will here treat only how these uncertainties affect the analysis efficiency on signal events, as the effect on the signal cross section is already included in what has been discussed in the previous section.

The uncertainty related on the proton parton distribution functions (PDFs) is evaluated following the PDF4LHC [5] recommendations. This is done by evaluating the selection efficiencies (and the relative error sets) of three different sets of PDFs: CT10 [6], MSTW2008NLO [7], NNPDF2.1 [8]. The corresponding uncertainty on signal efficiency is then taken as the envelope of the three error bands, and translates into a 2–4 % effect, with a dependance on the Higgs mass and on the b-tag category, as can be seen in Table 5.8 for 3 hypothetical Higgs boson masses (200, 400 and 600 GeV).

Missing higher orders in perturbation theory, which are not included in the POWHEG NLO computation, may modify the Higgs production kinematics, and

Table 5.8 Summary of systematic uncertainties on signal acceptance deriving from parton distribution functions

	$m_H = 200$ GeV			$m_H = 400$ GeV			$m_H = 600$ GeV		
	0-tag (%)	1-tag (%)	2-tag (%)	0-tag (%)	1-tag (%)	2-tag (%)	0-tag (%)	1-tag (%)	2-tag (%)
CT10	$^{+0.7}_{-3.7}$	$^{+0.9}_{-3.7}$	$^{+1.3}_{-4.6}$	$^{+0.8}_{-3.5}$	$^{+0.5}_{-3.3}$	$^{+0.9}_{-4.0}$	$^{+0.4}_{-2.6}$	$^{+0.4}_{-3.4}$	$^{+0.7}_{-4.3}$
all categories		$+1.0/-3.9$			$+0.7/-3.4$			$+0.6/-4.1$	
MSTW2008NLO	$^{-0.6}_{-0.6}$	$^{+0.5}_{-0.8}$	$^{+0.9}_{-0.8}$	$^{+0.5}_{-0.5}$	$^{+0.4}_{-0.3}$	$^{+0.8}_{-0.3}$	$^{-0.30}_{-0.0}$	$^{+0.3}_{-0.2}$	$^{+0.5}_{-0.3}$
all categories		$+0.6/-0.8$			$+0.5/-0.4$			$+0.4/-0.3$	
NNPDF2.1	$^{+1.8}_{+0.3}$	$^{+1.7}_{+0.1}$	$^{+2.2}_{+0.0}$	$^{+1.4}_{+0.0}$	$^{+1.4}_{+0.3}$	$^{+1.5}_{+0.1}$	$^{+0.7}_{+0.1}$	$^{+1.1}_{+0.3}$	$^{+1.6}_{+0.3}$
all categories		$+1.8/+0.1$			$+1.3/+0.2$			$+1.4/+0.3$	
Total	$^{+1.8}_{-3.7}$	$^{+1.7}_{-3.7}$	$^{+2.2}_{-4.5}$	$^{+2.3}_{-3.5}$	$^{+1.4}_{-3.3}$	$^{+1.5}_{-3.9}$	$^{+0.7}_{-2.6}$	$^{+1.1}_{-3.4}$	$^{+1.6}_{-4.3}$
all categories		$+1.8/-3.9$			$+1.3/-3.4$			$+1.4/-4.1$	

The table reports, in the different *b*-tag categories, for three hypothetical mass points (200, 400, and 600 GeV) for the considered PDF sets (CT10, MSTW2008NLO, NNPDF2.1), the expected effect on signal efficiency when modifying the PDFs by ± 1 standard deviation. The total uncertainty, at the bottom, is taken as the envelop of the three sets

therefore affect the selection efficiency. The related uncertainty was quantified through the use of the HQT [9] program, which includes NNLL effects, and exhibits a modified Higgs transverse momentum spectrum, as can be seen in Fig. 5.4, where the generated Higgs p_{T} as obtained in POWHEG and HQT are compared, for three Higgs boson masses. The POWHEG sample is reweighed in order to match the HQT spectrum, and the corresponding deviation from the nominal signal efficiency is taken as uncertainty. The effect was maximal for $m_H = 200$ GeV, where the reweighing translated into an efficiency drop of 2, 5, 3 %, respectively for the 0-, 1-, and 2-tag categories. At higher masses, POWHEG and HQT are found to be in better agreement, and the deviation is found to be within 1 %. Conservatively, the observed effect at $m_H = 200$ GeV was taken as systematic uncertainty for all masses.

Finally, the contribution of the Vector Boson Fusion (VBF) process to the Higgs production is considered. Only the contribution of gluon fusion was considered during the tuning of the analysis and the interpretation of the results, as it contributes to about 90 % of the total cross section over most of the mass range. A real signal, though, would contain the correct mixture of all the production processes, and therefore the VBF channel, which has in general different final state kinematics, may modify the selection efficiency on signal. We evaluated the corresponding uncertainty as the difference in acceptance between the two production processes, and multiplied it by the expected VBF fractional contribution to the total cross section. The results of this procedure are summarized in Table 5.9, for the considered masses of 200, 400, and 600 GeV. It must be noted that the increase in uncertainty at high masses is driven by the increasing contribution of the VBF production process (recall Fig. 1.5).

