Observability of a Heavy Higgs Boson with the CMS Detector at the LHC in the Channel $qq + H$ and $H \rightarrow l^+l^-q\bar{q}$

Zur Erlangung des akademischen Grades eines
DOKTORS DER NATURWISSENSCHAFTEN
von der Fakultät für Physik der
Universität Karlsruhe (TH) genehmigte

DISSERTATION

von

Dipl. Phys. Claus Ulrich Felzmann
aus Mannheim

Tag der mündlichen Prüfung: 15.06.2007

Referent: Prof. Dr. G. Quast
Institut für Experimentelle Kernphysik

Korreferent: Priv. Doz. Dr. W. Wagner
Institut für Experimentelle Kernphysik

Je pense, donc je suis (cogito, ergo sum).

René Descartes, Meditationes de prima philosophia (1641)

Observability of a Heavy Higgs Boson with the CMS Detector at the LHC in the Channel $qq + \mathrm{H}$ **and** $\mathrm{H} \to l^+l^-q\bar{q}$**.** The Higgs mechanism is a cornerstone of the Standard Model of particle physics. It is one of the main goals of the Large Hadron Collider (LHC) to prove the mechanism of electroweak symmetry breaking through the detection of the so far un-observed Higgs boson. The potential of the CMS experiment – currently being built at the LHC – to discover a Higgs boson decaying via two Z bosons subsequently into two quarks and either two muons or two electrons is investigated. While the Higgs boson production at the LHC is dominated by the gluon fusion, the characteristics of the the weak boson fusion as second largest contribution can be exploited to separate significantly the Higgs decays from contributing background events. An analysis strategy to separate Higgs boson decays from background processes is developed using a parameterized simulation of particle interactions with the material of the CMS detector. The expected statistical significance for the discovery of a heavy Higgs boson with respect to theoretical and experimental uncertainties is expressed. It is shown that the data volume taken in the low-luminosity phase of the LHC ($\mathcal{L} = 60\,\mathrm{fb}^{-1}$), will allow to discover a Higgs boson in the mass range between 500 GeV and 700 GeV with a statistical significance of more than $S > 3\sigma$.

Studien zum Entdeckungspotential eines schweren Higgs-Bosons mit dem CMS-Detektor am LHC im Zerfallskanal $qq + \mathrm{H}$ **mit** $\mathrm{H} \to l^+l^-q\bar{q}$**.** Die Bestätigung des spontanen Symmetriebrechung in der elektroschwachen Theorie des Standard-Models ist eine der Hauptaufgaben des neuen Teilchenbeschleunigers "LHC". Das Potential des sich am LHC im Bau befindenden CMS-Detektors, ein schweres Higgs-Boson in Zerfällen in zwei Quarks und zwei isolierte Myonen oder Elektronen zu entdecken, wird diskutiert. Den zweitgrößten Beitrag zum Wirkungsquerschnitt zur Produktion von Higgs-Bosonen liefert die Fusion von elektroschwachen Eichbosonen. Die Ausnutzung der für diesen Prozess typischen Ereignistopologie erlaubt eine wirksame Unterdrückung vieler kritischer Untergründe. Mit Hilfe einer einer schnellen Detektorsimulation von hypothetischen Signal- und Untergrund-Ereignissen wird untersucht, wie gut die oben angeführten Endzustände mit dem CMS-Detektor für verschiedene Massen des Higgs-Bosons rekonstruiert und identifiziert werden können. Es wird gezeigt, dass das Higgs-Boson im Massenbereich zwischen 500 GeV und 700 GeV nach der Datennahme von $\mathcal{L} = 60\,\mathrm{fb}^{-1}$ mit einer statistischen Signifikanz von $S > 3\sigma$ nachgewiesen werden kann. Der Einfluss von Unsicherheiten auf die vorhergesagte Signifikanz wird abgeschätzt.

Contents

Chapter 1

Introduction

The Standard Model (SM) of elementary particle physics explains precisely three of the four fundamental interactions of nature. Combining special relativity with quantum mechanics, it predicts the dynamics of matter particles and forces consistently with all the experimentally available data (apart from gravity).

All elementary particles described by the Standard Model can be divided into two groups: Ordinary particles that make up matter (fermions) and force mediating particles (bosons). At present, twelve different fermions are known. Six of these are classified as quarks and the other six as leptons. Every particle has a corresponding anti-particle. These particles carry charges (electric charge, weak charge, color charge) which make them susceptible to the fundamental forces. Three different types of force mediating particles are explained by the Standard Model: the photon for electromagnetism, the neutral Z and two charged W bosons for the weak force, and eight gluons for the strong interaction.

The mathematical framework of the Standard Model is based on group symmetries with a local parameter. Bosons are described in terms of mathematical fields and fermions as wave functions. The absolute phase of such a wave function has no physical meaning and cannot be determined in measurements. The fact that the phase is invariant under certain symmetry transformations changing the wave function and the coupling field concurrently is called gauge invariance. Therefore, the mediating particles are often referred to as gauge bosons since they carry the field.

In its original formulation, the Standard Model predicted massless gauge bosons for the force mediating particles of the electroweak theory which

unifies electromagnetic and weak interaction. But the discovery of the W and Z bosons at CERN in 1983 proved definitely that they are massive as it was already expected due to the limited range of the weak force. An additional scalar field which is named Higgs field after one of its proposers, was already introduced in 1960 to explain the massive weak gauge bosons. The Higgs field breaks the symmetry of the electroweak theory by giving masses to the W and Z bosons whereas the photon remains massless.

The particle which is associated with the Higgs field, referred to as Higgs boson, has not been discovered yet. It is one of the main goals of the Large Hadron Collider (LHC), currently being constructed at CERN near Geneva in Switzerland, to search in head-on proton-proton collisions for decay signatures of the Higgs boson. While the Standard Model predicts the existence of the Higgs boson, there are no predictions about its mass. Hence, the LHC needs to cover a large energy range to ensure that the Higgs boson can be observed. Even if no Higgs particles is discovered, the investigation of electroweak gauge boson scattering can be used to prove or to reject the theory of the Higgs mechanism.

This PhD thesis has been prepared during the installation and commissioning phase of the LHC in scope of the CMS experiment. The LHC will produce proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 14\,\text{TeV}$ taking place 40 million times per second. The CMS experiment is one of the two general purpose particle detectors which will be able to reconstruct the physics processes that produce some thousand particles which emerge from the proton-proton collisions. The Higgs boson is searched in decays into two neutral Z bosons subsequently decaying into two charged leptons (either electrons or muons) and two hadronic particle jets. The study is performed using a fast detector simulation for the mentioned decay mode in the high Higgs boson mass range. To extract the Higgs boson resonance from the overwhelming background processes which show the same signature in the detector, selection cuts on the kinematic and mass variables of the decay mode and the production process are studied.

This thesis starts with an introduction to the underlaying theory to give an idea of the electroweak theory in CHAPTER 2. The field equation, referred to as Lagrangian density function, for the Electroweak Theory is deduced and the mechanism of spontaneous symmetry breaking (the Higgs mechanism) is explained. It follows a description of the predicted properties of the Higgs boson and the experimental searches which have already been performed at the Large Electron-Positron collider at CERN and the Teva-

tron collider at Fermilab in CHAPTER 3. These experiments have not been able to discover the Higgs boson and provide only a lower exclusion limit on the Higgs boson mass. Besides of the experimental lower limit on the mass, theoretical constraints on an upper bound are also presented. Although there are arguments for expecting the Higgs mass to be relatively low, these expressions cannot be taken as real restrictions, since there is no experimental confirmation up to which energy scale the electroweak theory is valid. The CMS experiment at the LHC is then presented in CHAPTER 4 in which the collider itself and the subdetectors of CMS are explained. CHAPTER 5 focuses on the CMS software framework which is used to carry out this analysis. The Monte Carlo generation of hypothetical physical events as they will occur in real proton-proton collisions at the LHC is described together with the simulation of their interaction with the detector. The reconstruction of physics quantities and physics objects from data recorded by the data acquisition systems of the subdetectors is outlined as well. The strategies for the Higgs boson search in the dilepton-dijet final-state is presented in CHAPTER 6. It is shown that after the optimization of all selection cuts a statistical significance of 3.7 is expected to observe a Higgs boson with a mass of $600\,\mathrm{GeV}$ in the CMS experiment for an integrated luminosity of $\mathcal{L} = 60\,\mathrm{fb}^{-1}$. The conclusion made in CHAPTER 7 summarizes the gained knowledge which can be drawn from this study.

Chapter 2

The Standard Model of the Electroweak Interaction

The electroweak theory of the Standard Model describes precisely the interaction of the elementary fermions with the force carrying particles of the electromagnetic and the weak fields. In case of the electromagnetic interaction, the force is mediated by the photon γ, in case of the weak interaction by the weak bosons W and Z. This chapter outlines the derivation of the theoretical framework which is needed to explain the mass generation of the weak gauge bosons. The derivation starts with the construction of the Lagrangian density function of a Dirac particle interacting with the electromagnetic field which is a vector field. It is derived that the Lagrangian density function is invariant under local gauge transformations. An additional scalar field besides the vector field is necessary to create the masses of the W and Z bosons. It turns out, that the additional scalar field exhibits a massive particle. This so far unobserved particle is called Higgs boson named after the Scottish physicist Peter Ware Higgs who proposed a mechanism to explain the electroweak symmetry breaking [1]. This thesis is dedicated to the experimental search for the Higgs particle. The following discussion of the electroweak gauge theory has been adopted from [2] and [3].

2.1 Lagrangians in relativistic field theory

In classical mechanics the equations of motion for a dynamical system can be obtained from the Euler-Lagrange equation

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial \dot{q}_i} = 0, \tag{2.1}$$

where q_i are generalized coordinates and $\dot{q}_i$ are their time derivatives. The Lagrangian

$$L(q_i, \dot{q}_i, t) = T - V \tag{2.2}$$

is the difference of the kinetic energy T and the potential energy V of the system.

A particle is a localized entity, a so-called discrete system. A field on the other hand occupies some region of space as a function of the continuous space-time coordinates x_μ. Relativistic field theories aim at calculating one or more functions $\phi_i(x_\mu)$ of position and time instead of just the particle's position as a function of time. Hence, the Lagrangian of classical mechanics is replaced by a Lagrangian density

$$L(q_i, \dot{q}_i, t) \to \mathcal{L}\left(\phi_i, \partial_\mu \phi_i, x_\mu\right), \quad L = \int d^3 x \, \mathcal{L} \tag{2.3}$$

and is a function of the fields ϕ_i and their derivatives $\partial_\mu \phi_i = \frac{\partial \phi_i}{\partial x^\mu}$.

In a relativistic theory, the Euler-Lagrange equation generalizes to:

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi_i)}\right) - \frac{\partial \mathcal{L}}{\partial \phi_i} = 0. \tag{2.4}$$

The Lagrangian $\mathcal{L}$ is usually taken as axiomatic whereas in classical mechanics it can be derived.

Examples for fields with no sources or interactions are:

- The Klein-Gordon Lagrangian for a scalar (Spin-0) field ψ

$$\mathcal{L} = \frac{1}{2}\left[(\partial_\mu \psi)(\partial^\mu \psi) - m^2 \psi^2\right], \tag{2.5}$$

  has been specifically designed to reproduce the Klein-Gordon equation

$$\partial_\mu \partial^\mu \psi + m^2 \psi = 0. \tag{2.6}$$

- The Dirac Lagrangian for a spinor (Spin-$\frac{1}{2}$) field ψ: Substituting

$$\mathcal{L} = i\bar{\psi}\gamma_\mu \partial^\mu \psi - m\bar{\psi}\psi \qquad (2.7)$$

into EQUATION 2.4 (where each of the four components of ψ and of its adjoint spinor $\bar{\psi}$ are treated as independent field variables) gives the Dirac equation

$$(i\gamma^\mu \partial_\mu - m)\psi = 0. \qquad (2.8)$$

- The Proca Lagrangian for a vector (Spin-1) field A:

$$\mathcal{L} = -\frac{1}{4}(\partial^\mu A^\nu - \partial^\nu A^\mu)(\partial_\mu A_\nu - \partial_\nu A_\mu) + \frac{1}{2}m^2 A^\nu A_\mu. \qquad (2.9)$$

It is useful to introduce the shorthand $F^{\mu\nu} = (\partial^\mu A^\nu - \partial^\nu A^\mu)$. Then the Lagrangian reads

$$\mathcal{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \frac{1}{2}m^2 A^\nu A_\mu \qquad (2.10)$$

and the field becomes

$$\partial_\mu F^{\mu\nu} + m^2 A^\nu = 0. \qquad (2.11)$$

If the mass is set to $m = 0$, the Lagrangian describes an electromagnetic field in vacuum.

2.2   Local gauge invariance

An electron is described by a complex spinor whose absolute phase is not an observable. The phase transformation

$$\psi' = \exp[iq\alpha]\psi, \qquad (2.12)$$

where q is the charge of the particle involved and α is an arbitrary phase, leaves the Dirac Lagrangian and all expectation values unchanged. This is called global gauge invariance and can be easily checked by substituting ψ' into EQUATION 2.8. All the phase factors cancel out. The family of phase transformations $U(\alpha) = \exp(i\alpha)$, where α is a single real parameter, forms a unitary Abelian group know as $U(1)$ group. The conservation of

the electromagnetic charge q is a consequence of the $U(1)$ phase invariance.

The Dirac Lagrangian is not invariant under a local phase transformation $\alpha(x)$ if $\alpha(x)$ is an arbitrary function of space and time

$$\psi' = \exp\left[iq\alpha(x)\right]\psi, \tag{2.13}$$

since an extra term is picked up by the derivative

$$\partial_\mu \exp[iq\alpha(x)] = iq\exp[iq\alpha(x)]\partial_\mu\alpha(x). \tag{2.14}$$

It is an axiom of quantum field theory that the complete Lagrangian of a system has to be invariant under local phase transformations. To soak up the extra term

$$iq\bar{\psi}\gamma^\mu\partial_\mu\alpha(x)\psi \tag{2.15}$$

in the transformed Lagrangian, the derivative ∂_μ in EQUATION 2.8 needs to be replaced by

$$\partial_\mu \to D_\mu = \partial_\mu + iqA_\mu \tag{2.16}$$

where A_μ transforms as follows:

$$A_\mu \to A'_\mu = A_\mu + \partial_\mu\alpha(x). \tag{2.17}$$

By demanding local phase invariance to the Lagrangian

$$\begin{aligned}
\mathcal{L} &= i\bar{\psi}\gamma^\mu D_\mu\psi - m\bar{\psi}\psi \\
&= \bar{\psi}(i\gamma^\mu\partial_\mu\psi - m)\psi - q\bar{\psi}\gamma^\mu\psi A_\mu
\end{aligned} \tag{2.18}$$

a local vector field A_μ is introduced, called gauge field, which couples to the Dirac particle with coupling q. The vector field can be regarded as the photon field and according to EQUATION 2.10, the Lagrangian term corresponding to the energy of the free photon field has to be added since the photon is known to be a spin-1 particle.

This leads to the final Lagrangian of Quantum Electrodynamics

$$\mathcal{L}_{\text{QED}} = \bar{\psi}(i\gamma^\mu\partial_\mu\psi - m)\psi - q\bar{\psi}\gamma^\mu\psi A_\mu - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} \tag{2.19}$$

which describes the interaction of a charged fermion ψ with a vector field A_μ. The photon field A_μ is introduced as a consequence of the requirement of local gauge invariance of the Lagrangian. An extra mass term $1/2m^2A^\nu A_\mu$ for the photon would spoil local gauge invariance, thus the gauge field A_μ must be massless.

2.3 The mass term

The principle of local gauge invariance does not work for the Proca Lagrangian (EQUATION 2.10), since the presence of a mass term destroys the phase invariance. Therefore, this concept cannot be applied to the weak interaction without any modifications, since the interaction is mediated by the $W^\pm$ and Z^0 bosons with masses of about 80 and 91 GeV, respectively. Furthermore, the Lagrangian is not renormalizable anymore, if an extra mass term $m^2 W^\nu W_\mu$ is added.

A possible way out is to introduce the electroweak bosons first as massless particles into the Lagrangian and explain their masses as a consequence of their interaction with a scalar background field. This is called the Higgs mechanism.

2.4 Spontaneous symmetry breaking

The Lagrangian

$$\mathcal{L} = T - V = \frac{1}{2}(\partial_\mu \phi)(\partial^\mu \phi) - \frac{1}{2}\mu^2\phi^2 - \frac{1}{4}\lambda\phi^4 \tag{2.20}$$

for a scalar field $V(\phi)$ with $\lambda > 0$ is invariant under the symmetry which replaces ϕ by $-\phi$. Depending on the sign of μ^2, it gives one or three extrema for $V(\phi)$:

$$\frac{dV}{d\phi} = 0 \quad \Rightarrow \quad \phi_1 = 0, \quad \phi_{2,3} = \pm\sqrt{-\mu^2/\lambda}. \tag{2.21}$$

These both cases are depicted in FIGURE 2.1.

According to EQUATION 2.5, for $\mu^2 > 0$ the Lagrangian simply describes a scalar particle with mass $m = \mu$. The higher-order term ϕ^4 shows a self-interacting part with coupling λ.

However, for $\mu^2 < 0$ the mass term cannot be easily identified as m would be imaginary. Since the Feynman calculus is a perturbation theory which treats fields as fluctuation above the ground state, the Lagrangian needs to be expanded around the minimum $\phi = v$ or $\phi = -v$ where $v = \sqrt{-\mu^2/\lambda}$. Substituting

$$\phi(x) = v + \eta(x) \tag{2.22}$$
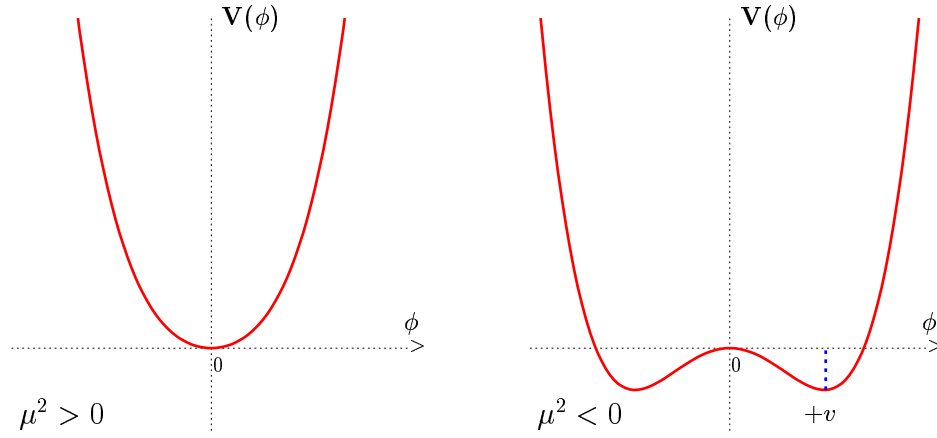
Figure 2.1. The potential $V(\phi) = \frac{1}{2}\mu^2\phi^2 + \frac{1}{4}\lambda\phi^4$ for $\mu^2 > 0$ (left) and $\mu^2 < 0$ (right), and $\lambda > 0$.

into EQUATION 2.20 having chosen to transform the field to $\phi = +v$, the Lagrangian becomes

$$\mathcal{L}' = \frac{1}{2}(\partial_\mu\eta)(\partial^\mu\eta) - \lambda v^2\eta^2 - \lambda v\eta^3 - \frac{1}{4}\lambda\eta^4 + \frac{1}{4}\lambda v^4 \qquad (2.23)$$

and the term $\lambda v^2\eta^2$ can be identified as mass term:

$$m_\eta = \sqrt{2\lambda v^2} = \sqrt{-2\mu^2}. \qquad (2.24)$$

The higher-order terms in η represent the interaction of the η field with itself.

As a transformation described by EQUATION 2.22 cannot change physics, the Lagrangians $\mathcal{L}$ (EQUATION 2.20) and $\mathcal{L}'$ (EQUATION 2.23) must be completely equivalent. Unfortunately, particle physics is not able to solve the Lagrangians exactly. Instead, perturbation theory is used which calculates the fluctuation around the minimum energy. Using EQUATION 2.20 to perform the Feynman calculus will yield a non-converging series, since it is developed around an unstable point.

By re-expressing the Lagrangian $\mathcal{L}$ (EQUATION 2.20) around the stable vacuum $\phi = +v$ as a function of the deviation η, a mass for the scalar particle is created. But the reformulated Lagrangian $\mathcal{L}'$ (EQUATION 2.23) is not even in η anymore, the symmetry has been broken (or rather hidden) by selecting a particular vacuum state. This way of mass generation without an external source is called spontaneous symmetry-breaking.

2.5 The Goldstone theorem

The Lagrangian

$$\mathcal{L} = (\partial_\mu \phi)^* (\partial_\mu \phi) - \mu^2 \phi^* \phi - \lambda (\phi^* \phi)^2, \tag{2.25}$$

where $\lambda > 0$ and $\mu^2 < 0$, for a complex scalar field ϕ

$$\phi = \frac{\phi_1 + i\phi_2}{\sqrt{2}}$$
$$\phi^* \phi = \frac{\phi_1^2 + \phi_2^2}{2} \tag{2.26}$$

of two real fields ϕ_1 and ϕ_2 is invariant under the transformation $\phi \rightarrow \exp[i\alpha]\phi$. The minimum of the potential $V(\phi)$ is represented by a circle of radius

$$v = \sqrt{-\frac{\mu^2}{\lambda}} \tag{2.27}$$

in the complex plane ϕ_1, ϕ_2. Expanding the Lagrangian (EQUATION 2.25) around the vacuum state in terms of fields η and ξ by substituting

$$\phi(x) = \frac{1}{\sqrt{2}} [v + \eta(x) + i\xi(x)] \tag{2.28}$$

gives

$$\mathcal{L}' = \frac{1}{2} (\partial_\mu \xi)^2 + \frac{1}{2} (\partial_\mu \eta)^2 + \mu^2 \eta^2 + \text{const.} + \text{higher order terms.} \tag{2.29}$$

The third term has the form of a mass term with $m_\eta = \sqrt{-2\mu^2}$ for the η-field. The second term represents the kinetic energy of the η-field. But there is no corresponding mass term for ξ, the ξ-field is massless. So this Lagrangian $\mathcal{L}'$ contains a massless scalar particle which is also called Goldstone boson. That is the conclusion stated by the Goldstone theorem: Whenever a physical theory is not symmetric in the ground state, at least one massless scalar particle occurs [4].

2.6 The Higgs mechanism

The Lagrangian $\mathcal{L}$ (EQUATION 2.25) is invariant under global phase transformations of the type

$$\phi \rightarrow \exp[i\alpha]\phi. \tag{2.30}$$

To gain local phase invariance as well, the derivative ∂_μ has to be replaced by its covariant counterpart

$$D_\mu = \partial_\mu + iqA_\mu, \tag{2.31}$$

where the field A transforms as following:

$$A_\mu \to A_\mu - \frac{1}{q}\partial_\mu\alpha. \tag{2.32}$$

The gauge invariant Lagrangian becomes thus

$$\mathcal{L} = (\partial^\mu - iqA^\mu)\phi^*(\partial_\mu + iqA_\mu)\phi + \mu^2\phi^*\phi - \lambda(\phi^*\phi)^2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \tag{2.33}$$

For $\mu^2 > 0$, this corresponds to the QED Lagrangian (EQUATION 2.19).

For $\mu^2 < 0$ the field ϕ has to be translated to the ground state:

$$\begin{aligned}\mathcal{L}' =& \frac{1}{2}(\partial_\mu\xi)^2 + \frac{1}{2}(\partial_\mu\eta)^2 - v^2\lambda\eta^2 + \frac{1}{2}q^2v^2A_\mu A^\mu \\ &+ qvA_\mu\partial^\mu\xi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \text{interaction terms.}\end{aligned} \tag{2.34}$$

This Lagrangian describes a system with a massless Goldstone boson ξ, a massive scalar field η with mass $m_\eta = \sqrt{2\lambda v^2}$ and a massive vector field A with mass $m_A = qv$. By giving mass to A, the polarization degrees of freedom have been raised from 2 to 3, because longitudinal polarization is now allowed. However, this Lagrangian predicts a massless boson which has never shown up in any high-energy physics experiment in the form of missing momentum. Moreover, the fourth term $qvA_\mu\partial^\mu\xi$ describes an interaction in which ξ turns into A.

To get a more physical Lagrangian, EQUATION 2.28 can be rewritten as

$$\begin{aligned}\phi(x) &= \frac{1}{\sqrt{2}}\left[v + \eta(x) + i\xi(x)\right] \\ &\simeq \frac{1}{\sqrt{2}}\left[v + \eta(x)\right]\exp[i\frac{\xi(x)}{v}]\end{aligned} \tag{2.35}$$

in lowest order in ξ. Substituting a different set of fields h and θ into

$$\phi(x) = \frac{1}{\sqrt{2}}\left[v + h(x)\right]\exp[i\frac{\theta(x)}{v}] \tag{2.36}$$

where $h(x)$ is real and the gauge field A transforms as

$$A_\mu \to A_\mu - \frac{1}{qv}\partial_\mu\theta \tag{2.37}$$

results in

$$\begin{aligned}
\mathcal{L}'' = &\frac{1}{2}(\partial_\mu h)^2 - \lambda v^2 h^2 + \frac{1}{2}q^2 v^2 A_\mu^2 - \lambda v h^3 - \frac{1}{4}\lambda h^4 \\
&+ \frac{1}{2}q^2 A_\mu^2 h^2 + v q^2 A_\mu^2 h - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}.
\end{aligned} \tag{2.38}$$

Here, the Goldstone boson does not appear in the Lagrangian. It just describes two massive particles, a vector boson A and a scalar boson h which is called the Higgs boson. The massless Goldstone particle has been replaced by an additional degree of freedom for the vector field.

2.7 Yang-Mills theory

The strong interaction does not distinguish between the neutron and the proton. Both are just nucleons and can therefore be expressed as a product of an identical wave-function ψ_{Dirac} and an isospinor χ:

$$\psi_{\text{nucleus}} = \psi_{\text{Dirac}}\chi, \qquad \chi_P = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \chi_N = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{2.39}$$

Proton and neutron form a so-called isospin doublet. Transformations between these two eigenstates can be mediated by an 2×2 unitarian matrix U:

$$\chi' = U\chi, \quad \text{where} \quad UU^\dagger = U^\dagger U = 1. \tag{2.40}$$

The group of all such matrices U is the $SU(2)$, which is an anti-Hermitian group. Infinitesimal rotations in $SU(2)$ can be expressed as

$$U = \lim_{n\to\infty}\left(1 + \frac{i}{2}\cdot\frac{\boldsymbol{\tau}\cdot\boldsymbol{a}}{n}\right)^n = \exp\left(\frac{i}{2}\boldsymbol{\tau}\cdot\boldsymbol{a}\right), \tag{2.41}$$

where

$$\tau_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \tau_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \tau_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{2.42}$$

are the Pauli matrices and the dot product $\boldsymbol{\tau} \cdot \boldsymbol{a}$ is shorthand for $\tau_1 a_1 + \tau_2 a_2 + \tau_3 a_3$.

Local phase invariance can be derived for the $SU(2)$ group as well. Considering the Lagrangian

$$\mathcal{L} = i\bar{\psi}_1 \gamma^\mu \partial_\mu \psi_1 - m_1 \bar{\psi}_1 \psi_1 + i\bar{\psi}_2 \gamma^\mu \partial_\mu \psi_2 - m_2 \bar{\psi}_2 \psi_2 \qquad (2.43)$$

for two particle fields ψ_1 and ψ_2. Assuming that the two particles have the same mass $m = m_1 = m_2$, the Lagrangian can be written in a more compact form by combining ψ_1 and ψ_2 into a two-component column vector:

$$\mathcal{L} = i\bar{\psi}\gamma^\mu \partial_\mu \psi - m\bar{\psi}\psi, \qquad \psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}. \qquad (2.44)$$

Local phase transformations $\alpha(x)$ for the two-element column vector

$$\psi \to \psi' = \exp\left[i\frac{g}{2}\tau\alpha(x)\right]\psi \qquad (2.45)$$

with coupling g are applied analogous to the electric charge q. The covariant derivative can be written as

$$D_\mu = \partial_\mu + i\frac{g}{2}\tau \cdot W_\mu \qquad (2.46)$$

with three gauge fields W_μ. It is not a trivial matter to deduce the transformation rules for W_μ. At first oder the transformation can be written as:

$$W_\mu \to W'_\mu \simeq W_\mu - \partial_\mu\alpha - g\left(\alpha \times W_\mu\right). \qquad (2.47)$$

The last term arises from the non-Abelian character of this group.

The conclusive gauge invariant (Yang-Mills) Lagrangian reads as follows:

$$\mathcal{L} = i\bar{\psi}\gamma^\mu \partial_\mu \psi - m\bar{\psi}\psi - \frac{1}{4}W_{\mu\nu}W^{\mu\nu} - (g\bar{\psi}\gamma^\mu \tau\psi) \cdot W_\mu \qquad (2.48)$$

with $W_{\mu\nu}$ defined by

$$W_{\mu\nu} = \partial_\mu W_\nu - \partial_\nu W_\mu - g(W_\mu \times W_\nu). \qquad (2.49)$$

Analog to the QED Lagrangian, three currents

$$J = g(\bar{\psi}\gamma^\mu \tau\psi) \qquad (2.50)$$

which act as sources for the gauge fields can be defined.

Historically, the Yang-Mills theory turned out to be of little use when applied to the strong interaction, but the $SU(2)$ symmetry is very useful in the context of the electroweak interaction. This can be seen in the following section.

2.8   The spontaneous symmetry breaking of a local SU(2) group

The Lagrangian

$$\mathcal{L} = (\partial_\mu \phi)^\dagger (\partial^\mu \phi) - \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2, \tag{2.51}$$

where ϕ is an $SU(2)$ doublet of complex scalar fields:

$$\phi = \begin{pmatrix} \phi_\alpha \\ \phi_\beta \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \tag{2.52}$$

The gauge invariant Lagrangian is then

$$\mathcal{L} = (\partial_\mu \phi + i\frac{g}{2}\tau W_\mu \phi)^\dagger (\partial^\mu \phi + i\frac{g}{2}\tau W^\mu \phi) - V(\phi) - \frac{1}{4} W_{\mu\nu} W^{\mu\nu}, \tag{2.53}$$

with

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2. \tag{2.54}$$

In the case $\mu^2 < 0$ and $\lambda > 0$, the potential term $V(\phi)$ then has its minimum at

$$\phi^\dagger \phi = |\phi| = \frac{1}{2}(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2) = -\frac{\mu^2}{2\lambda}. \tag{2.55}$$

The expansion of $\phi(x)$ about an arbitrary chosen minimum ϕ_0

$$\phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}, \qquad \phi_1 = \phi_2 = \phi_4 = 0, \quad \phi_3 = v = \sqrt{-\frac{\mu^2}{\lambda}} \tag{2.56}$$

breaks spontaneously the $SU(2)$ symmetry and only one field, the Higgs field $h(x)$, of the four fields $\phi_{1...4}$ remains:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}. \tag{2.57}$$

To determine the masses for the gauge bosons W_μ^a generated by $h(x)$, it is sufficient to substitute ϕ_0 from EQUATION 2.55 into EQUATION 2.53:

$$(i\frac{g}{2}\tau W_\mu \phi_0)^\dagger (i\frac{g}{2}\tau W^\mu \phi_0) = \frac{g^2 v^2}{8} \left[(W_\mu^1)^2 + (W_\mu^2)^2 + (W_\mu^3)^2 \right]. \tag{2.58}$$

This gives $m_{W^a} \simeq \frac{1}{2}gv$ for the masses of the three gauge fields.

2.9   The Weinberg-Salam model

The three generations of leptons can be arranged into a left-handed weak isospin doublet ψ_L

$$T = 1/2 : \quad \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L \quad \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L \quad \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}_L \qquad \begin{matrix} T_3 \\ +1/2 \\ -1/2 \end{matrix} \qquad (2.59)$$

with hypercharge $Y_L = -1$ and a right-handed weak isospin singlet ψ_R

$$T = 0 : \quad e_R^-, \quad \mu_R^-, \quad \tau_R^- \qquad (2.60)$$

with hypercharge $Y_R = -2$ where the weak hypercharge Y is defined by

$$Y = 2(Q - T_3) \qquad (2.61)$$

with electric charge Q and weak isospin T. Combining the left- and right-handed components of ψ together with the electromagnetic interaction into a single Lagrangian leads to the electroweak unification.

The left-hand component of ψ_L is invariant under local phase transformations

$$\psi_L \rightarrow \psi_L' = \exp\left[i\left(\alpha(x) \cdot T + \beta(x)Y\right)\right] \psi_L \qquad (2.62)$$

where T and Y are the generators for the $SU(2)_L$ and $U(1)_R$ groups, if a triplet of vector fields W^1, W^2 and W^3 is introduced. To gain local phase invariance for the right-hand component of ψ_R,

$$\psi_R \rightarrow \psi_R' = \exp\left[i\beta(x)Y\right] \psi_R, \qquad (2.63)$$

only a single vector field B is necessary. The covariant derivative is then

$$D_\mu = \partial_\mu + ig\, T \cdot W_\mu + i\frac{g'}{2} Y B_\mu, \qquad (2.64)$$

where g and g' are the coupling constants alike to the electromagnetic case.

For left-hand leptons the derivative is

$$T = \frac{\tau}{2}, \quad Y = -1 \quad \Rightarrow \quad D_\mu = \partial_\mu + i\frac{g}{2}\tau \cdot W_\mu - i\frac{g'}{2} Y B_\mu \qquad (2.65)$$

and for right-handed leptons it reads

$$T = 0, \quad Y = -2 \quad \Rightarrow \quad D_\mu = \partial_\mu - ig'B_\mu. \tag{2.66}$$

The resulting electroweak Lagrangian satisfying $SU(2) \times U(1)$ local gauge invariance is:

$$\begin{aligned}
\mathcal{L}_1 =& \bar{\psi}_L \gamma^\mu [i\partial_\mu - \frac{g}{2}\tau \cdot W_\mu + \frac{g'}{2}B_\mu]\psi_L \\
& + \bar{\psi}_R \gamma^\mu [i\partial_\mu + \frac{g'}{2}B_\mu]\psi_R - \frac{1}{4}W_{\mu\nu} \cdot W^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}.
\end{aligned} \tag{2.67}$$

$\mathcal{L}_1$ embeds the weak isospin and hypercharge interaction. The final two terms $B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$ and $W_{\mu\nu}$ represent the kinetic energy of the gauge fields and the self-coupling of the W boson (EQUATION 2.49).

The Higgs mechanism can be used to formulate the Lagrangian $\mathcal{L}_1$ in such a way that three vector bosons (meant to be the $W^\pm$ and the Z^0) become massive whereas one gauge boson (the photon γ) remains massless. The originally choice made in 1967 by Weinberg is to add to $\mathcal{L}_1$ a $SU(2) \times U(1)$ gauge invariant Lagrangian

$$\begin{aligned}
\mathcal{L}_2 = (i\partial_\mu\phi - \frac{g}{2}\tau \cdot W_\mu - \frac{g'}{2}YB_\mu\phi)^\dagger(i\partial^\mu\phi - \frac{g}{2}\tau \cdot W^\mu - \frac{g'}{2}YB_\mu\phi) \\
- \mu^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2 \quad (2.68)
\end{aligned}$$

for the scalar field ϕ

$$\phi = \begin{pmatrix} \phi_\alpha \\ \phi_\beta \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \tag{2.69}$$

with hypercharge $Y = 1$. The choice of the parameters $\lambda > 0$ and $\mu^2 < 0$ enforce the gauge bosons to be massive. To get the masses for the bosons, the vacuum expectation value

$$\phi_0 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v \end{pmatrix} \tag{2.70}$$

of the Higgs doublet is substituted into EQUATION 2.68. Carrying out the necessary calculations, the relevant term results in

$$\begin{aligned}
& \left| \left(-i\frac{g}{2}\tau \cdot W_\mu - i\frac{g'}{2}B_\mu \right) \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v \end{pmatrix} \right|^2 \\
& = \frac{1}{4}v^2 g^2 W_\mu^+ W^{-,\mu} + \frac{1}{8}v^2 \left[g^2(W_\mu^3)^2 - 2gg'W_\mu^3 B^\mu + g'^2 B_\mu^2 \right] \quad (2.71)
\end{aligned}$$

with $W^\pm = (W^1 \mp W^2)/\sqrt{2}$. Comparing the first term with the expected mass term for a charged gauge boson $M_W^2 W^+ W^-$ yields in

$$M_W = \frac{1}{2} v g. \tag{2.72}$$

Writing the last term as a sum of a massive and a massless term

$$\frac{1}{8} v^2 \left[g^2 (W_\mu^3)^2 - 2 g g' W_\mu^3 B^\mu + g'^2 B_\mu^2 \right] = \frac{1}{8} v^2 [g W_\mu^3 - g' B_\mu]^2 + 0 [g' W_\mu^3 + g B_\mu]^2$$

$$= \frac{1}{2} M_Z^2 Z_\mu^2 + \frac{1}{2} M_A^2 A_\mu^2, \tag{2.73}$$

helps to identify the corresponding mass terms for the neutral bosons

$$M_A = 0 \qquad\qquad \Rightarrow \qquad A_\mu = \frac{g' W_\mu^3 + g B_\mu}{\sqrt{g^2 + g'^2}}$$

$$M_Z = \frac{v}{2} \sqrt{g^2 + g'^2} \qquad \Rightarrow \qquad Z_\mu = \frac{g W_\mu^3 - g' B_\mu}{\sqrt{g^2 + g'^2}} \tag{2.74}$$

with the fields A_μ and Z_μ which are linear combinations of the W_μ^3 and B_μ fields. Since the fields A_μ and Z_μ are normalized, they can be re-expressed in terms of θ_W

$$\theta_W = \arctan \frac{g'}{g} : \qquad \begin{array}{rcl} A_\mu &=& \cos\theta_W B_\mu + \sin\theta_W W_\mu^3 \\ Z_\mu &=& -\sin\theta_W B_\mu + \cos\theta_W W_\mu^3 \end{array} \tag{2.75}$$

where θ_W is the so-called Weinberg-mixing angle between the W_μ^3 and B_μ fields. The Higgs doublet is constructed with the requirement that the photon field A_μ is massless whereas the triplet W^+, W^- and Z^0 acquires mass to be consistent with the experimental results. Comparing EQUATION 2.72 and EQUATION 2.75 leads to the mass relation

$$\frac{M_W}{M_Z} = \cos\theta_W \tag{2.76}$$

which is a prediction for the Standard Model with a Higgs doublet. On the other hand, it fixes the relative strength ρ of charged and neutral current weak interaction to equality:

$$\rho = \frac{M_W^2}{M_Z^2 \cos^2\theta_W} = 1. \tag{2.77}$$

2.10 Masses of the fermions

In a similar way as outlined in the previous section, the spontaneous symmetry breaking of the Higgs field can be used to give masses to the leptons. In the original Lagrangian (EQUATION 2.67), a fermion mass term $m\bar{\psi}\psi$ was excluded. To generate masses for the leptons, the $SU(2) \times U(1)$ gauge invariant term

$$\mathcal{L}_3 = -\tilde{g}_f \left[\bar{\psi}_L \phi \psi_R + \bar{\psi}_R \phi^\dagger \psi_L \right] \tag{2.78}$$

has to be included in the Lagrangian. The coupling $\tilde{g}_f$, also called Yukawa-coupling, is different for each fermion and cannot be calculated. However, substituting EQUATION 2.56 into $\mathcal{L}_3$ generates the required mass of the fermions:

$$m_f = \frac{\tilde{g}_f v}{\sqrt{2}}. \tag{2.79}$$

The quark masses are generated in a similar way; a detailed derivation can be found in [5].

The actual masses of the fermions are not predicted and are just parameters. However, since the Higgs coupling to the fermions is proportional to their masses

$$\tilde{g}_f = \frac{\sqrt{2}m_f}{v} \quad \Rightarrow \quad \tilde{g}_\tau = \frac{m_\tau}{m_\mu} \tilde{g}_\mu, \quad \ldots, \tag{2.80}$$

the electroweak theory can be tested if the Higgs particle is finally observed. The Higgs bosons' property to couple to fermions in proportion to their masses causes experimental challenges. From the experimental point of view, the decay into light fermions like electrons and up- and down-quarks or even muons could be detected easily but they couple to the Higgs boson only weakly. In contrast, the decay into heavier fermions like taus, charm-, bottom- and top-quarks is much more complicated to detect.

Chapter 3

Higgs Phenomenology

In order to understand the considerations which have to be taken into account while the search for the Higgs boson is planned, this chapter summarizes the predicted properties of the Higgs boson. First, the unitarity, triviality and stability constraints for the allowed mass range are introduced. Then the different production and decay modes together with their relative contributions are presented. Finally, the results of the searches for the Higgs boson which have already been conducted are reviewed.

