Universität Karlsruhe (TH)

ANALYSE DES KANALS $H \to 4\,\mu$
IN DER OBJEKTORIENTIERTEN
SOFTWAREUMGEBUNG VON CMS

JOANNA WENG

DIPLOMARBEIT

bei Prof. Dr. G. Quast
Institut für Experimentelle Kernphysik

Korreferent Prof. Dr. M. Feindt
Institut für Experimentelle Kernphysik

AN DER FAKULTÄT FÜR PHYSIK
DER UNIVERSITÄT KARLSRUHE

7. JANUAR 2003

Contents

Introduction

The Standard Model is among the greatest intellectual achievements of physics. All particles either being constituents of matter or mediating the forces between them have been found – except for one special particle, the Higgs boson, which is responsible for the generation of mass of the other particles. It is therefore a fundamental ingredient of the Standard Model, which actually stands and falls with the existence or nonexistence of the Higgs boson. Moreover, many open questions of modern physics, like neutrino masses or gravity or the Grand Unification Theorie, do involve the Higgs Field [Wil]. Many efforts have been made to find it, and a large possible mass range has already been excluded. Today, the lower limit for the Higgs mass has been set by the direct search at the LEP[1] experiments at CERN[2] to 114.4 GeV [3] whereas the upper limit was derived from the electroweak precision measurements to be less than 193 GeV at 95% confidence level as shown in figure 1. While this is not a proof that the Standard Model Higgs boson actually exists, it does serve as a guideline in what mass range to look for it.

The aim of this thesis is to examine the mass region of 125 - 140 GeV, which is well compatible with the predicted region, and the Higgs discovery potential of the channel $H \rightarrow 4\mu$ in this mass region with the CMS[4] detector at the LHC[5]. In order to perform the simulation as realistic as possible, the newest available, technically challenging, full detector simulation is used for the signal and background samples.

This study consists of eight chapters. In the first chapter, a brief introduction into the Standard Model is given, pointing out the essential role of gauge invariance and, as a consequence, the essential role of the Higgs mechanism.

The second chapter describes how the Higgs boson can actually be produced and what its decay products are at the LHC. A quantitative overview of the expected processes is given.

In the third chapter the LHC and the CMS detector are introduced and relevant technical details are provided. Because of its importance for this analysis, the Trigger System and the Muon System of CMS are described in more detail.

Chapter four outlines the complex hardware and computing infrastructure at the University of Karlsruhe that was set up and configured in 2002. It includes an

---

[1] Large Electron-Positron collider
[2] Conseil Européen pour la recherche Nucléaire
[3] throughout this study $\hbar$ and c have been set to 1
[4] Compact Muon Solenoid
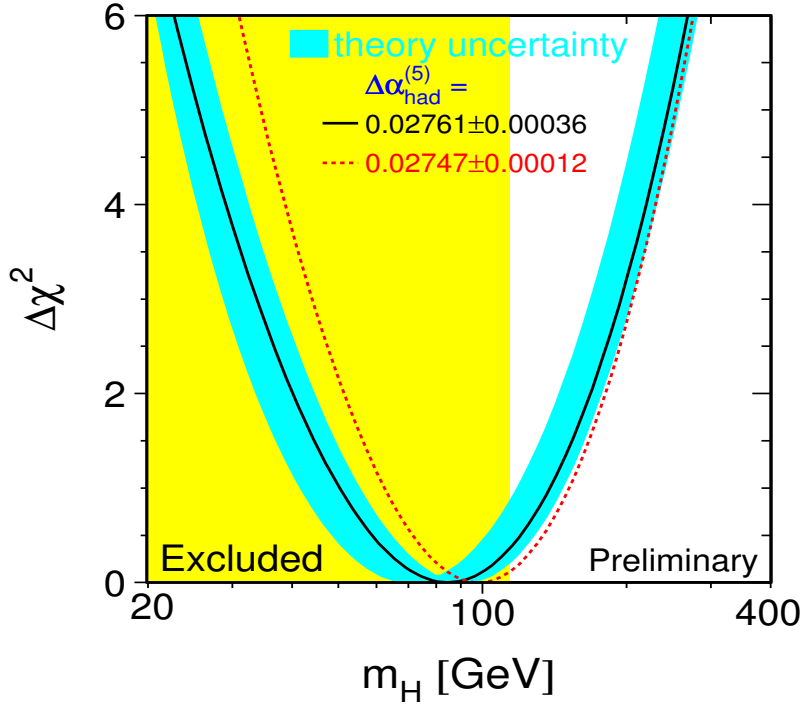[5] Large Hadron Collider