5.8 Higgs Width Modeling

In our study the cross section for on-shell Higgs production and decay was made in the zero-width approximation, and acceptance estimates are obtained with Monte Carlo simulations that are based on *ad-hoc* Breit-Wigner distributions for describing the Higgs boson propagation. Recent analyses show that the use of a QFT-consistent Higgs propagator, allowing also for the off-shellness of the Higgs boson, dynamical QCD scales and interference effects between Higgs signal and backgrounds will result, at Higgs masses above 300 GeV, in a sizable effect on conventionally defined but theoretically consistent parameters (mass and width) that describe the propagation of an unstable Higgs boson [4, 10, 11]. These effects are estimated to amount to an additional uncertainty (U) on the theoretical cross section which depends on the Higgs boson mass (m_{H}), and we evaluate it using the following formula:

$$U(m_{\mathrm{H}}) = 150\% \cdot (m_{\mathrm{H}}[\mathrm{TeV}])^3$$

Table 5.9 Summary of systematic uncertainties due to the vector boson fusion production mechanism

	$m_{\mathrm{H}} = 200$ GeV			$m_{\mathrm{H}} = 400$ GeV			$m_{\mathrm{H}} = 600$ GeV		
	0-tag (%)	1-tag (%)	2-tag (%)	0-tag (%)	1-tag (%)	2-tag (%)	0-tag (%)	1-tag (%)	2-tag (%)
Δ_{eff}	8	4	1	7	14	15	10	20	10
total		7			11			13	
Uncertainty	1	0.5	0.2	0.5	1	1	2	4	2
total		0.9			0.9			2.4	

Both the difference in signal efficiency (Δ_{eff}) between VBF and gluon fusion is reported, and the corresponding uncertainty, which depends on the VBF fractional contribution to the total Higgs cross section

Table 5.10 Adopted uncertainty due to the Higgs boson width modeling at four specific mass points

Mass (GeV)	Uncertainty (%)
300	4
400	10
500	19
600	32

As can be seen this uncertainty is neglibigle for masses inferior to 300 GeV, but grows rapidly with mass. Table 5.10 reports the value of this uncertainty at four specific mass points.

5.9 LHC Luminosity

The uncertainty on the measured integrated luminosity is taken from the official LHC recommendation [12]. The latest recommendation corresponds to an uncertainty of 4.5 %.

References

1. Khachatryan, V., et al.: Measurements of inclusive W and Z cross sections in pp collisions at $\sqrt{s} = 7$ TeV. J. High Energy Phys. 2011, 1. doi:10.1007/JHEP01(2011) 080. Available from http://dx.doi.org/10.1007/JHEP01(2011) 080 (2011)
2. CMS Collaboration.: Performance of CMS muon reconstruction in pp collisions at $\sqrt{s} = 7$ TeV (2010).
3. CMS Collaboration.: Jet energy resolution in CMS at $\sqrt{s} = 7$ TeV. CMS Physics Analysis Summary, CMS-PAS-JME-10-014. Available from http://cdsweb.cern.ch/record/1339945 (2010)
4. LHC Higgs Cross Section Working Group., Dittmaier, S., Mariotti, C., Passarino, G., Tanaka R (eds.) Handbook of LHC Higgs cross sections: 1. Inclusive observables. CERN-2011-002. http://arxiv.org/abs/1101.0593 (CERN, Geneva, 2011)
5. Alekhin, S., et al.: The PDF4LHC Working Group Interim Report. Available from http://www.citebase.org/abstract?id=oai:arXiv.org:1101.0536 (2011)
6. Lai, H.-L., Guzzi, M., Huston, J., Li, Z., Nadolsky, P. M., Pumplin, J., Yuan, C.P.: New parton distributions for collider physics. Available from http://www.citebase.org/abstract?id=oai:arXiv.org:1007.2241 (2010)
7. Martin, A.D., Stirling, W.J., Thorne, R.S., Watt, G.: Parton distributions for the LHC. (2009). Cite arxiv:0901.0002 Comment: 157 pages, 70 figures. Code can be found at http://projects.hepforge.org/mstwpdf/ and in LHAPDF V5.7.0. v3: final version published in EPJC with extended Section 12. Available from http://arxiv.org/abs/0901.0002
8. Ball, R.D., Bertone, V., Cerutti, F., Del Debbio, L., Forte, S., Guffanti, A., Latorre, J. I., Rojo, J., Ubiali, M.: Impact of heavy quark masses on parton distributions and LHC phenomenology. Available from http://www.citebase.org/abstract?id=oai:arXiv.org:1101.1300 (2011)
9. Bozzi, G., Catani, S., de Florian, D., Grazzini, M.: Transverse-momentum, resummation and the spectrum of the Higgs boson at the LHC. Nucl. Phys. B 737, 73. Available from doi:10.1016/j.nuclphysb.2005.12.022 (2006).