3.1   Constraints on the Higgs boson mass

The electroweak coupling constant g is related to the Fermi coupling constant $G_\mathrm{F} = 1.166 \times 10^{-5}\,\mathrm{GeV}$ found in the β decay by the low energy relation

$$\frac{G_\mathrm{F}}{\sqrt{2}} = \frac{g^2}{8M_\mathrm{W}^2}.$$

(3.1)

Combined with the W mass relation (EQUATION 2.72) the value v of the ground state of the Higgs field (also called vacuum expectation value) can be derived:

$$v = \sqrt{\frac{1}{\sqrt{2}G_\mathrm{F}}}$$

(3.2)

$$\simeq 246\,\mathrm{GeV}.$$

31

Therefore the typical range for electroweak phenomena, defined by the masses of the massive gauge bosons W and Z is of order 100 GeV.

The mass of the Higgs boson is determined from the vacuum expectation value

$$M_{\mathrm{H}}^2 = \lambda v^2 \qquad (3.3)$$

with the coupling constant λ which is an unknown parameter. So the mass of the Higgs boson cannot be predicted either. Nevertheless, the Higgs boson is required to be in a particular mass range to be consistent with the Standard Model.

Constraints on the Higgs boson mass can be obtained for example from the cross section for longitudinally polarized W pair production in $f\bar{f}$ annihilation. The process is mediated by t-channel neutrino exchange and s-channel photon and Z boson exchanges. An additional contribution arises from the s-channel Higgs boson exchange (FIGURE 3.1).
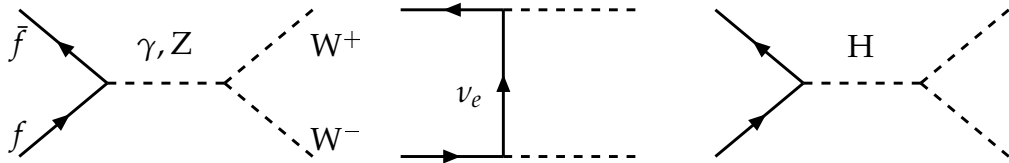

Figure 3.1. Leading order Feynman diagrams for the W pair production in e^+e^- collisions.

Without a Higgs boson canceling out the high-energy divergences, the cross section grows infinitely which is called violation of unitarity. This rise is only prevented if the Higgs boson coupling to fermions is proportional to their masses. The cancellation is only operative, if the Higgs mass is smaller than [6]

$$M_{\mathrm{H}} < \sqrt{\frac{8\pi\sqrt{2}}{3G_{\mathrm{F}}}} = 1 \, \mathrm{TeV}. \qquad (3.4)$$

As can be seen in FIGURE 3.2, the cross section for W pair production measured by the experiments of the Large Electron-Positron Collider (LEP) is in perfect agreement with the theoretical prediction.

A more restrictive upper bound on the value of the Higgs boson mass follows from hypothetical assumptions on the energy scale Λ up to which the Standard Model may be valid before new physical phenomena may
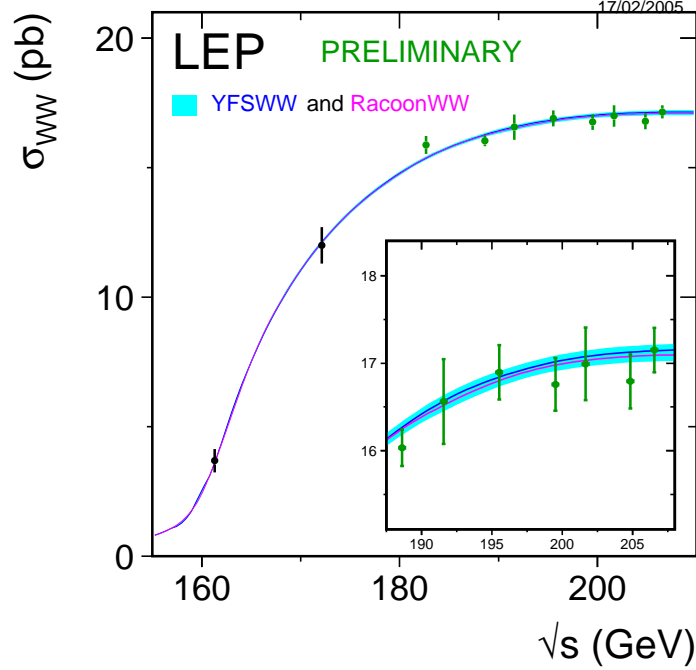
Figure 3.2. Cross section for the reaction $e^+e^- \to W^+W^-$ measured by the four LEP experiments up to a center-of-mass energy of $\sqrt{s} = 209\,\mathrm{GeV}$ together with the prediction of the electroweak theory (solid line). Taken from [7].

appear. Quantum fluctuations modify the self-interaction of the Higgs boson (FIGURE 3.3). This can be described by higher-order terms of λ in the Higgs potential. In order to avoid the coupling to increase indefinitely, the upper bound on the Higgs mass is [8]

$$M_{\mathrm{H}}^2 < \frac{8\pi^2 v^2}{3\log(\Lambda^2/v^2)}. \tag{3.5}$$

For the minimal cut-off parameter $\Lambda \sim 1\,\mathrm{TeV}$ up to which the Standard Model is assumed to be valid, the maximum value is $M_{\mathrm{H}} \sim 750\,\mathrm{GeV}$.

A lower bound can be based on the requirement for the Higgs potential to be stable. If the quantum corrections cause λ to become negative, the ground state is not stable anymore. One-loop corrections to the Higgs
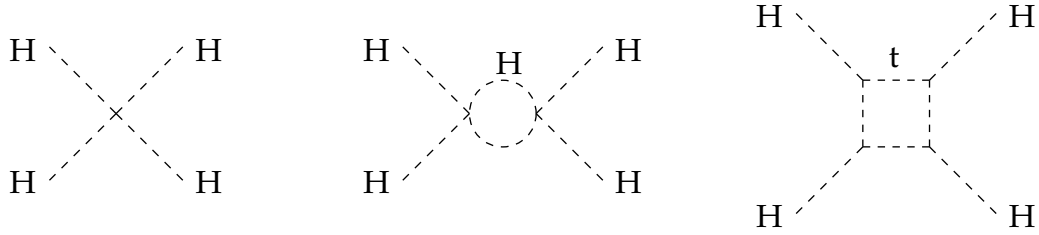
Figure 3.3. Feynman diagrams showing the self-interaction of the Higgs field [8].
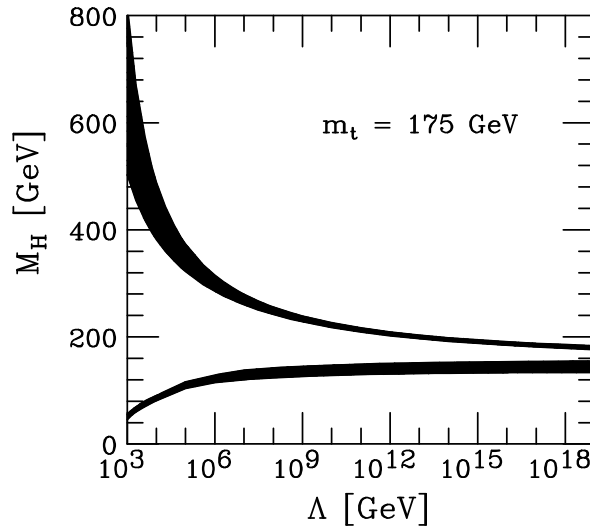
Figure 3.4. Constraints on the Higgs boson mass as a function on the energy scale Λ at which the Higgs boson would become strongly interacting. The allowed region lies between the bands (the shaded bands illustrate the impact of various uncertainties). From [8].

potential yield to the relation [9]:

$$M_H^2 > \frac{3G_F\sqrt{2}}{8\pi^2} \left(2M_W^4 + M_Z^4 - 4m_t^4\right) \log\left(\frac{\Lambda^2}{v^2}\right). \qquad (3.6)$$

Due to the very large mass of the top quark, this bound is negative and therefore not a real constraint. However, two-loop corrections to the Higgs potential are sizable and provide a positive-definite lower bound on the mass.

Taking all constraints into account, the Higgs boson mass is restricted to a narrow mass windows between 130 and 190 GeV (FIGURE 3.4), if the electroweak theory is indeed valid all the way up to a grand unification scale 10^{19} GeV.

3.2 Higgs boson decays

Once a value for the Higgs boson mass is assumed, it is straightforward to calculate, using the given Higgs coupling (EQUATION 2.80), the decay rates to massive fermions. In the Born approximation (neglecting any QCD corrections for hadronic decays), the partial decay width into a pair of fermions with color N_c is [10]

$$\Gamma(H \to f\bar{f}) = N_c \frac{G_F M_H}{4\pi\sqrt{2}} m_f^2 \beta_f^3, \qquad \beta_f = \left(1 - \frac{4m_f^2}{M_H^2}\right)^{\frac{1}{2}} \tag{3.7}$$

with β_f being the velocity of the fermions in the final state. The branching ratio is proportional to M_H in the limit of a large Higgs mass. In the lepton case ($N_c = 1$), only decays into $\tau^+\tau^-$ pairs and, to a much lesser extent, decays into muon pairs are relevant.

The partial decay widths for Higgs boson decays into W and Z pairs are [9]

$$\Gamma(H \to WW) = \frac{G_F M_H^3}{8\pi\sqrt{2}} \sqrt{1 - 4x} \left(1 - 4x + 12x^2\right), \quad x = \frac{M_W^2}{M_H^2} \tag{3.8}$$

and

$$\Gamma(H \to ZZ) = \frac{G_F M_H^3}{16\pi\sqrt{2}} \sqrt{1 - 4x'} \left(1 - 4x' + 12x'^2\right), \quad x' = \frac{M_Z^2}{M_H^2}. \tag{3.9}$$

The decay rates into gauge boson pairs are asymptotically proportional to M_H^3 and $1/2 M_H^3$, respectively. The factor $1/2$ arises from the fact that the W bosons are distinguishable. For large Higgs boson masses, the decay width into WW bosons is two times larger than the decay width into ZZ bosons. Below the WW threshold, the total width is rather small, typically much less than $1\,\text{GeV}$ (FIGURE 3.5).

In the final factors of EQUATION 3.8 and EQUATION 3.9, $8x^2$ and $8x'^2$ arise from decays into transversely polarized gauge bosons. For large Higgs boson masses, the weak bosons are mostly longitudinally polarized [11]:

$$\frac{\Gamma(H \to V_L V_L)}{\Gamma(H \to V_L V_L) + \Gamma(H \to V_T V_T)} = \frac{1 - 4x + 4x^2}{1 - 4x + 12x^2} \xrightarrow{M_H \gg M_V} 1. \tag{3.10}$$
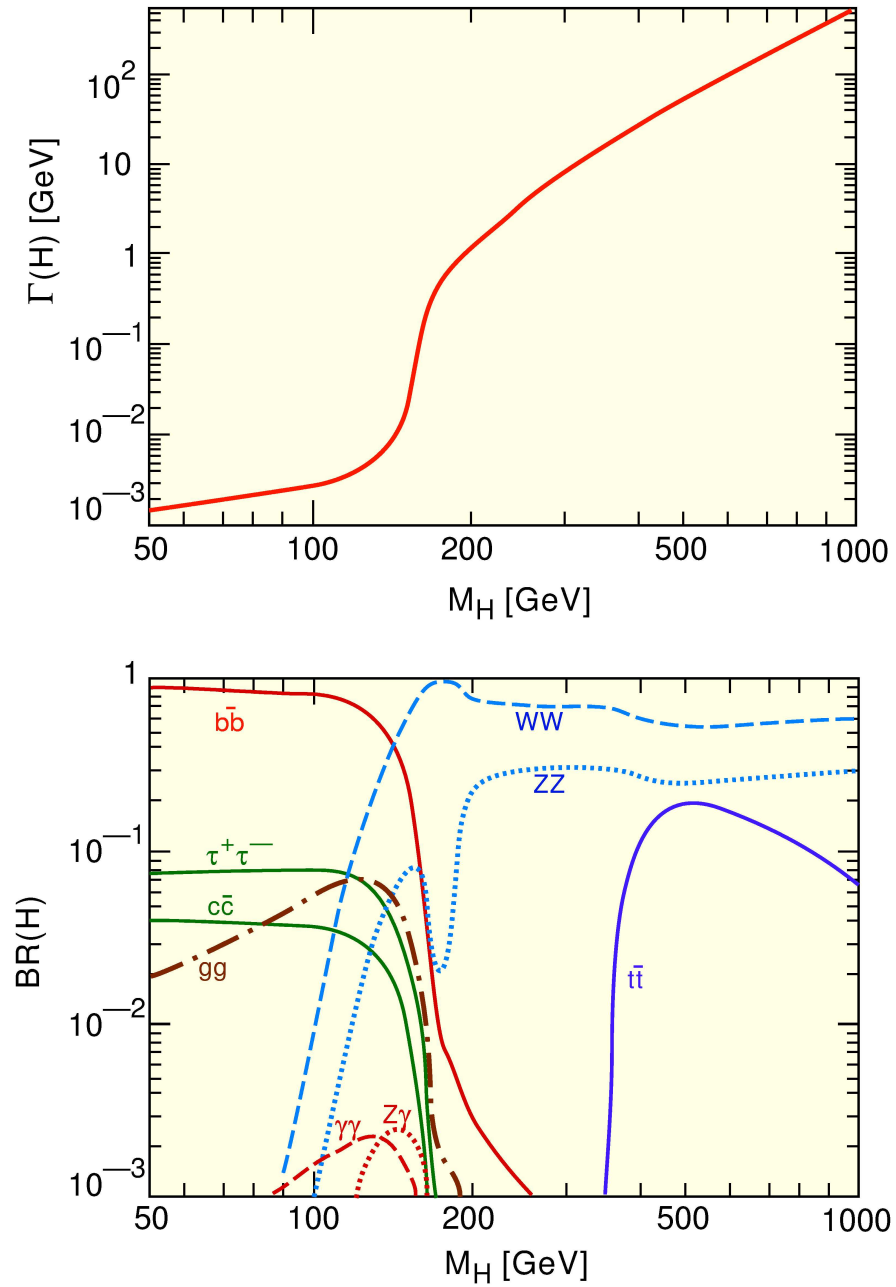
Figure 3.5. Total Higgs decay width Γ (top) and branching ratios (BR) of the main decay channels (bottom) for a Standard Model Higgs Boson as a function of its mass. Taken from [12].

Decays into massless photons are only possible through virtual intermediate states. The decay rate is

$$\Gamma(\text{H} \to \gamma\gamma) = \alpha \, \frac{G_\text{F} M_\text{H}^3}{8\pi^3 \sqrt{2}} \, |\mathcal{M}|^2, \qquad (3.11)$$

where $\mathcal{M}$ is the matrix element calculated by adding all the contributing loop diagrams involving massive particles.

The branching fractions for the Standard Model Higgs Boson which have been calculated with the HDECAY program [13] by its author are shown in FIGURE 3.5 as a function of the Higgs boson mass. For Higgs boson masses below 130 GeV, the H $\to$ b$\bar{\text{b}}$ decay channel dominates. The H $\to$ $\tau\tau$ decay rate is also sizable. For larger masses, the Higgs boson decays almost exclusively into WW$^{(*)}$ and ZZ$^{(*)}$ pairs.

The LHC has the capabilities to search for the Higgs boson over the full mass range from the current lower experimental limit of 114.4 GeV (see SECTION 3.4) up to 1 TeV. Generally spoken, the Higgs boson decays most of the time into the heaviest particles which are kinematically allowed. In the intermediate mass range $(140 - 180\,\text{GeV})$ and the high mass range $(180 - 800\,\text{GeV})$, the Higgs boson decays almost exclusively into weak boson pairs. Apart from a small gap between 160 GeV and 180 GeV, the most promising decay mode with a very clean experimental signature is the "Gold Plated Channel" H $\to$ $(\text{Z} \to ll)(\text{Z} \to ll)$ which features four isolated leptons coming from a common vertex.

In the low mass region, the decay channels H $\to$ $\gamma\gamma$, H $\to$ b$\bar{\text{b}}$ and H $\to$ $\tau\tau$ offer the possibility of finding the Higgs boson even if these final-states are more difficult to separate from background processes. Combining several decay final states, the Standard Model Higgs boson is expected to be found at the LHC with an accumulated luminosity of 30 fb^{-1} up to a mass $M_\text{H} \leq$ 800 GeV [14].

3.3 Higgs production

The main production mechanisms for Higgs particles at hadron colliders make use of the preferential coupling to heavy particles, which are basically W and Z bosons and the top quark. The four main production processes at LHC which are depicted in FIGURE 3.6 are explained in more detail in the following.

a) Gluon fusion, $gg \to$ H: The gluon fusion is mediated by heavy quark triangle loops which are mainly top quarks and, to lesser extent, bottom quarks. The gluon fusion mechanism is the dominant process at the LHC in the entire relevant mass range from 100 GeV up to 1 TeV.

b) Weak boson fusion, $qq \to VV \to qq +$ H: The second largest contribution to the cross section comes from the fusion of two weak bosons which are radiated from the initial-state quarks. The relative contribution for this production mechanism becomes more and more important as the Higgs boson mass increases and becomes comparable to the gluon fusion for masses beyond 600 GeV. For lower masses the cross section is smaller by one order of magnitude. The characteristics of the Higgs boson production via the vector boson fusion can be exploited to suppress the large backgrounds. These characteristics are the accompanying jets originating from the incoming initial-state quarks which emit the weak bosons and the lack of color flow between the initial-state quarks.

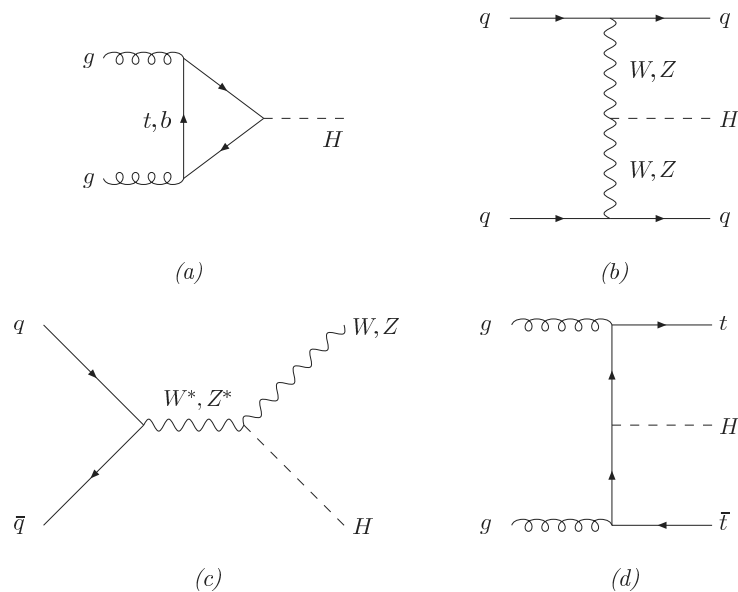**Figure 3.6.** Typical leading order Feynman diagrams for Higgs boson production mechanisms relevant at the LHC: Gluon fusion (a), weak boson fusion (b), Higgs-strahlung (c), associated production (d). From [15].

c) Associated production with W and Z, $q\bar{q} \to V^* \to V +$ H: In the associated production together with a W or a Z boson, the Higgs boson is radiated by an off-shell weak boson which was produced in a

fermion anti-fermion annihilation. Therefore, this production mode is also called Higgsstrahlung. Higgsstrahlung shows only a minor contribution to the production cross section at the LHC, whereas it was the dominating one at LEP.

d) Associated production with heavy quarks, $gg, q\bar{q} \rightarrow Q\bar{Q} + H$: Especially the Higgs boson production together with top quark pairs plays a significant role at the LHC for light Higgs bosons below 130 GeV. The branching ratio of t quark decays into b quarks reaches almost 100 % and jets formed out of b quarks can be identified in the detector due to the large lifetime of b quarks. This makes the observation of $H \rightarrow b\bar{b}$ and $H \rightarrow \gamma\gamma$ possible. The fact that the cross section of this production mode is directly proportional to the square of the Yukawa-coupling of the top quarks to the Higgs boson, attracts additional attention.

The Higgs boson production in association with bottom quark pairs is only interesting to study in supersymmetric scenarios where the Higgs coupling to b quarks can be strongly enhanced.

The cross section for Higgs boson production at the LHC is shown in FIGURE 3.7 for the four mentioned production modes.

3.4 Searches for the Higgs boson

First direct searches for the Higgs boson have already been carried out by the Large Electron-Positron Collider (LEP) experiments at CERN. During the phase LEP1, electron-positron collisions at energies near the Z boson resonance $\sqrt{s} \simeq M_Z$ were investigated. The predominant production mode at this energy is the Bjorken process where the generated Z boson decays into a real Higgs boson and an off-shell Z boson which decays into two light fermions. It can also be produced in the Z decay $Z \rightarrow H\gamma$ together with an associated photon through triangular loops mediated by heavy fermions and the W boson. The Higgs boson has been searched in final-states which do not suffer from the e^+e^- background. These are the final-state topologies

- $Z \rightarrow (H \rightarrow hadrons)(Z^* \rightarrow \nu\bar{\nu})$ and

- $Z \rightarrow (H \rightarrow hadrons)(Z^* \rightarrow e^+e^-, \mu^+\mu^-)$
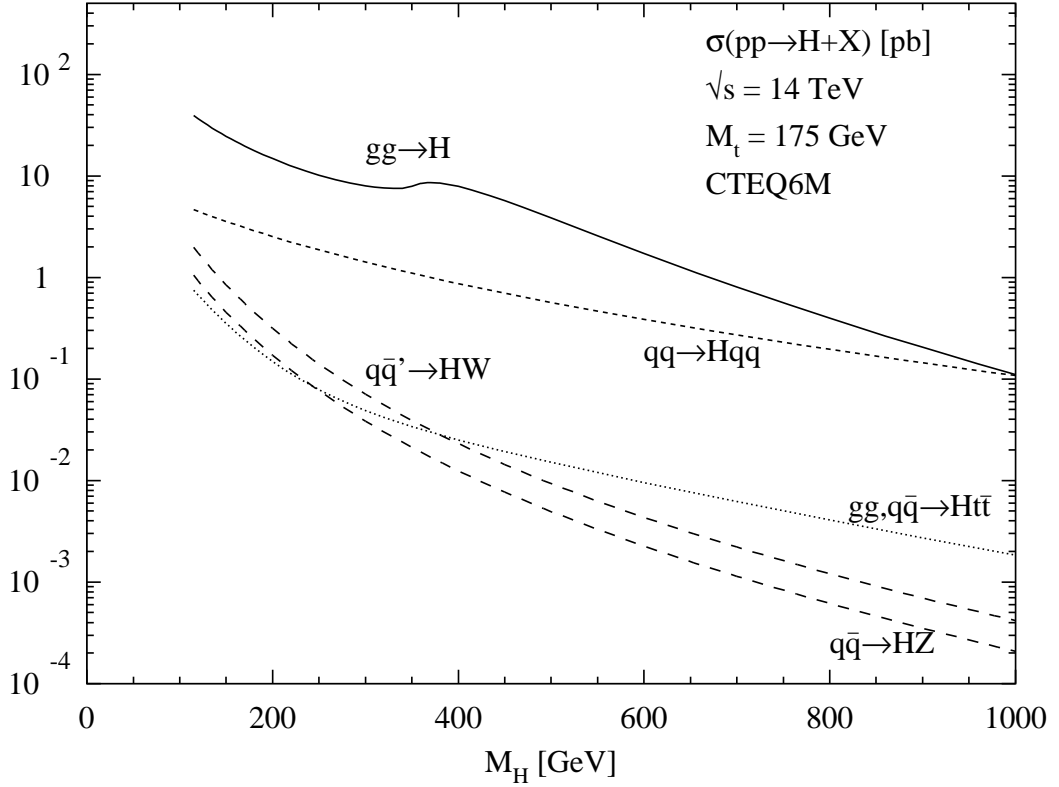
Figure 3.7. Production cross section σ for a scalar Higgs boson produced at the LHC as a function of its mass. Gluon fusion (solid line), weak boson fusion (short dashed line), associated production with a top quark pair (dotted line) and Higgs-strahlung (long dashed lines) are shown. Taken from [15].

with two energetic leptons. The absence of any Higgs boson signal by the four LEP collaborations, allowed to exclude a Higgs boson mass below $M_H < 65.2\,\mathrm{GeV}$ at 95 % confidence level [16].

The search has then been extended at LEP2 up to center-of-mass energies of $\sqrt{s} = 209\,\mathrm{GeV}$. The dominant production mode is Higgsstrahlung where the Higgs is radiated as Bremsstrahlung by an off-shell Z^* boson (see above). The final-states which were investigated at LEP2 are

- $Z \to (H \to b\bar{b})(Z^* \to \nu\bar{\nu})$,

- $Z \to (H \to b\bar{b})(Z^* \to e^+e^-, \mu^+\mu^-)$,

- $Z \to (H \to b\bar{b})(Z^* \to q\bar{q})$,

- $Z \to (H \to \tau^+\tau^-)(Z^* \to b\bar{b})$ and
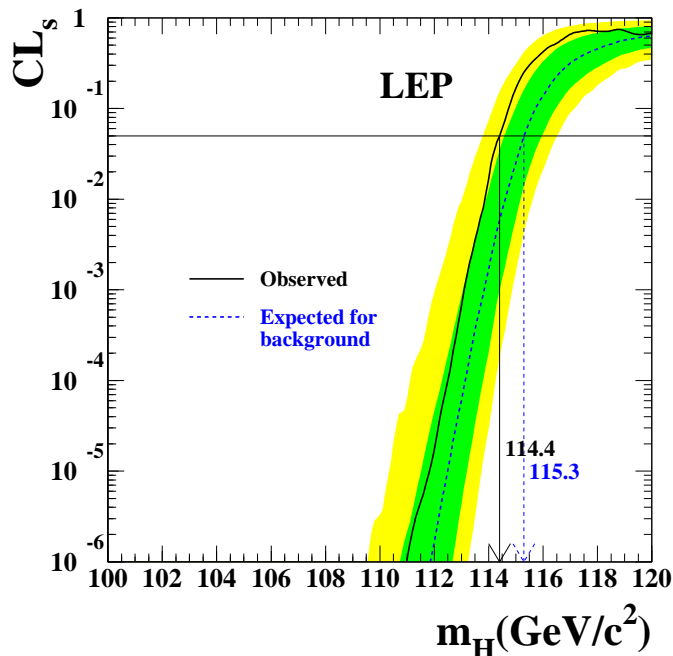
Figure 3.8. The confidence level ratio $CL_s = CL_{s+b}/CL_b$ for signal plus background hypothesis, as a function of the test mass M_H. Observation (solid) and median background expectation (dashed). The dark and light shaded areas around the background expectation corresponds to the 68 % and 95 % probability bands. The intersection of the horizontal line defines the 95 % confidence level lower bound on the Higgs boson mass [17].

- $Z \rightarrow (H \rightarrow b\bar{b})(Z^* \rightarrow \tau^+\tau^-)$.

Combining the results of the four LEP experiments did not show any significant excess above the expected background after having collected $2461\,pb^{-1}$ of data. The exclusion limit $M_H > 114.4\,GeV$ at 95 % confidence level was established [17] by the four LEP collaborations (FIGURE 3.8).

Measurements have also been performed by the TEVATRON experiments CDF and D∅ in a variety of channels [18]. In Run II the TEVATRON accelerator provides $p\bar{p}$ collisions at $\sqrt{s} = 1.96\,TeV$. In the Higgs boson mass region below $135\,GeV$, $b\bar{b}$ final-states in which the Higgs is produced in association with a W or Z boson ($p\bar{p} \rightarrow WH, p\bar{p} \rightarrow ZH$) are the most important ones. Therefore, studies have been conducted in the

- $W^* \rightarrow (H \rightarrow b\bar{b})(W \rightarrow l\nu)$,

Figure 3.9. Limits on the Higgs boson production cross section for the combined TEVATRON experiments as a function of the Higgs boson mass M_H [18].

- $Z^* \rightarrow (H \rightarrow b\bar{b})(Z \rightarrow \nu\nu)$ and

- $Z^* \rightarrow (H \rightarrow b\bar{b})(Z \rightarrow ee, \mu\mu)$

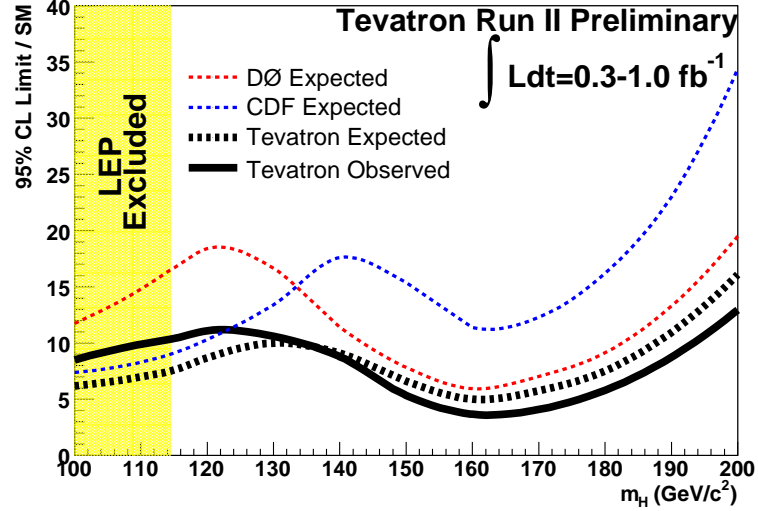decay channels. For higher masses above $135\,\text{GeV}$, the Higgs boson decays mostly into W pairs. Hence, the search is extended to the decay channel $WH \rightarrow (W \rightarrow l^+\bar{\nu})(W \rightarrow l^-\nu)(W^* \rightarrow q\bar{q})$. Considered leptons are again either electrons or muons. If the Higgs is produced via gluon fusion, leptonic decay channels of the W bosons $H \rightarrow (W \rightarrow l\nu)(W \rightarrow l\nu)$ can be used to suppress the large QCD background. After having combined the analyses of both TEVATRON experiments with luminosities in the range from $0.2 - 1.0\,\text{fb}^{-1}$, an upper limit on the Standard Model Higgs boson cross section has been established. At the 95 % confidence level, the cross section upper limit is a factor of 10.4 higher for a Higgs boson with $M_H = 115\,\text{GeV}$ and a factor of 3.8 higher for a Higgs particle with $M_H = 160\,\text{GeV}$ than the expected cross section predicted by the Standard Model (FIGURE 3.9).

Besides direct searches, further constraints on the Higgs boson mass can be obtained from high precision measurements of electroweak parameters at the Z pole. These observables include the Z boson mass, the hadronic and the leptonic partial decay width and parameters determined from the angular distributions of the decay products, for example. The masses of

the top quark, the W boson and the Higgs boson contribute to the calculation of these measurable quantities through radiative corrections. Whereas the top quark and the W boson introduce a quadratic dependence on their masses, the Higgs mass dependence is only logarithmic. The LEP electroweak working group (LEPEWWG [19]) has combined all the electroweak variables obtained from the data taken at LEP, TEVATRON and the Standford Linear Collider (SLC) into a constraint fit to the Standard Model treating the Higgs boson mass as a free parameter (FIGURE 3.10). The $\Delta\chi^2$ fit leads to a preferred value for the mass of the Higgs boson of 76 GeV and to an upper limit of 144 GeV at 95 % confidence level. This limit increases to 182 GeV if the exclusion limit given by LEP2 is taken into account.
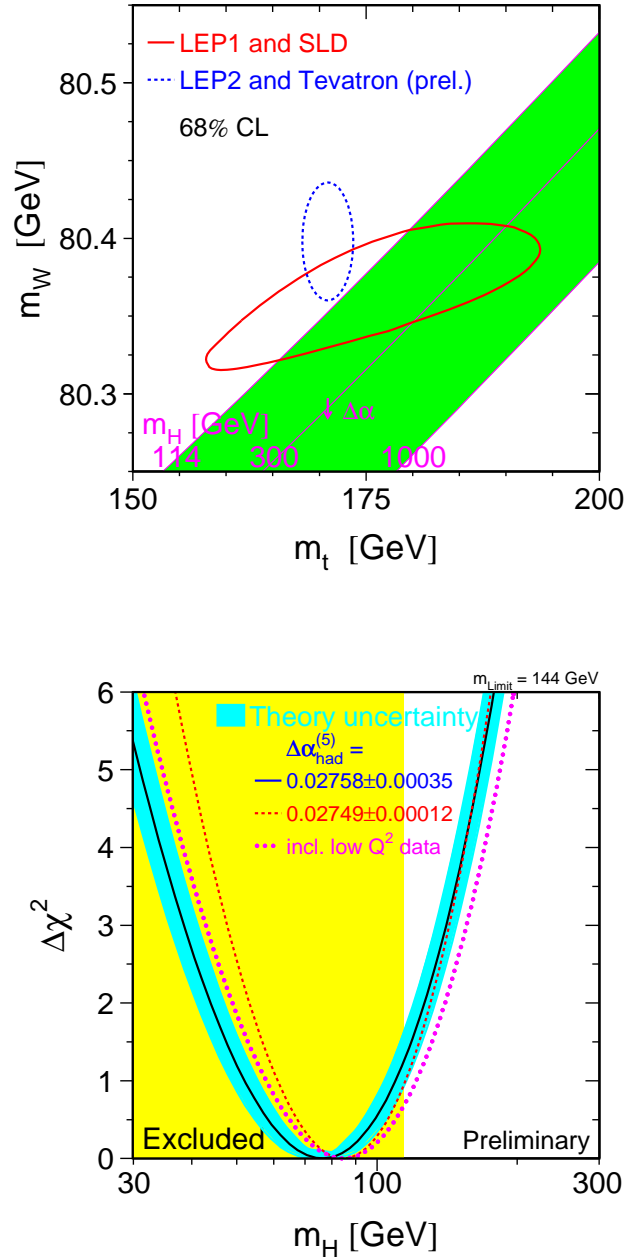
Figure 3.10. Standard Model relationship of the W boson and the top quarks mass as a function of the Higgs boson mass (top) and the $\Delta\chi^2$ curve derived from the fit to electroweak observables as a function of the Higgs mass assuming the Standard Model to be the correct theory of nature (bottom) [7].

Chapter 4

The CMS Experiment at the Large Hadron Collider (LHC)

The Large Hadron Collider (LHC) currently being under construction at CERN near Geneva is expected to answer several unsolved questions of Particle Physics. Among other tasks which are listed in this chapter, its main goal is to explore the nature of electroweak symmetry breaking by the discovery of the Higgs particle and the investigation of its properties. The CMS experiment which is one of the two large general purpose particle detectors being built on the LHC is introduced in this chapter as well.

4.1   The Large Hadron Collider (LHC)

In the last twenty years the Standard Model of Particle Physics has been tested experimentally to excellent precision and is well established as the description of nature at subatomic scales. Many of these experiments have been carried out at the Large Electron-Positron collider at CERN. The Large Electron-Positron collider has been operational from 1989 until 2000 and has provided accelerated electrons and positrons with an energy between 45 GeV at the beginning and 103 GeV at the end leading to center-of-mass energies of 90 and 209 GeV respectively. LEP precisely measured the quantities of the Z and W bosons and confirmed accurately the predictions of the Standard Model.

To explore new physics up to 1 TeV and to answer a number of fundamental questions which the Standard Model fails to predict, the next gen-

eration particle accelerator, the Large Hadron Collider, was planned and is currently being built in the 27 km long former LEP tunnel at CERN. The potential of LEP to find undiscovered phenomena was limited by its energy. At a center-of-mass energy of about 200 GeV most of the energy loss

$$-\Delta E = \frac{4\pi\alpha}{3R}\beta^3\gamma^4 \quad \text{with} \quad \beta = \frac{v}{c} \approx 1 \quad \text{and} \quad \gamma = \frac{E}{mc^2}, \qquad (4.1)$$

was caused by synchrotron radiation. There were two possibilities to reduced this effect. The radius of the collider could be increased (a linear one would be optimal) or heavier particles could be used. For obvious financial reasons, there was a strong motivation to re-use the LEP tunnel. Therefore, it was decided to build a proton collider. Since protons are 2000 times heavier than electrons, the energy loss in synchrotron radiation is about 10^{13} times smaller than for electrons. This eliminates synchrotron radiation as limiting factor of the energy of the collider. Since protons are not elementary particles like leptons but composite objects of quarks and gluons which carry only a fraction of the proton momentum, beam energies considerably above the scale up to which one would like to observe new phenomena are needed. The maximal energy of a proton beam in the LEP tunnel is limited by the maximal strength of the magnetic dipole field which is necessary to compensate the centrifugal force of the charged particles. For the LHC project, dipole magnets with a 8.33 T nominal magnetic field are assembled which correspond to a beam energy of 7 TeV. In order to provide such high magnetic fields, new types of superconducting magnets have been developed. The magnet coils are made of copper-clad niobium-titanium (NbTi) cables and will be cooled down for operation to superfluid liquid helium temperatures of about 1.9 K.

Since the cross sections for the most physics processes the LHC is looking for is very small, only a tiny fraction of the proton collisions produces interesting events at the interaction points. The event rate $\dot{N}$ for a certain physics process is proportional to its cross section σ and the luminosity $\mathcal{L}$ of the collider:

$$\dot{N} = \mathcal{L} \cdot \sigma. \qquad (4.2)$$

The luminosity of an accelerator which collides two bunches containing n particles at a frequency f is given by

$$\mathcal{L} = f \frac{n^2}{4\pi\sigma_x\sigma_y}, \qquad (4.3)$$

where σ_x and σ_y characterize the approximately Gaussian beam profiles in horizontal and vertical directions.

Since the cross section decreases proportional to E^2 for parton-parton collisions, an increase of the luminosity with that amount is required to obtain the same statistical significance. For hadron colliders the situation is even more complicated as the structure function of the proton has to be taken into account. This leads to the requirement of a luminosity of $10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ (commonly referred to as high luminosity running). The first three years of operation, the LHC will run at a reduced luminosity of $2 \times 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ (referred to as low luminosity running) which already exceeds the luminosity of the TEVATRON collider by more than an order of magnitude. The TEVATRON collider is today's most powerful proton collider, located at the Fermi National Accelerator Laboratory (FermiLab) near Chicago in the USA.

EQUATION 4.3 is valid for head-on collisions only. The luminosity for one bunch at the LHC is given by:

$$\mathcal{L} = f\,\frac{\gamma n^2}{4\pi \epsilon_n \beta^*}\,F, \tag{4.4}$$

where γ is the Lorentz factor, ϵ_n the normalized transverse emittance (with a design value of $3.75\,\mu$m), β^* is the amplitude function at the interaction point (0.55 m) and F is the reduction function due to the crossing angle. Inserting these numbers with the nominal bunch intensity of 1.15×10^{11} protons leads to a luminosity of $3.5 \times 10^{30}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ for one bunch. In order to obtain the required luminosity, each beam will have 2808 bunches with an nominal bunch spacing of 25 ns. This corresponds to a collision frequency of 40 MHz, which imposes stringent requirements on the response time of the LHC detectors. Due to the large number of protons per bunch, between 10 and 20 inelastic proton-proton collisions will occur per bunch crossing. This leads to increasingly more difficult operational conditions detecting the interesting events among 20 other superimposed (so-called pile-up) events.

In contrast to most hadron colliders in the past, the LHC investigates pp collisions instead of p$\bar{\text{p}}$ collisions since it is very difficult to provide enough anti-protons to achieve the LHC design luminosity. But at LHC energies of 7 TeV, gluons carry most of the longitudinal momentum and are therefore the most active components. This justifies the choice of pp beams but leads to experimental consequences for the design of the col-

Table 4.1. Most relevant LHC machine parameters for proton-proton collisions.

Energy at collision	7.0 TeV
Energy at injection	450 GeV
Circumference	26.658 km
Dipole field at 7 TeV	8.33 T
Design luminosity	$10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$
Number of particles per bunch	1.15×10^{11}
Number of bunches per beam	2808
Nominal bunch separation	25 ns
RMS beam radius at IP	16.7 μm
Frequency	11246 Hz
Stored energy per beam	362 MJ
Crossing angle	300 μrad
Number of collision / crossing	≈ 20

lider requiring two beam pipes with opposite magnetic field configurations.

The protons are injected into the LHC (FIGURE 4.1) with an energy of 450 GeV by the Super Proton Synchrotron (SPS). Before entering the SPS, the protons are first accelerated by the Linac, the Booster and the Proton Synchrotron (PS). 24 SPS cycles are needed before the protons are transfered to the LHC. After the LHC is filled up, which takes about seven minutes, data can be taken at the interaction points for about 15 hours before the luminosity decreases too much.

Five detectors are currently being installed and will observe collisions at four different interaction points. There are two general purpose detectors, CMS and ATLAS and three more specialized experiments, ALICE, LHCb and TOTEM. The latter one shares the intersection point 5 (SX5) with CMS.