Figure 1: The figure shows the $\Delta\chi^2$ curve derived from the precision electroweak measurements, performed at LEP and others, as a function of the Higgs-boson mass, assuming the Standard Model to be the correct theory of nature. The preferred value for its mass, corresponding to the minimum of the curve, is around 81 GeV. The precision electroweak measurements tell us that the mass of the Standard-Model Higgs boson is lower than about 193 GeV (one-sided 95 percent confidence level upper limit derived from $\Delta\chi^2 = 2.7$ for the blue band, thus including both the experimental and the theoretical uncertainty). The yellow band is the mass region excluded in direct search by LEP, that means that $m_H > 114.4$ GeV.

overview over the network architecture of the computing cluster and a description of the temperature monitoring tool, that was developed during this work.

In chapter five, the most important software components and the setup of the analysis tools as well as the full detector simulation chain are described. The new analysis program developed here is explained in more detail, as well as the relevant reconstruction algorithms and the developed interface to the analysis package ROOT.

The sixth chapter presents efficiency and resolution studies which were performed with the full detector simulation. Chapter seven finally presents the results of the analysis of the considered signal and background channels. In particular, the results on generator level and after the full simulation are compared. In the end, a conclusion and an outlook is given.

Chapter 1

The Higgs Boson in the Standard Model

1.1 The Standard Model of Particle Physics

In the Standard Model(SM) of Particle Physics particles of matter are divided into two classes: six so called "leptons" and six "quarks" which are all "fermions" (i.e. particle with half integer spin quantum numbers e.g. 1/2). Particles with a spin of an integer number i.e 0,1,2 are called "bosons". Because of similar properties one can sort the fermions into three generations (corresponding to the 3 columns of tab. 1.1). Particles that are made of quarks, are called "hadrons" (for example proton). In the Standard Model there are three kinds of interactions between these particles of matter: the electromagnetic, weak and strong interaction. The integration of the interaction from everyday life into the Standard Model, gravitation, has not been successful yet. However, gravitational interactions are extremely weak compared to the other interactions and therefore can be ignored. The Standard Model of Particle

ν_e	ν_μ	ν_τ	leptons	
e	μ	τ		particles of matter
u	c	t	quarks	
d	s	b		
γ		g	$W^\pm, Z$	gauge particles
electro-magnetic		strong	weak	

Table 1.1: Particles of matter and interaction in the Standard Model.

Physics is the relativistic quantum field theory of these strong, electromagnetic, and weak interactions. In such theories, each type of interaction has a characteristic set of force carrier particles associated with quantum excitation of the force field related

to that interaction. The carrier particles either appear in intermediate stages or are produced during all particle processes involving that type of interaction. Forces between particles can be described in terms of static force fields and exchanges of force by carrier particles, which are always bosons, between the affected particles.