10. Anastasiou, C., Buhler, S., Herzog, F., Lazopoulos, A.: Total cross-section for Higgs boson hadroproduction with anomalous standard model, interactions. http://arxiv.org/abs/1107.0683 (2011)
11. Passarino, G., Sturm, C., Uccirati, S.: Higgs pseudo-observables, second Riemann sheet and all that. Nuclear Physics B, 834, 77 (2010). Available from http://www.sciencedirect.com/science/article/pii/S0550321310001549, doi:10.1016/j.nuclphysb.2010.03.013
12. CMS Collaboration.: Measurement of CMS luminosity. CMS Physics Analysis Summary, CMS-PAS-EWK-10-004 2010. Available from http://cdsweb.cern.ch/record/1279145 (2010)

Chapter 6
Statistical Interpretation of Results

Abstract The strategy adopted by the CMS collaboration is to search for the Standard Model Higgs boson in a total of 173 mass points across the 114 − 600 GeV invariant mass range. This analysis has limited statistical power for masses below the ZZ production threshold, therefore we will focus on the high-mass range 200 − 600 GeV, which comprises of a total of 73 mass points. In this chapter we will describe how we model the presence of a hypothetical Higgs signal, and the statistical methods adopted to convert the analysis outcome into a statement on the Higgs boson's existence.

6.1 Modeling of the Signal

In each of the six analysis categories, the same m_{lljj} invariant mass spectrum is analyzed for numerous hypothetical Higgs boson signals, varying its postulated mass. Because of computing limitations, though, we are not able to generate Monte Carlo samples at each mass point in which we intend to perform a search. Rather, samples equivalent to high integrated luminosities have been generated at a number of pivotal mass points, where the behaviour of the expected signal is studied, and results are interpolated at every intermediate mass point.

Two quantities need to be parametrized as a function of the Higgs boson mass: the selection efficiency, and the shape of the expected signal. Figure 6.1 shows the expected efficiency on signal events, for the three b-tag categories (0-tag on the top, 1-tag in the middle, 2-tag on the bottom) and the two lepton flavours (muons on the left, electrons on the right). The red points represent the actual values obtained on the generated Monte Carlo samples, the dashed blue line shows the result of a polinomial fit to the points.

The modeling of the signal shape is done by subdividing signal events which pass the analysis selection in two categories: those which have jets which are correctly matched to the quarks originated in the Higgs boson decay ('matched' events), and those in which an incorrect jet pair has been chosen by the selection algorithm,

F. Pandolfi, *Search for the Standard Model Higgs Boson in the $H \to ZZ \to \ell^+\ell^- q\bar{q}$ Decay Channel at CMS*, Springer Theses, DOI: 10.1007/978-3-319-00903-2_6,

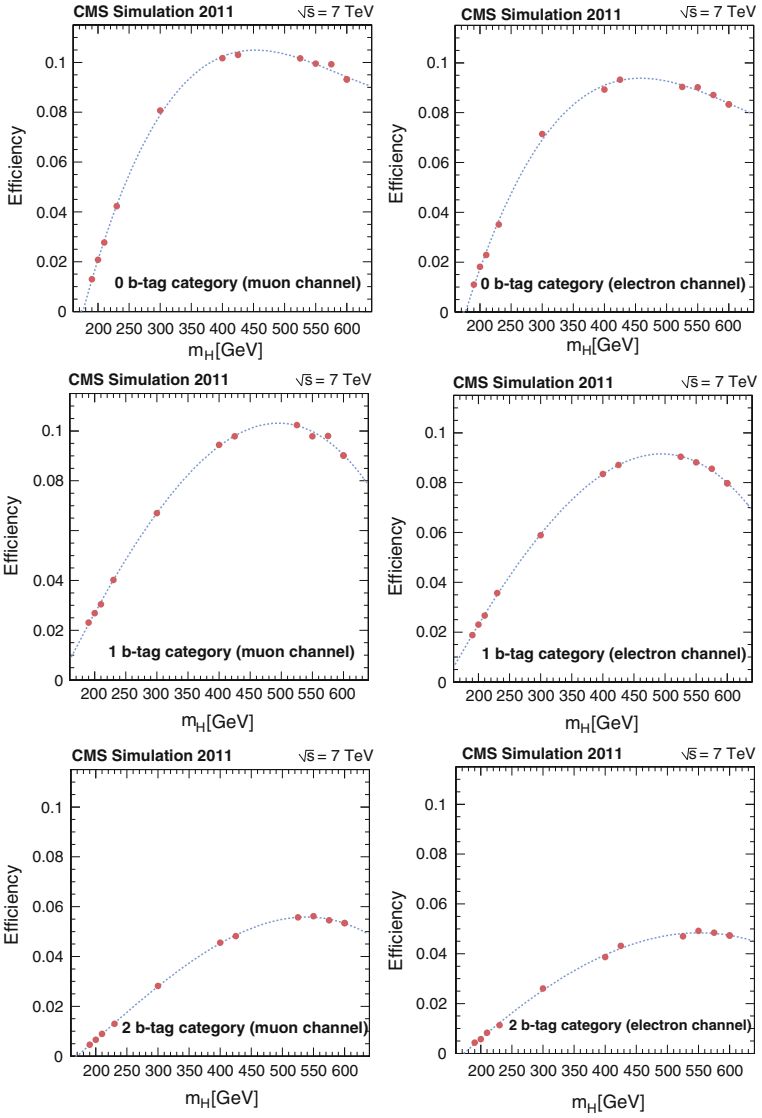

Fig. 6.1 Parameterization of signal efficiency as a function of Higgs mass hypothesis in 0 *b*-tag (*top*), 1 *b*-tag (*middle*), 2 *b*-tag (*bottom*) categories and in the muon (*left*) and electron (*right*) channels. The *blue dashed curves* represent the result of a polinomial fit to the efficiency values

because of signal self-combinatorics ('unmatched' events). This is done by accessing the generator information in signal samples, and performing a matching between the reconstructed jets and the generator quarks produced in the Higgs decay. The reconstructed invariant mass distribution of matched events is parametrized with