The LHC will also be used to collide heavy ions such as lead (Pb) with a collision energy of 1148 TeV.

4.2   Coordinate conventions

The coordinate system adopted by CMS has the origin centered at the nominal collision point inside the experiment, the *y*-axis pointing verti-
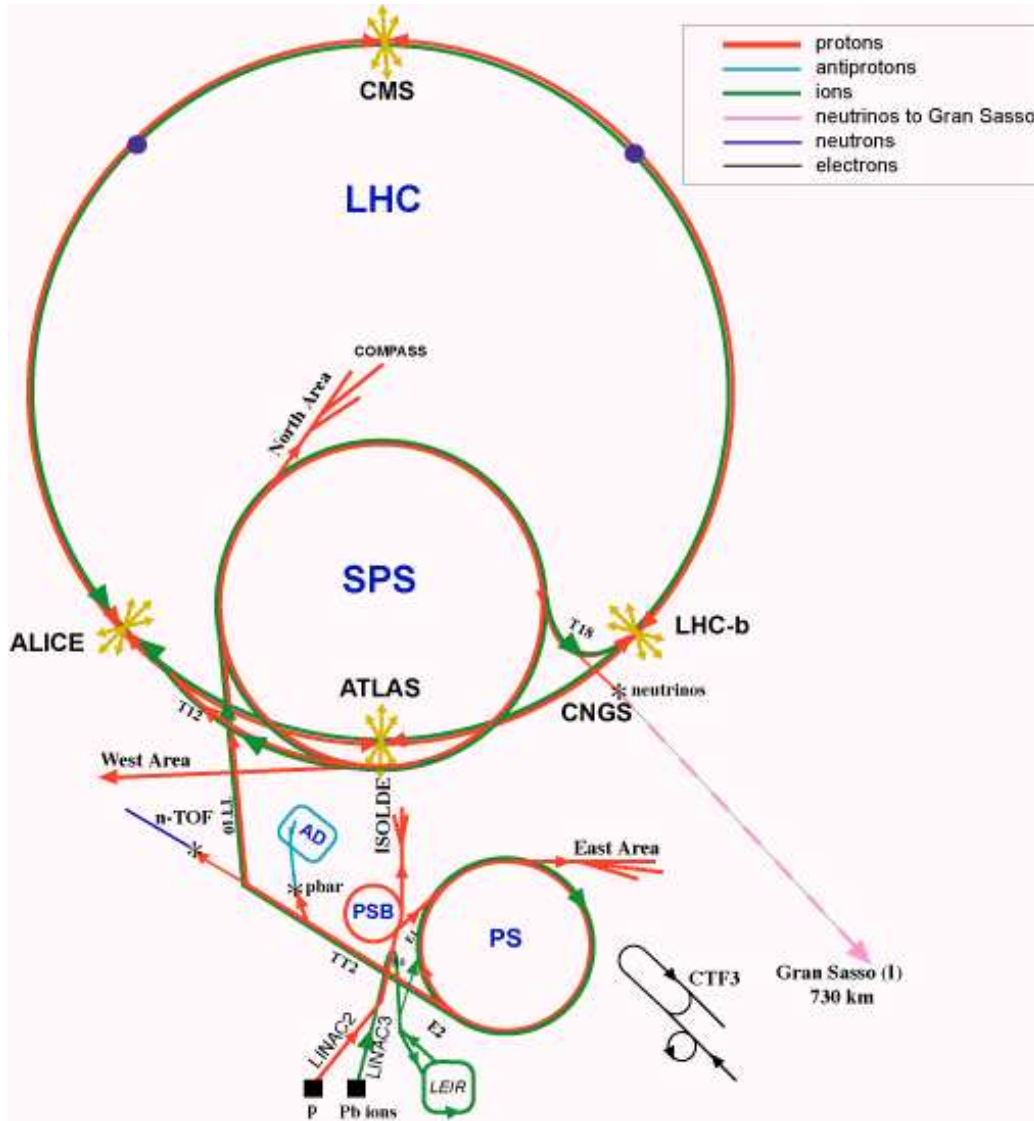
Figure 4.1. Schematic overview of the CERN accelerator complex.

cally upward, and the x-axis pointing radially inward toward the center of the LHC. Thus, the z-axis points along the beam direction toward the Jura mountains from the CMS pit. The azimuthal angle ϕ is measured from the x-axis in the x-y plane. The polar angle θ is measured from the z-axis. Instead of θ it is often more handy to use the rapidity y defined as

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right) \qquad (4.5)$$

which is a Lorentz invariant quantity. But in the majority of cases the pseudorapidity defined as

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right] \tag{4.6}$$

is used. For massless particles the rapidity y is equal to the pseudorapidity η.

Thus, the angular distance between two point objects, as observed from the origin of the CMS detector, is expressed as the size

$$R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}. \tag{4.7}$$

The momentum and energy measured transverse to the beam direction, denoted by p_T and E_T, respectively, are computed from the x and y components.

4.3   The Compact Muon Solenoid (CMS)

The design of the CMS experiment is driven by the cylindrical solenoid shape with the objective to provide a hermetic coverage around the collision point. Several subdetectors are positioned in concentric layers around each other (onion structure), each of them designed to measure different types of particles. In the design phase of CMS in the early 90s, the following main physics topics were defined as benchmarks for the CMS experiment:

- Proof the electroweak symmetry breaking through the detection of one or more Higgs bosons or study alternative theories in case the Higgs boson is not found. It is a particularly appropriate benchmark since there is a wide range of decay modes depending on the mass of the Higgs boson (SECTION 3.2).

- Search for supersymmetric particles in decays which are supposed to lead to a significant amount of missing transverse energy in the final-state since the lightest supersymmetric particle is expected to interact only very weakly.

- Search for new massive vector bosons in leptonic decays. Ways of distinguishing between different models for heavy objects like the

Z′ boson involve the measurement of the forward-backward asymmetry and the natural decay width which requires a good lepton reconstruction resolution.

- Investigate the existence of extra dimensions leading to signals involving the emissions of gravitons that escape into extra dimensions showing large missing energy.

- Explore the production of mini black holes showing democratic distributions of fundamental particles with very high momentum.

- Carry out detailed studies of Standard Model physics like multiple jet and top-quark production, CP-violation in processes involving B-hadrons, proof of lepton flavor conservation and precise measurement of the CKM mixing matrix.

- Probe the properties of strongly interacting nuclear matter (quark gluon plasma) in heavy ion collisions. This demands high-resolution calorimeters and tracking devices as well as flexible triggers.

These research interests result in the detector requirements for the CMS experiment which can be summarized as follows:

- Efficient muon identification and good muon momentum resolution up to $p_{\mathrm{T}} < 1\,\mathrm{TeV}$ over a wide geometric coverage.

- The best possible electromagnetic calorimeter – consistent with the muon requirement – providing an accurate diphoton and dielectron invariant mass resolution (1 % at 100 GeV) and efficient π^0 rejection.

- A inner tracking detector with good momentum resolution and reconstruction efficiency of charged particles. Efficient triggering and offline vertex reconstruction, requiring pixel detectors close to the interaction point.

- A hadron calorimeter with almost hermetic (4π) coverage to allow an accurate $E_{\mathrm{T}}^{\mathrm{miss}}$ determination. In addition, a fine lateral segmentation for a good dijet mass resolution is required.

The required performance of the muon system, and hence the bending power drives the design of the detector layout through the choice of the magnetic field configuration. A superconducting solenoid was chosen as
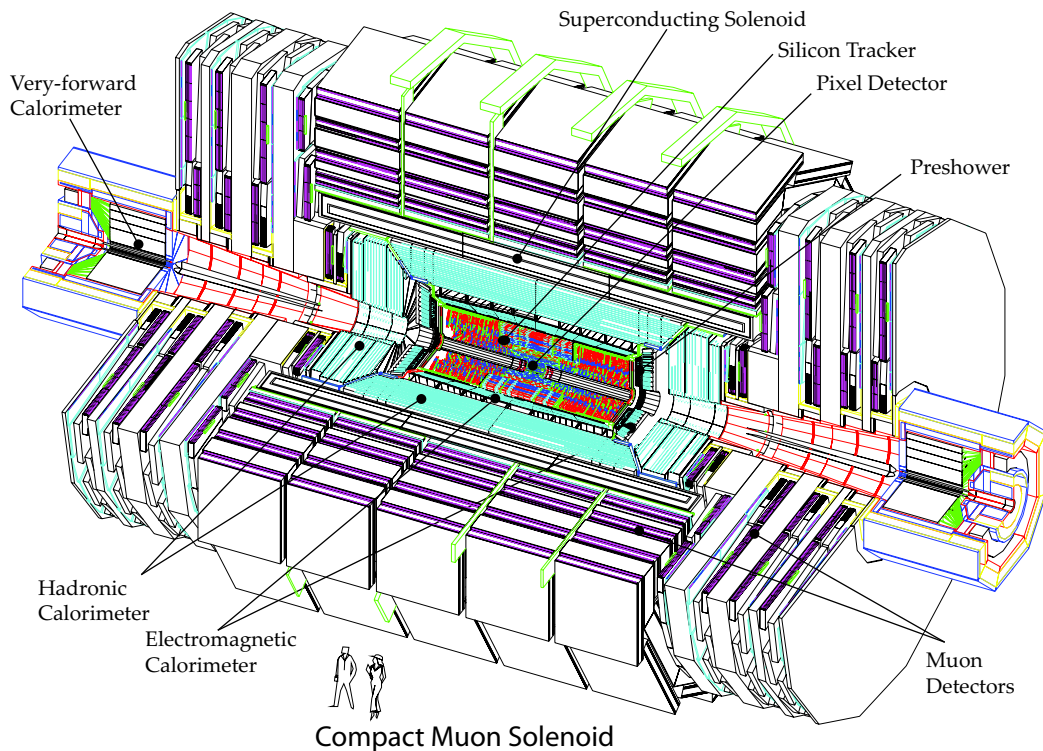
Figure 4.2. A schematic view of the CMS detector. From [20].

heart of the detector which produces a magnetic field of $B = 4\,\text{T}$ along the beam axis. The solenoid is 12.9 m long and 5.9 m in diameter, and its refrigerated superconducting niobium-titanium coil stores a total energy of 2.7 GJ. The return field is large enough to saturate 1.5 m of iron in the return yoke, allowing four muons stations to be inserted. The bore of the magnet is large enough to host the tracker, the electromagnetic and the hadronic calorimeter in the barrel part of the detector. Together with its two endcaps accommodating calorimeters and muon chambers as well, the CMS detector is nearly 22 m long with a width of 14.6 m which results in a total weight of about 12500 tons. The emphasis on muons in the CMS experiment relies on their large penetrating power which allows an efficient and precise identification and reconstruction of these particles even at high luminosities. The various components of the detector are depicted in FIGURE 4.2 and are briefly described in the following sub-chapters.
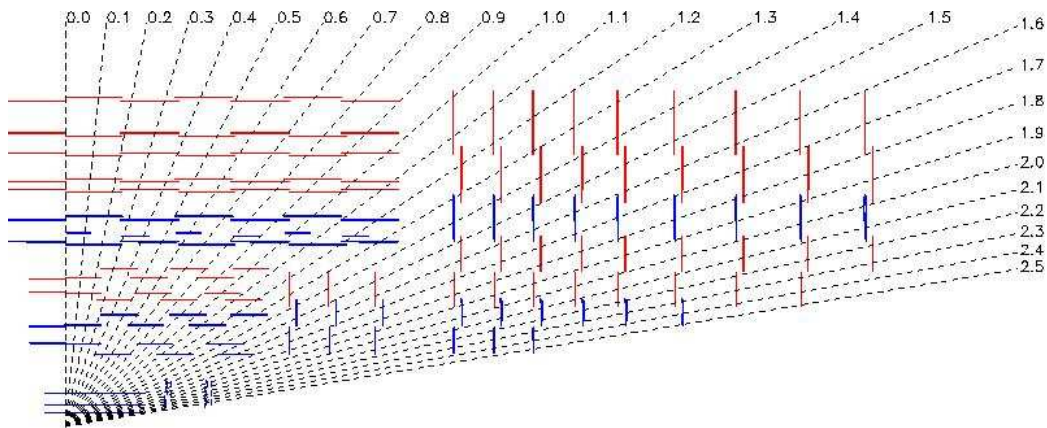
Figure 4.3. Layout of one quarter of the CMS tracker. Taken from [20].

4.3.1 Tracker

The capability to reconstruct charged particle tracks and measure with high resolution their momentum and their vertex of origin is a crucial part of the CMS physics program. High reconstruction efficiencies for particles with low transverse momentum ($1\,\text{GeV} < p_\text{T} < 5\,\text{GeV}$) play an important role to define isolated objects coming from gauge boson decays. An accurate vertex reconstruction helps in tagging particles coming from secondary vertices (arising from decays of b quarks and τ leptons) as well as in suppressing particles from pile-up events.

The CMS tracker comprises a silicon pixel detector and a silicon microstrip detector. The tracker volume is given by a cylinder of length 5.8 m and diameter 2.6 m. The pixel detector consists of three barrel layers and two layers in the endcap disks on each side. An almost square pixel shape of $100 \times 150\,\mu m^2$ has been adopted for an optimal vertex position resolution. In the barrel part, the silicon microstrip detectors can be divided into two parts: the tracker inner barrel (TIB) with four shorter layers and the tracker outer barrel (TOB) with the remaining 6 layers of microstrips. The forward region has nine microstip layers in each of the two endcaps. In the transition region between barrel and endcap, there are additional three inner disks. The total area of the pixel detector is about $1\,\text{m}^2$, whereas the area of the microstip detector is about $200\,\text{m}^2$, covering the region up to $|\eta| = 2.4$. The tracker comprises 66 million pixels and 9.6 million silicon strips.
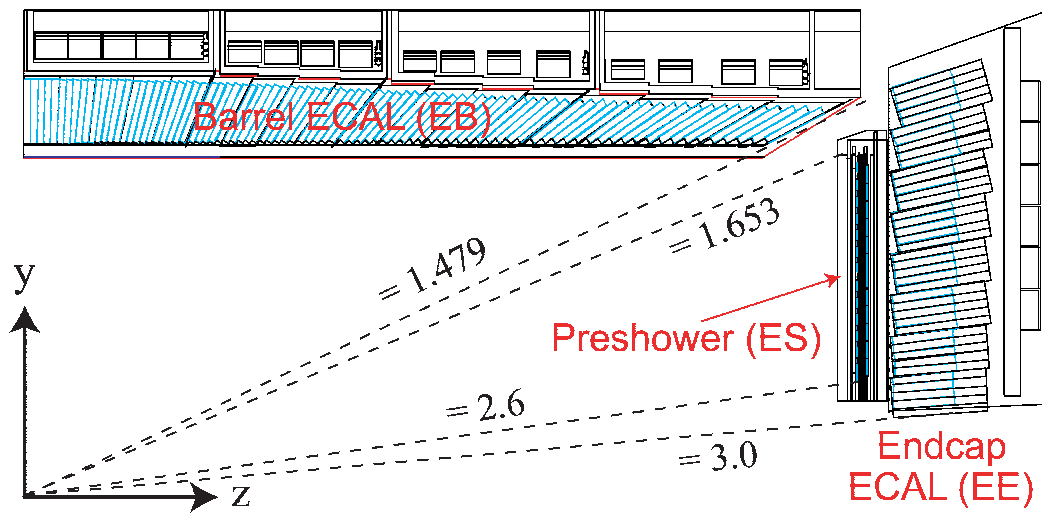
Figure 4.4. Schematic layout of the electromagnetic calorimeter. From [20].

4.3.2   Electromagnetic calorimeter (ECAL)

The electromagnetic calorimeter is hermetically placed around the silicon tracker (FIGURE 4.4) to measure the energy of electrons, photons and, in conjunction with the hadron calorimeter, jets with high precision. The design of the electromagnetic calorimeter was driven by the requirement to provide an excellent diphoton angular and energy resolution for the potential Higgs boson discovery mode H $\to \gamma\gamma$ (SECTION 3.2). To achieve such a good angular resolution, which is important for the separation of two photons coming from π^0 decays as well, a highly granular design is needed. Another requirement is to measure electrons coming from gauge boson decays with high precision and efficiency requiring a fast response from the system to ensure a fast first level trigger decision. Lead tungstate PbWO$_4$ has been chosen for the scintillating crystals because of its short radiation length of $X_0 = 0.89$ cm and its small Moliere radius of 2.2 cm. Another advantage of using PbWO$_4$ is its fast response: 80% of the light is emitted within 25 ns matching the LHC bunch crossing time. Negative aspects are the low light yield demanding the use of photodetectors with intrinsic gain. Silicon avalanche photodetectors (APD) are used in the barrel and vacuum phototriodes in the endcaps.

The geometrical coverage of the electromagnetic calorimeter extends up to $|\eta| = 3$ comprising about 76,000 crystals. The crystals in the barrel region of the detector ($|\eta| < 1.479$) are grouped into 36 supermodules, each of them consisting of 4 modules of crystals, covering an area of $\Delta\eta \times \Delta\phi = 0.0174 \times 0.0174$, corresponding to a front face cross section of about

HCal Barrel: HB

HCal Endcaps: HE

Forward HCal: HF

Outer HCal: HO

C.M.S PARAMETERS
Longitudinal View – Field Off
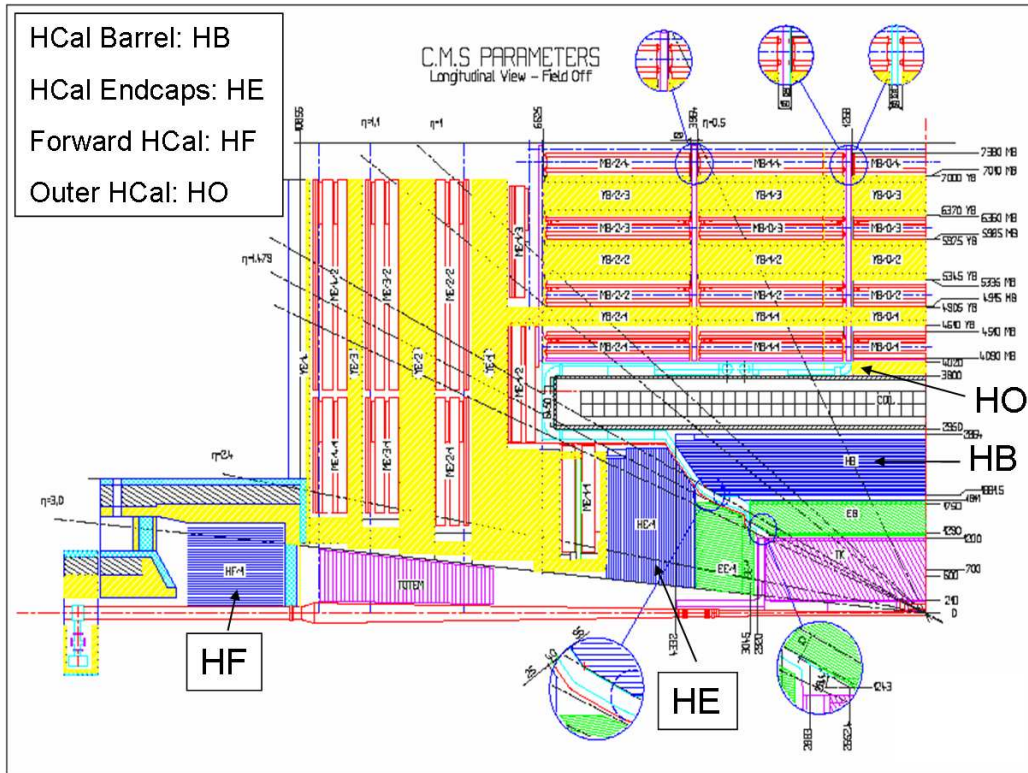
Figure 4.5. Schematic layout of one quarter of the CMS detector. From [20].

$22 \times 22 \, \text{mm}^2$ and a length of $230 \, \text{mm}$. The crystals in each of the two endcaps ($1.479 < |\eta| < 3.0$) are grouped into units of 5×5, referred to as supercrystals. The cross section and the length of the crystals in the endcap corresponds to $24.7 \times 24.7 \times 220 \, \text{mm}^2$ and are arranged in an x-y-grid. Over much of the crystals in the endcaps, a preshower device based on silicon sensors is placed. More detailed information about the electromagnetic calorimeter can be found in [21].

4.3.3 Hadronic calorimeter (HCAL)

The hadronic calorimeter surrounds the electromagnetic calorimeter completely and is used in conjunction with the electromagnetic calorimeter to measure the energy and direction of hadronic particle jets as well as to estimate the missing transverse energy. Indirectly, the hadron calorimeter also helps in the identification of electrons and photons. The hadron barrel (HB) part ($|\eta| < 1.4$) and each of the two endcaps (HE) of the hadron calorimeter ($1.4 < |\eta| < 3.0$) are placed inside the magnet coil

(FIGURE 4.5). The HCAL consists of plastic scintillator tiles read out with embedded wavelength-shifting (WLS) fibers interleaved with overlapping brass plates as absorber material. To reduce the tails in the energy resolution function and to improve the missing transverse energy resolution, an additional layer of scintillators, referred to as hadron outer detector (HO), is lined outside of the coil ($|\eta| < 1.26$). Outside the volume of the CMS inner detector, two forward calorimeters (HF) in steel/quartz fiber technology extend the coverage of the hadron calorimeter up to $|\eta| < 5$. Apart from improving the measurement of missing transverse energy, the forward calorimeters allow to tag or veto jets in the forward direction. The segmentation for the hadron calorimeter is finer granulated in the barrel ($\Delta\eta \times \Delta\phi = 0.087 \times 0.087$) than in the endcap ($\Delta\eta \times \Delta\phi = 0.175 \times 0.175$) to provide a good dijet separation for jets coming from highly boosted gauge bosons. More detailed information such as HCAL tower segmentation and transverse energy resolution from test beam data can be found in [20].

4.3.4 Muon system

Muons are present in most of the physics processes which the LHC is meant to explore. The ability to identify, to trigger and to reconstruct muons with high precision at high luminosities is a central concept of the CMS experiment. Because of their high mass and long lifetime, muons provide the cleanest experimentally measurable signatures. Muons coming from the interaction vertex are measured three times: in the tracker, behind the coil and in the return flux.

The muon system uses three different technologies to detect muons: drift tubes (DT) in the barrel region ($|\eta| < 1.2$), cathode strip champers (CSC) in the endcap region ($0.9 < |\eta| < 2.4$) and resistive plate chambers (RPC) in both, the barrel and the endcaps (FIGURE 4.6). The latter type provides a lower spacial resolution but a faster response than the former ones. The DTs or CSCs and the RPCs provide two independent and complementary sources of information for the first level trigger to ensure a robust, flexible and precise trigger decision. In the initial stage of the experiment, the RPCs will cover the region up to $|\eta| < 1.6$ and the coverage will be extended to $|\eta| < 2.1$ later. The described layout reaches a reconstruction efficiency of 90% for muons with a transverse momentum larger than 100 GeV in the entire range. The momentum resolution for muon tracks depending on the pseudorapidity can be found in [22].
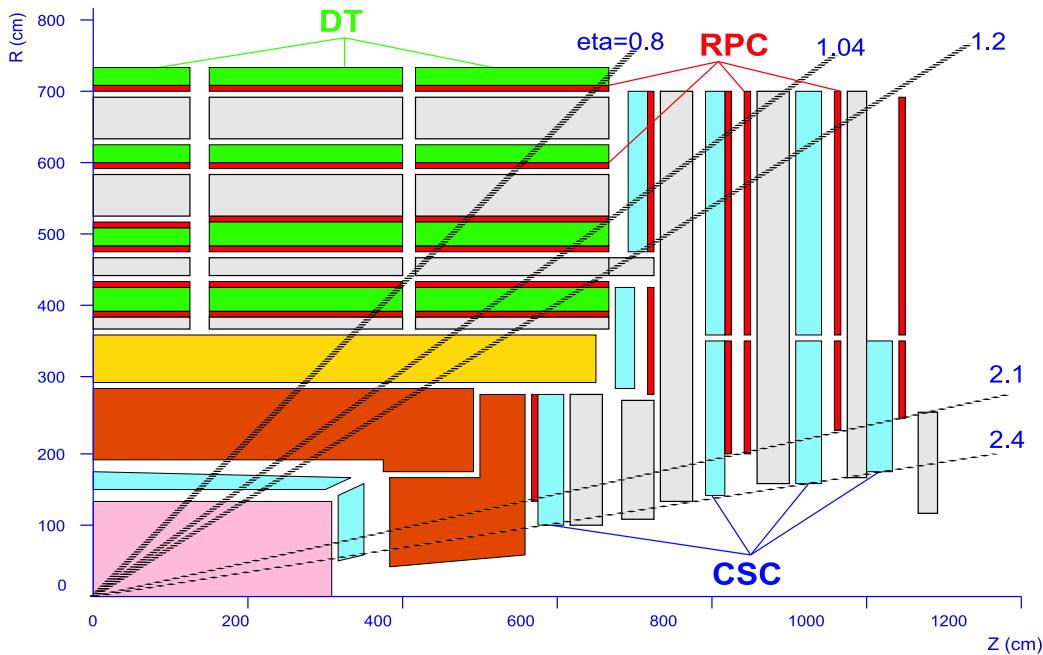
Figure 4.6. Schematic layout of the muon system. Taken from [20].

4.3.5   Data acquisition (DAQ)

The entire data acquisition system consists of the detector electronics, the first level trigger processors (Level-1 trigger), the readout network and the online filter system (CPU cluster) that executes the high-level trigger (HLT) software. At the nominal LHC luminosity of $10^{34}\,\mathrm{cm^{-2}\,s^{-1}}$, the bunch crossing rate of 40 MHz will lead to 10^9 interactions each second resulting in about 100 TB of data. The challenge for the CMS trigger system is to reduce this enormous amount of data to a manageable level by selecting the most interesting events (containing signatures from possible new phenomena, see SECTION 4.3) and rejecting the overwhelming background events. The decision taken by the Level-1 trigger algorithms involve measurements from the calorimeters and the muon system. While the signals are being transmitted from the front-end electronics to the trigger logics and a decision is being taken, the precise event data is stored in pipeline buffers for about $3.2\,\mu$s. At startup of the LHC, the event rate will be reduced to 16 kHz after the first level trigger, reaching 100 kHz at high luminosity running [23].

A further reduction of the event rate down to 100 Hz at the startup phase is achieved by the high-level trigger which already exploits a full event reconstruction if needed. The HLT filter farm consists of some thousand
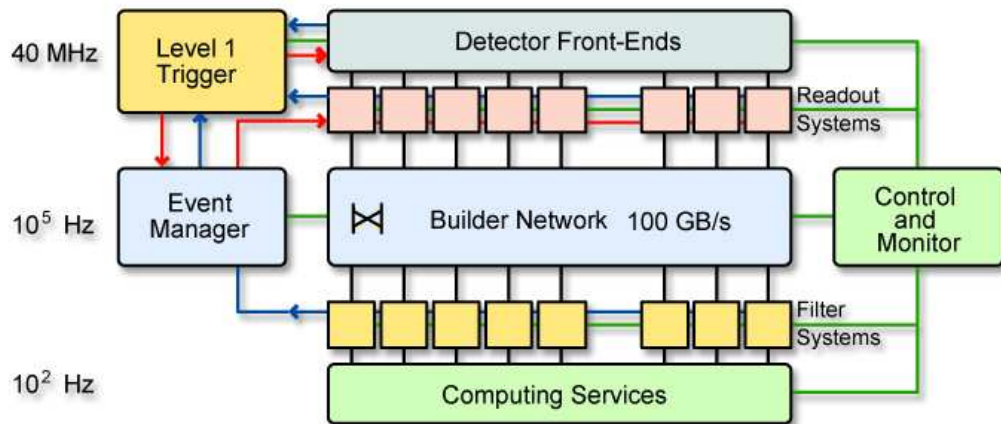
Figure 4.7. Overview of the CMS data acquisition system [20].

standard CPU boxes allowing maximum flexibility to optimize selection algorithms and thresholds during the LHC lifetime as well as profiting from the fast evolution of computing technology. The event data are delivered upon receipt of the L1 trigger after further signal processing to a given processor of the HLT farm by the data acquisition system. The HLT farm then writes selected raw events with the size of 1.5 MB to the mass storage system [24]. FIGURE 4.7 gives an overview of the systems which are involved in the data aquisition.

Chapter 5

CMS Software Components

The overall goal of the CMS software is to process and select events inside the high-level trigger farm, to deliver the processed results to physicists and to provide tools for them to analyze the obtained information so they can study the physics processes which took place during the collision.

To fulfill these requirements, the following structure has been defined for the CMS software:

- An application framework which facilitates the development and deployment of reconstruction and analysis software in different computing environments. In particular, the framework takes care of accessing the detector data which belongs to triggered physics events as well as all additional information needed to process the event.

- Physics software modules which reconstruct the physics processes that took place inside the CMS subdetectors and control the whole primary data-processing tasks like calibration and alignment. These modules can be plugged into the application framework which controls the access to the data independently of other modules.

- Several utility toolkits for physics calculation like fitting or displaying physics results. Computing integration services to allow physics applications software [25] to operate in the LHC computing environment [26].

The present analysis was carried out using basic software packages for event simulation including event generation and event reconstruction in

the former and now obsolete CMS software framework. During the year 2006, the so far developed framework COBRA [27], including the depending software packages, namely the detector simulation OSCAR [28] and the collection of reconstruction algorithms ORCA [29], have undergone a reorganization to eliminate a couple of identified drawbacks concerning in particular data flow, data handling and data management. At the time of writing this thesis, the new CMS software framework referred to as CMSSW [30] was still under construction and not fully completed. Furthermore, no simulated data was available in the new software framework to perform realistic physics analysis studies. The reconstruction algorithms are supposed to be converted into the new framework nearly untouched. Hence, the description of the simulation and reconstruction software according to the Physical Technical Design Report (TDR) Vol. 2 [15] given in this chapter remain valid within the new CMS software framework once finished.

It would have been impossible to carry out this analysis without using extensively the computing resources available on the LHC Computing Grid. About 20 million hypothetical physics events were generated on the Grid to obtain a realistic estimation of the background contributing of the final mass spectrum. Therefore, the CMS computing model is briefly described in this chapter as well. During the preparation of this thesis, a Tier-2/3 prototype center for the LHC Computing Grid at the Institut für Experimentelle Kernphysik at Karlsruhe University was set up where the author participated in. The setup of this Tier-2/3 prototype center is described in the last paragraph of this chapter.

5.1   Event generation

Event generators (also called Monte Carlo generators because of their underlying method) produce randomly hypothetical events with kinematic and topological distributions predicted by the theory. Several packages have been developed by theory groups for a wide range of collider experiments and each concentrates on different purposes and has therefore different advantages or disadvantages for a particular task.

The details of the implementation of the underlying physics process are different in each of these generators, but the fundamental idea is the same. The cross section and the differential kinematic distribution for almost all

processes in hadronic collisions are calculated as follows:

$$\sigma(pp \to CX) = \sum_{ij} \int f_i^p(x_1, Q^2) f_j^p(x_2, Q^2) \hat{\sigma}(ij \to C) dx_1 dx_2 \qquad (5.1)$$

where $f_i^p(x, Q^2)$ denotes the Parton Distribution Function (PDF) of the ith parton carrying a fraction x of the initial proton momentum at scale Q^2, and $\hat{\sigma}(ij \to C)$ is the cross section for the interaction of two partons i, j (the hard process).

The hard process is calculated by the evaluation of the scattering amplitude which is given by the square of the matrix element $|\mathcal{M}|^2$ for the studied process. The matrix element describes the collision at the smallest scales in time and distance when the colliding partons can be considered as free. Whereas the perturbative expansion provides a reliable prediction on these scales, the event enters the non-perturbative regime as the partons move further apart and confining effects of QCD become important. The perturbative evolution stops at the so-called factorization scale, assuming that the following hadronization step is independent of the hard scattering process. In this factorization assumption, the resulting partons are grouped together into color-singlet hadrons and unstable particles are decayed. The intermediate step of parton showering when the partons split into pairs of other partons can be included either in the calculation of the hard subprocess or the fragmentation step depending on the radiation patterns.

Finally, the underlying structure of the event such as beam remnants, interactions from other partons in the hadrons (multiple interactions), and collisions between other hadrons in the colliding beams (called pile-up) is added. The described consecutive evolution steps for event generation are depicted in FIGURE 5.1.

The event generators used to produce samples for this study are PYTHIA [32] and ALPGEN [33, 34] in combination with PYTHIA. These programs are introduced briefly below:

- PYTHIA is a generator for partonic events in pp, ee and ep collisions. It contains the matrix elements of about 240 subprocesses at leading order for the hard subprocess, initial- and final-state parton showers, hadronization and decay as well as underlying event inclusion. It belongs therefore to the group of general-purpose generators providing a full generation of the entire event.
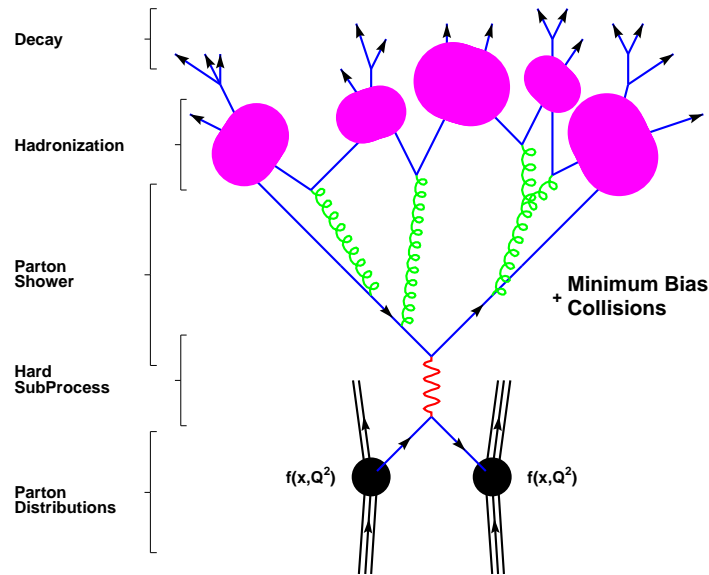
Figure 5.1. Basic structure of event generation in a simulated proton-proton collision. Taken from [31].

PYTHIA uses the Lund string fragmentation model [35] to describe hadronization. This is a phenomenological model based on a picture with linear confinement, where (anti-) quarks or other color (anti-) triplets are located at the ends of the string, and gluons are energy and momentum carrying kinks on the string. The string breaks by the production of new $q\bar{q}$ pairs, and a quark from one break can combine with an anti-quark from an adjacent one to form a color singlet meson.

The matrix elements included are mostly for $2 \rightarrow 2$, $2 \rightarrow 1$ and $2 \rightarrow 3$ subprocesses. Since multi-partonic configurations are obtained during the parton shower evolution, events which radiate multiple hard gluons at large angles are not well described by PYTHIA. The parton shower is most effective when the extra emissions are soft or collinear.

PYTHIA has been used by the CMS collaboration to perform most of the studies published in the Physics TDR Vol. 2 [15].

- ALPGEN is a tree-level matrix element calculator for a fixed number of partons (legs) in the final-state for hadronic collisions with emphasis on configurations with high jet multiplicities. It describes a specific final-state at higher leading order (LO) of α_s (without virtual
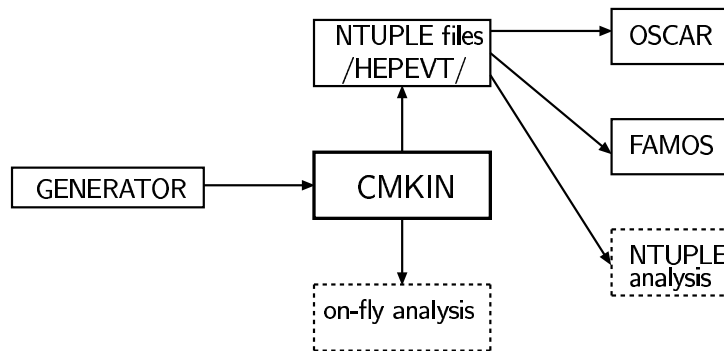
Figure 5.2. Illustration of the wrapper program CMKIN. Taken from [15].

loops) in perturbation theory and is based on the exact evaluation of the relevant Feynman diagrams in QCD and EW interactions.

ALPGEN generates only the hard process and does not include any form of hadronization. Thus, the output consists of bare quarks and gluons only. The hadronization has to be done in a separate step with routines as they are implemented in general-purpose generators like PYTHIA or HERWIG. Interfacing the ALPGEN output to an external showering package involves the risk that the same parton which was already generated in the matrix element calculation is added once again during the shower evolution. Although the partons generated in the fragmentation step are generally softer than or collinear to the former ones, this leads to the problem of double counting some portions of the phase space. A solution for this problem is known as the CKKW matching procedure for matrix elements and parton showers [36] discarding events that appear twice. In case of ALPGEN, the approach to remove double counted jet configurations is slightly different and implemented as so-called MLM matching [37].

CMS provides the program CMKIN [38] to interface various event generator outputs to the CMS detector simulation. For each supported event generator, CMKIN invokes the desired generator with the selected options describing the investigated physics process and converts the generator internal event structure to a standard one, based on the common block HEPEVT, a HEP standard to store particle information and kinematics for one event [39]. The output is stored in HBOOK ntuple files which are then used as input by the detector simulation program such as OSCAR and FAMOS (see SECTION 5.2 and SECTION 5.3). CMKIN supports almost every generator used in CMS like the general-purpose packages PYTHIA,

HERWIG [40] and ISAJET [41] as well as matrix element generators like ALPGEN, CompHEP [42] and MadGraph [43]. For the latter ones, it offers also the possibility to evolve the hard process through the showering and hadronization routines from PYTHIA taking care of proper matrix element and parton shower matching to avoid the interfering double-counting.

5.2   Event simulation

Theoretical predictions have formed an integral part for the planning stage of the CMS experiment. They help to define the experimental strategies and therefore the experimental design of the detector including the software needed for the operation. To serve such a purpose, these predictions need to reproduce as closely as possible the interaction processes taking place in hadron collisions and the interaction of the emerging particles with the various detector systems.

The detailed CMS detector simulation OSCAR [28] is based on the simulation toolkit GEANT4 [44]. GEANT provides a rich set of physics processes describing electromagnetic and hadronic interactions in detail. The detector geometry is described by OSCAR and GEANT models the propagation of charged and uncharged particles through the detector material and their interaction in presence of the magnetic field. The interaction effects include energy loss through ionization, multiple scattering of charged particles, electron bremsstrahlung, photon conversion and electromagnetic and hadronic showering. The simulation of the electronic response of the various data acquisition units is done in a subsequent step by ORCA. It starts with the simulated hits and simulated deposited energy in the subdetector components which are produced by OSCAR and generates the "digitized" output for tracker, muon detectors and electromagnetic and hadronic calorimeters. During the digitization, pile-up events coming from other inelastic pp collisions are also considered. From the computing point of view, it is much faster to simulate the pile-up events independently and merge them with the signal events afterwards. This procedure also allows to have the same sample for different run conditions corresponding to low and high luminosity operation of the LHC.
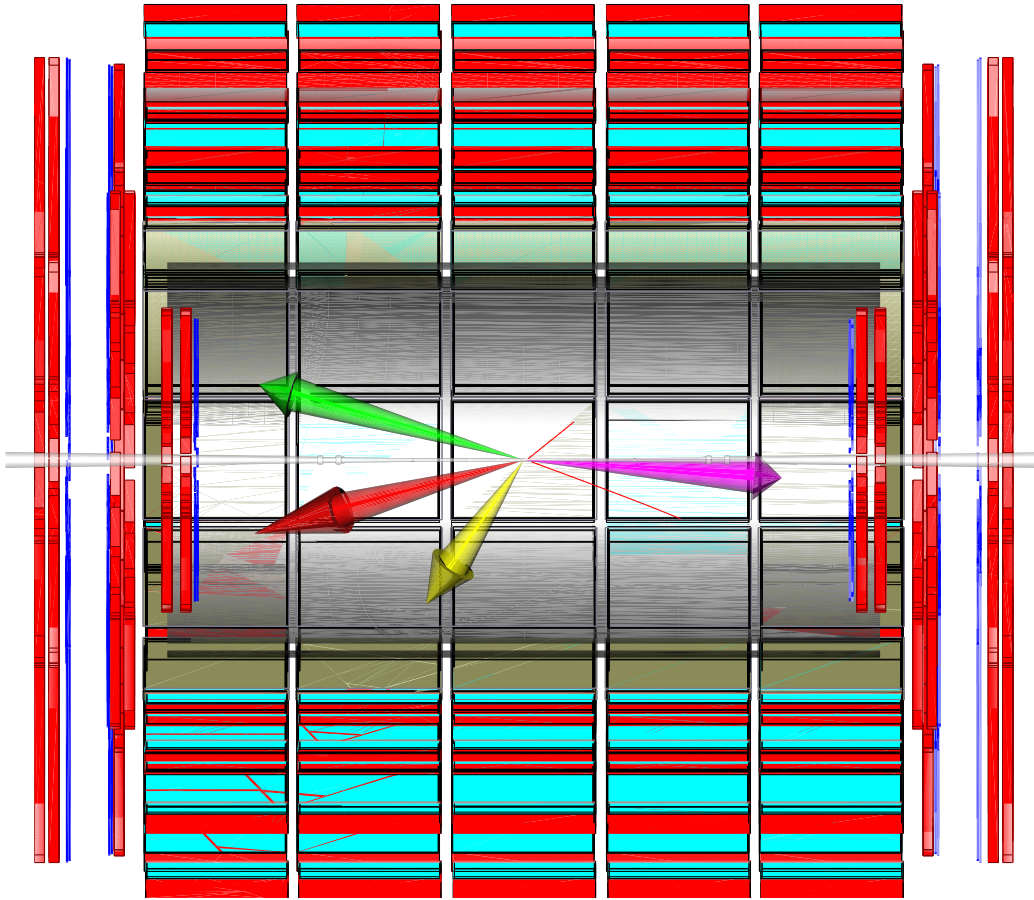
Figure 5.3. Full simulation (using OSCAR) and reconstruction (using ORCA) of the process qqH, H $\rightarrow$ (Z $\rightarrow$ qq)(Z $\rightarrow$ $\mu\mu$) visualized with IGUANA [45]. The direction of the four reconstructed jets is represented by plastic arrows, the reconstructed muon tracks are represented by thin lines.