As will be explained later, the bosons in table 1.1 are called "gauge" bosons, because in the Standard Model the force fields mentioned above are a consequence of how the terms describing particles (e.g. wave function) behave under certain transformations, so called "gauge" transformations. The central point is whether symmetries do exist and whether therefore physical quantities are conserved. This will be discussed in more detail in the following sections. A description of the Standard Model can be found for example in [HaMa] or [Berg]. The 12 particles of matter and the bosons are shown in tab. 1.1, a nice overview is also given below in Figure 1.1.
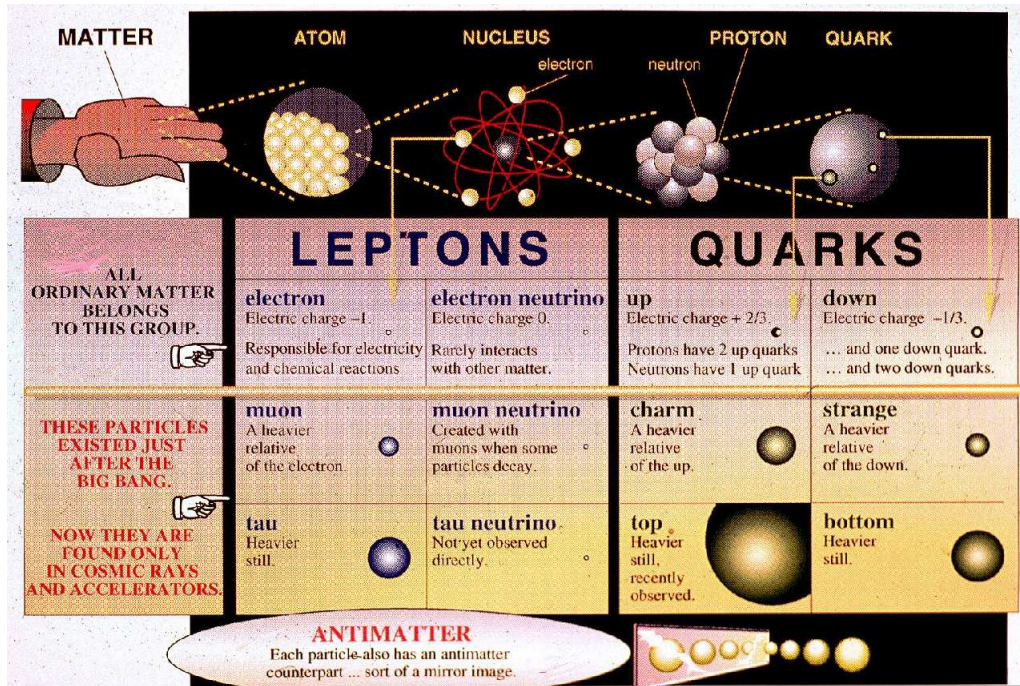

Figure 1.1: Overview over particles of matter

1.2 Gauge invariance and massive gauge bosons

Our present belief is that all particle interactions are dictated by symmetry principles and the behavior under local gauge transformations.

Emmy Noether's theorem states, that for every continuous symmetry, that means, invariance of the Lagrangian (or the action) under a certain transformation, there must exist a conservation law of a physical quantity. For example the conservation

law corresponding translational symmetry in space is the Law of Conservation of Momentum.

The connection between symmetries , invariance under certain transformations and conservation laws can be shown best in the framework of Lagrangian theory. One considers therefore the Euler-Lagrange equation ($\mathcal{L}$ is called Lagrangian):

$$\frac{\partial \mathcal{L}}{\partial \phi(x)} - \partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi(x))} = 0 \tag{1.1}$$

The Lagrangian for non-complex scalar (Dirac field) is :

$$\mathcal{L}^{Dirac} = \bar{\psi}(x)(i\gamma^\mu \partial_\mu - m)\psi(x) \tag{1.2}$$

Equation 1.2, a complex filed describing an electron, is invariant under the phase transformation $\psi(\chi) \to e^{i\alpha}\psi(\chi)$. The family of phase transformations $U(\alpha) = e^{i\alpha}$, where a single parameter α my run continously over real numbers, form a unitary Abelian group, the U(1) group. The fact, that $\mathcal{L}^{Dirac}$ is U(1) invariant, implies the existence of a conserved current $j^\mu = (\rho, \vec{j})$ that means $\partial_\mu j^\mu = 0$. It follows than that then charge $Q = \int d^3x j^0$ must be conserved. The at first sight rather unimportant observation of U(1)-invariance implies thus an important conservation law: the conservation of charge. The existence of a symmetry implies that some quantity is unmeasurable. For example, translation invariance means that we cannot determine an absolute position in space. Or, in our example above, invariance under the phase transformation means that the phase α is unmeasurable and has no physical meaning. Because α, once fixed, remains constant, we speak of global gauge transformation. More general, one can consider a local gauge transformations, where α differs from point to point, that is, $\alpha = \alpha(x)$.

Studying the behavior of equation 1.2 under the local gauge transformation $\psi(\chi) \to e^{i\alpha(\chi)}\psi(\chi)$, one discovers, that the invariance of $\mathcal{L}^{Dirac}$ is broken, broken by the term $\partial_\mu \alpha$. If we want to insist on the invariance of $\mathcal{L}^{Dirac}$, the derivate has to be modified that way, that it transforms covariantly under phase transformations, like ψ itself : $D_\mu \psi \to e^{i\alpha(x)} D_\mu \psi$ Therefore a vector field A_μ is introduced, with transformation properties such that the redundant term is canceled. This aim can be achieved by constructing:

$$D_\mu \psi := \partial_\mu - ieA_\mu, A_\mu \to A_\mu + \frac{1}{e}\partial_\mu \alpha \tag{1.3}$$