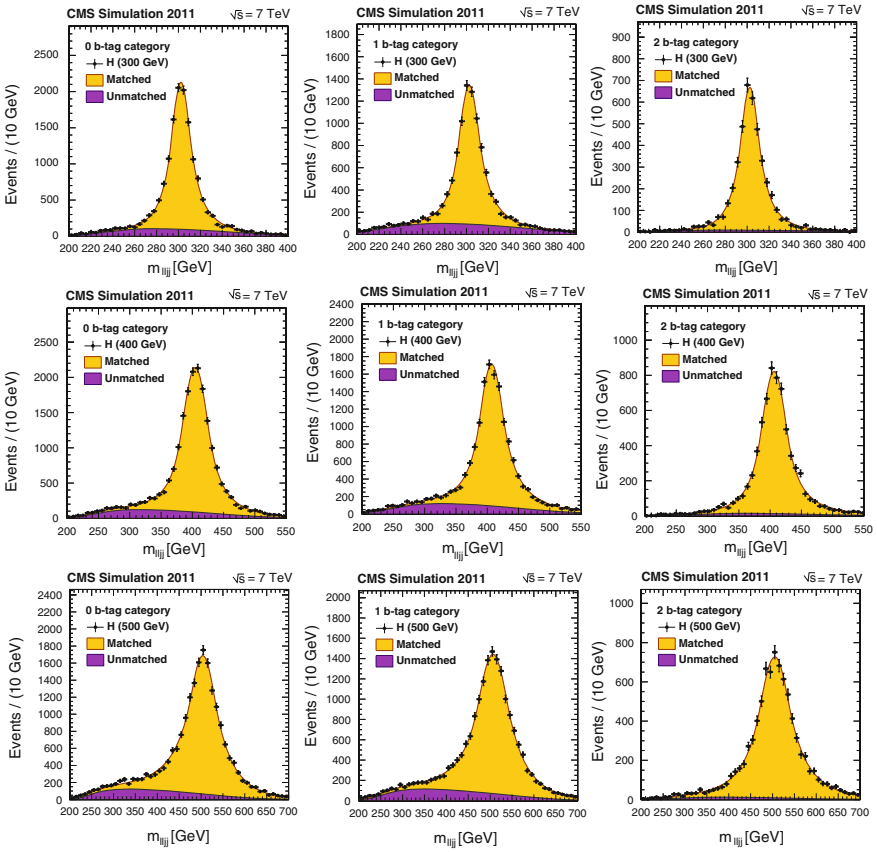

Fig. 6.2 Distribution of the reconstructed dilepton-dijet invariant mass (m_{lljj}) for three hypothetical Higgs boson signals (300 GeV on *top*, 400 GeV in the *middle* and 500 GeV on the *bottom*) and in the three *b*-tag categories (0 *b*-tag on the *left*, 1 *b*-tag in the *center* and 2 *b*-tag on the *right*). The spectra obtained on the simulation are shown with *black markers*, and the results of the fit is shown which a *continuous line*. Contributions from matched and unmatched events are shown separately

a double Crystal-Ball function, in order to take into account detector resolutions. Unmatched events are described with a triangle function smeared with a Crystal-Ball, which was empirically found to adequately describe the shape observed in the simulation. The sum of these two functions defines the adopted parametrization of the shape of the invariant mass distribution of signal events.

In order to have a signal parametrization valid for any given Higgs mass, unbinned maximimum-likelihood fits are performed to the invariant mass spectra obtained on the simulation, for all the available mass points and separately in the three *b*-tag categories. Examples of the results of such fits are shown in Fig. 6.2, for three hypothetical masses (300 GeV on top, 400 GeV in the middle and 500 GeV on the bottom) and for the three *b*-tag categories (0 *b*-tag on the left, 1 *b*-tag in the center

and 2 b-tag on the right). An overall good agreement between the fit results and the shape of the spectra is observed. Once the fits are performed at all of the mass points made available by the Monte Carlo production, the values of the fit parameters are studied as a function of the Higgs mass (m_{H}), and are fitted with linear or quadratic functions, so as to obtain a smooth dependance on m_{H}.

6.2 Statistical Analysis

The observed dilepton-dijet invariant mass spectra on 4.6 fb^{-1} of 2011 data, in the six analysis categories, are shown in Fig. 6.3: 0-tag on the top, 1-tag in the center, 2-tag on the bottom. The contributions of the electronic (left) and muonic (right) channels are shown separately. The data-driven estimate of the background contribution, as extrapolated from the dijet invariant mass sideband region events (as described in Sect. 4.9), is overlaid as a blue line. The expected contribution, in the simulation, of the dominant backgrounds, as well as of a 400 GeV signal enhanced by 2, is shown as a comparison.

For each mass hypothesis, we perform a simultaneous likelihood fit of the six m_{lljj} distributions using the statistical approaches discussed in [1]. As the prime method for reporting limits we use the $\mathrm{CL_s}$ modified frequentist technique [2]. All results are validated by using two independent sets of software tools, the RooStats package [3] and L&S [4].