5.3 Fast simulation

The huge complexity of the CMS detector brings up the need for a fast detector simulation. A package for a parameterized simulation of particle interactions, called FAMOS [46], has been developed to enable physics studies which require a big amount of simulated events to be considered. It was intended to be used for most physics analyses foreseen for the Physics TDR Vol. 2 [15] and beyond.

The input of FAMOS is a list of particles per event (as generated by an event generator listed above) characterized by their kinematic variables and originating vertices. The particles are propagated through the differ-

ent layers of various subdetectors regarding the magnetic field. Unstable particles are allowed to decay. Particles resulting from interactions with the detector material and pile-up events are added to the list and propagated in the same way. The interactions contained in FAMOS are: Electron bremsstrahlung, photon conversion, energy loss through ionization and multiple scattering in the tracker, and showering in the calorimeter cells. The response of the muon chambers is simply parameterized to reproduce the efficiencies and resolutions obtained from the full simulation. Muon synchrotron radiation is not implemented.

The output of FAMOS is a collection of high-level objects such as reconstructed hits in the tracker, deposited energy in the calorimeters which can then be used as input for the same algorithms as in the full reconstruction software package. This has been exploited to tune FAMOS successively to produce the same results as obtained in the detailed detector simulation. The parameterization of the detector simulation yields a speed-up of about three orders of magnitude for the needed computing time to simulate the detector interaction of an event.

5.4   Event selection and reconstruction

Reconstruction means the creation of physics quantities and physics objects from either raw data measured by the data acquisition systems of the subdetectors or from the simulated electronic response. These data in either case are called "digis". The algorithms performing the reconstruction are collected in CMS in the software project ORCA.

The reconstruction process can be divided into three steps: Firstly, a local reconstruction in individual subdetector modules is done, producing reconstructed hits (shortly "rec hits") representing position measurements in case of the muon and tracker system or calorimeter clusters representing deposited energy in case of the calorimeters. Secondly, the information from different modules of a particular subdetector is combined in the global reconstruction step. For example, the rec hits in the muon system are used to produce a reconstructed charged particle track. Finally, the information received from different subdetectors is used to form physics objects such as electrons from tracks and calorimeter clusters.

The following sections describe the specific reconstruction algorithms that have been used for the present analysis.

5.4.1 Muon reconstruction

The first step in the muon reconstruction, the Level-1 reconstruction, is a local reconstruction of track segments using aligned hits independently in the three muon subsystems (DT, CSC and RPC). The Level-2 tracking algorithm combines then reconstructed track segments and hits using a Kalman filter technique and provides local muon trajectories. The global muon reconstruction algorithm, called Level-3 reconstruction, performs a matching between hits in the tracker and the muon system. The muon trajectory is extrapolated from the innermost muon station to the outermost tracker surface and to the nominal interaction point to search for compatible tracks in the silicon tracker. A global fit is finally performed combining compatible hits in the tracker with hits from the stand-alone muon system. A selection on the fit result is made and finally delivers the muon candidates. The obtained trajectories are then refit using only hits in the silicon tracker and the innermost muon station to detect muon bremsstrahlung and other mechanisms of energy loss. This procedure improves significantly the resolution for muons with very high transverse momenta [47].

5.4.2 Electron reconstruction

An electron is measured and identified in the CMS detector by its single track emerging from the primary interaction vertex and its deposited energy in the electromagnetic calorimeter cells. While traversing the silicon layers of the tracker, electrons radiate bremsstrahlung photons which are spread in ϕ-direction due the bent trajectory of the electrons in presence of the magnetic field. These photons might convert to soft electron-positron pairs which can get partly trapped in the magnetic field loosing most of their energy before reaching the calorimeter cells.

The building of electron candidates starts with forming the energy in the calorimeter crystals above a given threshold into clusters by applying a pattern recognition method after having subtracted electronic noise. For barrel and endcap, two different clustering algorithms are used [48]. Next, the radiated energy is collected by making super-clusters, clusters of clusters, along the ϕ direction. The super-cluster is build by searching for the most energetic cluster and collecting the other clusters based on a geometric criterion.

The electromagnetic super-clusters initiate the search for two compatible

hits in the pixel detector. The two hits found serve as seeds for reconstructing the electron trajectory with the full tracker. The default reconstruction method relies on a Kalman filter algorithm which gives reasonable results for electrons carrying a high transverse momentum. For less hard electrons with $5\,\text{GeV} < p_\text{T} < 30\,\text{GeV}$, more accurate results can be obtained from a nonlinear filter approach which considers the energy loss of the electrons during track building [49].

The reconstructed electron trajectory is then matched to the super-cluster based on loose geometrical criteria and on the relation of super-cluster energy to tracker momentum. The energy deposited in the hadronic calorimeter must not exceed a given ratio of the electromagnetic seed cluster. The fraction of radiated bremsstrahlung is calculated as difference between the momentum measured at the last point and the origin of the electron trajectory. Depending on a set of four different electron classes [50], the energy measured in the super-cluster is then corrected by the amount of expected radiated bremsstrahlung. To suppress the contribution of particles faking the electron signature, further isolation and identification methods can be applied on these reconstructed electron candidates (see section SUBSECTION 6.2.3).

5.4.3   Jet reconstruction

A jet is defined as a collimated spray of high energetic hadrons formed out of quarks and gluons during hadronization of the scattered parton. Jet reconstruction algorithms are defined to group particles that are supposed to come from the same scattered parton into jets. Final-state radiation of the scattered parton can lead to a splitting of the jets in the detector. The particles contained in a jet are absorbed by the electromagnetic and the hadronic calorimeter cells. Calorimeter towers ("calo towers") representing the deposited energy of the particles are formed by allocating the energy from the electromagnetic calorimeter cells with higher spatial granularity to the corresponding hadronic calorimeter cells. The energy associated with a tower is the sum of all readout cells exceeding a given threshold. The jet finding algorithm has to undertake the task of reconstructing the energy and the direction of the parton scattered in the initial hard interaction out of these calorimeter towers.

The first step in jet reconstruction is to apply noise suppression and pile-up subtraction on the input towers. The jet clustering algorithm is then

invoked to form high-level objects out of the input towers. The fundamental principals of a jet clustering algorithm are a distance measurement to define the separation between input towers and a procedure specifying when and how towers should be recombined. The energy of the original scattered parton is systematically underestimated by these high-level objects due to effects described below. Therefore, the total reconstructed energy associated with a jet is finally corrected.

Two principal techniques are usually used: Cone type algorithms where objects are clustered together which are close in angle around an high-energetic seed calorimeter tower and clustering algorithms where objects are combined that have the smallest distance of all pairwise combinations possible. In CMS, the latter is implemented as iterative k_{T} algorithm whereas the iterative cone and the midpoint cone algorithm are provided for traditional cone-based reconstruction [51]:

- The Iterative Cone algorithm starts with a list of calorimeter towers ordered by its transverse energy. A simple cone of configurable size $R^2 = (\Delta\eta)^2 + (\Delta\phi)^2$ in η-ϕ-space is cast around the input tower with the largest transverse energy above a specified seed threshold. The objects inside the cone are merged to a "proto-jet" which serves to seed a new proto-jet. The energy and direction are calculated according to the E_{T} recombination scheme:

$$
\begin{aligned}
E_{\mathrm{T}} &= \sum_i E_{\mathrm{T}}^i \\
\eta &= \frac{1}{E_{\mathrm{T}}} \sum_i E_{\mathrm{T}}^i \eta_i \\
\phi &= \frac{1}{E_{\mathrm{T}}} \sum_i E_{\mathrm{T}}^i \phi_i
\end{aligned}
\tag{5.2}
$$

This procedure is iterated until the energy changes by less than 1% or the direction of the proto-jet changes by less than $\Delta R < 0.01$ between two iterations. After having found a stable proto-jet, the associated towers are removed from the list of input objects and the proto-jet is added to the list of jets. This procedure is repeated, until the input list contains no more towers above a given seed threshold. When the algorithm finishes, different recombination schemes may be applied to the jet constituents to define kinematic properties. An overview about recombination schemes is given below.

- The Midpoint Cone algorithm is designed to address a couple of disadvantages of the iterative cone algorithm. In a first step, it also starts with an iterative procedure to find stable proto-jets in a cone of size R. In contrast to the iterative cone algorithm, constituents belonging to the proto-jets are not removed from the list of input calorimeter towers. This results in overlapping jets. The kinematic properties of the proto-jets are calculated according to the energy recombination scheme (EQUATION 5.4).

  In a second step, for all stable proto-jets which are closer than the cone diameter, a midpoint between these two proto-jets is calculated as the direction of the combined momentum. This midpoint is used as seed to find more stable proto-jets. When all stable proto-jets are found, the splitting and merging of proto-jets is done to ensure collinear and infrared safety. This second step starts with the proto-jet in the list, having the largest transverse energy. If the proto-jet does not overlap with other proto-jets, it is defined as jet and removed from the list. Otherwise, the energy shared with the highest E_T neighbor proto-jet is compared. If the shared energy is larger than a configurable fraction f (typically 50% or 75%), these two proto-jets are merged. Otherwise, the shared constituents are removed and assigned to the jet that is closest in $\eta - \phi-$space. This procedure is repeated until all proto-jets are defined as jets. After the splitting and merging is terminated, the final jet kinematic properties are determined following a configurable recombination scheme.

- The Inclusive k_T algorithm starts with a list of calorimeter towers. For each object i in the input list and each pair of objects (i, j) the following distances are calculated:

$$d_i = (E_T^i)^2 R^2$$
$$d_{ij} = \min\{(E_T^i)^2, (E_T^j)^2\} R_{ij}^2 \quad \text{with} \quad R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2,$$

$$(5.3)$$

  where R^2 is a dimensionless configurable parameter normally set to $R^2 = 1$. The algorithms looks for the smallest value of d_i or d_{ij}. If the distance d_{ij} is the smallest, the two objects (i, j) are removed from the list, merged and added to the list of input objects again. Otherwise, the object d_i is removed from the list and defined as final jet. The procedure is repeated until all objects are included in jets. The algorithm merges successively all objects which have a distance $R_{ij} < R$. The merging of objects is performed using the E_T recombination scheme.

The recombination scheme which is used by the jet reconstruction algorithms described above to add the constituents and to calculate the direction of the jets is one of three ways:

- In the energy scheme, constituents are simply added as four-vectors:

$$E = \sum_i E^i, \qquad p_{x,y,z} = \sum_i E^i_{x,y,z}. \tag{5.4}$$

  This produces massive jets.

- In the E_T recombination scheme, massless jets are produced by summing up the transverse energy of the constituents and calculating the direction afterwards:

$$E_T = \sum_i E^i_T, \qquad p_{x,y,z} = \sum_i E^i_{x,y,z},$$
$$\sin\theta = \frac{p_T}{|\vec{p}|}, \qquad E = \frac{E_T}{\sin\theta}. \tag{5.5}$$

- The inclusive k_T algorithm calculates the energy and the direction directly during merging of objects:

$$E_T = \sum_i E^i_T, \qquad p_{x,y,z} = \sum_i E^i_{x,y,z},$$
$$\eta = \frac{\sum E^i_T \eta_i}{\sum E^i_T}, \qquad \phi = \frac{\sum E^i_T \phi_i}{\sum E^i_T}. \tag{5.6}$$

The energy reconstructed by the jet algorithms does not correspond to the true energy of the initial scattered parton. The factors influencing the identification of the correct jet energy can be divided into two groups. The first one follows from the detector performance and includes electronic noise, neighboring clusters not kept by the jet reconstruction algorithm for low energy particles ("out-of-cone" effects), absorber ("dead") material and non-linear calorimeter response. The second one is connected with the jet as physical object and includes the fragmentation model, initial- and final-state radiation, missing energy from muons and neutrinos, underlying event and particles coming from pile-up events. So-called calibration procedures are supposed to restore the original jet energy.

As long as the CMS detector does not collect jet data from real collisions, only Monte Carlo calibration techniques can be used which assume a perfectly described detector. Two types of Monte Carlo calibration are available: particle-level and parton-level calibrations. The first method applies
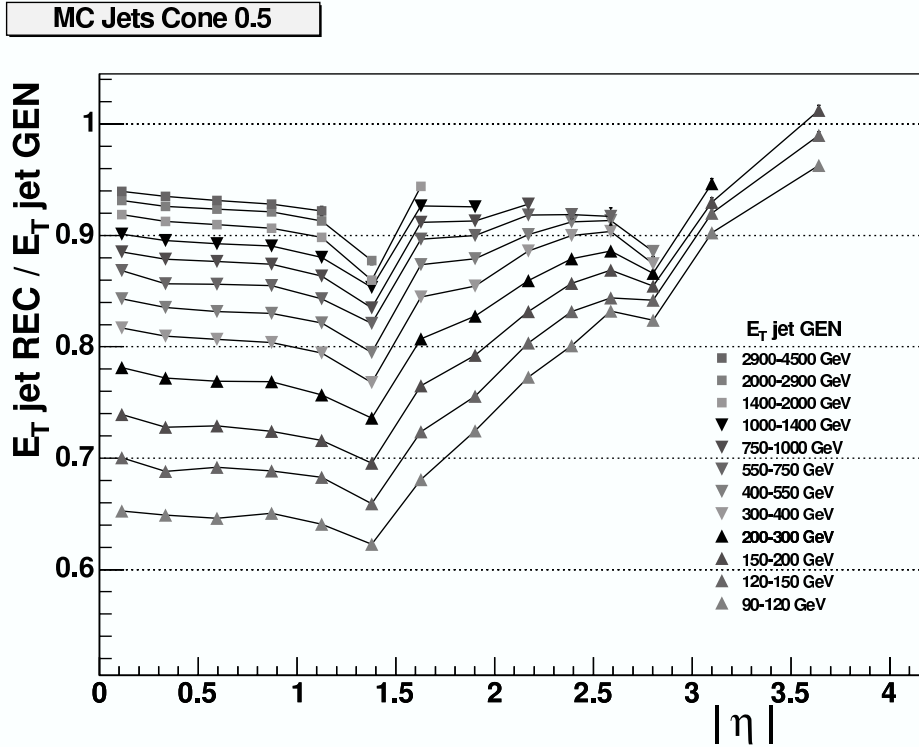
Figure 5.4. Response map for reconstructed transverse jet energies for 18 different p_T ranges as a function of the pseudorapidity η of generated jets from a QCD dijet sample. The iterative cone algorithm with a cone size of $R = 0.5$ and a matching criterion based on the distance $R^2 < \Delta\eta^2 + \Delta\phi^2$ was used to obtain the ratio E_{rec}/E_{gen}. Taken from [20].

the identical jet reconstruction algorithm on simulated calorimeter towers and on generated particles excluding muons and neutrinos. Based on geometrical criteria, the reconstructed jets are then matched to the clustered generated particles. Then the energy difference between the matched objects is parametrized as a function of η and p_T (FIGURE 5.4). This method cannot quantify the exact correction factors as, depending on the cone size, out-of-cone effects still occur when clustering generated particles. The parton-level calibration matches the reconstructed jets to initial scattered partons in the hard process before any showering. This kind of calibration depends on the fragmentation and the hadronization model and the type of origination parton. Both Monte Carlo calibration techniques depend on the physics channel and might give wrong results for other channels if the event topology diverges strongly from the ones used to calculate the correction factors. Furthermore, the identical input parameter set as used in

the calculation of the correction factors is required for the jet reconstruction algorithm to ensure the correct restoration of the jet energy.

A number of data-driven calibration approaches exist to improve the understanding of the jet energy scale as soon as the CMS detector starts to record real LHC data, namely QCD dijet and γ+jet calibration exploiting the transverse momentum conservation [52] and W boson mass calibration using a fit to the W mass distribution in semi-leptonic decaying $t\bar{t}$ events [53].

5.4.4 Primary vertex reconstruction

Reconstructed tracks with transverse momentum $p_T > 1.5\,\text{GeV}$ and transverse impact parameter significance $d_0/\sigma_{d_o} < 3$, based on their distance of closest approach to the beam line, are grouped together according to their separation in the z direction, if the maximum separation between two successive tracks is $\Delta z < 1\,\text{mm}$. A vertex candidate fit is iterated by discarding incompatible tracks until all kept tracks are compatible to more than 5 % for this vertex candidate. For all tracks originating from a vertex candidate, the sum of p_T^2 is calculated. The vertex with the largest sum is considered as primary vertex [54].

5.5 Physics analysis tools

5.5.1 ROOT

ROOT [55] is a well-known and widely spread object-oriented framework within the high-energy physics community. It is written in C++ and contains a collection of classes which are aimed at solving the data analysis challenges arising in todays high-energy physics experiments. The project was started in 1995, and in 2003 CERN decided to support ROOT officially.

ROOT provides classes for histograming and fitting data, for physical vectors, a geometry package, a high-performance input/output system, several mathematics libraries, supports network communication and parallel data processing.

5.5.2 PAX

The Physics Analysis eXpert (PAX) class collection [56, 57] is a C++ toolkit for high-energy physics analyses with particular emphasis on complex and ambiguous event topologies. It provides additional functionality on top of the physics vector classes of ROOT or CLHEP [58]. Three types of generalized physics objects, together with a persistent event container and relation management, allow the definition of an abstraction layer beyond the detector reconstruction software and protect the physics analysis code from changes in the underlying framework.

The three types of generalized physics objects provided by the PAX kernel are particles, vertices and collisions. An event container holds the complete information about a multi-collision event and connects the reconstructed particles and vertices to a complete decay tree according to a distinct event interpretation. The subsequent evolution of an event interpretation allows the easy management and testing of different combinatorial ambiguities and different physics hypotheses of an event. At the final step of an analysis, one out of a numerous number of evolved hypothesis is selected based on standard decision techniques.

5.6 Modern computing in high-energy physics

5.6.1 The LHC Computing Grid (LCG)

The mission of the LHC Computing Grid (LCG) project is to design, build and maintain a distributed computing infrastructure to fulfill the needs of the four LHC experiments as defined in their offline computing models. As soon as the LHC starts operation, the LCG will have to cope with roughly 15 PetaBytes of data annually. These experimental data has to be stored safely over its entire lifetime and provided worldwide to the physicists who are participating in the LHC project. Because of the tremendous requirements of storage resources and CPU power, a globally distributed model was chosen: a computing Grid [26]. Further key benefits of this approach are reduced costs of maintaining and upgrading the necessary resources for which national organizations are responsible and increased system stability round-the-clock by spanning all time zones. Challenges arising from a distributed system include ensuring high data throughput networks, maintaining coherence of software versions installed, coping

with heterogeneous hardware, providing authentication and accounting mechanisms for users and groups and protecting the data so that it is neither lost nor corrupted over the lifetime of the LHC and beyond.

The LCG architecture consists of a defined set of fundamental services and applications running on the computing infrastructure of the participating national LCG partners. The Grid itself is implemented as a hierarchal four-tiered model:

- One single Tier-0 center located at CERN records the raw event data as emerging from the experiment high-level triggers. These data is written to tape after having performed a first-pass reconstruction. A second copy is transfered to one of the Tier-1 centers associated with the experiment. According to the policy of the experiment, reconstructed data are also transfered to the Tier-1 center.

- The Tier-1 centers are interconnected with CERN using high-speed networks (10 Gbit). Their role differs slightly for the different LHC experiments, but in general they perform the reconstruction of the raw data and manage the permanent storage of raw, reconstructed and simulated data.

- The role of the Tier-2 centers is mostly to provide sufficient computational resources for theoretical simulations and physics analyses. Reconstructed data are received from allocated Tier-1 centers and simulated data are sent to these Tier-1 centers for permanent storage.

- Other laboratories or universities might take part in analyzing LHC data as Tier-3 facilities. Their role is not defined in the scope of the LHC Project, although access to the data in assigned Tier-2 centers has to be provided but direct support is not guaranteed. Nevertheless, these sites might make significant contributions to the experiment's needs as they provide computing resources for interactive user analyses or theoretical simulations.

During the past few years, numerous Grid middleware products have been developed and released by different Grid projects, namely the Enabling Grids for E-SciencE (EGEE) [59], the Open Science Grid (OSG) [60] and the Nordic Data Grid Facility (NDGF) [61]. Each infrastructure must provide the essential Grid services such as job submission and management, Grid data handling and Grid information services conforming to the

agreed set of interfaces. The main services that should be made available to the LHC experiments and their functionality as defined by the Baseline Services Working Group [62] are briefly described in the following:

- The Computing Element (CE) is a set of services that provide access to the local job queuing system at a site. These services include job submission, job management, authentication/authorization mechanisms and publication of information describing their current status and the resources (basically batch queue configuration) available at this site.

- A Storage Element (SE) provides services for authentication / authorization, accessing disk pool and tape library, transferring data in and out of the storage element and input/output facilities to allow applications to read the data stored on the storage element.

- The Workload Management provides a mechanism through which the specific job can express its resource requirements (e.g. software version installed, data streams available at the storage system, permitted CPU running time). The service submits the job to any site which fulfills those requirements. Additional job monitoring tools monitor and trace these submitted jobs.

- Grid Catalog Services include mapping of logical file names to their global identifier and storage location (storage element and physical file path), determination of replicas for a certain file, a directory like structured hierarchical namespace, access control policies and interfaces for the workload management services.

- Virtual Organization Membership Services: These services are deployed to manage the membership of registered users within a specific virtual organization (VO), an abstract entity group, and issues certificates for them which contain information about their authorized use of resources for the organization.

- Information Services: These services publish and maintain labels about resources and their current status in the Grid.

The requirements of the LHC experiments have been defined in individual Computing Technical Design Reports and sum up to a CPU capacity of 110 million SPECint2000 [64], to about 40 PetaBytes of disk storage and 40 PetaBytes of mass storage (tape) for the year 2008. An overview over the CMS computing model is given in FIGURE 5.5.
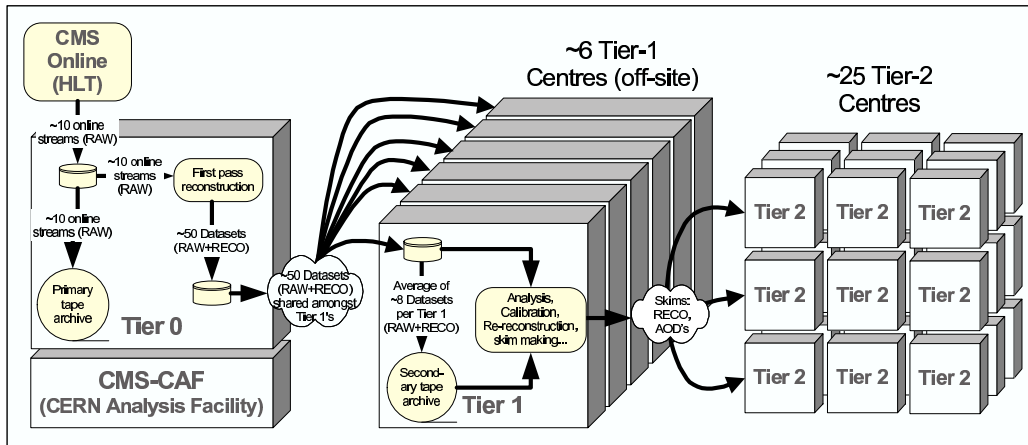
Figure 5.5. Illustrated data flow according to the CMS computing model. Taken from [63].

5.6.2 The CMS computing model

The CMS offline computing environment relies on Grid services for data processing, data archiving and event simulation and makes use of the hierarchy of the four tiers as foreseen by the LCG. In addition, a CMS analysis facility at CERN (CMS-CAF) combines services foreseen for typical Tier-1 and Tier-2 centers to enable short turnarounds for critical data processing which needs to be carried out to control the stable and efficient operation of the CMS detector.

The event data flow among the different Tier centers involved is organized as follows: The high-level trigger farm writes raw event data with a size of 1.5 MB at 150 Hz. Depending on their trigger history, the raw events are classified into primary datasets. These datasets must be transfered to the Tier-0 center at CERN in real-time at a rate of 225 MB/s where a first reconstruction is performed. The reconstructed data (RECO) with a size of 0.25 MB per event is further stripped down to Analysis Object Data (AOD) which contains all high-level objects necessary for physics analysis like calorimeter energies and positions and hits belonging to reconstructed tracks.

AOD, RECO and RAW events are transfered to a Tier-1 center, transfers to other Tier-1 centers are foreseen in case of additional available network bandwidth. The Tier-1 centers produce new AOD versions as soon as a new software version of the reconstruction code is released and distribute these new versions among themselves. Subsets of datastreams are foreseen to be copied to Tier-2 centers for calibration and alignment studies. Other functions supported by a Tier-2 center are the event generation and
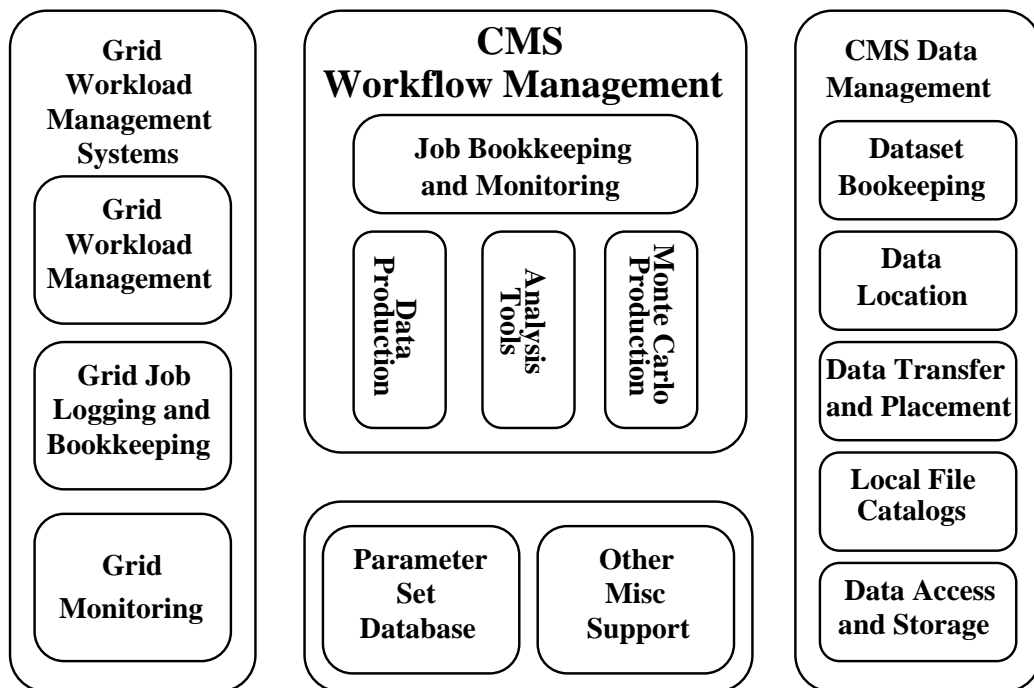
Figure 5.6. Schematic overview of services supporting the workflow management system. Taken from [63].

simulation (Monte Carlo production) and data processing for user analyses. Tier-3 centers will provide local user communities interactive access to execute analysis on smaller physics group data samples created by Tier-2 centers.

The CMS computing system relies mostly on services provided by the Grid projects. However, some services and applications needed for the workflow management (figure FIGURE 5.6) are specific for the CMS community and therefore are developed within CMS using standard Grid services and interfaces. This includes in particular services running at the sites which know about the CMS data structure and include applications to manage the large amount of data produced, reconstructed and analyzed by the CMS computing system. The basic data management architecture consists of the following services:

- The Data Bookkeeping System provides information describing the event data such as existence and names of produced datasets and their packaging units.

- The Data Location Service determines the replicas in the distributed

environment belonging to existing data and provides the names of
the Tiers hosting the data.

- The Data Placement and Transfer System keeps track on files, data
transfer requests, and replicating and moving individual files among
the Tier centers. This service is implemented by the PhEDEx [65]
project (APPENDIX C).

- Local File Catalogs provide the physical location of data files hosted
by the individual site. A Local File Catalog presents a POOL inter-
face [66] which returns the full physical filename (PFN) of a logical
file name (LFN) which is defined by CMS through a Global Unique
IDentifier (GUID).

- The Data Access and Storage System interfaces stored data to run
CMS application through a POSIX-like interface.

Further services needed for a coherent CMS computing system support
the Grid workload management and include job bookkeeping and log-
ging, Grid monitoring and parameter set management. The computing
system is expected to operate in a heterogeneous Grid environment and
therefore provides interfaces to most of the implementations making the
details invisible to the CMS physicist. All these services together support
the workflow for large data processing tasks.

The basic steps of a typical user analysis, event simulation, creation of
physics group data (skimming) or AOD production is summarized in the
following:

- Task formulation: The desired application together with the input
parameters and the input datasets has to be configured. This takes
place on the User Interface (UI) where all the needed Grid and CMS
specific services are installed as well as the software packages for
developing and testing the application.

- Data discovery: A query to the Dataset Bookkeeping system returns
the data which needs to be accessed for the specified task.

- Job splitting: The task might be split for computational reasons into
several jobs, each of which will access a non-overlapping subset of
the selected dataset.

- Job configuration: The workload management creates two configurations, the first for the CMS software framework and the second for the Grid job management, for every job which is to be submitted.

- Job submission: The prepared jobs are then submitted to the Grid. Up to this stage, all tasks take place on the User Interface.

- Job scheduling: The Grid workload management decides about the Computing Element (CE) which matches the job requirements and dispatches them to the CE.

- Job run-time: The job arrives at the Computing Element with its still independent configuration and is forwarded to the Worker Node (WN) where the job is executed. If real-time monitoring is used, the job contacts the job monitoring database.

- Job completion: Once the job is completed, the output must be stored somewhere. Either it is encapsulated in the Output Sandbox to be retrieved by the User Interface or it is copied to a Storage Element.

- Task monitoring: During the job scheduling, execution and completion, the progress of the individual jobs belonging to the production or analysis task, are monitored by the User Interface.

- Task completion: As the entire set of jobs finishes, the output can be collected from the destination sites and retrieved to the User Interface. In case of large data output of common interest like event simulation or AOD production, the output is published to the Data Bookkeeping System and Data Location Service and handed off to an agent of the Data Transfer System for duplicating or moving to some other Storage Element.

5.6.3   The Tier-2/3 prototype center at the IEKP

Within the LCG, the Institut für Experimentelle Kernphysik (IEKP) [67] at Karlsruhe University (TH) offers the full grid functionality for CMS, such as data transfers, CMS software installation and grid based analyses. The installation of the Grid middleware has been adapted according to the local Linux computing cluster. The cluster itself (FIGURE 5.7) consists of two parts, the inner network and the outer network. It comprises portal machines, file servers, computing nodes and a controlling machine. The

cluster is internally called "EKPplus" and its setup is described in [68, 69]. Since May 2005, the LCG components Computing Element, Storage Element, User Interface and Monitoring Box have been successfully integrated into the EKPplus cluster. The network architecture and the cluster component specifications are described in the following paragraphs.
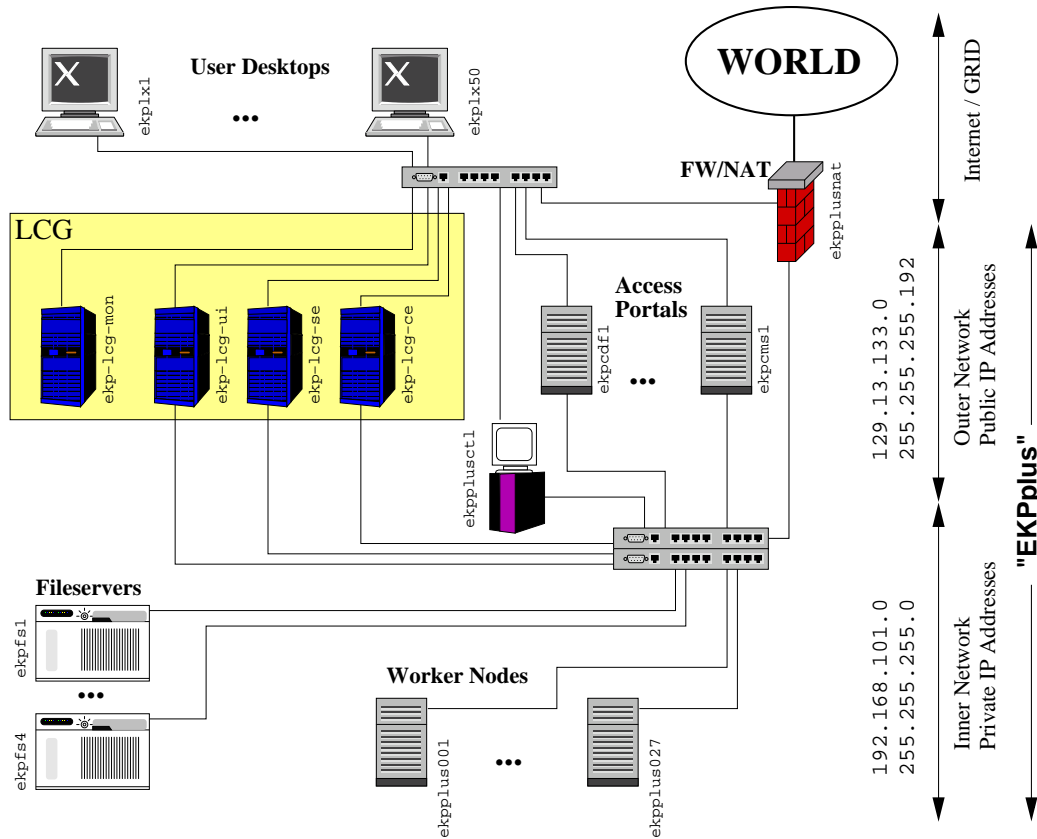

Figure 5.7. A schematic overview of the architecture of the EKPplus cluster and the integration of the LCG components CE, SE, UI and MON.

The inner network of the EKPplus cluster consists of the computing nodes "ekpplusXXX", several file servers "ekpfsX" and a dedicated cluster control machine named "ekpplusctl". This control machine takes care of the local users, manages the job queues for the batch system and provides the root file system for the computing nodes.

The outer network consists of publicly accessible portals which serve as testbeds for the development of analysis software. To accomplish this task for local users of the CDF [70], CMS and AMS [71] working groups, experiment specific software is installed and kept up-to-date on the differ-

ent portals assigned to the respective experiments. Via multiple ethernet cards, the portals are also connected to the inner private network. Thus, they offer access to the file servers and the usage of the Worker Nodes via the local batch system.

Apart from the monitoring host "ekp-lcg-mon" which is only connected to the outer network, all other LCG components – Computing Element "ekp-lcg-ce", Storage Element "ekp-lcg-se" and User Interface "ekp-lcg-ui" – are fully integrated into the EKPplus Linux computing cluster. Therefore, they are connected to the outer and the inner network, like all portal machines. The User Interface provides the grid access point for local users and allows them to submit jobs to the grid. The Storage Element stores the VO specific software and re-exports some file-space of the file servers reserved for LCG purposes. The Computing Element forwards the received LCG jobs to the cluster control machine "ekpplusctl" which enqueues the job in the local batch system. In this manner, the internal computing nodes "ekpplusXXX" need to be and are considered as LCG Worker Nodes since the Computing Element acts only as a gateway between the LHC Computing Grid and the existing local Linux cluster.

The whole EKPplus cluster is connected to the IEKP desktop network by a 1 GBit connection. It is currently protected against attacks from the outside by a the firewall named "ekpplusnat". FIGURE 5.7 depicts the overall architecture of the EKPplus cluster and the integration of the LCG components. A full description about this Tier-2/3 prototype center can be found in [72].

Chapter 6

Sensitivity of CMS for $\mathrm{H} \to llqq$

One of the main goals of the CMS experiment is to prove electroweak symmetry breaking by discovering the Higgs particle which is connected with the isospin doublet of complex scalar fields in the Lagrangian density of the Standard Model (CHAPTER 2). Although there are several restrictions deduced from consistency constraints and earlier experimental searches, the Higgs boson mass cannot be predicted by the Standard Model as outlined in CHAPTER 3. The discovery potential for a Higgs boson with the CMS experiment has been investigated over a wide mass range from the current lower experimental limit of $M_\mathrm{H} = 114\,\mathrm{GeV}$ up to $M_\mathrm{H} = 600\,\mathrm{GeV}$ involving several decay and production modes which provide promising detectable event signatures [15].

The probability for a discovery of a signal peak in a mass spectrum is usually expressed in multiples of Gaussian standard deviations, denoting how unlikely it is that the observed peak might be a random fluctuation above the expected mean value of the background. This quantification of probability is called statistical significance or more often just significance. Most physics studies aim for maximizing the statistical significance of an observation by exploring the event characteristics to enhance the signal over the (expected) background contribution. Therefore, the decay channel with the highest branching ratio is not necessarily the one which gives the highest statistical significance for an observation.

The searches for a Higgs boson within CMS can be divided into two mass ranges. Below $M_\mathrm{H} < 135\,\mathrm{GeV}$, the decay into $b\bar{b}$ quarks which has by far the dominant branching ratio of 85% will be detectable but suffers from large uncertainties [73]. The decays into $\tau^+\tau^-$ leptons or two photons
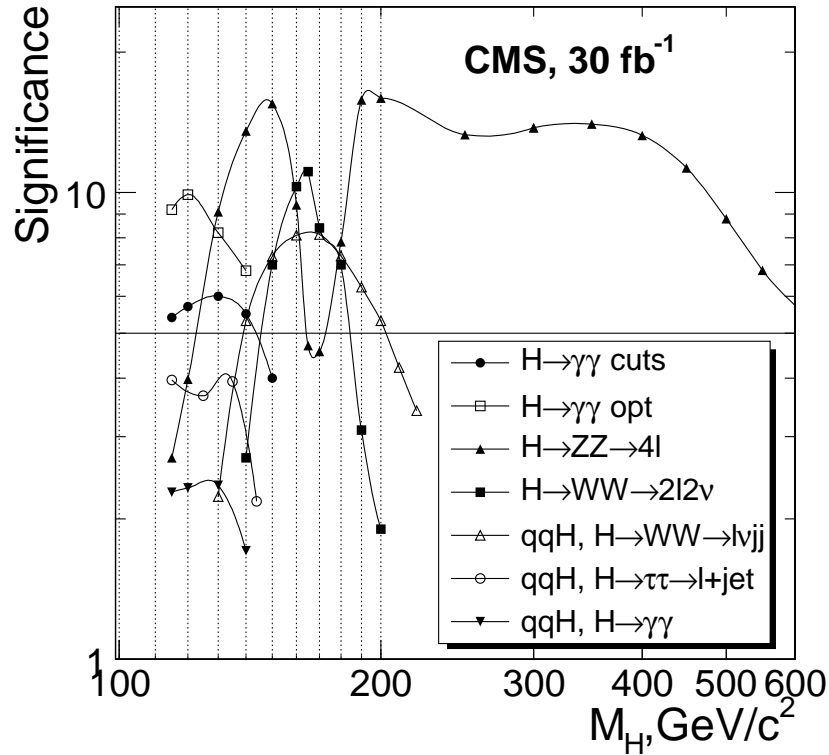
Figure 6.1. Expected statistical significance for the discovery of the Higgs boson as a function of the Higgs boson mass for an integrated luminosity of $30\,\mathrm{fb}^{-1}$ corresponding to the three years of data taking with low luminosity $\mathcal{L} = 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. Taken from [15].

with a branching ratio of 8 % and 0.2 %, respectively, provide a good discovery potential. Although the decay into two photons has a very small branching ratio, it yields a very clean signal signature with a good statistical significance, since the photons are highly isolated. This channel also profits from the excellent resolution of the electromagnetic calorimeter allowing a very narrow and precise reconstruction of the Higgs mass peak.