Now we can replace ∂_μ by our new derivate and see, that equation 1.2 remains invariant:

$$\begin{aligned} \mathcal{L}^{Dirac} &= \bar{\psi}(x)(i\gamma^\mu D_\mu - m)\psi(x) \\ &= \bar{\psi}(x)(i\gamma^\mu \partial_\mu - m)\psi + e\bar{\psi}(x)\gamma^\mu \psi A_\mu \end{aligned}$$

where the second term guarantees the invariance.

To resume, by demanding local phase invariance, we were forced to introduce a vector field A_μg local phase invariance, we were forced to introduce a vector field A_μ , a so called gauge field, which couples to the Dirac particle (e.g. e^-) in the same way

as the photon field. If we want to consider A_μ as the physical photon field, we must add a term corresponding to its kinetic energy that must be invariant under phase transformation ,too: the gauge invariant field strength tensor: $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$

This leads to the QED Lagrangian :

$$
\begin{aligned}
\mathcal{L} &= \mathcal{L}^{Dirac} + \mathcal{L}^{int} + \mathcal{L}^{em.field} \\
&= \bar{\psi}(x)(i\gamma^\mu \partial_\mu - m)\psi(x) + e\bar{\psi}\gamma^\mu A_\mu \psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}.
\end{aligned}
\tag{1.4}
$$

It is clear, that a new field has to be introduced, since otherwise local phase changes will create phase differences, which would be observable if not compensated in some other way.

Locale gauge invariance can be restored by the introduction of a (massless) photon field. This is also possbile for strong interaction by the introduction of massless gluons as gauge bosons. This is done in e.g. [HaMa]

Until now all gauge bosons, photons and gluons, are massless. This leads to a serious problem: How can we get a gauge invariant Lagrangiens for weak interactions, knowing that the bosons $(W^\pm, Z)$ do have a mass of the order of 100Gev ? We consider therefore $SU(2)$ transformations of the type :

$$
\boxed{\psi'(x) = e^{i\frac{\vec{\sigma}}{2}\cdot\vec{\alpha}(x)}\psi(x).}
\tag{1.5}
$$

$SU(2)$ once again is not Abelian. To apply the gauge principle on weak interactions, 3 fields have to be introduced (or a three-dimensional vector) $\vec{W}_\mu$. The Lagrangian is then :

$$
\begin{aligned}
\mathcal{L}_{weak} &= \mathcal{L}^{Dirac} + \mathcal{L}^{int} + \mathcal{L}^{field} + \mathcal{L}^{mass} \\
&= \bar{\psi}(x)(i\gamma^\mu \partial_\mu - m)\psi(x) - \bar{\psi}(i\gamma^\mu ig\frac{\vec{\sigma}}{2}\vec{W}_\mu)\psi \\
&\quad -\frac{1}{4}\sum_i F^{(i)}_{\mu\nu}F^{(i)\mu\nu} + \frac{1}{2}M^2_{gauge}\vec{W}^\mu\vec{W}_\mu,
\end{aligned}
\tag{1.6}
$$

$$
F^{(i)}_{\mu\nu} = \partial_\nu W^i_\mu - \partial_\mu W^i_\nu + g\epsilon_{ijk}W^j_\mu W^k_\nu,
\tag{1.7}
$$

where the factor $g\epsilon_{ijk}W^j_\mu W^k_\nu$ once again appears because of the non-Abelian character of $SU(2)$. We introduce the covariant derivate :

$$
D_\mu = \partial_\mu - ig\frac{\vec{\sigma}}{2}\vec{W}_\mu.
\tag{1.8}
$$

The gauge transformations are now $\psi \to e^{ig\frac{\vec{\sigma}}{2}\vec{\beta}(x)}$ and $\vec{W} \to \vec{W} - \partial_\mu\vec{\xi}$.

The first three terms of equation 1.6 remain invariant, but the mass term of the gauge bosons does not. The Lagrangian 1.6 is not invariant under SU(2) transformations anymore, gauge invariance is broken by the mass terms of the massive gauge bosons. A solution to this problem is presented in the section 1.4 .