Based on the expected normalization and shape of the m_{lljj} distribution, for signal and background, and the corresponding systematic uncertainties, we generate a large number of random pseudo-experiments. For each of them, the expected background distribution is generated, a likelihood fit is performed, and an exclusion limit is extracted. The median of the results is taken as central value of the expected statistical power of the analysis, and the distribution is integrated to define 68 and 95 % probability intervals about the median. These values are then compared to the observed limit, which is obtained by the fit to the analyzed data.

Observed (markers) and expected (dashed line) exclusion limits on the product of the Higgs boson production cross section and the branching fraction of $H \rightarrow ZZ$ are presented in Fig. 6.4 using the $\mathrm{CL_s}$ technique. The expected limit also shows the 68 and 95 % probability ranges, respectively marked by a green and a yellow shade. As a comparison, the expectation of the production cross section times branching fraction are shown for the Standard Model (SM), and for an extensions of the latter (SM4), in which a fourth generation of massive fermions is introduced [5–7]. The main difference from the SM Higgs production is that, due to the couplings introduced by the additional fermions, the signal production cross section is enhanced by a factor which varies between 8.3 and 4.8 for a Higgs boson in the $200 \div 600$ GeV mass range. We assume the main uncertainties on the SM4 Higgs production cross section to be the same as for the gluon-fusion mechanism in the Standard Model but with an additional 10 % uncertainty due to the electroweak radiative corrections. This

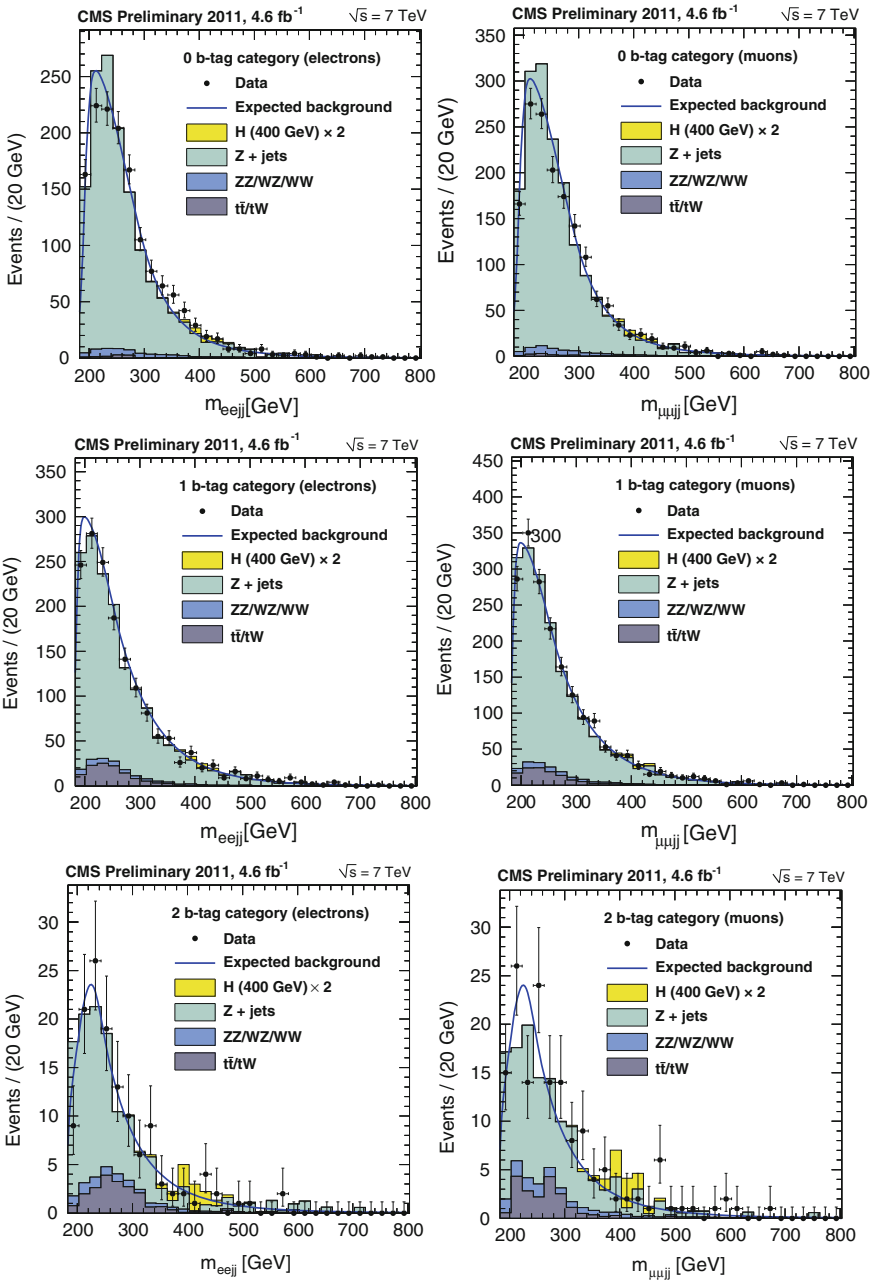

Fig. 6.3 The dilepton-dijet invariant mass after full selection in the six analysis categories: 0-tag on the *top*, 1-tag in the *center*, 2-tag on the *bottom*. The contributions of the electronic (*left*) and muonic (*right*) channels are shown separately. The data-driven estimate of the background contribution is overlaid as a *blue line*. The expected contribution, in the simulation, of the dominant backgrounds, as well as of a 400 GeV signal enhanced by 2 (*yellow*), is shown as a comparison