Above $M_\mathrm{H} > 135\,\mathrm{GeV}$, the observation of a Higgs boson is only possible in decays into weak boson pairs with at least one isolated electron or muon in the final-state. Apart from a drop between about 160 and 180 GeV, the most promising decay mode up to 500 GeV is the Gold Plated channel into four isolated leptons, although the branching ratio is only about 0.5%. The statistical significance for the observation of a Higgs boson in this channel in the first three years of LHC operation drops below 5σ at around 700 GeV due to the decreasing cross section for Higgs boson production. The

only way out is a decay mode which exhibits a higher branching ratio. An alternative is the decay into a pair of Z bosons where one Z decays into two jets and the other into either two muons or two electrons H → (Z → ll)(Z → qq). The branching ratio into this final-state is about 20 times higher than into four isolated leptons:

$$
\begin{aligned}
BR(ZZ \to 4l) &= [BR(Z \to ee) + BR(Z \to \mu\mu)]^2 \\
&= [3.37\,\% + 3.37\,\%]^2 \\
&= 0.45\,\%
\end{aligned}
\tag{6.1}
$$

$$
\begin{aligned}
BR(ZZ \to llqq) &= 2 \times [BR(Z \to ee) + BR(Z \to \mu\mu)] \times BR(Z \to \text{hadrons}) \\
&= 2 \times [6.74\,\% \times 69.91\,\%] \\
&= 9.42\,\%
\end{aligned}
\tag{6.2}
$$

The Higgs boson decay mode with two leptons and two jets in the final state will be impossible to detect if only the momenta of the decay products and the resonance in their mass distribution are used to extract the signal from the expected background. Those variables are just too few handles to separate the signal from many QCD induced processes with huge cross sections.

The weak boson fusion as second largest contribution becomes more and more competitive to the gluon fusion as the Higgs boson mass increases. The characteristics of the Higgs boson production via the weak boson fusion can be exploited to suppress the large backgrounds. These characteristics are the accompanying jets originating from the incoming initial-state quarks which emit the weak bosons. These additional forward jets at large pseudorapidities but in opposite hemispheres can be detected by the hadron endcap calorimeter which covers the region $1.5 \le |\eta| \le 3$ and the hadron forward calorimeter which covers $3 \le |\eta| \le 5$ (SUBSECTION 4.3.3). Furthermore, additional jet activity is suppressed in the central region due to the lack of color flow between initial- and final-state quarks. Therefore, jet tagging in the forward region of the detector [74, 75] together with jet vetoing in the central region [76] are expected to be useful tools to reduce the large background contributions.

Several CMS detector simulation studies have already been carried out for Higgs boson production in the low and intermediate mass range via weak boson fusion (FIGURE 6.1). The present analysis is the first study of
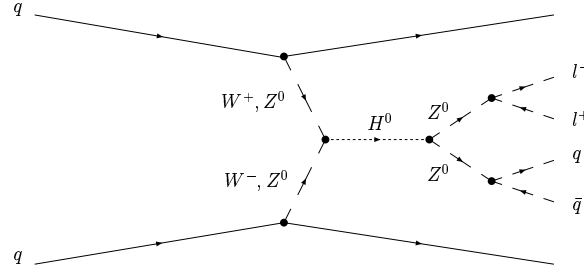
Figure 6.2. Feynman diagram at lowest order for the considered Higgs boson production and decay channel.

the discovery potential with two jets and either two muons or two electrons in the final-state (FIGURE 6.2) for the high mass range between 400 and 700 GeV. The following paragraphs explain first the signal and then the background simulation. The event reconstruction and selection with particular emphasis on the jet reconstruction is shown next. Finally, the expected number of signal and background events after all event selection cuts and the discovery potential of CMS for integrated luminosity of $60 \, \text{fb}^{-1}$ is given corresponding to three years of data taking.

Table 6.1. Next-to-leading order cross sections σ_{WBF} for Higgs boson production through weak boson fusion (as provided by M. Spira, see SECTION B.3), branching ratio BR(H $\rightarrow$ ZZ) (calculated with HDECAY [13]), cross section times branching ratio $\sigma_{\text{WBF}} \times$ BR(H $\rightarrow$ ZZ) $\times$ BR(ZZ $\rightarrow$ llqq) and expected number of produced events after three years of data taking for the studied decay channel.

M_{H}	σ_{WBF}	BR(H $\rightarrow$ ZZ)	$\sigma_{\text{WBF}} \times$ BR(H $\rightarrow$ ZZ) $\times$ BR(ZZ $\rightarrow$ llqq)	expected produced events for $60 \, \text{fb}^{-1}$
200 GeV	2.53 pb	26.13%	62.22 fb	3734
300 GeV	1.42 pb	30.75%	41.10 fb	2466
400 GeV	0.87 pb	27.42%	22.43 fb	1346
500 GeV	0.57 pb	26.07%	13.89 fb	834
600 GeV	0.39 pb	27.03%	9.82 fb	580
700 GeV	0.28 pb	28.20%	7.43 fb	446
800 GeV	0.20 pb	29.28%	5.40 fb	324
900 GeV	0.15 pb	30.20%	4.26 fb	256
1000 GeV	0.11 pb	30.90%	3.20 fb	192

6.1 Event generation and simulation

6.1.1 Signal samples

All signal samples used in this analysis were taken from the official Monte Carlo production done by the CMS collaboration [77]. They were generated with PYTHIA 6.223 interfaced to CMKIN 3.1.0. The Higgs boson was produced in the WW and ZZ fusion process and forced to decay to a ZZ boson pair. One Z boson was chosen to decay into quark anti-quark pairs, whereas the other one was forced to decay either to muons or electrons. Since no further preselection is applied after the generation of the partonic final-state, the samples are normalized to their effective cross sections listed in TABLE 6.1.

6.1.2 Event topology

The studied decay channel is characterized by two leptons and four jets in the final-state. Whereas it is straightforward to reconstruct the leptonically decaying Z boson, the identification of the jets originating from the hadronically decaying Z boson is ambiguous. In order to get an idea how to restore the four-vector of the Higgs boson from the particles in the final-state, the kinematic distributions of the partonic event are studied. For this purpose, the particles as generated in the hard subprocess from CMKIN ntuple files are taken before showering and hadronization was performed (particles with status identifier "3"). This simple approach gives reasonable results for investigating the interesting properties without the necessity of applying a jet clustering algorithm for evaluating final-state jets. The following depicted distributions which are shown in this paragraph are produced with PYTHIA for an assumed Higgs boson mass of 500 GeV only. However, the shape of the distributions remains almost the same if the Higgs boson mass is altered; just the location of the most probable value is slightly shifted.

Although several methods for separating the accompanying jets from the weak boson fusion exist in literature at varying levels of complexity, all these methods are expected to give similar results when studying real proton-proton interactions. For the present analysis, a simple geometrical criterion based on the absolute value of the pseudorapidity η is chosen to distinguish these jet pairs. The jets produced by the quarks originat-
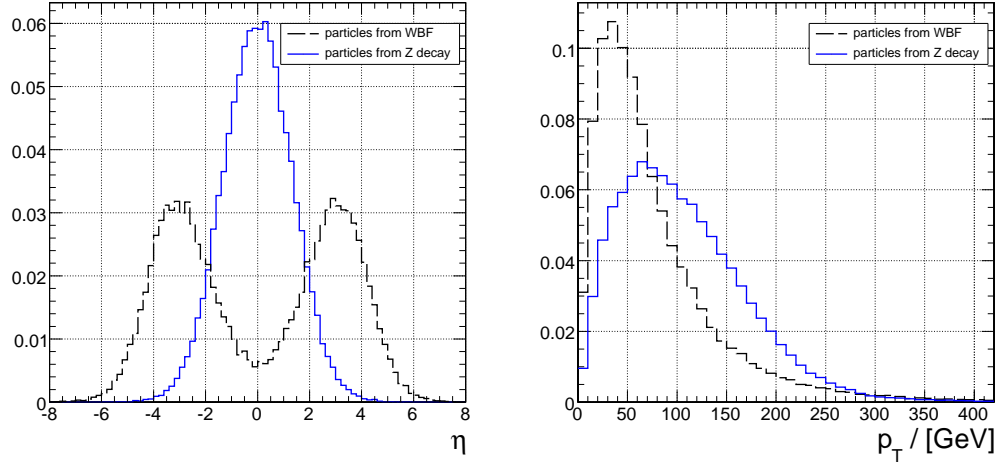
Figure 6.3. The distribution normalized to unit area in pseudorapidity (left) and transverse momentum (right) of the particles coming from the weak boson fusion (dashed line) and the Z boson decay (solid line), respectively. The process was simulated for a Higgs boson mass of 500 GeV.

ing from the Z decay, also referred to as signal jets, exhibit high transverse momenta due to the large Higgs boson mass and are mainly concentrated in the central region of the detector. However, the transverse momenta of the partons which emit the W or Z bosons are governed by the scale of the weak boson masses and are thus quite small. These partons produce jets at large pseudorapidities due to the large energy of the proton beam and are therefore referred to as forward jets or tagging jets. FIGURE 6.3 shows the normalized distributions for the mentioned quantities. As can be seen in the left-hand plot, a small fraction of the events (4 %) have forward jets outside the detector acceptance ($|\eta| > 5$).

FIGURE 6.4 demonstrates efficiency and purity of the signal and tagging jet-selection as a function of the pseudorapidity η. Here, all parton jets found below a specific absolute η threshold are considered as coming from the Z decay and all other parton jets above the threshold are treated as coming from the Higgs production process. Efficiency is defined as the number of jets (either signal jets or tagging jets) in the given interval divided by the total number of jets. Purity is defined as the number of jets (either signal jets or tagging jets) in the given range divided by the total number of signal and tagging jets in this range. Efficiency and purity are
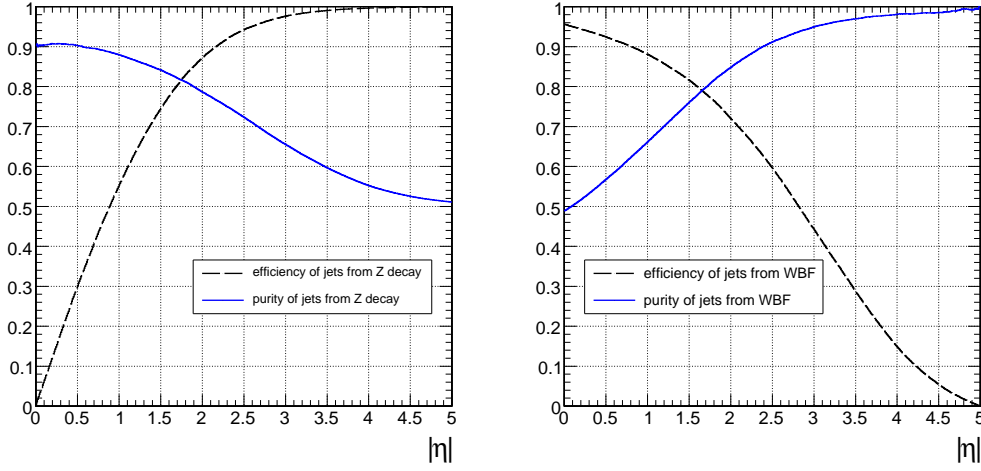
Figure 6.4. Efficiency and purity for the signal jet-selection (left) and forward jet-selection (right) as a function of the cut on $|\eta|$. The efficiency of the forward jet selection does not reach 100% due to the limited detector coverage (only jets up to $|\eta| \leq 5$ are considered here).

determined by summing over all generated events:

$$
\text{signal jet efficiency}(x) = \frac{\text{signal jets in } 0 \leq |\eta| \leq x}{\text{total number of signal jets}}
$$
$$
\text{signal jet purity}(x) = \frac{\text{signal jets in } 0 \leq |\eta| \leq x}{\text{total number of jets in } 0 \leq |\eta| \leq x}
$$

(6.3)

$$
\text{tagging jet efficiency}(x) = \frac{\text{tagging jets in } x \leq |\eta| \leq 5}{\text{total number of tagging jets}}
$$
$$
\text{tagging jet purity}(x) = \frac{\text{tagging jets in } x \leq |\eta| \leq 5}{\text{total number of jets in } x \leq |\eta| \leq 5}
$$

(6.4)

Already the investigation of the kinematic properties of the generated process directs the experimentalist's attention to the challenges that might arise:

- The forward tagging jets are typically located outside the tracker acceptance. This makes an assignment to the primary interaction vertex impossible in order to distinguish them from pile-up events which are typically concentrated in the forward area of the detector. Furthermore, the efficiency of reconstructing jets exceeds 95% only for jets with a transverse momentum higher than 40 GeV [51], but some 30% of the forward jets have a smaller transverse momentum.
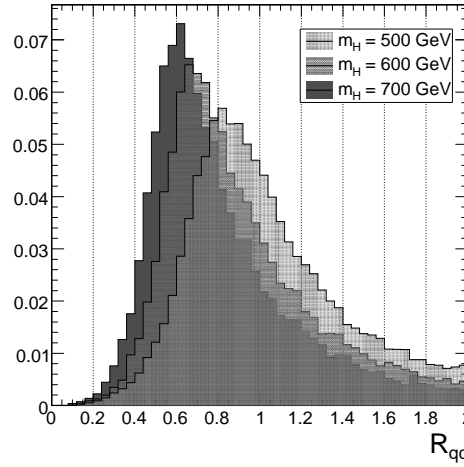
Figure 6.5. The distance $R^2 = (\Delta\eta)^2 + (\Delta\phi)^2$ between the two quarks emerging from the Z boson decay. The distributions which are normalized to unit area are shown for three different Higgs boson masses.

- The large Higgs boson mass leads to a small opening angle of the two quarks originating from the Z decay, causing overlapping signal jets in the detector (FIGURE 6.5). These jets might be difficult to separate and it might be even more difficult to restore the original energy of the two quarks.

6.1.3   Background processes

Physics channels that have signatures similar to the signal in the detector are considered as background processes. The following potential background processes which produce two isolated leptons of the same flavor but opposite sign together with additional hadronic jets are investigated:

- $t\bar{t}$ + multiple jets. Top quarks almost exclusively decay to a W boson and a b quark. If two oppositely charged leptons of the same flavor are produced in the decay chains of the W bosons, the event can have the same event topology in the final state. Despite of its huge cross section, it should be easy to suppress these process since the possibly reconstructed bosons do not show a resonance at the Z boson mass pole. Furthermore, this channel features missing transverse energy.

- ZZ + multiple jets. This process, one weak boson decays hadronically and another decays leptonically, fakes the signature of the signal. Since the Z bosons are produced on-shell, this background is supposed to be largely irreducible.

- WZ + multiple jets. If the Z boson decays leptonically and the W boson hadronically, this leads to the same event signature as the signal.

- Z + multiple jets. This is expected to be the most dangerous background because of its huge cross section. The leptonic decay of the Z boson into a dilepton pair together with additional jet production results in a similar event topology. The main discriminating variables between the signal and background processes are the forward tagging jet separation, the opening angle between the signal jets and the lack of additional jet activity in the central part of the detector for the signal.

TABLE 6.2 lists the expected cross sections for these processes at a center-of-mass energy of $\sqrt{s} = 14$ TeV.

Table 6.2. Expected cross sections for the contributing background processes at leading order (LO) at the LHC ($\sqrt{s} = 14$ TeV). All but the value for the $t\bar{t}$ cross sections were determined by the CMS collaboration using PYTHIA 6.323 with the CTEQ5L structure functions [78]. The cross section for top quark pair production at next-to-leading (NLO) order at the LHC quoted here was predicted by Bonciani et al. [79]. Cross sections for multiple gauge boson production at next-to-leading order can be found in [80].

process	Z+jets (LO)	WZ (LO)	ZZ (LO)	$t\bar{t}$ (NLO)
σ at $\sqrt{s} = 14$ TeV	≈ 14 nb	≈ 30 pb	≈ 15 pb	840 pb

The description of processes with high jet multiplicities obtained from Monte Carlo programs that use shower evolution is inaccurate because hard radiation is suppressed (SECTION 5.1). A more precise description can be achieved if the occurrence of additional partons is included in the calculation of the hard scattering amplitude. The ALPGEN library contains dedicated codes for all relevant processes mentioned above. Therefore the hard processes for all background samples were generated using ALPGEN 2.05 with parton density parameterization CTEQ5M [81]. All samples used in the analysis [82] were generated by the CMS Generator group [83] with the following preselection applied to the partonic events:

- Partons: $p_T > 20 \, \text{GeV}$, $|\eta| < 5$, $\Delta R_{jj} > 0.7$.

- Leptons: $p_T > 3 \, \text{GeV}$, $|\eta| < 3$.

The complete parameter set used by the CMS collaboration for the ALP-GEN sample generation can be found in SECTION B.2.

Showering and hadronization of the generated tree-level process was performed with PYTHIA 6.325 interfaced to CMKIN 6.0.1 (SECTION 5.1). To avoid the earlier mentioned issue of double counting some regions of the phase space, the jets produced in the showering routine are matched to the partons obtained from the matrix element calculation. For this purpose, a jet cone clustering algorithm with cone size $R = 0.7$ and minimum transverse energy $E_T^{\min} > 20 \, \text{GeV}$ is applied to the final-state particles. The event is kept if every hard parton in the event can be matched to a jet based on the distance in η-ϕ-space requiring $(\Delta\eta)^2 + (\Delta\phi)^2 < R^2$, otherwise it is rejected. The parton-level configuration for the samples is generated for a particular number of hard jets ("exclusive sample"). Only for the sample with the highest jet multiplicity, extra jets which do not match to hard partons are allowed to be present after the showering is performed ("inclusive sample").

The following listed processes, calculated at leading order in QCD and EW interactions by ALPGEN, provide the demanded final-states to cover the background examination:

- $(Z/\gamma^\star \to l^+l^-) + N$ jets, $N \leq 6$. The code includes all subprocesses with up to 2 additional light quark pairs. However, the emission of additional hard gluons can be calculated and the code works up to 6 final-state jets. Since heavy quarks are only produced in gluon splitting the code does not describe well enough the branching of gluons into – for example – b quark pairs at large angles. Since the identification of b quarks is not necessary in the present analysis, the required topology is most accurately described by the code [84].

  If the final-state of a charged lepton pair is selected, the interference between intermediate Z and $\gamma^\star$ is contained. Here, the invariant mass of the lepton pair is required to lie in the mass window $40 \, \text{GeV} < m_{ll} < 200 \, \text{GeV}$. Only event samples with at least 3 extra jets are analyzed. One further (soft) jet might arise from the showering step or from contribution due to pile-up events resulting in the signal signature with four final-state jets.

- $(Z \rightarrow f\bar{f}) + (Z \rightarrow l^+l^-) + N$ jets and $(W \rightarrow f\bar{f}') + (Z \rightarrow l^+l^-) + N$ jets, $N \leq 3$, where $f\bar{f}$ denotes all types of fermion pairs indicating full inclusive decays. The code generates two on-shell weak bosons plus up to 3 additional jets. All contributions from EW and QCD processes are included. Since the cross section is about the same order as the signal, all final-jet multiplicities are considered. The used ALPGEN version 2.05 was modified because of a bug reported by the author correcting the branching ratios in ZZ decays [84].

- $t\bar{t} + N$ jets, $N \leq 4$. The kinematical configuration was generated for a top mass of $m_t = 175\,\text{GeV}$. The number of events generated for studies performed in scope of the Physics TDR Vol. 2 [15] was rather small. Since the contribution arising from $t\bar{t}$ production is expected to be of secondary importance, only the available samples are analyzed.

The number of analyzed events for each background sample together with their corresponding cross section obtained after jet-parton matching are listed in SUBSECTION 6.5.1.

6.1.4 Detector simulation

All Monte Carlo samples were processed using the fast detector simulation FAMOS 1.4.0 (SECTION 5.3). In addition, the available full simulated signal samples (SUBSECTION 6.1.1) produced by the CMS collaboration were used to estimate the accuracy of the fast simulation. The detailed interaction of the generated particles with the detector components was performed with OSCAR 2.4.5 (SECTION 5.2). The electronic response of the different subcomponents was obtained from ORCA 7.6.1 (SECTION 5.2). Pile-up events according to a Poisson distribution with average $\mu = 5$ on top of the events, as expected for real collision data at low luminosity operation, was overlaid in both the detailed and the fast reconstruction chain. The trigger table listed in the Physics TDR Vol. 2 was used to simulate the response of the high-level trigger system in both chains but the simulation of the response of the first-level trigger is only technically implemented in the detailed simulation.

6.2 Reconstruction of basic detector objects

All detector objects as electrons, muons, jets and vertices are reconstructed using the standard ORCA and FAMOS algorithms. To be able to compare the reconstruction efficiency and the momentum resolution for fast and full detector simulation, exactly the same algorithms are used in fast and full simulation. The used FAMOS version 1.4.0 was the most recent release within the software framework in CMS. The implemented reconstruction algorithms were adapted from ORCA version 8.13.1. Therefore the results obtained in the fast simulation are compared with this ORCA version.

6.2.1 Online selection

At the first step of the event selection, the event is required to pass either the inclusive muon (stream #43) or the inclusive electron (stream #2) high-level trigger. The thresholds at low luminosity operation are $p_T > 19\,\text{GeV}$ and $p_T > 26\,\text{GeV}$ for muons and electrons, respectively. Taking the single lepton instead of the double lepton trigger results in a higher online selection efficiency. On the other hand, this also leads to higher contribution from events coming from background processes. During the event selection step these events are rejected by the requirement to contain an on-shell Z boson.

Table 6.3. High-level trigger efficiencies for different Higgs boson masses obtained from fast detector simulation.

m_H	single e	single μ	e OR μ
200	40.9% ± 0.3%	45.9% ± 0,3%	85.5% ± 0.4%
300	44.0% ± 0.4%	48.2% ± 0.4%	90.1% ± 0.5%
400	45.8% ± 0.4%	48.4% ± 0.4%	91.2% ± 0.6%
500	47.8% ± 0.4%	48.4% ± 0.4%	92.2% ± 0.6%
600	48.6% ± 0.4%	48.3% ± 0.4%	92.4% ± 0.8%
700	49.9% ± 0.5%	48.3% ± 0.4%	92.5% ± 0.7%

The efficiency for passing the high-level trigger is about 86% for a Higgs boson mass of 200 GeV and reaches asymptotically 93% for a Higgs boson mass of 700 GeV (TABLE 6.3). This is in good agreement with the transverse momentum distribution of the leptons. The online selection efficiencies for the considered background samples are given in TABLE 6.4.

Table 6.4. High-level trigger efficiencies for the studied background samples listed in SUBSECTION 6.1.3. The efficiencies for the samples containing at least one Z boson refer to inclusive samples containing all leptonic decays into charged leptons ($l = $ e, μ, τ). The numbers for the t$\bar{\text{t}}$ samples refer to inclusive top samples containing all possible W decay modes.

| Dataset		single e	single μ	e OR μ
Z + N jets	N=2	28.3% $\pm$ 0.0%	33.9% $\pm$ 0.0%	61.7% $\pm$ 0.0%
Z + N jets	N=3	29.5% $\pm$ 0.0%	34.0% $\pm$ 0.0%	62.7% $\pm$ 0.1%
Z + N jets	N=4	30.8% $\pm$ 0.1%	34.0% $\pm$ 0.1%	63.3% $\pm$ 0.1%
Z + N jets	N=5	32.2% $\pm$ 0.1%	33.9% $\pm$ 0.1%	64.1% $\pm$ 0.1%
Z + N jets	N$\geq$6	34.5% $\pm$ 0.1%	33.4% $\pm$ 0.1%	64.8% $\pm$ 0.2%
ZZ + N jets	N=0	26.8% $\pm$ 0.1%	32.5% $\pm$ 0.1%	57.5% $\pm$ 0.1%
ZZ + N jets	N=1	29.0% $\pm$ 0.1%	33.5% $\pm$ 0.1%	60.1% $\pm$ 0.2%
ZZ + N jets	N=2	31.3% $\pm$ 0.1%	34.3% $\pm$ 0.1%	62.4% $\pm$ 0.2%
ZZ + N jets	N$\geq$3	34.5% $\pm$ 0.1%	33.4% $\pm$ 0.1%	64.8% $\pm$ 0.2%
WZ + N jets	N=0	26.0% $\pm$ 0.4%	35.0% $\pm$ 0.5%	58.6% $\pm$ 0.6%
WZ + N jets	N=1	29.6% $\pm$ 0.5%	36.3% $\pm$ 0.6%	61.8% $\pm$ 0.7%
WZ + N jets	N=2	34.0% $\pm$ 0.6%	38.2% $\pm$ 0.6%	66.2% $\pm$ 0.8%
WZ + N jets	N$\geq$3	37.7% $\pm$ 0.3%	32.9% $\pm$ 0.3%	69.2% $\pm$ 0.2%
t$\bar{\text{t}}$+ N jets	N=0	15.4% $\pm$ 0.0%	17.0% $\pm$ 0.0%	30.7% $\pm$ 0.1%
t$\bar{\text{t}}$+ N jets	N=1	16.7% $\pm$ 0.0%	17.1% $\pm$ 0.0%	31.9% $\pm$ 0.1%
t$\bar{\text{t}}$+ N jets	N=2	18.2% $\pm$ 0.1%	17.0% $\pm$ 0.1%	33.1% $\pm$ 0.1%
t$\bar{\text{t}}$+ N jets	N=3	19.9% $\pm$ 0.1%	16.8% $\pm$ 0.1%	34.5% $\pm$ 0.2%
t$\bar{\text{t}}$+ N jets	N$\geq$4	23.6% $\pm$ 0.1%	16.7% $\pm$ 0.1%	37.6% $\pm$ 0.2%
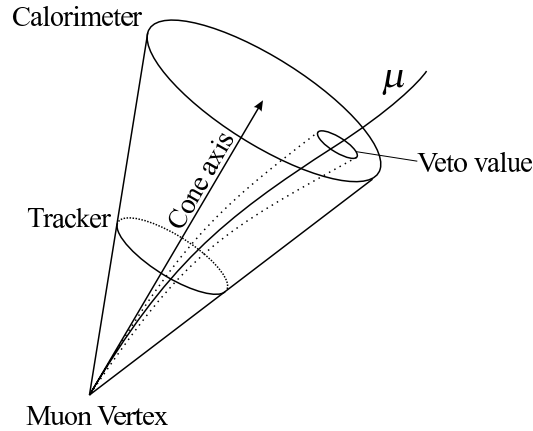
Figure 6.6. Schematic illustration of the isolation computation. All tracks of other particles that are found in the cone defined by the muon direction at the primary vertex are summed up. A small region around the reconstructed muon track is excluded. From [20].

6.2.2 Muon reconstruction

Offline reconstruction of muons is performed using the standard reconstruction algorithms (SUBSECTION 5.4.1). A muon candidate is formed by this algorithm if a track is found in the standalone muon system which can be attached to a track in the inner silicon tracker.

The analysis takes advantage of the fact that muons produced in jets are not isolated in contrary to muons coming from heavy objects such as W and Z bosons. The transverse momentum of all reconstructed tracks with $p_T > 0.9$ GeV falling in a cone of size $R = 0.25$ around the muon direction is summed up (FIGURE 6.6). In the calorimeter, the sum of $E_T > 0.2$ GeV in a cone of size $R = 0.3$ around the muon track is taken. To profit from the better discrimination performance of the electromagnetic calorimeter, the energy deposit in the cone is defined as weighted sum of the transverse electromagnetic and hadron calorimeter energies, using $E_T = 1.5\,E_T^{\mathrm{ECAL}} + E_T^{\mathrm{HCAL}}$. The contribution from by the muon itself is subtracted to improve the discriminating power of the isolation algorithm.

If the primary vertex is found in the event, the reconstructed muon track is extrapolated to the interaction vertex. The transverse impact parameter significance of the muon track is determined by dividing the transverse impact parameter d_T - the shortest distance in the transverse plane between the track and the interaction point - by its uncertainty σ_T.

6.2.3 Electron reconstruction

Electron candidates are reconstructed combining tracks in the inner silicon tracker with electromagnetic clusters in the electromagnetic calorimeter (SUBSECTION 5.4.2). To improve the estimate of the momentum at the interaction vertex, the momentum which is measured in the tracker $\vec{p}_{rec}$ is corrected by the energy measured in the electromagnetic calorimeter E_{rec}:

$$\vec{p}_{cor} = \vec{p}_{rec} \cdot \frac{E_{rec}}{|p_{rec}|} \tag{6.5}$$

This is motivated by the fact that energy and transverse momentum are affected differently by bremsstrahlung, and from the better fractional resolution of the electromagnetic calorimeter for electrons with high transverse momentum.

Again, if the primary vertex can be reconstructed in the event, the impact parameter and its significance for the electron track are calculated.

To distinguish real electrons from other misidentified candidates such as pions, a likelihood approach as implemented in the full simulation package [85] was adopted for the fast simulation. The likelihood is calculated for the real ψ_S and fake electron hypothesis ψ_B

$$L(\mathbf{x}, \psi) = \prod_i p_i(x_i, \psi) \tag{6.6}$$

where x_i are the following discriminating variables:

- Difference in pseudorapidity between the reconstructed track and super-cluster in the electromagnetic calorimeter:

$$x_1 = |\eta_{track} - \eta_{SC}|$$

- Ratio of deposited energy in electromagnetic calorimeter and uncorrected momentum (Eop):

$$x_2 = \frac{E_{SC}}{|p_{track}|}$$

- Ratio of deposited energy in hadronic and electromagnetic calorimeter (HoE):

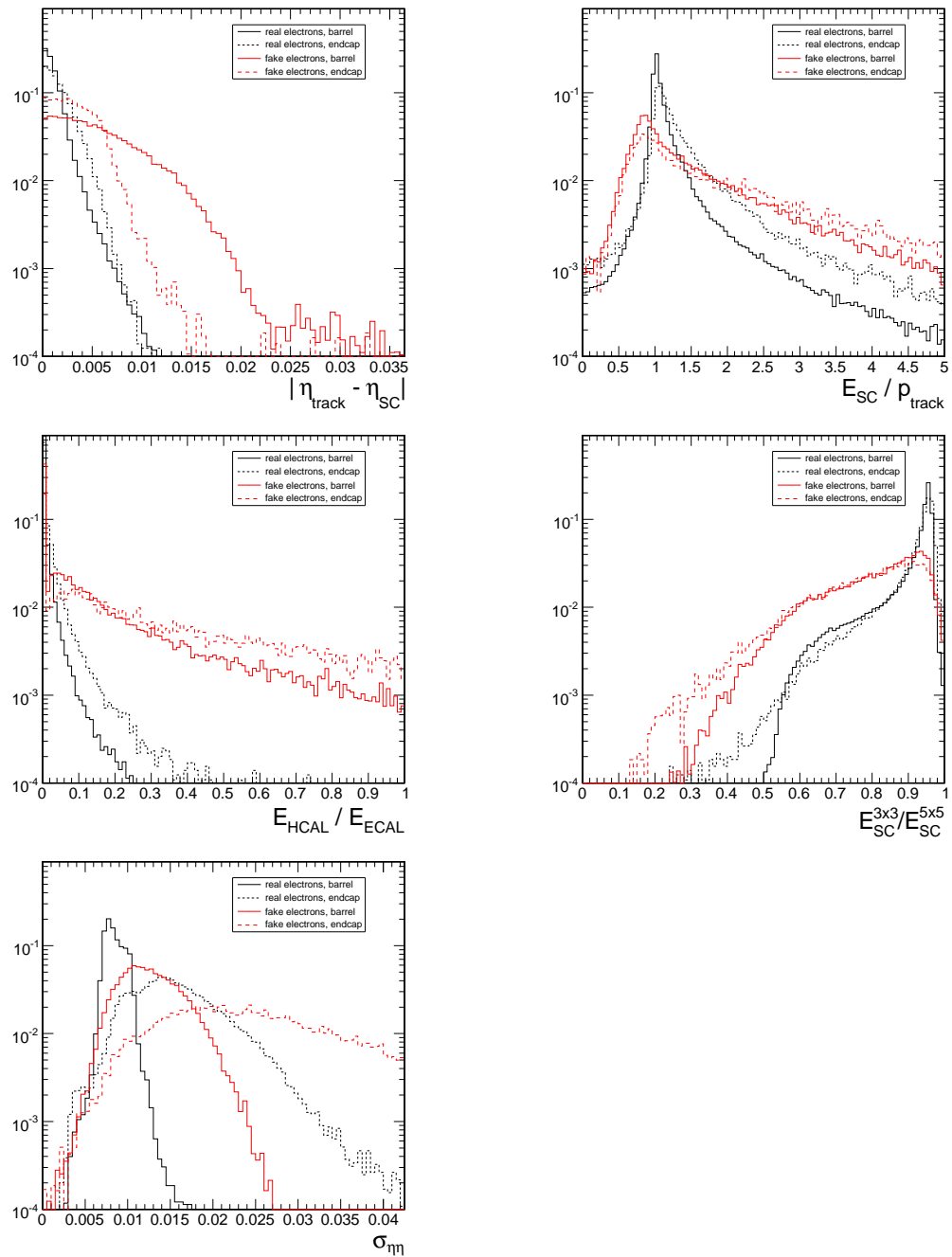$$x_3 = \frac{E_{HCAL}}{E_{ECAL}}$$

Figure 6.7. Probability density functions [29] used in the likelihood estimator for the electron identification. From top left to bottom right, the distribution for the discriminating variables x_1, x_2, x_3, x_4 and x_5 as described in this section for real and fake electrons.

- Ratio of deposited energy in a 3×3 array and a 5×5 array of ECAL crystals around the crystal with the highest measured energy:

$$x_4 = \frac{E_{\mathrm{SC}}^{3 \times 3}}{E_{\mathrm{SC}}^{5 \times 5}}$$

- Shower spread in longitudinal direction, given by the covariance matrix element for 5×5 array of ECAL crystals:

$$x_5 = \sum_i (\eta_i - \eta_{E\mathrm{max}})^2 \frac{E_i}{E_{\mathrm{SC}}}$$

The reference distributions p_i (FIGURE 6.7) for variable i, having value x_i, given hypothesis ψ, were constructed for both the barrel ($|\eta| < 1.5$) and the endcaps ($|\eta| > 1.5$).

The likelihood ratio used to discriminate between real and fake electrons is then given by:

$$\Lambda = \frac{L(\mathbf{x}, \psi_{\mathrm{S}})}{L(\mathbf{x}, \psi_{\mathrm{S}}) + L(\mathbf{x}, \psi_{\mathrm{B}})} \tag{6.7}$$
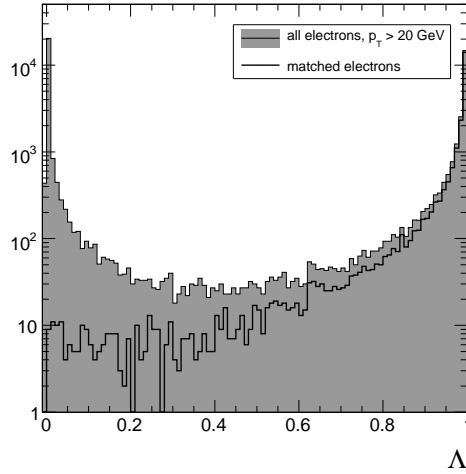

Figure 6.8. Distribution of the likelihood ratio of all reconstructed electrons (shaded) and these reconstructed electrons which could be matched to generated particles (solid line). Only electrons which carry a transverse momentum higher than $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ are shown.

The distribution of the likelihood ratio Λ of all objects which have been reconstructed as electrons is shown in FIGURE 6.8. The Λ distribution of reconstructed electrons which could be matched to an generated electrons is overlyed. The matching of reconstructed to generated particles is explained in detail in SECTION 6.4.

6.2.4   Jet reconstruction

Jet reconstruction is a crucial part of this analysis. Since the final-state consists of four jets, the event reconstruction strongly depends on the detector accuracy in finding and identifying jets with high efficiency and good energy resolution.

Jet finding is performed using the standard Iterative Cone algorithm with cone size $R = 0.5$ on calorimeter towers (SUBSECTION 5.4.3). The towers used for clustering are required to have a transverse energy in excess of $E_T > 0.5\,\text{GeV}$ and higher energy in excess of $E > 0.8\,\text{GeV}$ to suppress noise and jets produced by pile-up and underlying events. The tower constituents are added according to the E_T recombination scheme. Only jets which deposit more energy than $E_T > 10\,\text{GeV}$ in the calorimeter are considered. A jet calibration technique based on QCD dijet Monte Carlo samples is applied. All jet energies quoted in the current chapter are corrected jet energies. Due to the availability of jet energy correction factors for the selected cone size, the Iterative Cone algorithm is chosen for the present analysis.

As mentioned above, electrons deposit their energy in the electromagnetic calorimeter and are therefore reconstructed as jets by the jet clustering algorithms. In principle, this is appreciated, since the particles contained in jets always produce electrons when transversing matter and the energy loss in the electromagnetic calorimeter can be added to the total jet energy. On the other hand, isolated electrons originating from Z boson decays are misidentified as jets and the electron appears twice in the list of reconstructed high-level objects. Therefore, these double-counted objects need to be removed from the particle list for the subsequent analysis. To accomplish this task, jets are required to be well-separated from electrons identified by the likelihood approach described above. Furthermore, jets must deposit a minimum amount of their energy in the hadronic calorimeter cells.

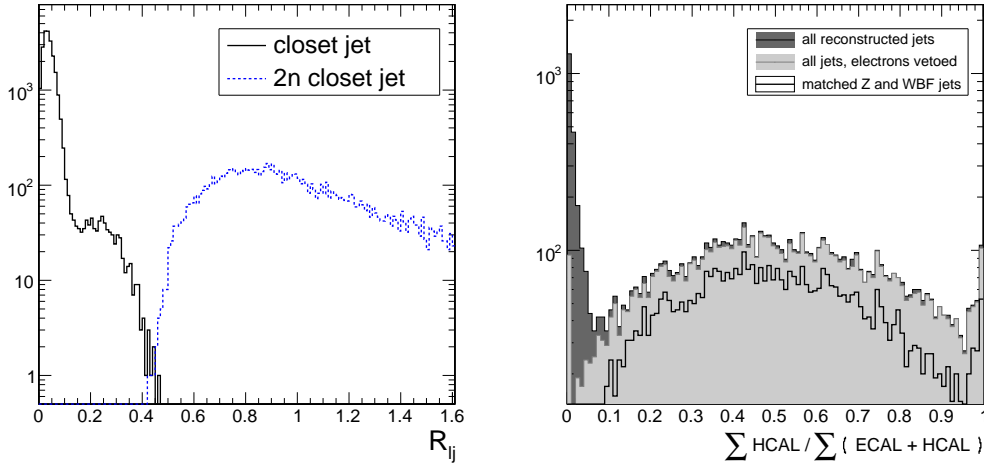FIGURE 6.9 shows the mentioned quantities for the jet identification. The

Figure 6.9. The distance $R^2 = \Delta\eta^2 + \Delta\phi^2$ between identified electrons and their closest jet and 2nd closet jet (left). Hadronic energy fraction of the jets reconstructed with the described iterative cone algorithm (right). The jets before (dark shaded area) and after excluding jets in a cone of size $R = 0.3$ around identified electrons (light shaded area) is shown together with jets matched to the generated partons (solid line). Only particles which carry at least a minimum transverse momentum of 10 GeV are considered.
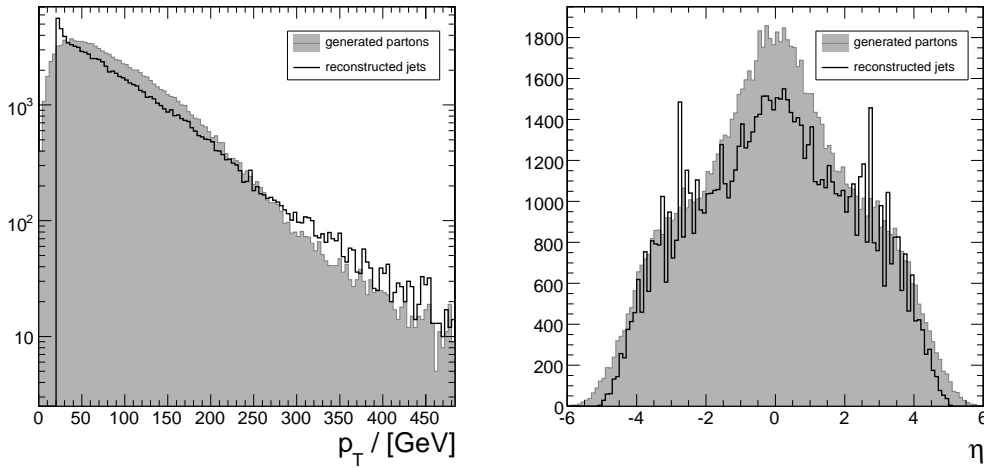

Figure 6.10. Transverse momentum spectrum for all reconstructed jets together with generated partons (left) and their distribution in pseudorapidity (right). Only partons which carry at least 20 GeV are shown in the η distribution. The enrichment of jets at $\eta = 2.7\ldots3.0$ arises from an unusually large HCAL readout tower (index 28) in the endcap region. The readout scheme has been changed recently to provide a finer granularity and obtain a higher precision. The detector is instrumented with hadronic calorimeter cells up to $\eta = 5.2$.

left-hand chart depicts the distance of electrons to their closest jet and next-to-closet jet found in η-ϕ-space. In order to demonstrate the effect of the electron veto, electrons are identified as coming from the Z boson decay by two methods, first by selecting electrons with a minimum likelihood ratio $\Lambda > 0.2$ and second by using only electrons which could be matched to the generated signal electrons. The plot shows clearly that each reconstructed electron obviously corresponds to a reconstructed jet.