1.3 Spontaneous breaking of a global gauge symmetry

The gauge bosons have to be massless to fullfill gauge invariance. On the other hand the mass of the gauge bosons of the weak interaction is known to be very large. Thus it seems impossible to hold on to the gauge principle in this case.

However, it is still possible if one supposes that the gauge bosons are actually massless but *seem to have a mass* because they interact with an invisible background field, the Higgs field [Hig].

Such a field ϕ has to be a background field everywhere in space and its expectation value $<\phi_0>$ should be different from zero.

If ϕ is a simple normed complex variable: with $\phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2)$, the easiest ansatz for the potential and the Lagrangian is :

$$
\begin{aligned}
V(\phi) & = -\mu^2|\phi|^2 + \lambda^2|\phi|^4, & (1.9)\\
\rightarrow \mathcal{L} & = T - V = (\partial_\mu\phi)(\partial^\mu\phi) - \mu^2|\phi|^2 + \lambda|\phi|^4, & (1.10)
\end{aligned}
$$

The potential is shown in 2D in fig. 1.2: The Lagrangian is invariant under $\phi \rightarrow e^{i\alpha}\phi$,
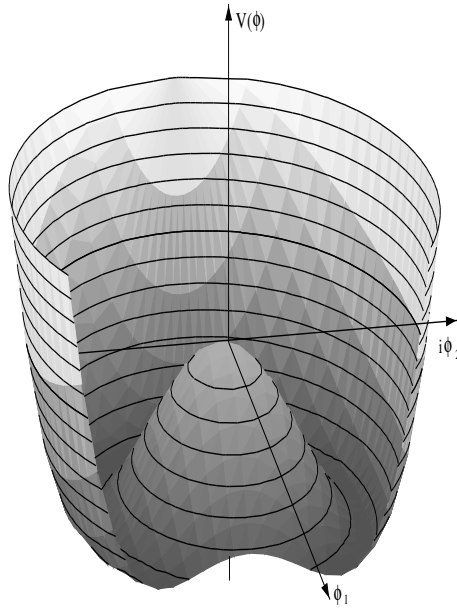

Figure 1.2: A Higgs potential-like function: $f(x) = -3x^2 + x^4$. Taken from [Kem].

i.e. $\mathcal{L}$ possesses a global U(1) gauge symmetry and becomes minimal on a circle around the origin: $|\phi_0| = \frac{\mu}{\sqrt{2\lambda}} = \frac{v}{\sqrt{2}}$. One can now expand ϕ about a minimum in terms of $\eta(x), \zeta(x)$:

$$
\phi(x) = \frac{1}{\sqrt{2}}(v + \eta(x) + i\zeta(x)), \qquad (1.11)
$$

We cannot expand around $\phi = 0$, because that is an unstable point and the perturbation series would not converge. Therefore we took $\phi_0 = \frac{v}{\sqrt{2}}$, which is a minimum

and a stable point and obtain for the Lagrangian:

$$V(\phi) = -\frac{\mu^2}{2}(v + \eta + i\zeta)(v + \eta - i\zeta) + \frac{\lambda}{4}[(v + \eta + i\zeta)(v + \eta - i\zeta)]^2$$

$$\rightarrow \mathcal{L} = \underbrace{\frac{1}{2}(\partial_\mu \eta)(\partial^\mu \eta) - \mu^2 \eta(x)^2}_{\text{particle with mass } m_\eta = \sqrt{2}\mu} + \underbrace{\frac{1}{2}(\partial_\mu \zeta)(\partial^\mu \zeta)}_{\text{massless particle, "Goldstone boson"}} + \cdots$$

The symmetry of the Lagrangian has now been broken by our choice of the ground state - this is called *spontaneous symmetry breaking*. For the $\eta(x)$ field there is a mass term and one term representing its kinetic energy but in the case of $\zeta(x)$ the mass term is missing. This means that we have one massive particle and a massless scalar particle, known as a *Goldstone Boson*. The Goldstone theorem states, that if the Lagrangian has an exact continuous symmetry which is broken in the ground state, a massless particle occurs. A known example for this theorem is a ferromagnet, which is described by a Lagrangian which is invariant under rotations in space. In the ground state, however, the elementary spins are aligned in a particular direction and the rotational symmetry is broken - the analogue of the Goldstone boson are the spin waves.

In the case of our Higgs potential, the potential is flat in the tangent direction ($\zeta(x)$ -direction). There is no resistance to excitations along the $\zeta(x)$ -direction (massless mode).