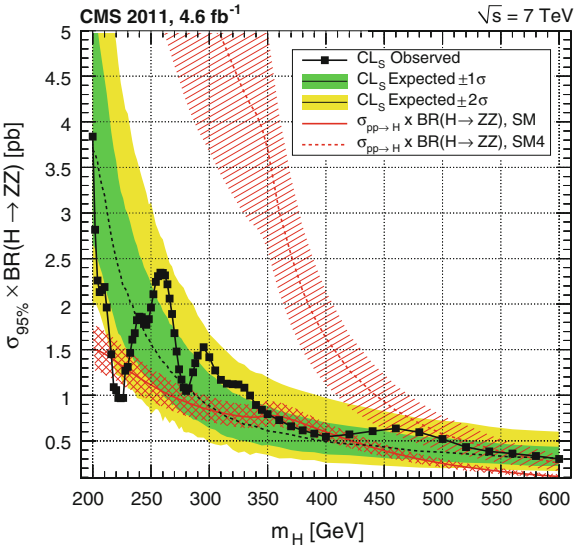

Fig. 6.4 Observed (markers) and expected (*dashed line*) 95 % confidence level *upper limit* on the product of the Higgs boson production cross section and the branching fraction of $H \rightarrow ZZ$ using 4.6 fb^{-1} of data obtained with the CL$_s$ technique. The 68 and 95 % ranges of expectation are also shown with *green* and *yellow* bands. The expected product of the SM Higgs production cross section and the branching fraction is shown as a *red solid curve* with a band indicating theoretical uncertainties at 68 %. The same expectation in the SM4 model are shown with the *upper red curve*, with a band indicating theoretical uncertainties

additional uncertainty is added linearly to uncertainties from QCD renormalization and factorization scales, PDFs, and α_s.

We further incorporate uncertainties on the Higgs production cross section and present a limit on the ratio of the SM Higgs boson production cross section to the SM expectation in Fig. 6.5: the observed limit (markers) is compared to the expected one (dashed line), and the latter is provided of 68 % (green) and 95 % (yellow) probability bands. This search alone, with 4.6 fb^{-1} of data, reaches the sensitivity for a 95 % confidence level exclusion of a Standard Model Higgs boson in two mass ranges: 224–226 and 360–400 GeV. As can be seen the observed exclusion presents a good degree of compatibility with expectations. The significant deviation from the expected trend observed around 225 GeV been deeply scrutinized, and was found compatible with a statistical fluctuation.

A similar limit on the ratio to the Higgs boson production cross section in the SM4 model is shown in Fig. 6.6. A range of SM4 Higgs mass hypotheses are excluded between 200 and 460 GeV at 95 % confidence level.

Figure 6.7 summarizes the results of Higgs searches performed at CMS with up to 4.7 fb^{-1} of data at $\sqrt{s} = 7$ TeV. The top plot shows the observed (solid) and expected (dashed) 95 % confidence level upper limit on the ratio of the SM Higgs boson production cross section to the SM expectation when combining all search

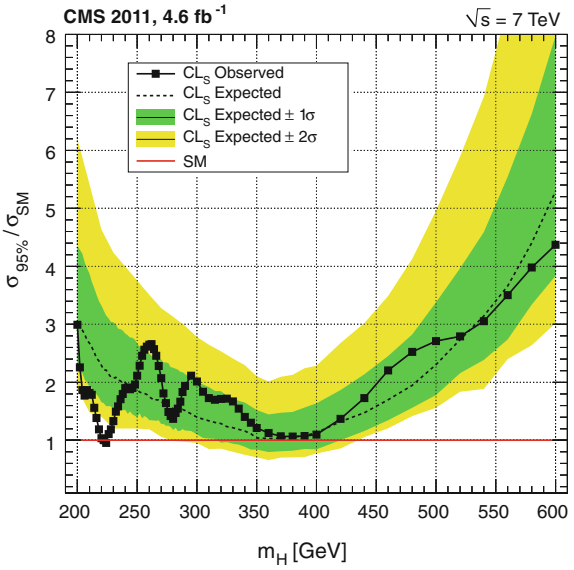

Fig. 6.5 Observed (*dashed*) and expected (*solid*) 95 % confidence level *upper limit* on the ratio of the Higgs boson production cross section to the SM expectation using 4.6 fb^{-1} of data obtained with the CL_s technique. The 68 and 95 % ranges of expectation are also shown with *green* and *yellow bands*. The *solid line* at unity indicates the SM expectation