On the right-hand plot of FIGURE 6.9 the hadronic energy fraction of the total deposited energy of reconstructed jets is shown. The first bin in the histogram corresponds to particles which deposit a negligible fraction of their energy in the the hadronic calorimeter like electrons do. Motivated by the previous consideration, jets in a cone size of $R = 0.3$ around identified electrons are excluded. This procedure almost eliminates the contribution by particles with little reconstructed hadronic energy. Jets outside the acceptance of the electromagnetic calorimeter deposit their energy exclusively in the hadronic calorimeter and appear in the last bin of the histogram. The distributions in η and p_T of the reconstructed jets is shown in FIGURE 6.10.

6.3 Event reconstruction

6.3.1 Offline reconstruction

After the event passed the high-level trigger for single muons or single electrons and a primary vertex was found, the offline reconstruction of the Higgs decay chain starts. Out of all reconstructed objects, two leptons and four jets that fulfill the following criteria are selected:

- The lepton with the largest transverse momentum and the lepton with the next to largest transverse momentum with same flavor but opposite charge are selected. These leptons must carry a transverse momentum larger than 20 GeV. Electrons have to be identified by the likelihood approach explained above by a likelihood ratio larger than $\Lambda > 0.2$ (SUBSECTION 6.2.3). Jets within a distance of $R < 0.3$ around a selected electron are assumed to be misidentified electrons (SUBSECTION 6.2.4) and are removed from the list of input objects for the subsequent analysis.

- In the central part of the detector up to $|\eta| < 2$, the most energetic and next to most energetic jet are selected, if their transverse momenta is larger than 20 GeV. This selected jet pair is assumed to come from the hadronically decaying Z boson.

- The accompanying forward jets are searched in the outer part of the detector. In the region $|\eta| \geq 2$, the most energetic jet and the next to most energetic jet in the opposite hemisphere $\eta_{tag1} \cdot \eta_{tag2} < 0$ are picked. Again, their transverse momenta needs to be larger than 20 GeV.

- All jets are asked to deposit more the 20% of their total energy in the hadronic calorimeter (jets which are beyond the acceptance of the electromagnetic calorimeter show a hadronic energy of 100% per definition).

6.3.2 Mass resolution of the leptonically decaying Z boson

The leptonic Z boson resonance can be cleanly reconstructed from the selected lepton pair as seen in FIGURE 6.11. The mass peak is slightly shifted to smaller masses due to bremsstrahlung and detector influences mentioned above. The probability to find two leptons fulfilling the lepton selection criteria mentioned above is 73.7 % with respect to all simulated signal events for a 500 GeV Higgs boson.

6.3.3 Mass resolution of the hadronically decaying Z boson

The limited performance of the hadronic calorimeter and the jet reconstruction influenced by the effects mentioned in SUBSECTION 5.4.3 cause a much broader reconstructed resonance for the hadronically decaying Z boson. The overlapping jets are very difficult to separate for the Iterative Cone algorithm. Furthermore, for higher Higgs boson masses, the Z mass resonance is shifted to higher values. This is caused by the correction for out-of-cone effects [1]. The individual jets interfere with each other

---

[1] Particles that are physicswise associated with the jet but fall out of the cone.
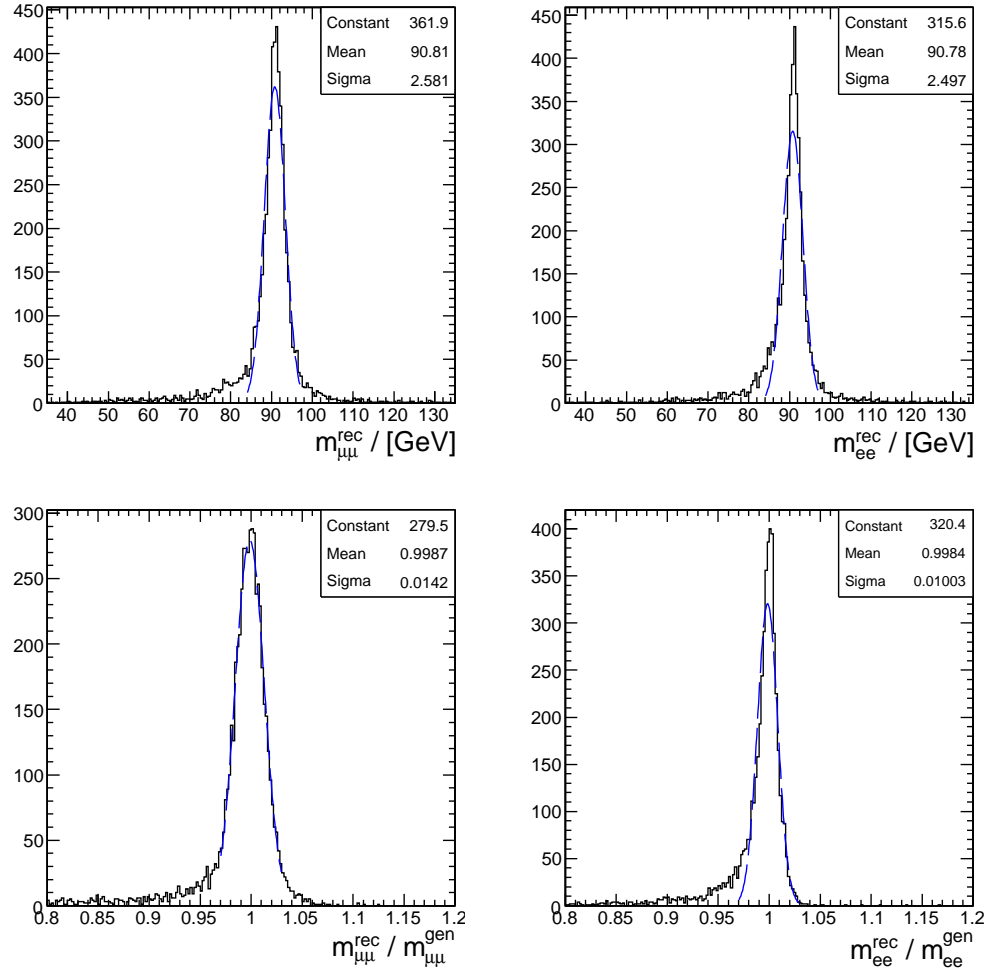
Figure 6.11. Reconstructed invariant masses and ratio for reconstructed to true invariant masses of the leptons, for muon (left) and electron (right) pairs. The selection criteria for leptons as described in the text are applied. The resolution obtained from a fit of a normalized Gaussian distribution for electrons and muons is $\sigma = 1.0\,\%$ and $\sigma = 1.4\,\%$ respectively.

by spreading into the neighboring cone and after applying the jet calibration the jet energy is over-estimated ("over-calibration of jets"). This effect increases with higher Higgs boson masses. The smaller the opening angle of the jets originating from the Z boson is, the more the jets are over-calibrated (TABLE 6.5). Here, one can see the strong dependence of the jet energy calibration on the reference channel used to evaluate the correction factors.

Table 6.5. Reconstructed invariant central jet pair mass as a function of the generated Higgs bosons mass. Mean value and variance was determined by fitting a Gaussian function in the mass interval from 60 to 140 GeV.

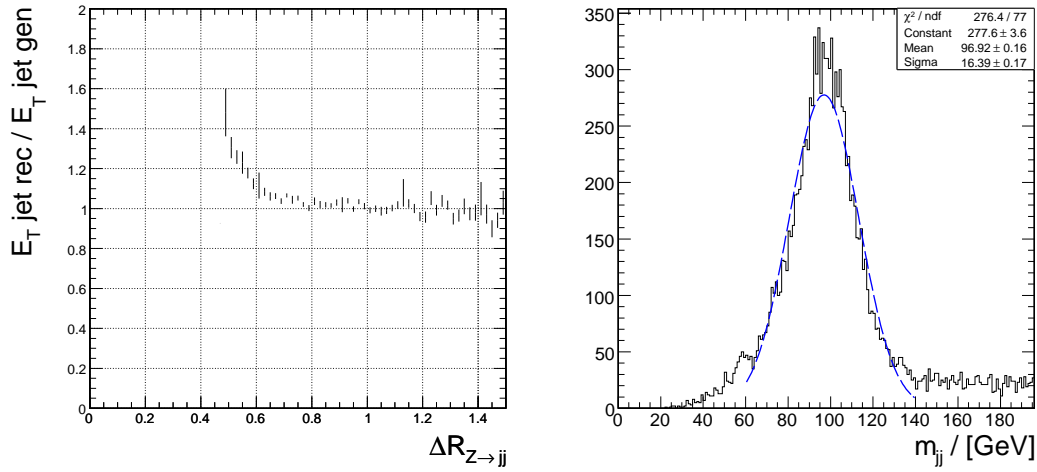| M_{H} / [GeV]	200	300	400	500	600	700
M_{Z} / [GeV]	91.0	92.7	94.5	96.9	97.7	99.2
σ_{Z} / [GeV]	29.8	21.1	18.1	16.4	16.2	15.7

Figure 6.12. Ratio of reconstructed and generated transverse energy of the leading jet as a function of the distance to the next to leading jet (left) and invariant mass of both jets (right) originating from Z boson decays for a Higgs boson mass of 500 GeV.

Incorrect selection of signal and forward jets broadens the invariant mass distribution additionally and leads to large tails as depicted in the right-hand plot of FIGURE 6.12.

6.3.4 Efficiency of the forward tagging jet reconstruction

The reconstruction efficiency for forward tagging jets is only moderate as already discussed above. Nevertheless, for a reasonable suppression of the various background processes the occurrence of both forward jets is strongly required. Since the partons which initiate the weak boson fusion come out of the proton beam, the produced forward jets show a huge energy. These jets are further on nearly back-to-back so that the calculation of an invariant mass from their Lorentz vectors has a meaningful representation and can be used as distinguishing feature.
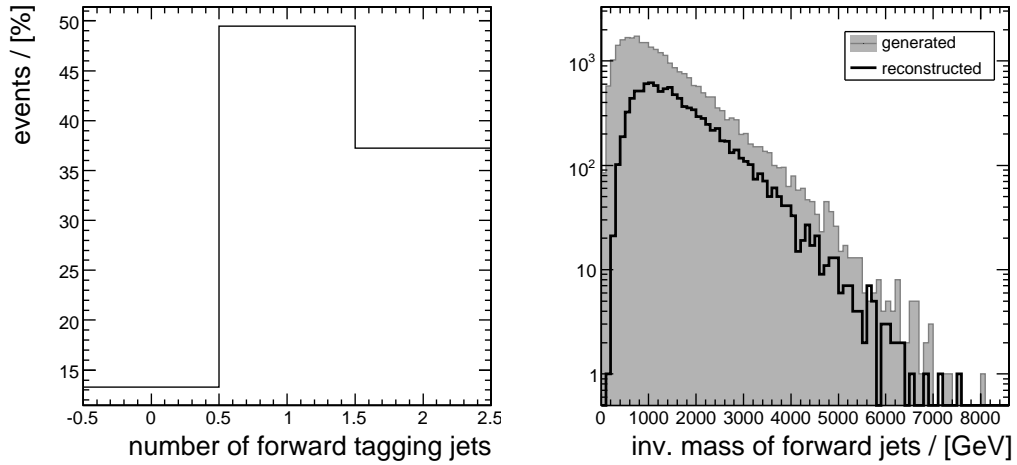
Figure 6.13. Number of forward jets found behind $|\eta| > 2$ (left) and invariant mass of reconstructed forward jets (right).

Apart from their very large invariant mass of the order of the beam energy, the forward jets balance the event in the transverse plane of the detector. If neutrinos, which occur in semi-leptonic decays of particles in hadronic jets are neglected, the transverse momentum of all reconstructed particles is conserved.

Furthermore, jets and leptons produced in background processes typically give higher pseudorapidities compared to objects coming from Higgs boson decays. The location of the forward tagging jets can be exploited to define a central region of the detector. Events which show decay products located outside of the central region defined by the forward tagging jets can be suppressed by this means.

Unfortunately, the probability of finding simultaneously two forward jets with transverse momenta larger than 20 GeV is only ≈ 35% (FIGURE 6.13). The missing ones are either in the central region of the detector or not reconstructed at all. Therefore the reconstruction and selection efficiency of Higgs boson decays is mostly dominated by the forward jet reconstruction efficiency.

6.3.5 Mass resolution of the reconstructed Higgs boson

For a significant suppression of the contributing background processes, a complete exploitation of the characteristic event topology is strictly de-
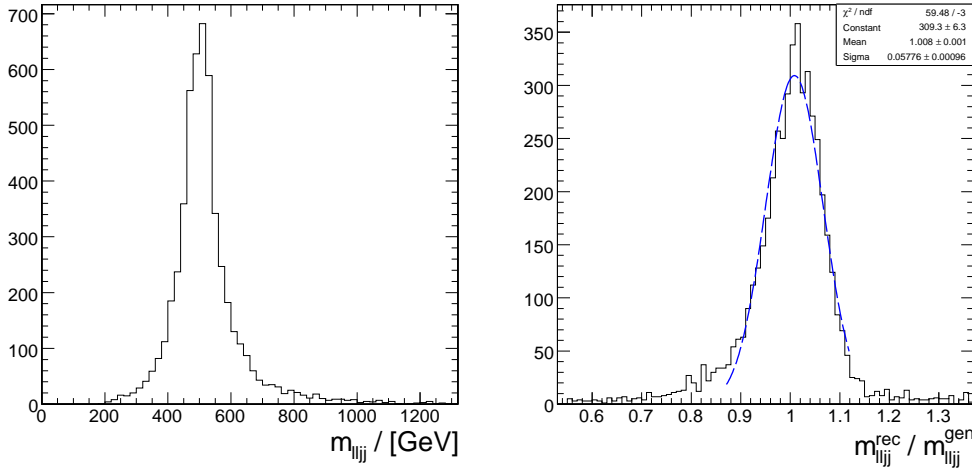
Figure 6.14. Invariant mass distribution m_{lljj} for the fully reconstructed Higgs events (left). The signal events were generated for a Higgs boson mass of 500 GeV. The resolution obtained from a fit of a Gaussian distribution to the ratio of reconstructed and generated Higgs boson mass (right) is $\sigma = 5.8\,\%$.

manded. All reconstructed events which do not possess a reconstructed Higgs boson together with two identified accompanying forward jets are discarded for the subsequent analysis and do not enter the event selection procedure.

FIGURE 6.14 shows the invariant mass of the two identified leptons and the two selected signal jets for a Higgs boson mass of 500 GeV. Only 18.4 % of the simulated events show the event topology (referred to as "weak boson fusion topology" in the following) required by the analysis to get accepted for the further event selection step.

6.4 Comparison between full and fast simulation

To determine the accuracy of the fast simulation, 15.000 identical signal events which contained a Higgs boson with a mass of 600 GeV were processed both with the full simulation (OSCAR/ORCA) and the fast simulation chain (FAMOS). The reconstructed objects are matched to the corresponding Monte Carlo particles based on a combined criterion taking into account their absolute resolutions in ϕ, η and their relative resolutions in p_{T} (TABLE 6.6).
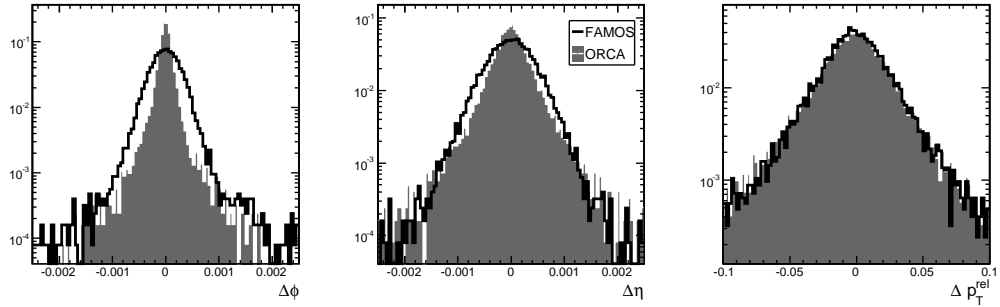
Figure 6.15. Resolutions for muons in ϕ (left), η (center) and the relative p_T resolution (right). The agreement between the fast (FAMOS) and the full detector simulation (ORCA) is acceptable for η and p_T. However, the ϕ distribution is not correctly determined by the fast simulation. This is only of minor importance for the present analysis since the Higgs boson mass resolution is dominated by the jet resolution anyway.
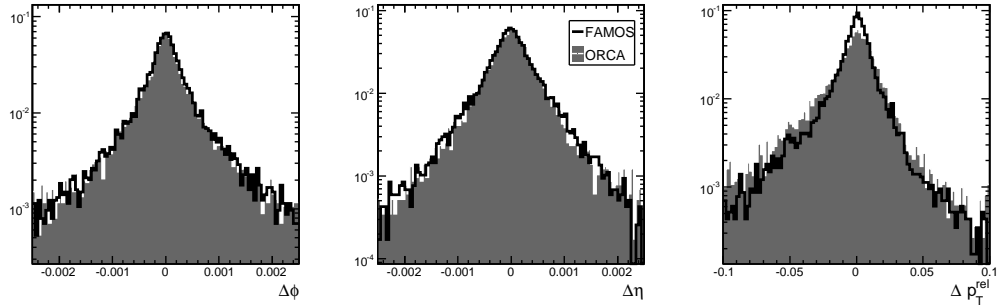

Figure 6.16. Resolutions for electrons in ϕ (left), η (center) and the relative p_T resolution (right). There is only a slight difference between the fast and the full simulation.
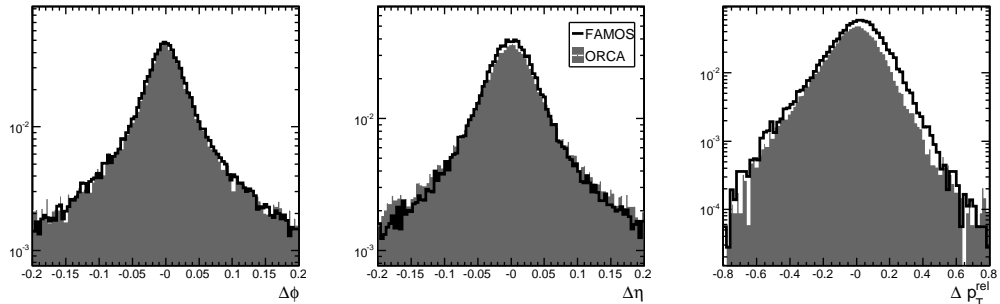

Figure 6.17. Resolutions for jets in ϕ (left), η (center) and the relative p_T resolution (right). There is a shift in the relative p_T distribution between the fast and the full simulation. However, the shape of the distribution is almost the same so the mass spectra will just be slightly shifted to higher values in the fast simulation.

As can be seen in FIGURE 6.15, FIGURE 6.16 and FIGURE 6.17, the resolutions reflected by the fast simulation are in general consistent with the resolutions obtained in full simulation. This is in good agreement with the design values of the fast simulation package which has been tuned subsequently to reproduce the same results as the full simulation.

The reconstructed objects whose quantities are heavily used as discriminating variables are compared alongside. TABLE 6.7 lists the mass resolutions and reconstruction efficiencies for the Z bosons, the reconstructed Higgs boson, the invariant mass of the two forward jets (referred to as dijet) and the overall event reconstruction efficiency with respect to all 15.000 simulated events. The particles were required to pass the preselection cuts as described in SUBSECTION 6.3.1. Furthermore, the entire event must contain a reconstructed primary vertex and trigger decision must be present. In case of the full simulation, the Level-1 trigger decision was not taken into account.

A good agreement of the reconstruction efficiencies and the mass resolutions is found between the fast and the full simulation. In case of the overall event acceptance, the reconstruction efficiencies obtained in fast and full simulation differ by less than 1 %.

Table 6.6. Overview of the absolute and relative resolutions for the reconstructed objects used in this analysis.

high-level object		FAMOS	ORCA
muon	$\Delta\phi$	$4.4 \cdot 10^{-04}$	$1.5 \cdot 10^{-04}$
	$\Delta\eta$	$4.1 \cdot 10^{-04}$	$2.7 \cdot 10^{-04}$
	$\Delta p_{\mathrm{T}}^{\mathrm{rel}}$	1.5 %	1.5 %
electron	$\Delta\phi$	$5.8 \cdot 10^{-04}$	$5.5 \cdot 10^{-04}$
	$\Delta\eta$	$3.8 \cdot 10^{-04}$	$3.5 \cdot 10^{-04}$
	$\Delta p_{\mathrm{T}}^{\mathrm{rel}}$	1.8 %	2.0 %
jet	$\Delta\phi$	$4.6 \cdot 10^{-02}$	$4.5 \cdot 10^{-02}$
	$\Delta\eta$	$4.2 \cdot 10^{-02}$	$4.1 \cdot 10^{-02}$
	$\Delta p_{\mathrm{T}}^{\mathrm{rel}}$	14.7 %	14.1 %

Table 6.7. Comparison of the reconstruction efficiency ϵ_{rec} and the mass resolution σ_{res} for both the fast (FAMOS) and the full simulation chain (ORCA). The reconstruction efficiencies for both Z bosons, the Higgs boson and the invariant mass of the two forward jets ("dijet") are listed with respect to all simulated events. The total acceptance denotes the fraction of events which possess a reconstructed Higgs boson accompanied by two identified forward jets.

	FAMOS		ORCA	
	ϵ_{rec}	σ_{res}	ϵ_{rec}	σ_{res}
$Z \rightarrow qq$	68.9 %	15.4 %	72.2 %	15.5 %
$Z \rightarrow ll$	75.5 %	1.4 %	73.0 %	1.3 %
Higgs	56.5 %	6.6 %	59.7 %	6.5 %
dijet	37.0 %	13.8 %	34.3 %	14.4 %
total acceptance	19.3 %		20.3 %	

6.5 Event selection

6.5.1 Kinematic event selection cuts

The $qq \rightarrow qqH, H \rightarrow (Z \rightarrow ll)(Z \rightarrow qq)$ signal presents a characteristic event topology which allows a powerful rejection of events arising from the background processes listed above (SUBSECTION 6.1.3). The selection cuts applied to the events which were accepted in the first analysis step (SUBSECTION 6.3.1) are optimized to maximize the significance of the signal while conserving a reasonable predictable shape of the background distribution. The maximization of the significance was performed by hand for a sample which contained Higgs bosons with a mass of 600 GeV. Selection cuts on discriminating variables which exhibit mass information are only loosely tightened. An event selection which is sensitive to the mass leads to an enrichment of background events in the mass region where the Higgs boson is searched for and makes a prediction of the background shape from generated background events almost impossible.

The events are selected according to the following sequential cuts:

a) To reject events whose selected leptons do not seem to emerge from the primary vertex, the impact parameter significance in the transverse plane for the leptons has to be smaller than $d_{\text{T}}/\sigma < 3.5$. Furthermore, the muons are required to be isolated in the tracker by means of showing a smaller contribution of charged particle tracks of

Table 6.8. Summary of the applied selection cuts. These cuts are optimized for the discovery of a Higgs boson with a mass of 600 GeV.

Selection	Configuration
event selection	high-level trigger decision, primary vertex found
lepton selection	$p_T > 20$ GeV , $d_T/\sigma < 3.5$
	muons: $\sum_{R=0.25}^{\text{tracker}} p_T < 1.6$ GeV, $\sum_{R=0.3}^{\text{ECAL}} E_T < 4.0$ GeV
	electrons: $\Lambda > 0.2$
jet selection	$p_T > 20$ GeV, $|\eta| < 2.0$, $\Delta R_{lj} > 0.3$
	$E_{\text{HCAL}}/(E_{\text{HCAL}} + E_{\text{ECAL}}) > 20\%$
forward jet	$p_T > 20$ GeV, $|\eta| > 2.0$, $\Delta R_{lj} > 0.3$
tagging	$\eta_{\text{tag1}} \cdot \eta_{\text{tag2}} < 0$
leptonic Z	84 GeV $< m_{ll} < 97$ GeV, $0.3 < \Delta R_{ll} < 1.6$
	$|\tilde{\eta}_l| = |\eta_l - (\eta_{\text{tag1}} + \eta_{\text{tag2}})/2| < 1.8$
hadronic Z	75 GeV $< m_{jj} < 115$ GeV, $\Delta R_{jj} < 1.6$
	$|\tilde{\eta}_j| = |\eta_j - (\eta_{\text{tag1}} + \eta_{\text{tag2}})/2| < 1.8$
dijet	$m_{jj} > 900$ GeV
	$p_T^{\text{balance}} < 50$ GeV
Higgs	$2.2 < \Delta R_{ZZ} < 3.8$

$\sum_{\Delta R=0.25} p_T < 1.6$ GeV and a smaller energy deposit of $\sum_{\Delta R=0.3} E_T < 4.0$ GeV in the calorimeters (SUBSECTION 6.2.2).

b) The invariant mass of the two leptons must be in the mass window of 84 GeV $< m_{ll} < 97$ GeV. In addition, the opening angle is requested to fulfill $0.3 < \Delta R < 1.6$.

c) The invariant mass of the two jets in the central region of the detector must lie in the mass window of 75 GeV $< m_{jj} < 115$ GeV. Furthermore, the opening angle between these two jets must be smaller than $\Delta R < 1.6$.

d) The forward jets are used to identify the region of the detector which is preferred by the Higgs decay products. For all reconstructed particles the shifted pseudorapidity $\tilde{\eta}_i = \eta_i - 1/2(\eta_{\text{tag1}} + \eta_{\text{tag2}})$ as introduced in SUBSECTION 6.3.4 is determined. Events that have decay products outside $|\tilde{\eta}_i| > 1.8$ are rejected.

e) Momentum conservation is required in the transverse plane. The event is required to be balanced by the restriction $p_T^{\text{balance}} < 50$ GeV

Table 6.9. Analyzed events of the signal Monte Carlo samples containing decays of the topology $qq \to qqH, H \to (Z \to ll)(Z \to qq)$. Here, l refers to electrons and muons. The numbers given in the row "events with WBF topology" refer to accepted events which show a reconstructed Higgs boson together with two identified forward jets. The weight is defined as the ratio of expected events and analyzed Monte Carlo events: $w = N_{exp}/N_{MC}$.

$M_H/[\text{GeV}]$	400	500	600	700
σ / [fb]	22.43	13.89	9.82	7.43
events analyzed	30,000	27,999	18,999	18,999
events with WBF topology	4,706	5,153	3,681	3,435
σ WBF topology / [fb]	3.52	2.56	1.90	1.34
lepton selection (a)	81.79%	81.89%	82.07%	81.78%
$Z \to ll$ cut (b)	59.16%	64.41%	67.78%	66.90%
$Z \to jj$ cut (c)	34.98%	43.39%	45.42%	44.57%
$\tilde{\eta}$ cut (d)	29.49%	36.74%	37.92%	37.06%
p_T balance (e)	26.18%	31.40%	32.17%	30.71%
dijet mass cut (f)	20.53%	25.40%	25.81%	25.53%
Higgs ΔR cut (g)	19.32%	24.55%	25.32%	25.15%
remaining events	909	1265	932	864
σ_{eff} / [fb]	0.68	0.63	0.48	0.34
weight for 60 fb^{-1}	0.04	0.03	0.03	0.02
exp. events for 60 fb^{-1}	41	38	29	20

where $\vec{p}^{\,\text{balance}} = \vec{p}_{\text{Higgs}} + \vec{p}_{\text{tag1}} + \vec{p}_{\text{tag2}}$ is the vector sum of the momenta of all final-state particles belonging to the event.

f) The invariant mass of the forward tagging jets must exceed $m_{\text{dijet}} > 900\,\text{GeV}$ (SUBSECTION 6.3.4).

g) The two centrally produced Z bosons tend to emerge back-to-back from the Higgs boson decay due to the large Higgs boson mass. Therefore, the opening angle R_{ZZ} between the Z bosons must lie between 2.2 and 3.8 in η-ϕ-space.

TABLE 6.8 gives a summary of the quantities used for the offline event reconstruction and selection. The normalized distributions of the listed discriminating variables used to distinguish between signal and background events are gathered in APPENDIX A.

Table 6.10. Analyzed events of the $(Z \rightarrow ll)$+jets ALPGEN samples separated by the jet multiplicity as generated in the hard subprocess. Here, the Z boson is allowed to decay into a charged lepton pair ($l = e, \mu, \tau$). ALPGEN 2.05 did not support the generation of samples that reflect the typical WBF topology up to 6 extra jets. Therefore, 37 million Monte Carlo events needed to be generated to retain enough events after the preselection.

$(Z \rightarrow ll) + N$ jets	$N = 3$	$N = 4$	$N = 5$	$N \geq 6$
σ / [fb]	66,000	17,000	4,000	3,100
events analyzed	26,981,336	6,200,697	2,630,063	785,649
events with WBF topology	306,858	138,857	94,971	61,688
σ WBF topology / [fb]	751	381	144	243
lepton selection (a)	77.97%	77.17%	75.21%	70.55%
$Z \rightarrow ll$ cut (b)	11.68%	15.59%	18.45%	25.47%
$Z \rightarrow qq$ cut (c)	1.23%	1.42%	1.57%	1.58%
$\tilde{\eta}$ cut (d)	0.76%	0.88%	0.95%	0.78%
p_T balance (e)	0.52%	0.54%	0.44%	0.21%
dijet mass cut (f)	0.12%	0.17%	0.16%	0.09%
Higgs ΔR cut (g)	0.11%	0.14%	0.13%	0.08%
remaining events	331	193	119	48
σ_{eff} / [fb]	0.81	0.53	0.18	0.19
weight for 60 fb^{-1}	0.15	0.16	0.09	0.24
exp. events for 60 fb^{-1}	49	32	11	11

In Higgs production via weak boson fusion no color is exchanged between the initial-state quarks which leads to a suppressed hadronic activity of the central region. Therefore, a veto on extra jets could be used to reject background events whose gluons produce jets with a transverse momentum of about $20\,\text{GeV} < p_T < 40\,\text{GeV}$ [76]. A detailed investigation of extra jets found in the barrel region with transverse momenta $p_T > 20\,\text{GeV}$ was performed. However, it was revealed by the fast detector simulation and confirmed by the full detector simulation that additional jets are also present in reconstructed Higgs boson decays. These jets arise from out-of-cone effects, underlying and overlapping events and show almost identical kinematical and spacial distributions as the background events. This prevented the implementation of a central jet veto in the event selection.

The cut flow for the Higgs events is tabulated in TABLE 6.9 for the four investigated mass values between 400 GeV and 700 GeV. The background

samples are split among TABLE 6.10, TABLE 6.11 and TABLE 6.12. All listed tables are divided in three parts. First, the number of total events analyzed together with their assumed cross section is given. Second, the cut flow with respect to the number of events which show the typical characteristics of Higgs decays produced in weak boson fusion as mentioned in SUBSECTION 6.3.1 is listed. Third, the remaining events with their effective cross section and the number of expected events for an integrated luminosity of $60 \, \mathrm{fb}^{-1}$ is quoted. In case that no events survived the selection cuts, an upper limit at the 95 % confidence level on the expected contribution is given.

Table 6.11. The $t\bar{t}$ process is not expected to give any significant contribution to the background. To ensure that no contribution arises from these events, the samples which have been available in CMS are analyzed. No events pass the selection cuts and the upper limit at the 95 % confidence level on the expected number of events is therefore only a very pessimistic estimate.

$t\bar{t}+ N$ jets	$N = 0$	$N = 1$	$N = 2$	$N = 3$	$N \geq 4$
σ / [fb]	190,000	170,000	100,000	40,000	61,000
events analyzed	730,837	885,139	289,288	105,395	117,443
events with WBF topology	270	739	473	291	979
σ WBF topology / [fb]	70.19	141.93	163.50	110.44	508.49
lepton selection (a)	49.63%	55.62%	50.95%	56.70%	43.72%
$\mathrm{Z} \to ll$ cut (b)	0.37%	0.41%	0.42%	0.69%	1.02%
$\mathrm{Z} \to qq$ cut (c)	0.00%	0.00%	0.00%	0.34%	0.10%
$\tilde{\eta}$ cut (d)	0.00%	0.00%	0.00%	0.00%	0.00%
p_T balance (e)	0.00%	0.00%	0.00%	0.00%	0.00%
dijet mass cut (f)	0.00%	0.00%	0.00%	0.00%	0.00%
Higgs ΔR cut (g)	0.00%	0.00%	0.00%	0.00%	0.00%
remaining events	0	0	0	0	0
σ_eff / [fb]	0	0	0	0	0
weight for $60 \, \mathrm{fb}^{-1}$	15.6	11.52	20.74	22.77	31.16
exp. events for $60 \, \mathrm{fb}^{-1}$	≤ 47	≤ 35	≤ 62	≤ 68	≤ 93

6.5.2 Signal extraction

A variety of methods exists to express the probability of a putative signal in the presence of the background. To quantify this probability in general, the concept of statistical significance has been embedded consistently in

Table 6.12. Analyzed events of the diboson ALPGEN samples, $(Z \to f\bar{f}) + (Z \to l^+l^-) + N$ jets and $(W \to f\bar{f}') + (Z \to l^+l^-) + N$ jets, . The first boson (either W or Z) is allowed to decay inclusively, the second weak boson (Z in both cases) is forced to decay into charged lepton pairs ($l = $ e, μ, τ). ALPGEN 2.05 was patched to correct an error in the branching ratio of the second boson. In case that no simulated event passed the selection, an upper upper limit at the 95 % confidence level is quoted.

ZZ + N jets	$N = 0$	$N = 1$	$N = 2$	$N \geq 3$
σ / [fb]	1,450	680	270	230
events analyzed	417,605	198,317	243,525	44,912
events with WBF topology	1,337	1,229	3,247	1,987
σ WBF topology / [fb]	4.64	4.21	3.60	10.18
lepton selection (a)	78.38%	76.73%	76.87%	74.64%
$Z \to ll$ cut (b)	8.68%	15.95%	22.08%	32.21%
$Z \to qq$ cut (c)	2.62%	2.85%	3.05%	2.72%
$\tilde{\eta}$ cut (d)	1.42%	1.46%	2.06%	1.66%
p_T balance (e)	1.12%	0.90%	1.17%	0.50%
dijet mass cut (f)	0.15%	0.16%	0.46%	0.05%
Higgs ΔR cut (g)	0.15%	0.16%	0.43%	0.05%
remaining events	2	2	14	1
σ_{eff} / [fb]	0.01	0.01	0.02	0.01
weight for 60 fb^{-1}	0.21	0.21	0.07	0.31
exp. events for 60 fb^{-1}	≤ 1	≤ 1	≤ 1	≤ 1

WZ + N jets	$N = 0$	$N = 1$	$N = 2$	$N \geq 3$
σ / [fb]	1,700	1,140	610	831
events analyzed	17,000	12,000	227,660	258,892
events with WBF topology	44	63	2,867	12,001
σ WBF topology / [fb]	4.40	5.99	7.68	38.52
lepton selection (a)	84.09%	76.19%	77.75%	72.86%
$Z \to ll$ cut (b)	15.91%	12.70%	25.18%	33.76%
$Z \to qq$ cut (c)	2.27%	0.00%	3.56%	2.62%
$\tilde{\eta}$ cut (d)	2.27%	0.00%	2.48%	1.52%
p_T balance (e)	0.00%	0.00%	1.33%	0.53%
dijet mass cut (f)	0.00%	0.00%	0.21%	0.21%
Higgs ΔR cut (g)	0.00%	0.00%	0.17%	0.18%
remaining events	0	0	5	22
σ_{eff} / [fb]	0	0	0.01	0.07
weight for 60 fb^{-1}	6	5.7	0.16	0.19
exp. events for 60 fb^{-1}	≤ 18	≤ 17	≤ 1	≤ 4

a statistical framework. In general, a result is called significant if it is unlikely to have occurred by chance. The concept of statistical significance is formally introduced in the context of hypothesis testing. The significance level of a test is defined as the maximum probability, assuming the null hypothesis that the statistic would be observed. Hence, the significance level is the probability that the null hypothesis is rejected in error although it is true.

In high-energy physics, the significance S of observing a signal above an expected background distribution is counted in multiples of Gaussian standard deviations. This is motivated by the fact that S follows a standard Gaussian distribution with a mean of zero and standard deviation one. In this context, two different hypothesis are considered: the null hypothesis, assuming that the observed distribution is caused by background fluctuations only, and an alternative hypothesis, assuming the presence of signal events above a background distribution. In the following, the likelihood ratio approach as suggested in [86] is used to estimate the significance of the observed signal distribution and compared with the simple event counting method.

A binned likelihood fit is used to determine the contribution from signal and background events in the final invariant mass distribution. The observed distribution is modeled as a sum of signal p_s and background p_b probability density functions

$$f(M_{lljj}, m_0, \Gamma, \sigma) = N_b \times p_b(M_{lljj}) + N_s \times p_s(M_{lljj}, m_0, \Gamma, \sigma) \qquad (6.8)$$

with the free parameters N_s and N_b to determine the absolute level of signal and background. The signal distribution p_s is taken to follow a convolution of a Breit-Wigner (mean m_0, FWHM[2] Γ) with a Gaussian (mean m_0, standard deviation σ) accounting for the mass resolution smearing due detector effects. The dependence of the mass resolution and the position of its maximum on the Higgs boson mass is included and obtained from the detector simulation of the various signal samples (FIGURE 6.18).

The probability density function for the background p_b distribution is currently determined from fits to the invariant mass distribution which contains only background events. The function p_b is modeled as a sum of four

---

[2]Full width at half maximum: Difference between the two values of the independent variable at which the dependent variable is equal to half of its maximum.
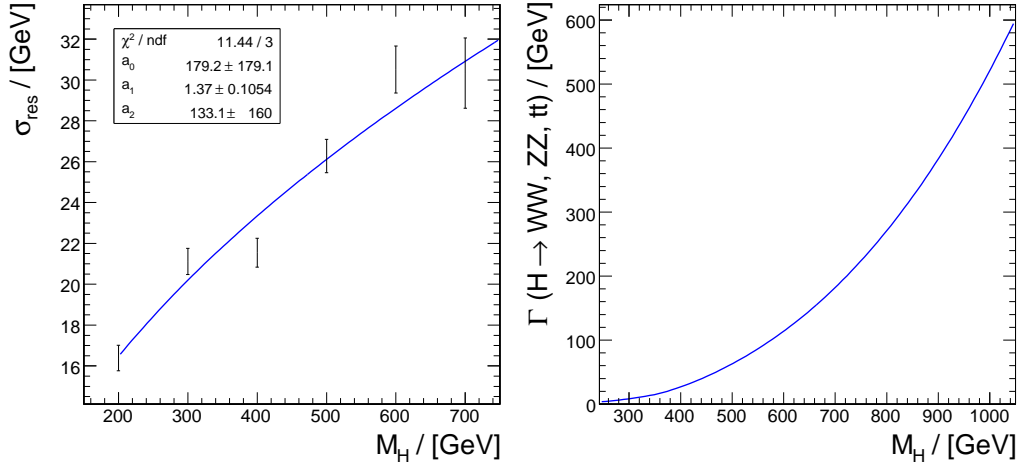
Figure 6.18. Parameterization of the Higgs boson mass resolution (left): The function $\sigma(M_\mathrm{H})^2 = p_0 + p_1 * (M_\mathrm{H} - p_2)$ is fitted to the mass resolution obtained from fast detector simulation. Calculation of the decay width as a function of the Higgs boson mass (right): Partial decay widths into weak boson and top pairs are added according to EQUATION 3.7, EQUATION 3.8 and EQUATION 3.9.

Gaussian distributions $\mathcal{N}_i$

$$p_\mathrm{b} = C \cdot \sum_{i=1}^{4} a_i \times \mathcal{N}_i(m_i, \sigma_i) \tag{6.9}$$

where a_i, m_i and σ_i are the parameters accounting for the amplitude, mean and width, respectively. The restriction of p_b to be normalized to unity with the normalization constant

$$C = \int_{m_{ZZ}}^{\infty} dx \left\{ \sum_{i=1}^{4} a_i \times \mathcal{N}_i(m_i, \sigma_i) \right\}^{-1} \tag{6.10}$$

results in 11 free parameters for the background modeling (FIGURE 6.19). In the final analysis, the background shape and its normalization can be extracted from collision data using relaxed selection cuts. The technique to get a reliable description of the background contribution from real data will be discussed in SUBSECTION 6.5.3.