1.4   Higgs mechanism

Now there is a massless Goldstone boson instead of massive gauge bosons. Finally, we consider now spontaneous symmetry breaking of local gauge symmetry again in the simpler case of U(1) gauge symmetry (for weak interaction SU(2) symmetry should be considered). First, we make the Lagrangian 1.10 invariant under U(1) transformations $\phi \rightarrow e^{i\alpha(x)}\phi$ and introduce once again a covariant derivate:

$D_\mu = \partial_\mu - ieA_\mu$ with a gauge field that transforms as $A_\mu \rightarrow A_\mu + \frac{1}{e}\partial_\mu \alpha$.

The gauge invariant Lagrangian is then:

$$\mathcal{L} = (\partial^\mu + ieA^\mu)\phi^*(\partial_\mu - ieA_\mu)\phi - \mu^2|\phi|^2 - \lambda|\phi|^2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{1.12}$$

We consider the case $\mu^2 > 0$, since only then there is a spontaneous symmetry breaking of the Lagrangian and repeat now the same procedure as we did in the section 1.3 to translate the field ϕ the ground state:

$$\mathcal{L}' = \frac{1}{2}(\partial_\mu \zeta)^2 + \frac{1}{2}(\partial_\mu \zeta)^2 - v^2\lambda\eta^2 + \frac{1}{2}e^2v^2 A_\mu A^\mu - evA_\mu\partial^\mu\zeta - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \dots \tag{1.13}$$

The particle spectrum is here a massless Goldstone boson ζ, a massive scalar η and a massive vector A_μ. That means, mass is dynamically generated for the gauge field :

$$m_\zeta = 0, m_\eta = \sqrt{2\lambda v^2}, m_A = ev \tag{1.14}$$

But there is still the Goldstone boson left. The clue is now to notice, that by giving mass to A_μ, the polarisation degrees of freedom raised from 2 to 3. Simple variable transformation, however, cannot create a new degree of freedom. By a particular choice of gauge the Goldstone boson can be turned into the degree of freedom of the massive particle. If we take a different set of real fields h, θ, A_μ with

$$\phi \to \sqrt{\frac{1}{2}}(v + h(x))e^{i\theta(x)/v} \text{ and } _\mu \to A_\mu + \frac{1}{ev}\partial_\mu\theta \text{ we obtain the Lagrangian :}$$

$$\mathcal{L}'' = \frac{1}{2}(\partial_\mu h)^2 - \lambda v^2 h^2 + \frac{1}{2}e^2 v^2 A_\mu^2 - \lambda^3 - \frac{1}{4}\lambda h^4 + \frac{1}{2}e^2 A_\mu^2 h^2 + v e^2 A_\mu^2 h - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad (1.15)$$

The Goldstone boson does not appear in the Lagrangian - there are only two interacting massive particles, a vector gauge boson A_μ and a massive scalar h, which is called Higgs particle ($m_h = \sqrt{\lambda}v$). The apparent extra degree of freedom coresponds only to the freedom to choose a gauge transformation and can be turned into the longitudinal polarisation of the massive gauge particle - this is called Higgs mechanism.

For weak interaction one has to repeat the procedure for an SU(2) gauge symmetry. This is done, for example in [HaMa]. There are then 4 Goldstone bosons interacting with 3 three massless gauge bosons - by a particular choice of gauge they can be turned again in the longitudinal degrees of freedom of the three gauge bosons , which are then massive ; These are the three gauge bosons of weak interaction : $W^\pm, Z^0$. The remaining boson is the Higgs Boson. This model is called "Weinberg-Salam Model"[Wei]. However, one should remember, that the motivation for introducing the Higgs scalar was entirely theoretical, and there is at present no evidence, that this particle actually exists. Its importance arises from the fact, that it allows us to generate the masses of the weak bosons without spoiling the renormalizability of th electroweak gauge theory. The renormalizability of the theory is not trivial. It was eventually demonstrated by 't Hooft and Veltmann in 1973. Another attractive feature of the model is that the Higgs field which generates $W^\pm$, Z^0 masses is also sufficient to give mass to the leptons and quarks. The mass of the Higgs boson cannot be calculated, because it depends on the unknown coupling λ, as well as the masses of the fermions are just parameter of the theory and are not predicted -their empirical values must be the input.