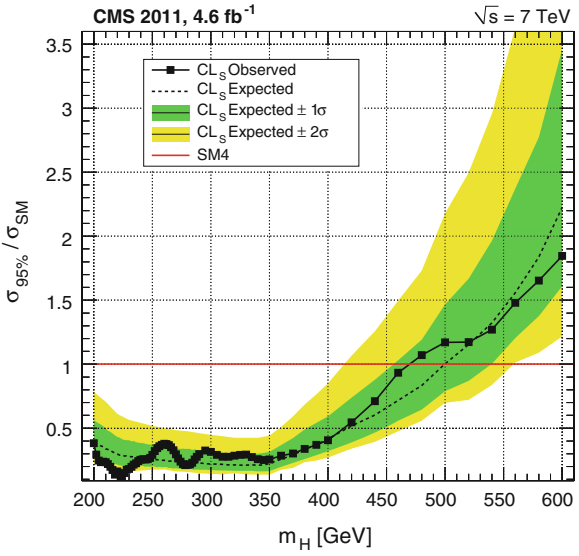

Fig. 6.6 Observed (*dashed*) and expected (*solid*) 95 % confidence level *upper limit* on the ratio of the Higgs boson production cross section to the expectation with the SM4 model using 4.6 fb^{-1} of data obtained with the CL_s technique. The 68 and 95 % ranges of expectation are also shown with *green* and *yellow bands*. The *solid line* at unity indicates the SM4 expectation

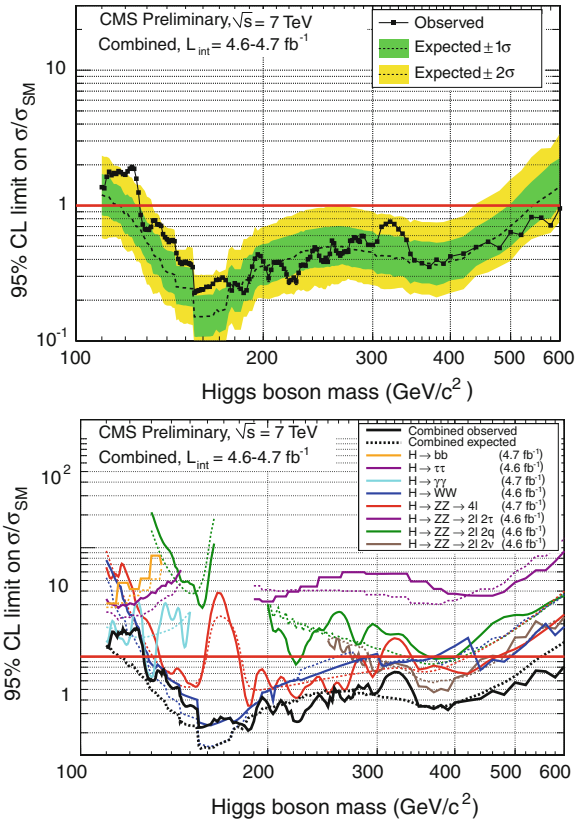

Fig. 6.7 Higgs search combination at CMS, with up to 4.7 fb^{-1} of data. *Top*: observed (*dashed*) and expected (*solid*) 95 % confidence level *upper limit* on the ratio of the Higgs boson production cross section to the SM expectation using 4.6 fb^{-1} of data obtained with the CL_s technique. The 68 and 95 % ranges of expectation are also shown with *green* and *yellow* bands. The *solid line* at unity indicates the SM expectation. *Bottom*: the contribution of single channels is shown separately with coloured lines, and the combined exclusion is marked in *black*

analyses conducted at CMS. The 68 and 95 % ranges of expectation are also shown with green and yellow bands. As can be seen, by combining all its search channels, CMS excludes the presence of a Higgs boson in the 127–600 GeV mass range, at 95 % confidence level. The intriguing excess of events observed around 125 GeV proved [8, 9], once the searches were extended to $\sqrt{s} = 8$ TeV data, to be caused by a narrow resonance with properties compatible with those of a Standard Model Higgs boson.

The bottom plot emphasizes the role of the individual analyses: the sensitivity of each search channel is shown separately, and the combination of them is shown with

a solid black line. In the high mass region, the $H \to ZZ \to \ell^+\ell^- q\bar{q}$ channel (shown in green) gives significant contributions to the exclusion limit, integrating the effort of the $H \to ZZ \to \ell^+\ell^-\ell^+\ell^-$, $H \to ZZ \to \ell^+\ell^-\nu\bar{\nu}$ and $H \to WW \to \ell^+\nu\,\ell^-\bar{\nu}$ analyses.

References

1. CMS Collaboration. SM Higgs Combination. CMS Physics Analysis Summary, CMS-PAS-HIG-11-011 (2011). Available from: http://cdsweb.cern.ch/record/1370076
2. Read, A.L.: Presentation of Search Results: the CLs Technique. J. Phys. G: Nuclear Particle Phys. **28** (2002). 2693. Available from:http://stacks.iop.org/0954-3899/28/i=10/a=313
3. Moneta, L., Belasco, K., Cranmer, K., Lazzaro, A., Piparo, D., Schott, G., Verkerke, W., Wolf, M.: The RooStats Project. (2010). Available from: http://arxiv.org/abs/1009.1003, http://arxiv.org/abs/1009.1003
4. Chen, M., Korytov, A.: Limits and Significance. Available from: https://mschen.web.cern.ch/mschen/LandS/
5. Schmidt, N.B., Cetin, S.A., Istin, S., Sultansoy, S.: The Fourth Standart Model Family and the Competition in Standart Model Higgs Boson Search at Tevatron and LHC. Eur. Phys. J., C66 (2010), 119. http://arxiv.org/abs/0908.2653, http://dx.doi.org/10.1140/epjc/s10052-010-1238-1
6. Li, Q., Spira, M., Gao, J., Li, C.S.: Higgs Boson Production via Gluon Fusion in the Standard Model with four Generations. Phys. Rev. D83 (2011), 094018. http://arxiv.org/abs/1011.4484, http://dx.doi.org/10.1103/PhysRevD.83.094018
7. Anastasiou, C., Buhler, S., Herzog, F., Lazopoulos, A.: Total cross-section for Higgs boson hadroproduction with anomalous Standard Model interactions (2011). http://arxiv.org/abs/1107.0683
8. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. Phys. Lett. B, 716 (2012), 30. Available from: http://www.sciencedirect.com/science/article/pii/S0370269312008581, http://dx.doi.org/10.1016/j.physletb.2012.08.021
9. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. Phys. Lett. B, **716** (2012), 1. Available from: http://www.sciencedirect.com/science/article/pii/S037026931200857X, http://dx.doi.org/10.1016/j.physletb.2012.08.020