In order to test for the existence of a signal, the modeled distribution (EQUATION 6.8) is fitted twice to the observed events. First, the different contributions for N_s and N_b are identified for the alternative hypothesis H_1, assuming signal in the presence of background. In the second fit with N_s fixed to zero the null hypothesis, assuming that the observed shape is
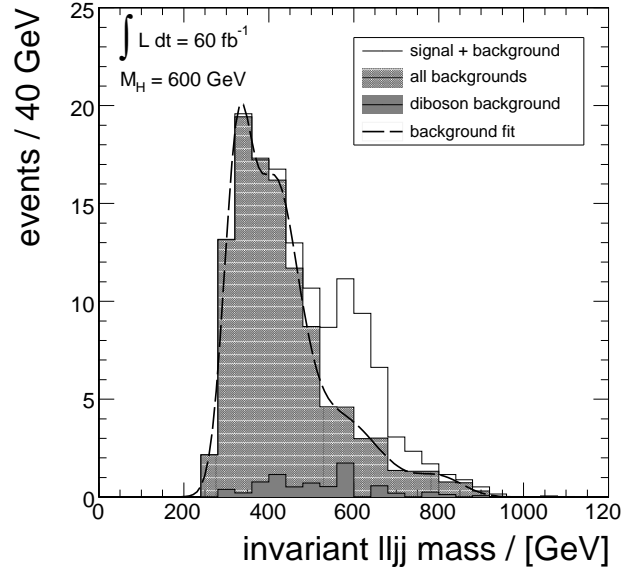
Figure 6.19. The different contributions to the background and its parameterization. Only a minor fraction of the overall background contribution (shaded histogram) arises from the WZ+jets and ZZ+jets events (dark histogram). The background is substantially dominated by Z+jets events. The shape of the background is very well described by the sum of four Gaussian distributions.

formed by background events only, is tested. This procedure explores the difference in shape between the signal and the background and is therefore not sensitive to uncertainties in the expected signal and background levels.

The likelihood ratio

$$Q = \frac{L_{s+b}}{L_b} \tag{6.11}$$

is used as a test statistic to distinguish between these two hypotheses. Here, L_{s+b} is the maximum likelihood value obtained in the full signal-plus-background fit, testing the alternative hypothesis H_1, and L_b is the maximum likelihood from the background-only fit, testing the null hypothesis H_0.

According to the theorem proved by S.S. Wilks [87], the quantity $-2 \ln Q$ is expected to follow a χ_n^2-distribution with the number of degrees of freedom n equal to the difference in the free parameters between the two likelihood fits. If x_i are n independent normally distributed random variables

with mean zero and variance one, then the random variable

$$X = \sum_{k=1}^{n} x_i^2 \qquad (6.12)$$

is distributed according to the χ_n^2-distribution. Thus, if the number of degrees of freedom n is one,

$$S_{\mathrm{L}} = \sqrt{-2 \ln Q} \qquad (6.13)$$

is directly the difference between the alternative hypothesis H_1 and the null hypothesis H_0 expressed in number of standard deviations. The condition that the difference of free parameters is one is obtained by fixing m_0, Γ and σ and the shape of the background in the fits. This justifies the choice of S_{L} to estimate the statistical significance for the supposed signal of the Higgs boson in the present final-state distribution.

The likelihood estimator can also be used to express the statistical significance for an excess of observed events N_{obs} in a defined signal region. The probability of finding exactly N events for an expected occurrence of λ events is given by the Poisson distribution

$$P(N; \lambda) = \frac{\lambda^N}{N!} \exp(-\lambda). \qquad (6.14)$$

Applying the likelihood-ratio to the Poisson probabilities for the alternative hypothesis $P(N_{\mathrm{obs}}; N_{\mathrm{s}} + N_{\mathrm{b}})$ and the null hypothesis $P(N_{\mathrm{obs}}; N_{\mathrm{b}})$ leads to

$$Q = \left(1 + \frac{N_{\mathrm{b}}}{N_{\mathrm{s}}} \right)^{N_{\mathrm{s}} + N_{\mathrm{b}}} \exp\left(-N_{\mathrm{s}}\right). \qquad (6.15)$$

Setting the expectation value of N_{obs} to $N_{\mathrm{s}} + N_{\mathrm{b}}$ gives the counting estimator for the significance definition based on the likelihood-ratio:

$$S_{\mathrm{cL}} = \sqrt{2 \left((N_{\mathrm{s}} + N_{\mathrm{b}}) \ln \left(1 + \frac{N_{\mathrm{s}}}{N_{\mathrm{b}}} \right) - N_{\mathrm{s}} \right)} \qquad (6.16)$$

In the limit of large numbers $N > 100$ for N_{s} and N_{b}, the Poisson probabilities can be replaced by Gaussian probabilities and EQUATION 6.16 results in

$$S_{\mathrm{cG}} = \sqrt{\frac{N_{\mathrm{s}}^2}{N_{\mathrm{b}}} + \ln \left(\frac{N_{\mathrm{b}}}{N_{\mathrm{s}} + N_{\mathrm{b}}} \right)}. \qquad (6.17)$$

For $N_b \gg N_s$ this simplifies to the widely used significance definition

$$S_c = \frac{N_s}{\sqrt{N_b}}. \tag{6.18}$$

Here, the definition of significance is illustrated once more. S_c expresses directly the excess of a number of events N_s on the statistical error of the background level $\sqrt{N_b}$ in Gaussian deviations.

Due to the small number of background events in the signal region, the counting significance estimator S_{cL} from EQUATION 6.16 is used to compare its significance expectation with the one obtained from the likelihood ratio. For the simple event counting method, the signal region in the mass spectrum is defined as $[m_0 - 1\sigma; m_0 + 1\sigma]$ where σ includes both the natural decay width and the mass resolution. The final mass spectra after all selection cuts together with the expected significances can be seen in FIGURE 6.20.

Table 6.13. Expected significance obtained from the log-likelihood based estimator S_L for different Higgs boson masses for the integrated luminosities of $60\,\text{fb}^{-1}$.

M_H / [GeV]	400	500	600	700
S_L for $60\,\text{fb}^{-1}$	3.16	4.27	4.26	4.07

TABLE 6.13 quotes the predicted significances for the observation of a heavy Higgs boson with the CMS experiment. The values are obtained from the log-likelihood ratio of the full signal-plus-background and the background-only fit. FIGURE 6.21 shows the dependence of the result of the signal-plus-background fit on the number of signal events. Since the distribution follows nicely a parabola between zero signal events $N_s = 0$ and the minimum $N_s = N_s^{\text{min}}$, the translation of the difference $2 \ln Q$ into Gaussian standard deviations is appropriate.

It should be stressed explicitly that the significances given here are the prediction for an expected signal at a forthcoming experiment and not obtained in a real experiment. In a real experiment the total number of events N_{obs} is measured. The number of signal events is determined as $N_s = N_{\text{obs}} - N_b$ and compared with the average expected number of background events N_b. In order to express the significance of the number of measured signal events, the fluctuation of the average number of backgrounds events must be taken into account which leads in the Gaussian limit to EQUATION 6.18.
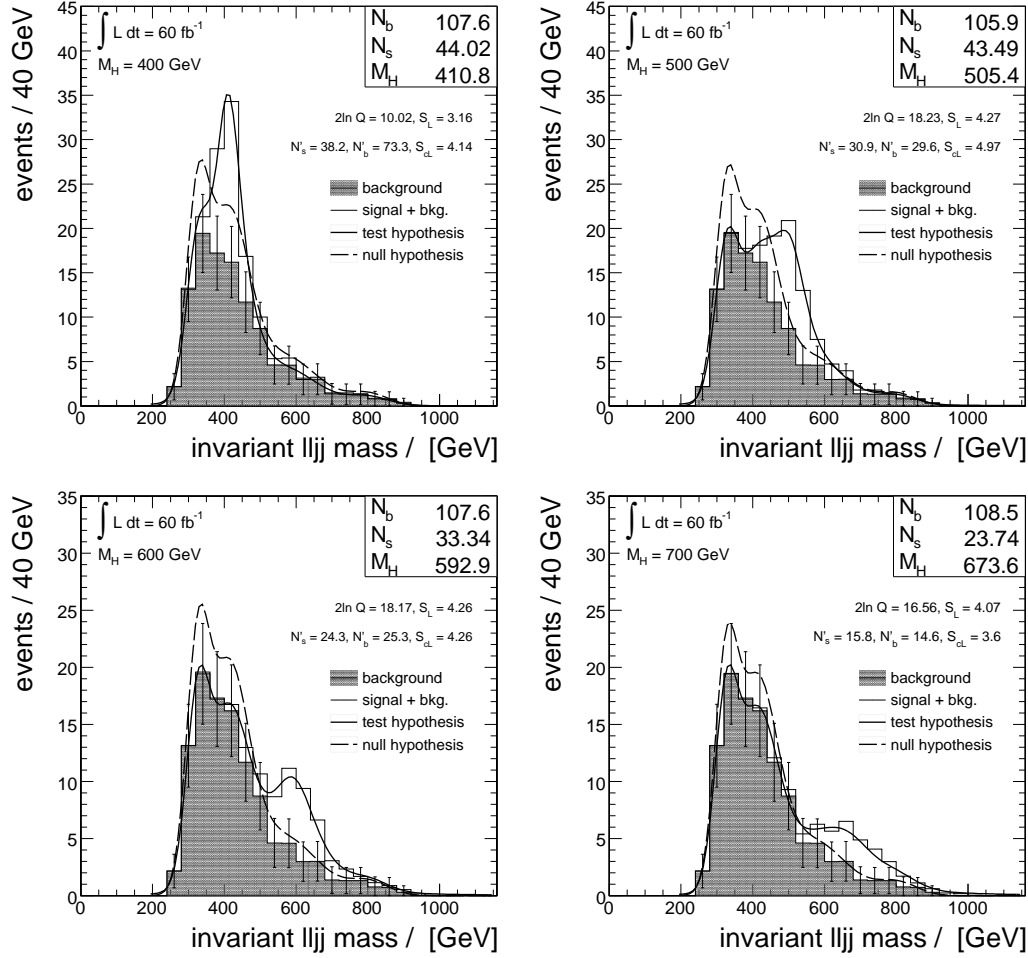
Figure 6.20. The invariant dilepton-dijet mass spectra after all selection cuts for four different Higgs boson masses between 400 and 700 GeV (open histograms) together with the background contributions (shaded histograms). The distributions are scaled to an integrated luminosity of $\mathcal{L} = 60\,\mathrm{fb}^{-1}$. The total number of expected events for the signal N_s and the background contributions N_b which result from the log-likelihood fit are written in the upper right corner of each diagram. The ratio Q of the two maximum likelihood values obtained in the full signal-plus-background fit (solid line) and the background-only fit (dashed line) is given below as $2\ln Q$ together with the corresponding significance S_L. For comparison, the counting estimator S_cL for the significance definition based on the likelihood-ratio is also given. S_cL is determined by counting the signal N'_s and background events N'_b separately in a 2σ mass window around the reconstructed Higgs boson resonance.
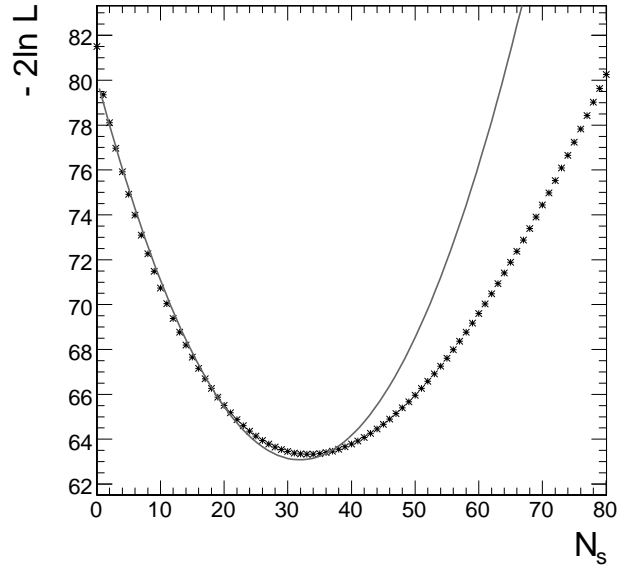
Figure 6.21. The dependence of the result $-2 \ln L_{s+b}$ of the log-likelihood fit. At each point on the curve, EQUATION 6.8 a was fitted to the mass distribution of FIGURE 6.19 for a fixed number of signal events N_s. The value of the likelihood at the minimum corresponds to the alternative hypothesis H_1, the value at the zero signal events corresponds to the null hypothesis H_0. Since the distribution follows nicely a parabola between $N_s = 0$ and $N_s = N_s^{\min}$, the difference $2 \ln Q$ can be directly translated into Gaussian standard deviations.

Here, the significances are average expected values reflecting the production rates and selection efficiencies according to the present knowledge. The treatment of the systematic and statistical uncertainties in the significance due to non-exact knowledge of production rates and unknown detector performance will be discussed in detail in SUBSECTION 6.5.4.

6.5.3 Evaluation of background from data

The precise understanding of the background contribution is a crucial issue concerning the search for new phenomena. The direct use of Monte Carlo predictions, i.e to determine the total number of background events

$$N_b = L \times \sigma_b \times \epsilon_b, \tag{6.19}$$

where ϵ is acceptance of the event selection for background events, leads to high systematic uncertainties due to theoretical calculations and to experimental uncertainties (SUBSECTION 6.5.4). The most reliable approach

to address this problem is to measure the background directly from real data. The commonly used method is to extrapolate the background contribution from signal-free phase space regions ("sidebands") to the signal region.
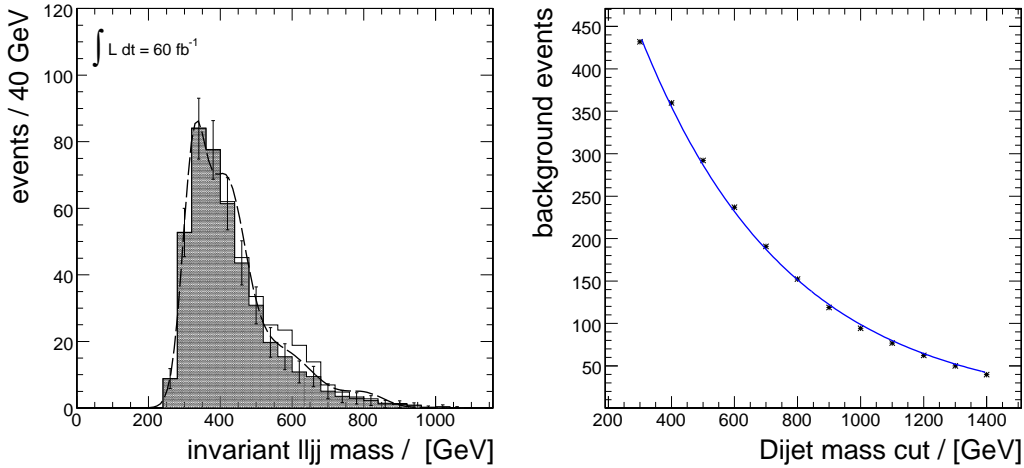
Figure 6.22. The invariant lljj mass spectrum at a second working point (left) using a relaxed dijet mass cut. The Higgs boson resonance (open histogram) on top of the background (shaded histogram) is almost negligible and the invariant dilepton-dijet spectrum allows a reliable determination of the background shape (dashed line). Tightening the cut on the invariant mass of the two forward jets results in an exponential decrease of background events which survive the selection (right).

In this analysis such a control region cannot be identified in the background distribution. However, the background shape and its normalization can by obtained from an invariant mass distribution where the background processes are enhanced. The selection cut on the invariant mass of the two forward jets (dijet mass cut, see SUBSECTION 6.5.1) can be relaxed insofar as the signal peak from the Higgs boson mass gets relatively small and can therefore be neglected in the overall mass spectrum.

The left-hand plot of FIGURE 6.22 shows the overall invariant mass distribution (background and signal events, Higgs boson mass of 600 GeV) for a relaxed forward dijet mass cut of $m_{\text{dijet}} > 300\,\text{GeV}$. This spectrum where the background processes are enriched allows the determination of the total number of background events at the second working point "loose cuts". The evolution of the background contribution as a function of the dijet mass cut is demonstrated in the right-hand plot of FIGURE 6.22. The background reduction follows a smooth decreasing exponential function
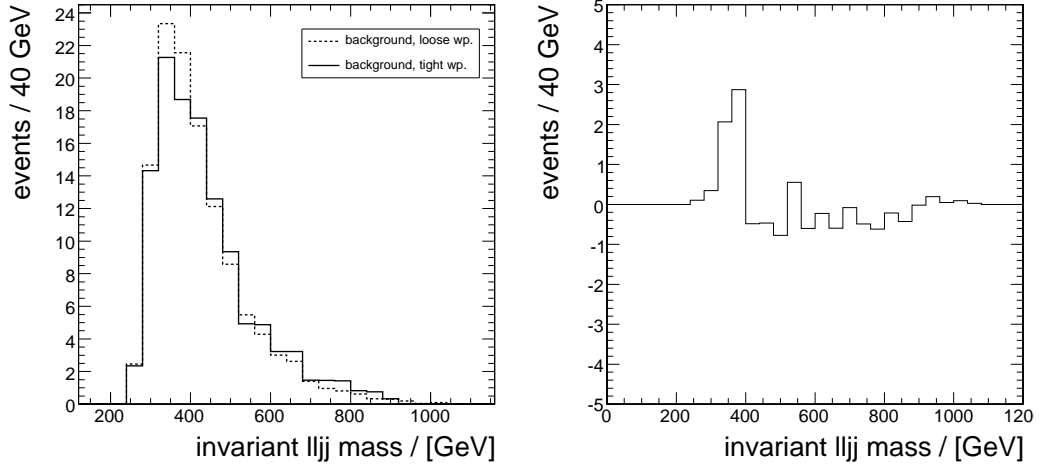
Figure 6.23. The background shapes (left) and the difference in the shape at both working points applying loose and tight selection cuts. The distributions are normalized to the number of events expected at the tight working point.

and therefore allows the reliable calculation of a reduction factor α from Monte Carlo events for both the two working points applying "loose" and "tight" selection cuts:

$$\alpha_{\mathrm{MC}} = \frac{N_{\mathrm{MC}}^{\mathrm{tight}}}{N_{\mathrm{MC}}^{\mathrm{loose}}}. \tag{6.20}$$

In contrast to the prediction of the exact cross sections for the background processes, the background shape can be predicted convincingly. This justifies the calculation of a reduction factor from Monte Carlo events.

The number of background events contributing to the measured distribution is then calculated as

$$N_{\mathrm{data}}^{\mathrm{tight}} = \alpha_{\mathrm{MC}} \cdot N_{\mathrm{data}}^{\mathrm{loose}}. \tag{6.21}$$

The background enriched mass spectra can also be used to measure the background shape at the second working point from collision data. The difference in the background shape for both working points is illustrated in FIGURE 6.23. The influence on the falling spectrum is rather small if the shape from the loose working point is scaled by the factor α to the tight working point.

The uncertainty in the number of background events extrapolated to the

signal region is given by

$$(\Delta N_b)^2 = (\Delta N_b^{\text{stat}})^2 + (\Delta N_b^{\text{shape}})^2 \tag{6.22}$$

where

$$\Delta N_b^{\text{stat}} = \alpha_{\text{MC}} \sqrt{N_b^{\text{loose}}}$$

is the statistical uncertainty in the number of events at the loose working point and N_b^{shape} is the uncertainty in the background extrapolation. The effect of the background uncertainty on the overall significance is quantified in SUBSECTION 6.5.4 and included in the total error on the expected significance.

6.5.4 Systematic uncertainties

As briefly mentioned in SUBSECTION 6.5.2, the average expected significance is only a mean value which shows an error. The theoretical and experimental uncertainties affecting the significance of this analysis are covered in this section. In order to evaluate the effect of systematic and statistical uncertainties in the average expected significance, the counting estimator S_c is used. The error estimation based on EQUATION 6.18 benefits from the fact that it can be expressed analytically. However, S_c slightly overestimates the significance at small event numbers (FIGURE 6.24).

Two cases of systematic errors influencing the prediction of the average expected significance

$$< S_c > = \frac{N_s}{\sqrt{N_b}}. \tag{6.23}$$

must be distinguished:

A) If systematic uncertainties on the background distribution affect the determination of the absolute number of the signal events, the theoretical uncertainty ΔN_b must be convoluted with the statistical error on the background $\sqrt{N_b}$. If it is assumed that ΔN_b follows a Gaussian distribution, the denominator of EQUATION 6.23 has to be replaced with the quadratic sum of both terms:

$$< S_c > = \frac{N_s}{\sqrt{N_b + (\Delta N_b)^2}}. \tag{6.24}$$
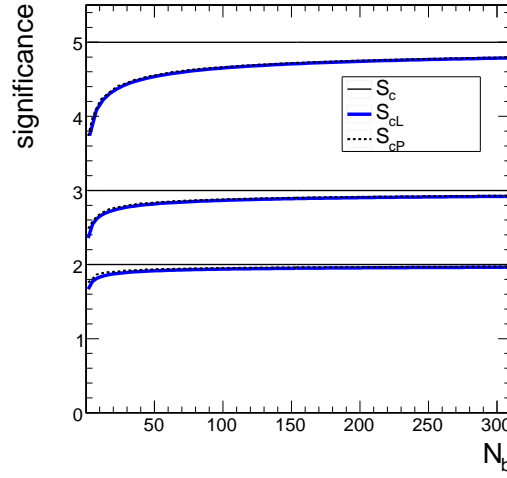
Figure 6.24. Comparison of the different significance estimators as a function of background events N_b. The number of signal events is taken as $N_s = S_c \times N_b$, hence the solid lines correspond to fixed values of S_c. S_{cL} agrees perfectly with S_{cP} which is calculated from the Gaussian tail for Poisson probabilities. S_c slightly overestimates the significance at small numbers ($N_b < 50$). Modified version taken from [88].

B) If the background contribution can be determined from measured data, the systematic error on the background influences only the predicted average expected significance. To take this case into account, error propagation can be applied to the formula above and $< S_c >$ reads

$$< S_c >= \frac{N_s}{\sqrt{N_b + (\Delta N_b)^2}} \pm \Delta S_c, \tag{6.25}$$

where ΔS_c is a function of ΔN_s and ΔN_b.

To express the error on the expected average significance, the uncertainties in the background need to be classified in errors either of type A or B. Systematic errors associated with detector measurements, luminosity measurements, selection efficiencies and cross section predictions are of type B since the absolute background contribution can be estimated from measured data. On the contrary, the errors on the background shape and the statistical fluctuation on the number of events affecting the calculation of the reduction factor α cannot be determined from measured data and are therefore type A errors. Uncertainties in the signal prediction are not critical for establishing an excess of events over the background. These

uncertainties influence only the ability to predict the expected significance and are always errors of type B.

The statistical error in the background prediction is given by the number of Monte Carlo events at the loose working point and yields 5%. The error in the background shape depends on the Higgs boson mass and varies between 12% at 400 GeV and 0% at the cross-over point at 600 GeV. A general uncertainty in the background shape of 10% is used in the following. Since the cross section for the background processes is known at leading order only, an overall uncertainty of 30% is assumed. The theoretical errors in the prediction of the expected number of Higgs boson events are dominated by uncertainties in the cross section. The statistical error in the selection efficiency is negligible due to the large number of Monte Carlo events which are analyzed. Two sources of uncertainties enter the cross section calculation of Higgs boson production in weak boson fusion: The parton density functions of the interacting particles and the definition of the Q^2 scale of the hard process. These effects have been investigated by the CMS collaboration and are taken into account in the following. The k-factor which is defined as the ratio of the cross section evaluated at next-to-leading (NLO) and leading order (LO)

$$ k = \frac{\sigma_{\mathrm{NLO}}}{\sigma_{\mathrm{LO}}}. \tag{6.26} $$

is taken as the error on the cross section itself. Higher order calculations (NNLO) are not supposed to give higher contributions to the cross sections than the correction at first order (NLO). The k-factors were determined with VBFNLO [89].

Apart from theoretical errors, there are uncertainties associated with detector measurements. In the following, the most dominant sources of errors contributing to analyses carried out in the period of first data taking up to $10\,\mathrm{fb}^{-1}$ are evaluated:

- The jet energy scale uncertainty (JES) will be approximately 3% for jets with a transverse momentum higher than $p_{\mathrm{T}} > 50\,\mathrm{GeV}$. This accuracy can be achieved by applying a W boson mass calibration in top quark pair production events [53]. The hadronic jets emerging from W boson decays feature a mean transverse momentum of 50 GeV. In the region excluded from the W boson mass calibration, the jet energy scale uncertainty will linearly increase to about 10% down to 20 GeV set by the γ+jet calibration [52]. To take the jet

energy scale into account, the reconstructed jet energy is shifted to lower and higher energies and the influence on the final mass spectrum is evaluated. The applied procedure follows the CMS prescription [90] which recommends to rescale the four-momentum of the reconstructed jet

$$p^{\mu}_{\text{scaled}} = (1 \pm \alpha) \cdot p^{\mu}_{\text{rec}}$$

where α is the uncertainty for different p_T ranges:

$$\alpha = \left\{ \begin{array}{ll} 10\% & p_T < 20\,\text{GeV} \\ 10\% - 7\% * (p_T - 20\,\text{GeV})/30\,\text{GeV} & 20 < p_T < 50\,\text{GeV} \\ 3\% & p_T > 50\,\text{GeV} \end{array} \right. .$$

- In addition to jet energy scale uncertainties, there are uncertainties in the jet resolution. It is expected that the determination of the dijet balancing resolution will limit the uncertainty in the jet resolution to 10%. Therefore, each reconstructed jet is smeared by

$$p^{\mu}_{\text{smeared}} = p^{\mu}_{\text{rec}} \cdot \mathcal{N}(1, 0.1),$$

  where the random numbers follow a Gaussian distribution $\mathcal{N}$ with mean one and standard deviation 0.1 reflecting the error in energy resolution of the hadronic calorimeter.

- The precision of the luminosity measurement is expected to be of about 3 % for 30 fb^{-1} based on W and Z boson measurements.

- The uncertainty in the muon and electron reconstruction efficiency is estimated to be around 1% [91, 92]. Another 1% arises in case of muons from the efficiency of the isolation cut. The uncertainty for the efficiency of the transverse impact parameter cut is assumed to be also around 1%. These uncertainties in the lepton selection accumulate to less than 2.5% and are therefore neglected.

TABLE 6.14 summarizes the errors in the predicted significance taking into account all sources of uncertainties listed above. In order to quantify the effect of the uncertainties of type A in the background prediction, S_L as estimated in SUBSECTION 6.5.2 is scaled by a factor f (see EQUATION 6.24):

$$S'_L = f \cdot S_L, \qquad f = \sqrt{\frac{N_b}{N_b + (\Delta N_b)^2}}. \qquad (6.27)$$

Table 6.14. Discovery potential of CMS for the investigated decay channel including all systematical and statistical errors.

M_{H} / [GeV]	**400**	**500**	**600**	**700**
N_{s} exp. for $60\,\mathrm{fb}^{-1}$	38.2	30.9	24.3	15.8
N_{b} exp. for $60\,\mathrm{fb}^{-1}$	73.3	29.6	25.3	14.6
S_{c} (Gaussian)	4.46	5.68	4.83	4.14
S_{cL} (Poisson)	4.14	4.97	4.26	3.60
S_{L} ($2\ln Q$)	**3.16**	**4.27**	**4.26**	**4.07**
type A errors				
$\Delta N_{\mathrm{b}}^{\mathrm{stat}}$	5 %			
$\Delta N_{\mathrm{b}}^{\mathrm{shape}}$	10 %			
S_{L}'	**2.28**	**3.65**	**3.71**	**3.74**
type B errors $\Delta S/S$				
σ_{WBF} (PDFs)	3.10%	3.30%	3.70%	4.00%
σ_{WBF} (Q^2 scale)	1.10%	0.70%	1.00%	1.50%
σ_{WBF} (k-factor)	4.20%	4.10%	4.30%	4.20%
$\sigma_{\mathrm{backgrounds}}$	15.00%	15.00%	15.00%	15.00%
jet energy scale	13.23%	14.61%	13.04%	12.59%
jet resolution	0.00%	0.11%	0.07%	0.14%
luminosity	0.10%	0.05%	0.05%	0.05%
total ΔS	20.70%	21.60%	20.69%	20.48%
significance S	**2.28 $\pm$ 0.67**	**3.65 $\pm$ 1.05**	**3.71 $\pm$ 0.87**	**3.74 $\pm$ 0.78**

All uncertainties which arise from unknown detector performance affect both the signal and the background distribution in the same way and are thus correlated. The re-calculated significance of the final mass spectra is obtained after applying the variations on both the signal and background events and counting the events in the signal region assuming Gaussian statistics:

$$S_{\mathrm{c}} = \frac{N_{\mathrm{s}}}{\sqrt{N_{\mathrm{b}}}}, \qquad S_{\mathrm{c}}' = \frac{N_{\mathrm{s}}'}{\sqrt{N_{\mathrm{b}}'}}, \qquad \Delta S = \frac{S_{\mathrm{c}}' - S_{\mathrm{c}}}{S_{\mathrm{c}}}. \tag{6.28}$$

The derived variation is taken as the relative error ΔS on the significance S_{L}'.

6.6 Results

The channel studied here is of particular interest during the LHC operation at low luminosity. The weak boson channels might suffer substantially under high luminosity running conditions from pile-up in the forward region of the detector making forward jet tagging inefficient.

FIGURE 6.25 presents the evaluated observability of the studied Higgs boson decay channel with the CMS experiment. CMS can detect a heavy Higgs boson with a mass larger than 500 GeV after data from LHC collisions corresponding to an accumulated luminosity of $60\,\mathrm{fb}^{-1}$ have been taken. A Higgs boson with a mass between $500 < M_{\mathrm{H}} < 700\,\mathrm{GeV}$ should yield in a significant excess of events in the expected invariant mass distribution and its resonance on top of a rapidly falling mass spectra should be clearly visible. For a Higgs boson with a mass of 600 GeV, the investigation of this decay channel should lead to the "weak evidence" of a heavy Higgs boson. Assuming that the reconstructed collision data will only correspond to $30\,\mathrm{fb}^{-1}$ after three years of LHC operation, this decay channel can confirm the discovery of a heavy Higgs boson in complement to the Gold Plated channel which features four isolated leptons in the final-state (FIGURE 6.1).
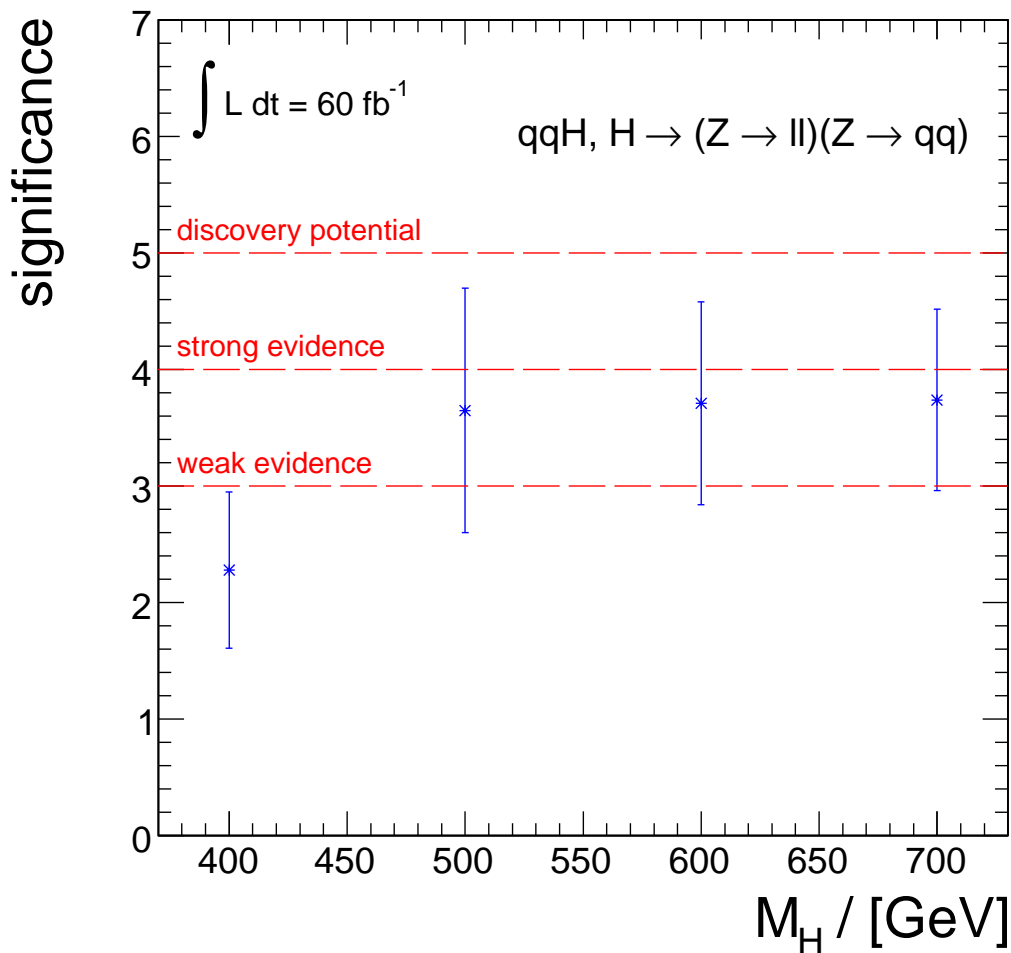
Figure 6.25. Observability of a heavy Higgs boson produced in weak boson fusion at the LHC in the decay channel H → $l^+l^-q\bar{q}$, investigated for the CMS experiment. The curve shows the average expected significance together with its error as a function of the Higgs boson mass. The dashed horizontal lines indicate the different significance levels (the statistical expressiveness) in terms of traditional frequentist hypothesis testing.

Chapter 7

Summary and Conclusion

The Higgs mechanism is a cornerstone of the Standard Model of particle physics. The discovery of the Higgs boson and the detailed investigation of weak boson scattering at highest energies are main ambitions of the CMS experiment at the Large Hadron Collider.

The potential of the CMS experiment to discover a Higgs boson decaying via two Z bosons subsequently into two quarks and either two muons or two electrons has been investigated in this thesis. Since the cross section for Higgs boson production decreases for larger Higgs boson masses, the detection of the Higgs boson in the high-mass range is difficult to explore by only considering four-leptons final-states. Although the decay into four isolated leptons shows the cleanest signature in the detector, it suffers from its low branching ratio in the high-mass region. The branching ratio of the Higgs boson decaying into a quark and a charged lepton pair is about 20 times larger than into four leptons. It therefore provides a promising alternative to the latter channel.

While gluon fusion dominates the Higgs boson production at the LHC, the weak boson fusion as second largest contribution becomes more and more competitive as the Higgs boson mass increases. The distinct features of Higgs boson production via electroweak boson fusion can be exploited to separate the Higgs decays from background contributions. The most important characteristic of the event topology is the presence of the accompanying jets originating from the incoming initial-state quarks which emit the electroweak bosons. The CMS detector provides a good hermetical coverage to detect these accompanying jets even at large pseudorapidities.

To express quantitatively the observability of a heavy Higgs boson in this particular decay channel, an analysis strategy to extract the Higgs boson decays from background processes was developed. To accomplish this task, hypothetical physics events for proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 14\,\text{TeV}$ were generated using Monte Carlo techniques. These events were processed using a parameterized simulation of the particle interactions with the detector material taking the response of the trigger system into account. Realistic reconstruction algorithms as they will be applied to collision data were used to reconstruct the trajectories of muons, electrons and the calorimeter response to hadronic particle jets in the presence of pile-up events. The results obtained in the fast detector simulation were verified with the full detector simulation. It was found that the reconstruction efficiency of the full Higgs boson event with respect to all produced Higgs bosons showing the investigated topology is limited to 20% by the CMS detector performance.

Performing a physics analysis in the commissioning phase of an experiment means also to gain valuable experience in order to to be well prepared for the start-up phase. Several issues have been discovered in the software framework and reported to the responsible developers. About 40 million hypothetical physics events had to be generated and simulated on the LHC Computing Grid. The derived operational experience will help to establish a Tier-2 center for the LHC Computing Grid at Karlsruhe University (TH) which is meant to provide local physicists direct access to the data taken by the CMS detector.

The suppression of background events that produce two isolated leptons together with additional hadronic jet activity was achieved by selecting only events whose topology showed well-specified characteristics. For this purpose, an optimization of the event selection based on kinematical quantities was studied. A fitting approach was presented to extract the number of events arising from Higgs boson decays from the final invariant mass spectra. The efficiency of the event selection of Higgs boson decays is expected to be 5% with respect to all Higgs decays into the studied final-state.

The precise understanding of the background contribution is a crucial issue concerning the search for Higgs decay events. Therefore, a method is provided to measure the total number of background events and the shape of their distribution directly from collision data.

Finally, the expected significance for the discovery of a heavy Higgs bo-

son in the evaluated decay channel with respect to theoretical and experimental uncertainties was expressed. It was shown that after the low-luminosity phase of the LHC ($\mathcal{L} = 60\,\mathrm{fb}^{-1}$), a heavy Higgs boson with a mass between $500 < M_\mathrm{H} < 700\,\mathrm{GeV}$ should yield a significant excess of events in the studied decay channel. The resonance of a heavy Higgs boson on top of the rapidly falling mass spectra will be clearly visible and will lead to the evidence of its existence with a statistical significance of more than 3σ. Even with a lower data volume ($\mathcal{L} \approx 30\,\mathrm{fb}^{-1}$), this decay channel can confirm the discovery of a heavy Higgs boson in the earlier mentioned four-lepton final-state in a complementary way. Once the Higgs boson is discovered, the coupling of the Higgs boson to the electroweak bosons can be determined. This will facilitate mandatory consistency checks with the predictions of the Standard Model in order to verify or even falsify the Higgs mechanism.