There are basically three Higgs coupling, which are important for the understanding of Higgs boson production and decay :

$$y_f = \frac{-igm_f}{2m_W} \quad (1.16)$$

$$g_{ZZH} = \frac{-igm_Z}{cos\theta_W}g^{\mu\nu} \quad (1.17)$$

$$g_{WWH} = igm_W g^{\mu\nu} \quad (1.18)$$

where m_f is the mass of the corresponding fermion, m_W the $W^\pm$ mass, m_Z the Z^0 mass, θ_W the weak mixing angle and g the standard SU(2) gauge group coupling. y_f is called Yukawa coupling (of the corresponding fermion f), g_{ZZH} is the ZZH coupling and g_{WWH} is the WWH coupling. After the masses and couplings of

fermions and Gauge bosons are determined experimentally, predictions of the Higgs mass are possible. The expected situation to produce the Higgs boson at the LHC is described in the next chapter.

Chapter 2

Higgs production and decay at the LHC

2.1 Definition of luminosity

The very basic of every collider experiment is formed by the equation

$$\frac{dN}{dt} = L\sigma \tag{2.1}$$

It relates the observed event rate dN/dt with the corresponding cross section σ. The machine dependent proportional factor L is called *luminosity* and has to be measured e.g. via comparison to a theoretically well-known reaction. In order to gather as many events of a certain kind as possible, one would like to have a large luminosity. For Gaussian beam profiles with horizontal and vertical widths $\sigma_{x_1}^*$, $\sigma_{x_2}^*$, $\sigma_{y_1}^*$, $\sigma_{y_2}^*$ is given by

$$L = \frac{1}{e^2 f_b} \frac{I_1 I_2}{2\pi(\sigma_{x_1}^* \sigma_{y_1}^* + \sigma_{x_2}^* \sigma_{y_2}^*)} \tag{2.2}$$

where I_1, I_2 are the beam currents and f_b is the bunch frequency. If the expected events for e.g. one year LHC are considered, the equation 2.1 has to be integrated over this time period and it follows

$$N = L_{int}\sigma \tag{2.3}$$

L_{int} is called the integrated luminosity and N is the number of expected events for e.g. one year. The expected design luminosity at the LHC is described in the next chapter.

2.2 Higgs production at the LHC

At hadron colliders there are several ways to produce Higgs bosons. The dominant process with the largest cross section is the gluon fusion process, shown at LO (Leading Order)in fig. 2.1.
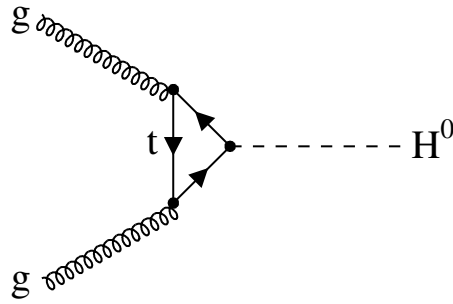
Figure 2.1: The Feynman graph of gluon fusion at LO (Leading Order). Is is the dominant process for Higgs production at the LHC. The Higgs boson couples to gluons via a (top)quark loop

Since gluons are massless they cannot couple directly to the Higgs boson, so that a quark loop connects the gluons with the colourless Higgs boson. Because of its heavy mass, it is a top quark loop in most cases. The ratio of higher order and leading order(LO) calculations is called K factor. For example, the K factor for NLO (next to leading order) is $K = \frac{\sigma_{nlo}}{\sigma_{LO}}$. As can be seen later in the analysis chapter, K factors for the cross section of the gluon fusion are close to 2, that means an enormous uncertainty. Another prominent source of uncertainty in theoretical calculations are the parameterization of the partons distribution functions(pdf).