Chapter 7
Conclusions

Abstract We have presented a search for a heavy Higgs boson in the decay channel $H \longrightarrow ZZ \longrightarrow \ell^+\ell^- q\bar{q}$. This channel presents jets in the final state, which threaten to degrade the analysis performance both by worsening its mass resolution and by increasing the possible sources of backgrounds. Therefore stringent requirements are imposed on the jet reconstruction performance, as both an accurate calibration and a good resolution on the measurement of their quadrimomenta are needed. This is achieved by using Particle Flow jet reconstruction, calibrated *in situ* with photon + jet events. The resolution on the Higgs invariant mass is further boosted by the application of a kinematic fit to the hadronic decay of the Z boson.

The analysis selection maximises the sensitivity to a presence of a Higgs boson signal by pursuing two main directives:

- an angular analysis, to discriminate events compatible with the decay of a scalar boson from non-resonant backgrounds;
- the use of jet parton flavour tagging as means of background rejection and sensitivity maximization.

After applying the full selection on 4.6 fb^{-1} of $\sqrt{s} = 7$ TeV data collected in 2011 by the CMS detector, no evidence for the presence of a Standard Model Higgs boson has been found, and we set upper limits on its production cross section, reaching sensitivity to the Standard Model prediction in two mass ranges: 224–226 and 360–400 GeV. We also constrain the presence of a Higgs boson in the context of an extended Standard Model, in which a fourth generation of massive fermions is introduced, by excluding it in the 200–460 GeV mass range, at 95 % confidence level. When combined to the other searches performed at CMS, the Standard Model Higgs boson is excluded in a broad mass range: between 127 and 600 GeV.

The procedure of jet energy scale and transverse momentum resolution measurements with photon+jet events developed in the context of this thesis has become the standard in CMS, and is currently still used to calibrate jets. Its results are now being confronted closely and integrated by the measurements made with Z + jet events, which suffer from a lower cross section but have orthogonal systematic uncertainties.

F. Pandolfi, *Search for the Standard Model Higgs Boson in the H→ZZ→ $\ell^+\ell^- q\bar{q}$ Decay Channel at CMS*, Springer Theses, DOI: 10.1007/978-3-319-00903-2_7,

By combining these two measurements, we expect to reach a jet calibration precision possibly better than 1 % within the year.

The likelihood discriminator capable of recognizing jets originating from gluonic or light quark partons also has been developed in the context of this thesis. It represents an innovative experimental technique, with a number of potential physics applications. It is currently being further improved and additional control samples on data are being scrutinized, in order to gain deeper insight on its performance on gluon jets.

The definition of such a likelihood discriminator has been possible only thanks to the Particle Flow jet reconstruction technique, which provides information on jet composition at particle level. Studies of jet particle composition such as this open new frontiers in the understanding of jets and QCD, and on an experimental side offer a broad range of possible applications: particle-level comparisons between data and the simulation are currently used in CMS as validations of the functioning of the Particle Flow full event reconstruction; the possibility to reconstruct individually all particles produced in a proton beam crossing enables the definition of jet energy corrections which take into account pile up effects on an event-by-event basis; the coordinate use of different subdetectors simplifies jet identification; the treatment of infra-detector correlations improves the performance of charged lepton isolation requirements.

The search for a massive Higgs boson in the $H \longrightarrow ZZ \longrightarrow \ell^+\ell^- q\bar{q}$ decay channel has given a significant contribution to the combined CMS effort in the high mass region. With the increase of luminosity delivered by the LHC, the sensitivity of this channel alone is expected to increase, but it will have to confront itself with the search conducted in the $H \rightarrow ZZ \rightarrow \ell^+\ell^-\ell^+\ell^-$ channel, which is soon expected to become the dominant search channel in the high mass range. A similar analysis strategy, though, could be exploited for the search of exotic, spin-2 particles, such as the case of the graviton predicted in the context of small extra-dimension models [1]. An adapted angular likelihood, to take into account the different spin correlations, would provide similar means of background discrimination.

Reference

1. Randall, L., Sundrum, R.: Large mass hierarchy from a small extra dimension. Phys. Rev. Lett. **83** (1999), 3370. Available from: http://link.aps.org/doi/10.1103/PhysRevLett.83.3370,http://dx.doi.org/10.1103/PhysRevLett.83.3370