Appendix A

Discriminating Variables
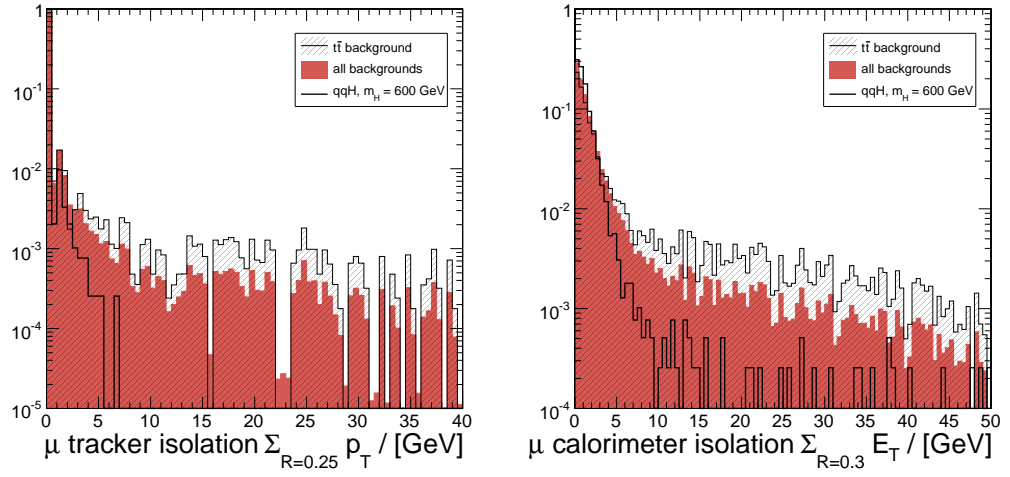
Figure A.1. Normalized distributions for the muon isolation measured in tracker (left) and calorimeter (right).
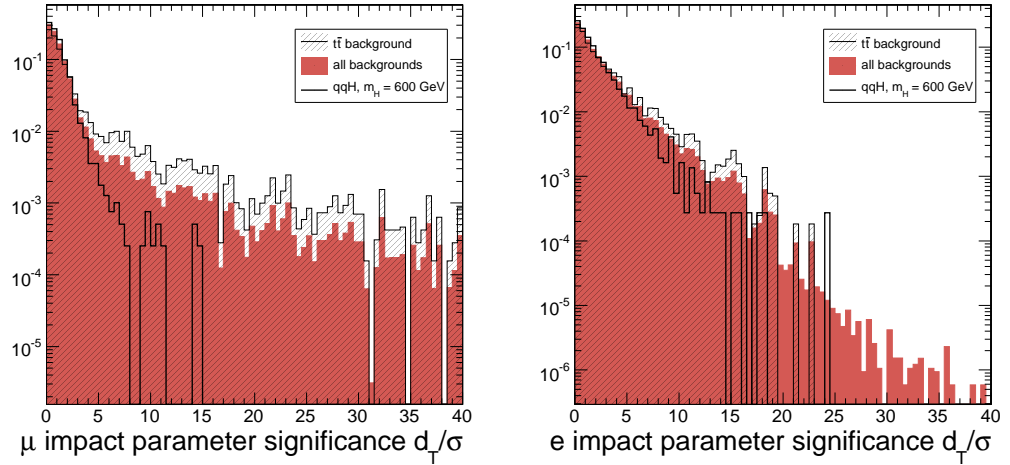
Figure A.2. Normalized distributions for the transverse impact parameter significance calculated for muons (left) and electrons (right).
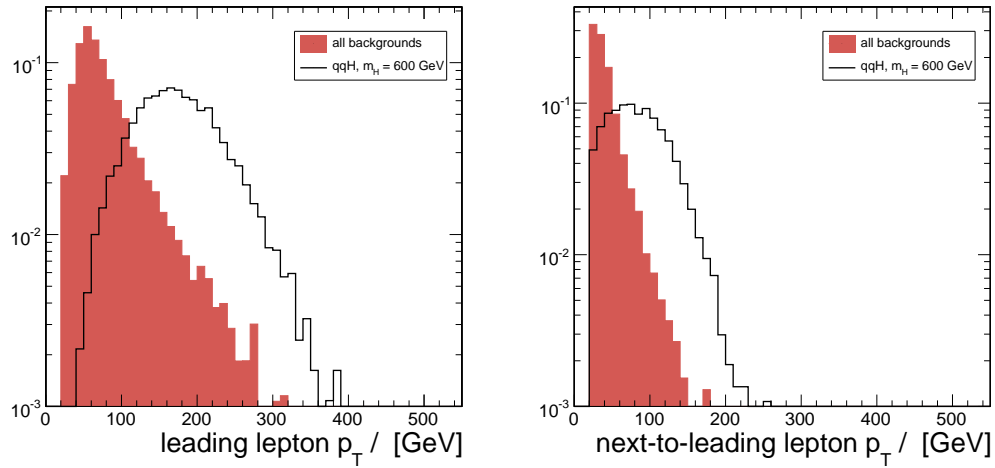
Figure A.3. Normalized distributions for the transverse momentum of the leading (left) and next-to-leading centrally produced lepton.
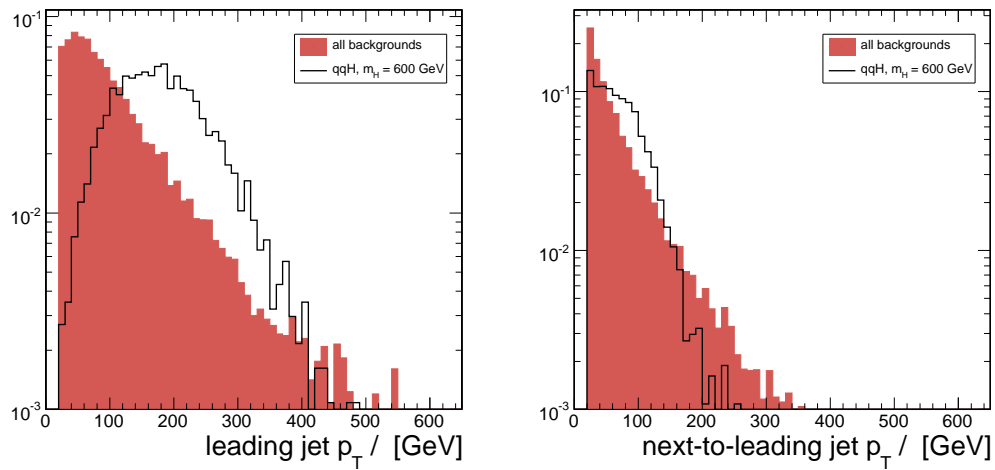
Figure A.4. Normalized distribution for the transverse momentum of the leading and next-to-leading centrally produced jet.
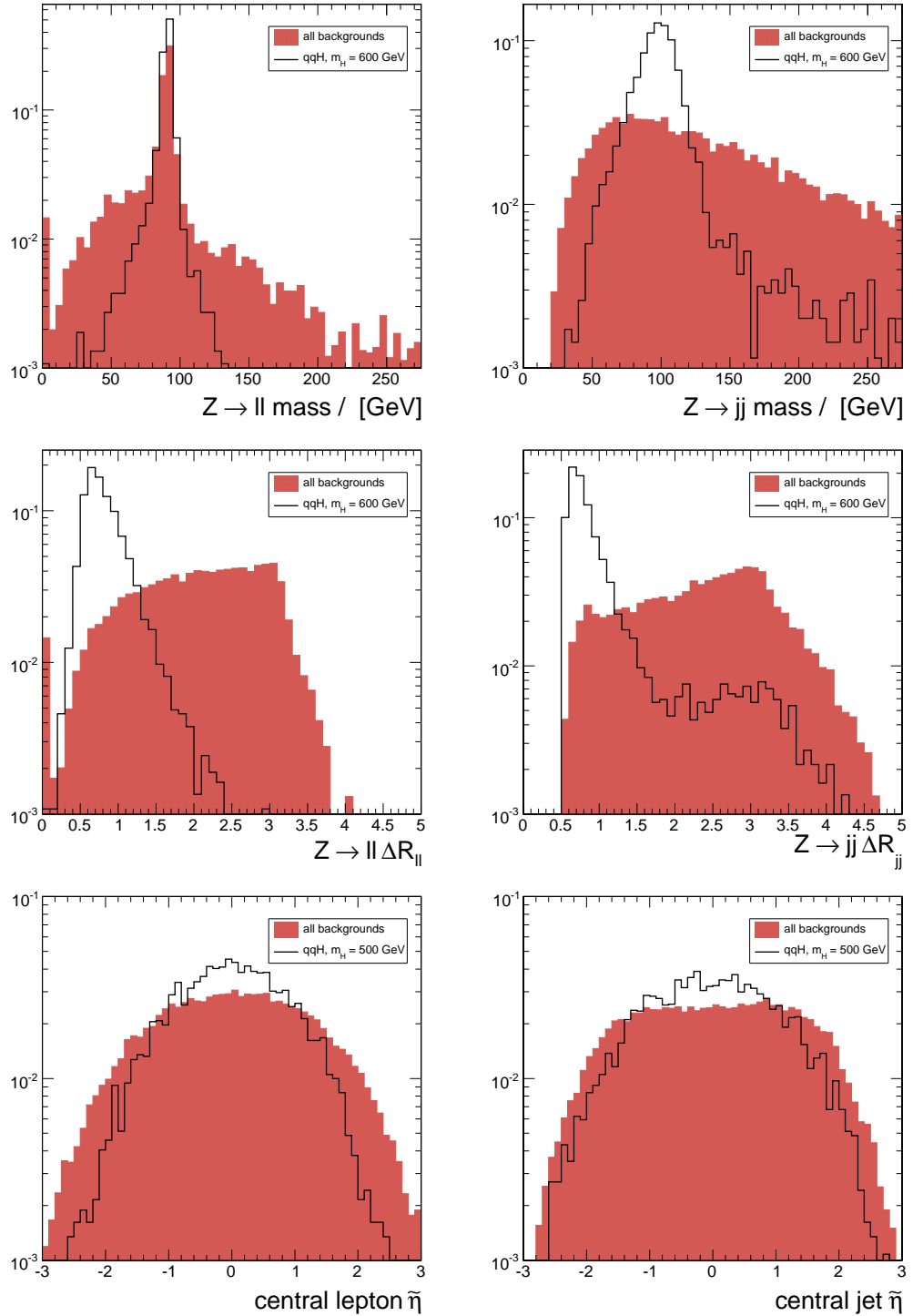
Figure A.5. Normalized distributions for the invariant mass, the opening angle and the shifted pseudorapidity for the two centrally produced leptons (left) and the two centrally produced jets (right).
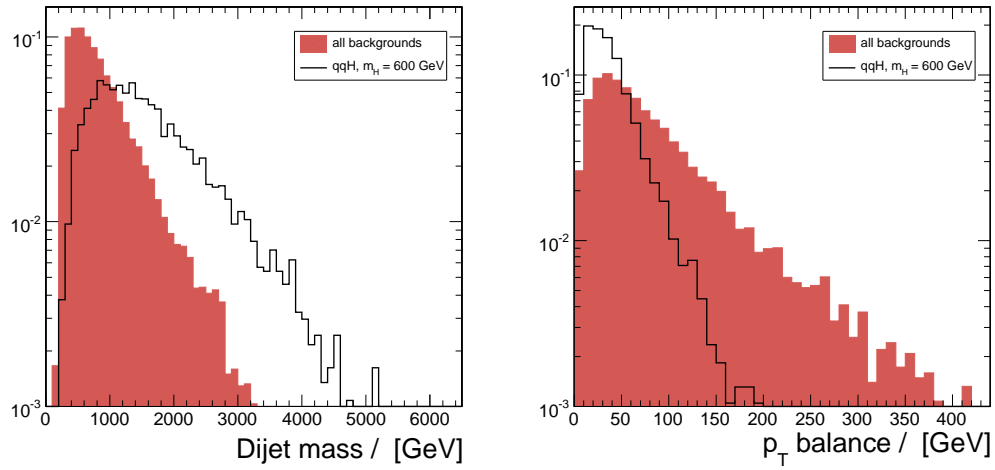
Figure A.6. Normalized distributions for the invariant mass of the two forward jets (right) and the balance of the transverse momenta of the final-state paricles (right).
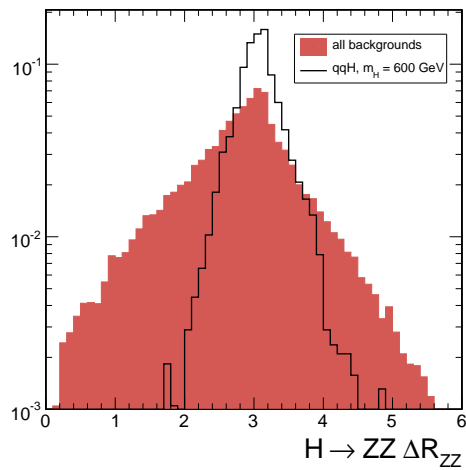
Figure A.7. Normalized distributions for the angular distance of the two Z bosons originating from the Higgs boson decay.

Appendix B

Monte Carlo Event Generation

B.1   PYTHIA parameters

```
C
C PYTHIA Particle Mass
C
PMAS 5,1 = 4.2         !mass of b quark
PMAS 6,1 = 175.        !mass of top quark
PMAS 23,1 = 91.187     !mass of Z
PMAS 24,1 = 80.22      !mass of W
PMAS 25,1 = 200.0      !mass of Higgs
C
C PYTHIA Process Selection
C
MSEL = 0               !full user control
MDME 174,1 = 5         !Z decay into d dbar
MDME 175,1 = 5         !Z decay into u ubar
MDME 176,1 = 5         !Z decay into s sbar
MDME 177,1 = 5         !Z decay into c cbar
MDME 178,1 = 5         !Z decay into b bbar
MDME 179,1 = 5         !Z decay into t tbar
MDME 180,1 = 0         !Z decay into b' b'bar
MDME 181,1 = 0         !Z decay into t' t'bar
MDME 182,1 = 4         !Z decay into e- e+
MDME 183,1 = 0         !Z decay into nu_e nu_ebar
MDME 184,1 = 4         !Z decay into mu- mu+
MDME 185,1 = 0         !Z decay into nu_mu nu_mubar
MDME 186,1 = 0         !Z decay into tau- tau+
MDME 187,1 = 0         !Z decay into nu_tau nu_taubar
MDME 188,1 = 0         !Z decay into tau'- tau'+
```

```
MDME 189,1 = 0        !Z decay into nu'_tau nu'_tau
MDME 210,1 = 0        !Higgs decay into dd
MDME 211,1 = 0        !Higgs decay into uu
MDME 212,1 = 0        !Higgs decay into ss
MDME 213,1 = 0        !Higgs decay into cc
MDME 214,1 = 0        !Higgs decay into bb
MDME 215,1 = 0        !Higgs decay into tt
MDME 218,1 = 0        !Higgs decay into e nu e
MDME 219,1 = 0        !Higgs decay into mu nu mu
MDME 220,1 = 0        !Higgs decay into tau nu tau
MDME 222,1 = 0        !Higgs decay into g g
MDME 223,1 = 0        !Higgs decay into gam gam
MDME 224,1 = 0        !Higgs decay into gam Z
MDME 225,1 = 1        !Higgs decay into Z Z
MDME 226,1 = 0        !Higgs decay into W W
MSTJ 11 = 3           !Choice of the fragmentation function
MSTJ 22 = 2           !Decay those unstable particles
MSTP 2 = 1            !Which order running alphaS
MSTP 33 = 0           !Inclusion of K factors
MSTP 51 = 7           !Structure function chosen
MSTP 81 = 1           !Multiple parton interactions
MSTP 82 = 4           !Defines the multi-parton model
MSTU 21 = 1           !Check on possible errors during execution
MSUB 123 = 1          !ZZ fusion
MSUB 124 = 1          !WW fusion
PARJ 71 = 10.         !For which ctau  10 mm
PARP 82 = 1.9         !pt cutoff for multiparton interactions
PARP 89 = 1000.       !Sqrts for which PARP82 is set
PARP 84 = 0.4         !Multiple interactions: matter distribution
PARP 90 = 0.16        !Multiple interactions: rescaling power
C
C GENERATOR
C
KSEL = 0
NSEL = 500            !Maximal number of subevents per AA event
TRIG = 1000000        !Maximum number of tries
ECMS = 14000.         !Energy per nucleon
```

B.2   ALPGEN parameters

```
------
hard process code (not to be changed):
ihrd= 7
------
Select pp (1) or ppbar (-1) collisions:
ih2= 1
------
beam energy in CM frame (e.g. 7000 for LHC):
ebeam=  7000.
------
parton density set:
ndns= 5
NDNS      Set         Lambda_4    Lambda_5_2loop    Scheme

   1      CTEQ4M        .298         .202                 MS
   2      CTEQ4L        .298         .202                 MS
   3      CTEQ4HJ       .298         .202                 MS

   4      CTEQ5M        .326         .226  (as=0.118)  MS
   5      CTEQ5L *      .192         .144  (asLO=0.127)MS
   6      CTEQ5HJ       .326         .226  (as=0.118)  MS

   7      CTEQ6M        .326         .226  (as=0.118)  MS
   8      CTEQ6L        .326         .226  (as=0.118)  MS
   9      CTEQ6L1       .215         .165  (asLO=0.130)MS
10-50     CTEQ6xx       .326         .226  (as=0.118)  MS

 101      MRST99 COR01 .321          .220                 MS

 102      MRST2001      .342         .239  (as=0.119 ) MS
 103      MRST2001      .310         .214  (as=0.117 ) MS
 104      MRST2001      .378         .267  (as=0.121 ) MS
 105      MRST2001J     .378         .267  (as=0.121)  MS
 106      MRST2002LO * .22           .167  (asLO=0.13) MS

  PDF sets followed by * are obtained from a 1-loop analysis,
  and the relative values of Lambda  refer to 1-loop.
  The MSbar scheme is used by default with 1-loop
  structure functions.
  In all cases the values of Lambda and loop order are set
  automatically by the code, The user only needs to input ndns

  ------
 scale option (process dependent):
 iqopt= 1
```

```
Options for Factorization/renormalization scale Q:
iqopt=0 => Q=qfac
iqopt=1 => Q=qfac*sqrt{sum(m_tr^2)}
iqopt=2 => Q=qfac*sqrt(x1*x2*S)
where:
- m_tr^2=m^2+pt^2, summed over heavy quarks and light jets
------
Q scale rescaling factor:
qfac=  1.
------
number of light jets:
njets= 0
------
heavy flavour type for procs like WQQ, ZQQ, 2Q, etc
ihvy= 5
------
2nd heavy flavour type for procs like 4Q:
ihvy2= 5
------
charm mass:
mc=  0.
------
bottom mass:
mb=  4.7
------
top mass:
mt=  175.
------
minimum pt for light jets:
ptjmin=  20.
------
ptmin for bottom quarks (in procs with explicit b):
ptbmin=  20.
------
ptmin for charm quarks (in procs with explicit c):
ptcmin=  20.
------
max|eta| for light jets:
etajmax=  5.
------
max|eta| for b quarks (in procs with explicit b):
etabmax=  5.
------
max|eta| for c quarks (in procs with explicit c):
etacmax=  5.
------
min deltaR(j-j), deltaR(Q-j) [j=light jet, Q=c/b]:
drjmin=  0.699999988
------
```

```
min deltaR(b-b) (procs with explicit b):
drbmin=  0.699999988
------
min deltaR(c-c) (procs with explicit charm):
drcmin=  0.699999988
------
first random number seed (5-digit integer):
iseed1= 12345
------
second random number seed (5-digit integer):
iseed2= 67890
------
first random number seed for unweighting (5-digit integer):
iseed3= 12345
------
second random number seed for unweighting (5-digit integer):
iseed4= 67890
```

B.3   Higgs boson cross sections

```
From Michael.Spira@psi.ch Thu Sep 29 09:55:41 2005
Date: Thu, 29 Sep 2005 09:42:29 +0200 (CEST)
From: Michael Spira <Michael.Spira@psi.ch>
Subject: Re: SM xsections / Br for mT = 175


Here are the new numbers for Mt=175 GeV:


SM Higgs cross sections at NLO in pb.
=====================================
cross section errors due to the scale variation (in per cent and different above and below -
thus e.g. +17-14 means +17% and -14%). The errors for VBF are unprecise, since they are of the
order of the integration errors in many cases. The scale dependence of VBF is of the order of +-2%.

MH   gg->H        VV->H*          WH**                ZH**                    ttH
---------------------------------------------------------------------------------------------------
115  39.3 (+18-14) 4.65 (+0.9-1.1) 1.98    (+0.3-0.0) 1.05     (+0.3-0.0) 0.749    (+6.1-9.2)
120  36.5 (+18-14) 4.47 (+0.8-1.1) 1.74    (+0.2-0.0) 0.922    (+0.2-0.0) 0.667    (+6.3-9.1)
125  33.9 (+18-14) 4.30 (+0.9-1.2) 1.53    (+0.1+0.1) 0.813    (+0.2+0.1) 0.596    (+6.0-9.1)
130  31.7 (+18-14) 4.14 (+0.8-1.2) 1.35    (+0.1+0.1) 0.719    (+0.1+0.1) 0.534    (+6.0-9.2)
135  29.6 (+18-14) 3.98 (+0.8-0.9) 1.19    (+0.1+0.1) 0.639    (+0.1+0.1) 0.479    (+5.9-9.2)
140  27.8 (+18-14) 3.83 (+0.7-0.9) 1.06    (+0.1+0.0) 0.569    (+0.1+0.1) 0.431    (+6.1-9.1)
145  26.1 (+18-14) 3.70 (+0.8-1.3) 0.944   (+0.4+0.3) 0.507    (+0.4+0.3) 0.389    (+5.9-9.1)
150  24.6 (+18-14) 3.56 (+0.9-1.2) 0.844   (+0.4+0.2) 0.454    (+0.4+0.3) 0.352    (+5.9-9.1)
160  21.9 (+18-14) 3.32 (+0.5-1.0) 0.681   (+0.4+0.0) 0.368    (+0.4+0.1) 0.291    (+5.9-9.2)
165  20.7 (+18-14) 3.20 (+0.7-0.9) 0.614   (+0.5-0.0) 0.332    (+0.4+0.0) 0.265    (+5.9-9.3)
170  19.7 (+18-14) 3.09 (+0.6-0.9) 0.554   (+0.5-0.1) 0.300    (+0.5-0.0) 0.243    (+5.8-9.4)
180  17.8 (+17-14) 2.88 (+0.5-0.7) 0.455   (+0.6-0.1) 0.247    (+0.6-0.1) 0.204    (+6.0-9.2)
190  16.2 (+17-14) 2.71 (+0.2-1.2) 0.377   (+0.6-0.2) 0.205    (+0.6-0.2) 0.174    (+5.8-9.4)
200  14.8 (+17-14) 2.53 (+0.4-1.1) 0.315   (+0.6-0.3) 0.171    (+0.6-0.2) 0.149    (+6.3-9.4)
250  10.2 (+17-14) 1.87 (+0.2-0.8) 0.140   (+0.9-0.6) 0.0766   (+0.9-0.6) 0.0785   (+6.7-9.8)
300  8.00 (+17-14) 1.42 (-0.0-0.9) 0.0709  (+0.1-1.0) 0.0385   (+1.0-0.9) 0.0490   (+6.9-10.5)
350  7.93 (+17-14) 1.10 (-0.2-0.6) 0.0391  (+1.2-1.0) 0.0212   (+1.2-1.0) 0.0340   (+7.7-10.8)
400  7.88 (+17-14) 0.869 (-0.2-1.1) 0.0231 (+1.5-1.1) 0.0124   (+1.5-1.1) 0.0251   (+8.0-11.3)
450  5.70 (+17-14) 0.698 (-0.8-0.9) 0.0144 (+1.6-1.4) 0.00768  (+1.6-1.4) 0.0193   (+8.6-11.2)
500  3.86 (+16-14) 0.566 (-0.6-0.7) 0.00932 (+1.7-1.5) 0.00495 (+1.7-1.5) 0.0151   (+8.5-11.8)
550  2.58 (+16-14) 0.464 (-0.2-0.6) 0.00626 (+1.8-1.6) 0.00330 (+1.8-1.6) 0.0120   (+8.9-12.0)
600  1.73 (+16-14) 0.386 (-0.8-1.0) 0.00432 (+1.9-1.7) 0.00226 (+1.9-1.7) 0.00956  (+8.5-12.2)
800  0.397 (+16-14) 0.196 (-0.7-1.5) 0.00121 (+2.2-2.3) 0.000616 (+2.2-2.3) 0.00408 (+9.2-12.5)


* The scale dependence of VBF is of the order of +-2%

** use +- 3 % for all masses for WH and ZH production.


===========================================================================
The scales are varied between 0.5 and 2 times the central scales with the
renormalization and factorization scales identified. The central scales are
given by:


gg -> H:  M_H
VV -> H:  Q_V  (square root of the absolute value of the V-momentum
                squared at each leg)
V* -> HV: M_HV (invariant mass of HV)
ttH:      m_t+M_H/2

Mt = 175 GeV

pdf: CTEQ6M
```

Appendix C

PhEDEx: CMS data transfer management

PhEDEx is a data transfer management system designed to cope with the large data transfers for the CMS experiment. All experiments in high energy physics which investigate particle collisions at high luminosities struggle with the large manpower needed to replicate data, to ensure data safety and to migrate collected data to mass storage systems such as tape libraries. PhEDEx provides a scalable infrastructure for managing these operations by automating many low level activities allowing the site responsibles to focus on handling data as logical entities rather than individual files.

PhEDEx itself does depend on the underlaying grid or other distribution tools. Designed as an additional abstract layer interfacing the different grid flavors LCG, OSG and NorduGrid, it uses existing tools for copying, migrating and replicating data sets. The workflow for data management and data placement is automated by a set of agents which are written in PERL. Each agent undertakes a particular task in a reliable way like file replication, routing decisions, tape migration and disk staging. The partitioning of functionality into subsets of tasks is the key element that makes PhEDEx robust and reliable. The information exchange among the individual agents running at the different site is managed by a central database located and maintained at CERN.

Bibliography

[1] P. W. Higgs, "Broken symmetries, massless particles and gauge fields,"
Phys. Lett. **12** (1964) 132–133.

[2] D. Griffiths, "Introduction to Elementary Particles". Wiley & Sons, 1987.

[3] F. Halzen and A. D. Martin, "Quarks and Leptons: Introductory Course in
Modern Particle Physics". Wiley & Sons, 1984.

[4] J. Goldstone, A. Salam, and S. Weinberg, "Broken Symmetries," *Phys. Rev.*
127 (1962) 965–970.

[5] V. A. Bednyakov, N. D. Giokaris, and A. V. Bednyakov, "On Higgs mass
generation mechanism in the standard model,"
`arXiv:hep-ph/0703280`.

[6] B. W. Lee, C. Quigg, and H. B. Thacker, "Weak interactions at very high
energies: The role of the Higgs-boson mass," *Phys. Rev. D* **16** (Sep, 1977)
1519–1531. `doi:10.1103/PhysRevD.16.1519`.

[7] **LEP Electroweak Working Group** Collaboration.
`HTTP://LEPEWWG.WEB.CERN.CH/LEPEWWG/`.

[8] H. Spiesberger, M. Spira, and P. M. Zerwas, "The Standard Model:
Physical Basis and Scattering Experiments," `arXiv:hep-ph/0011255`.

[9] C. Quigg, "The electroweak theory," `arXiv:hep-ph/0204104`.

[10] J. R. Ellis, M. K. Gaillard, and D. V. Nanopoulos, "A Phenomenological
Profile of the Higgs Boson," *Nucl. Phys.* **B106** (1976) 292.

[11] V. D. Barger, K.-m. Cheung, A. Djouadi, B. A. Kniehl, and P. M. Zerwas,
"Higgs bosons: Intermediate mass range at e+ e- colliders," *Phys. Rev.* **D49**
(1994) 79–90, `arXiv:hep-ph/9306270`.

[12] **CMS** Collaboration, "CMS Outreach."
HTTP://CMSINFO.CERN.CH/OUTREACH/.

[13] A. Djouadi, J. Kalinowski, and M. Spira, "HDECAY: A program for Higgs boson decays in the standard model and its supersymmetric extension," *Comput. Phys. Commun.* **108** (1998) 56–74, arXiv:hep-ph/9704448.

[14] **CMS** Collaboration, R. Kinnunen, "LHC potential for the Higgs boson discovery," *Acta Phys. Slov.* **55** (2005) 037–048.

[15] **CMS** Collaboration, "The CMS Physics Technical Design Report, Volume 2," *CERN/LHCC* **2006-021** (2006). CMS TDR 8.2
HTTP://CMSDOC.CERN.CH/CMS/CPT/TDR/.

[16] J. F. Grivaz, "New particle searches,". Rapporteur talk given at International Europhysics Conference on High Energy Physics (HEP 95), Brussels, Belgium, 27 Jul - 2 Aug 1995.

[17] **LEP Working Group for Higgs boson searches** Collaboration, R. Barate et al., "Search for the standard model Higgs boson at LEP," *Phys. Lett.* **B565** (2003) 61–75, arXiv:hep-ex/0306033.

[18] S. Soldner-Rembold, "Standard model Higgs searches and perspectives at the Tevatron," arXiv:hep-ex/0610014.

[19] The LEP Collaborations: ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration and the LEP Electroweak Working Group, "A combination of preliminary electroweak measurements and constraints on the standard model," arXiv:hep-ex/0511027.

[20] **CMS** Collaboration, "The CMS Physics Technical Design Report, Volume 1," *CERN/LHCC* **2006-001** (2006). CMS TDR 8.1
HTTP://CMSDOC.CERN.CH/CMS/CPT/TDR/.

[21] **CMS** Collaboration, "The Electromagnetic Calorimeter Technical Design Report," *CERN/LHCC* **97-033** (1997). CMS TDR 4.

[22] **CMS** Collaboration, "The Muon Project Technical Design Report," *CERN/LHCC* **97-32** (1997). CMS TDR 3.

[23] **CMS** Collaboration, "The TriDAS Project Technical Design Report, Volume 1: The Trigger Systems," *CERN/LHCC* **2000-38** (2000). CMS TDR 6.1.

[24] **CMS** Collaboration, "The TriDAS Project Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger," *CERN/LHCC* **2002-26** (2002). CMS TDR 6.2.

[25] "LCG Project Application Home Page." `HTTP://LCGAPP.CERN.CH/`.

[26] "LHC Computing Project (LCG)." `HTTP://LCG.WEB.CERN.CH/LCG/`.

[27] V. Innocente, L. Silvestris, and D. Stickland, "CMS software architecture: Software framework, services and persistency in high level trigger, reconstruction and analysis," *Computer Physics Communications* **140** (2001) 31–44. `doi:10.1016/S0010-4655(01)00253-3`.

[28] "OSCAR: CMS Simulation Package Home Page." `HTTP://CMSDOC.CERN.CH/OSCAR`.

[29] "ORCA: CMS Reconstruction Package." Site located at `HTTP://CMSDOC.CERN.CH/ORCA`.

[30] "CMS Software Page." `CMSDOC.CERN.CH/CMS/CPT/SOFTWARE/HTML/GENERAL/`.

[31] M. A. Dobbs et al., "Les Houches guidebook to Monte Carlo generators for hadron collider physics," `arXiv:hep-ph/0403045`.

[32] T. Sjostrand, L. Lonnblad, and S. Mrenna, "PYTHIA 6.2: Physics and manual," `arXiv:hep-ph/0108264`.

[33] F. Caravaglios, M. L. Mangano, M. Moretti, and R. Pittau, "A new approach to multi-jet calculations in hadron collisions," *Nucl. Phys.* **B539** (1999) 215–232, `arXiv:hep-ph/9807570`.

[34] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa, "ALPGEN, a generator for hard multiparton processes in hadronic collisions," *JHEP* **07** (2003) 001, `arXiv:hep-ph/0206293`. `doi:10.1088/1126-6708/2003/07/001`.

[35] T. Sjostrand and M. Bengtsson, "The Lund Monte Carlo for Jet Fragmentation and e+ e- Physics. Jetset Version 6.3: An Update," *Comput. Phys. Commun.* **43** (1987) 367.

[36] S. Catani, F. Krauss, R. Kuhn, and B. R. Webber, "QCD matrix elements + parton showers," *JHEP* **11** (2001) 063, `arXiv:hep-ph/0109231`.

[37] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, "Matching matrix elements and shower evolution for top-quark production in hadronic collisions," *JHEP* **01** (2007) 013, `arXiv:hep-ph/0611129`.

[38] V. Karimaki, D. Bourilkov, A. Nikitenko, and S. Slabospitsky, "User's Guide for CMS Physics Generators Interface CMKIN Version 4," *CMS Internal Note* **2004/052** (2004).

[39] T. Sjostrand et al., "High-energy-physics event generation with PYTHIA 6.1," *Comput. Phys. Commun.* **135** (2001) 238–259. `doi:10.1016/S0010-4655(00)00236-8`.

[40] G. Corcella et al., "HERWIG 6: An event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)," *JHEP* **01** (2001) 010, `arXiv:hep-ph/0011363`. `doi:10.1088/1126-6708/2001/01/010`.

[41] F. E. Paige, S. D. Protopopescu, H. Baer, and X. Tata, "ISAJET 7.69: A Monte Carlo Event Generator for pp, $\bar{p}p$, and e^+e^- Reactions," `arXiv:hep-ph/0312045`.

[42] A. Pukhov et al., "CompHEP - a package for evaluation of Feynman diagrams and integration over multi-particle phase space. User's manual for version 33," `arXiv:hep-ph/9908288`.

[43] F. Maltoni and T. Stelzer, "MadEvent: Automatic event generation with MadGraph," *JHEP* **02** (2003) 027. `doi:10.1088/1126-6708/2003/02/027`.

[44] **GEANT4** Collaboration, S. Agostinelli et al., "GEANT4: A simulation toolkit," *Nucl. Instrum. Meth.* **A506** (2003) 250–303.

[45] "IGUANA Web Site." `HTTP://IGUANA.WEB.CERN.CH/IGUANA/`.

[46] **CMS** Collaboration, D. Acosta et al., "CMS Physics TDR Volume 1, Section 2.6: Fast simulation," *CERN/LHCC* **2006-001** (2006) 55.

[47] G. Bruno, B. V. de Vyver, P. Cox, S. Lacaprara, N. Neumeister, S. Villa, and R. Wilkinson, "Local Reconstruction in the Muon Detectors," *CMS Note* **2002/043** (2002).

[48] E. Meschi, T. Monteiro, C. Seez, and P. Vikas, "Electron Reconstruction in the CMS Electromagnetic Calorimeter," *CMS Note* **2001/034** (2001).

[49] R. Frühwirth and T. Speer, "A Gaussian-sum filter for vertex reconstruction," *Nucl. Instrum. and Methods* **A534** (2004) 217–221. `doi:10.1016/j.nima.2004.07.090`.

[50] S. Baffioni, C. Charlot, F. Ferri, D. Futyan, P. M. I. Puljak, C. Rovelli, R. Salerno, and Y. Sirois, "Electron reconstruction in CMS," *CMS Note* **2006/040** (2006). CERN-CMS-NOTE-2006-040.

[51] A. Heister, O. Kodolova, V. Konoplianikov, S. Petrushanko, J. Rohlf, C. Tully, and A. Ulyanov, "Measurement of Jets with the CMS Detector at the LHC," *CMS NOTE* **2006/036** (2006).

[52] V. Konoplianikov, O. Kodolova, and A. A. Ulyanov, "Jet Calibration using γ+jet Events in the CMS Detector," *CMS Note* **2006/042** (2006).

[53] J. D'Hondt, S. Lowette, J. Heyninck, and S. Kasselmann, "Light quark jet energy scale calibration using the W mass constraint in single-leptonic $t\bar{t}$ events," *CMS Note* **2006/025** (2006).

[54] P. Vanlaer, L. Barbone, N. D. Filippis, T. Speer, and O. B. F.-P. Schilling, "Impact of CMS Silicon Tracker Misalignment on Track and Vertex Reconstruction," *CMS Note* **2006/029** (2006).

[55] "ROOT : an Object Oriented Data Analysis Framework." `HTTP://ROOT.CERN.CH/`.

[56] M. Erdmann, U. Felzmann, A. Flossdorf, S. Kappler, M. Kirsch, G. Mueller, G. Quast, C. Saout, A. Schmidt, and J. Weng, "Concepts, Developments and Advanced Applications of the PAX Toolkit," in *Computing in High Energy and Nuclear Physics (CHEP06).* Bombay (India), 2006. `arXiv:physics/0605063`.

[57] M. Erdmann, U. Felzmann, D. Hirschbühl, S. Kappler, M. Kirsch, G. Quast, A. Schmidt, and J. Weng, "The PAX Toolkit and Its Applications at Tevatron and LHC," *IEEE Transactions on Nuclear Science* **53** (April, 2006) 506–512. `HTTP://PAX.WEB.CERN.CH/PAX/`.

[58] "CLHEP - A Class Library for High Energy Physics." `HTTP://PROJ-CLHEP.WEB.CERN.CH/PROJ-CLHEP/`.

[59] The EGEE (Enabling Grids for E-sciencE) Project. `HTTP://WWW.EU-EGEE.ORG/`.

[60] Open Sience Grid. `HTTP://WWW.OPENSCIENCEGRID.ORG/`.

[61] **NDGF (Nordic Data Grid Facility)** Collaboration.
HTTP://WWW.NDGF.ORG/.

[62] B. S. W. Group. HTTP://LCG.WEB.CERN.CH/LCG/PEB/BS/.

[63] **CMS** Collaboration, M. D. Negra, A. Petrilli, A. Herve, and L. Fo, "CMS:
The computing project. Technical design report," *CMS/LHCC* **2005-023**
(2005). HTTP://CMSDOC.CERN.CH/CMS/CPT/TDR/.

[64] S. P. E. C. (SPEC). HTTP://WWW.SPEC.ORG/.

[65] "PhEDEx - high-throughput data transfer management system."
HTTP://CMSDOC.CERN.CH/CMS/APROM/PHEDEX/.

[66] D. Duellmann et al., "POOL Development Status and Plans," in
Proceedings of the CHEP'04 Conference. Interlaken, Switzerland,
September 27th - October 1st, 2004.

[67] "Institut für Experimentelle Kernphysik, Universität Karlsruhe."
HTTP://WWW-EKP.PHYSIK.UNI-KARLSRUHE.DE/.

[68] P. Schemitz and M. Stanitzki, "Aufbau eines Datenanalyse-Clusters,"
IEKP-KA **2003-12** (2003).

[69] Y. Kemp, "Identification of Electrons in the Forward Region of the CDF
Experiment for the Search for Electroweak Top Quark Production". PhD
thesis, Institut für Experimentelle Kernphysik, Universität Karlsruhe, 2006.
IEKP-KA/2006-001.

[70] "Karlsruhe CDF Group."
HTTP://WWW-EKP.PHYSIK.UNI-KARLSRUHE.DE/CDF/.

[71] "AMS02-Group at the IEKP."
HTTP://WWW-EKP.PHYSIK.UNI-KARLSRUHE.DE/~AMSWWW/.

[72] V. Büge, U. Felzmann, C. Jung, H. Stadie, and A. Vest, "Integrating the
IEKP Linux cluster as a Tier-2/3 prototype centre into the LHC Computing
Grid," *IEKP-KA* **2006-3** (2006).

[73] D. Benedetti, S. Cucciarelli, C. Hill, J. Incandela, S. A. Koay, C. Riccardi,
A. Santocchia, A. Schmidt, P. Torre, and C. Weiser, "Observability of Higgs
produced with top quarks and decaying to bottom quarks," *Journal of
Physics G: Nuclear and Particle Physics* **34** (2007), no. 5, N221–N250.

[74] R. Kleiss and W. J. Stirling, "Tagging the Higgs," *Phys. Lett.* **B200** (1988) 193.

[75] V. D. Barger, T. Han, and R. J. N. Phillips, "Improving the heavy Higgs boson two charged lepton - two neutrino signal," *Phys. Rev.* **D37** (1988) 2005–2008.

[76] V. D. Barger, R. J. N. Phillips, and D. Zeppenfeld, "Mini - jet veto: A Tool for the heavy Higgs search at the LHC," *Phys. Lett.* **B346** (1995) 106–114, `arXiv:hep-ph/9412276`.

[77] V. Lefebure and J. Andreeva, "RefDB: The Reference Database for CMS Monte Carlo Production," *CMS Conference Report* **2003/017** (2003).

[78] S. Slabospitsky, "CMS values for the major SM cross sections." http://cmsdoc.cern.ch/cms/PRS/gentools/www/xsec/cmsxsec.html.

[79] R. Bonciani, S. Catani, M. L. Mangano, and P. Nason, "NLL resummation of the heavy-quark hadroproduction cross- section," *Nucl. Phys.* **B529** (1998) 424–450, `arXiv:hep-ph/9801375`.

[80] S. Haywood et al., "Electroweak physics," `arXiv:hep-ph/0003275`.

[81] H. L. Lai et al., "Global QCD analysis and the CTEQ parton distributions," *Phys. Rev.* **D51** (1995) 4763–4782, `arXiv:hep-ph/9410404`.

[82] "CMS Alpgen sample generation." `HTTPS://TWIKI.CERN.CH/TWIKI/BIN/VIEW/MAIN/PROD06`.

[83] "CMS Generator Tools group." `HTTP://CMSDOC.CERN.CH/CMS/PRS/GENTOOLS/`.

[84] M. Mangano, "Private electronic communication," 2007.

[85] D. Futyan, D. Fortin, and D. Giordano, "Search for the Standard Model Higgs Boson in the Two-Electron and Two-Muon Final State with CMS," *CMS Note* **2006/136** (2006).

[86] V. Bartsch and G. Quast, "Expected Signal Observability at Future Experiments," *CMS Note* **2005/004** (2005).

[87] S.S.Wilks, "The Large-Sample Distribution of the Likehood Ratio for Testing Composite Hypotheses," *Annals of Math. Stat.* **9** (1938) 60–62.

[88] **CMS** Collaboration, G. Quast, "CMS Physics TDR Volume 2, Appendix A.1," *CERN/LHCC* **2006-021** (2006) 515.

[89] T. Figy, C. Oleari, and D. Zeppenfeld, "Next-to-leading order jet distributions for Higgs boson production via weak-boson fusion," *Phys. Rev.* **D68** (2003) 073005, `arXiv:hep-ph/0306109`.

[90] J. Rohlf and C. Tully, "Recommendations for Jet and Missing Transverse Energy Reconstruction Settings and Systematics Treatment," *CMS Internal Note* **2006/025** (2006).

[91] S.Abdullin, D.Acosta, P.Bartalini, R.Cavanaugh, A.Drozdetskiy, A.Korytov, G.Mitselmakher, Yu.Pakhotin, B.Scurlock, and A.Sherstnev, "Search Strategy for the Standard Model Higgs Boson in the H$\rightarrow$ ZZ$^{(\star)} \rightarrow 4\mu$ Decay Channel using $M(4\mu)$-Dependent Cuts," *CMS Note* **2006/122** (2006).

[92] S. Baffioni, C. Charlot, F. Ferri, N. Godinovic, P.Meridiani, I. Puljak, R. Salerno, and Y. Sirois, "Discovery potential for the SM Higgs boson in the H$\rightarrow$ ZZ$^{(\star)} \rightarrow$ e$^+$e$^-$e$^+$e$^-$ decay channel," *CMS Note* **2006/115** (2006).

Acknowledgements

Finally, I am very thankful to my family for their support and their care.