The second largest contribution to Higgs production at the LHC is vector boson fusion, where in addition to the Higgs boson two highly energetic quarks in forward and backward direction are produced ($qq\prime \to H^0qq\prime$). This is shown in fig. 2.2.
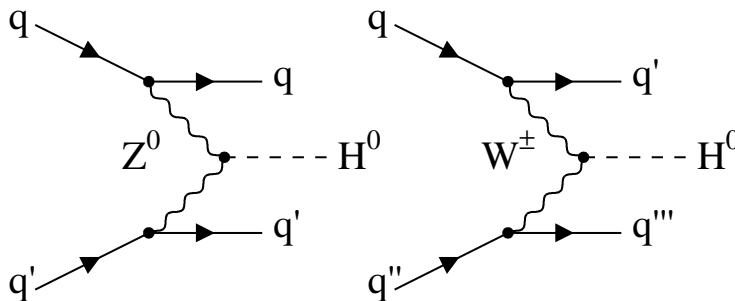
Figure 2.2: Vector boson fusion process at LO. After the gluon fusion the most important process for Higgs production at the LHC. The Z^0 fusion leaves the quark flavors unchanged, the $W^{\pm}$ changes the quark flavors from down to up type or vice versa

The next contribution is given by several processes with a cross section of similar order of magnitude and a Higgs production with one or more accompanying particles ($gg, q\bar{q} \to H^0t\bar{t}$, $q\bar{q} \to H^0Z^0$, $q\bar{q}\prime \to H^0W^{\pm}$). This is called associated production and can be seen in fig. 2.3 and in fig. 2.4.

An overview of the relevant processes for Higgs production and their cross sections at the LHC is given in fig. 2.9. The gluon fusion is the dominant production in the entire mass range. In our region of interest (110 − 150 GeV), we expect Higgs production cross sections σ in order of 20 − 40 pb. In case the Higgs is heavy ($M_H > 200$ GeV), not only the total he Higgs signal dissolves in the background. Production cross section σ goes down, but also its width increases rapidly. This is
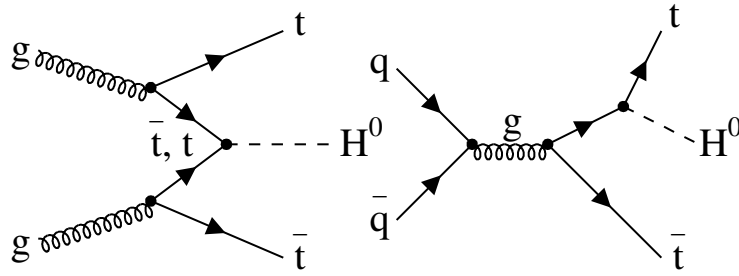
Figure 2.3: Examples of associated Higgs production at LO, here Higgs strahlung of heavy top quarks
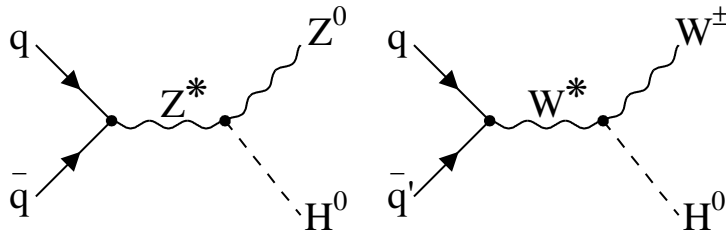

Figure 2.4: Associated Higgs production at LO, here Higgs strahlung of vector bosons. Z^0 and $W^\pm$ are produced off-shell and can therefore radiate a Higgs boson

shown in fig. 2.5. The experimental mass resolution is only important for a low mass Higgs ($M_H < 200$ GeV), where the width is smaller than 2 GeV. For large Higgs masses, the Higgs signal dissolves in the background.

2.3 Higgs decays at the LHC

In general, the Higgs boson prefers to decay to the heaviest particles kinematically possible, especially when the decay products are on shell. Therefore the branching ratio of each Higgs decay channel changes dramatically with the assumed Higgs mass M_H. The heavier the Higgs mass, the more decays into relatively light particles lose importance and new channels open up. In the SM the Higgs can decay either to massive fermion or boson pairs, shown in fig. 2.8 and in fig. 2.6, or to massless particles (gluons, photons) via loops (once again mostly via top loops), as shown in fig. 2.7.

An overview is given in fig. 2.10, where the branching ratio for the main decay channels is shown as a function of the Higgs mass.

The decay channel discussed in this work, $H^0 \to ZZ \to 4\mu$ opens up at around 100 GeV, when the decay to one off shell and one on shell Z becomes possible. An off shell particle means a virtual particle with a mass different from the measured values for the real particle (called on shell particle). Virtual particles can exist a certain time according to $\Delta E \Delta t \geq \hbar/2$, but they cannot be observed and are, in fact, only a part of a quantum probability calculation for Feynman diagrams. Their production cross section is smaller than the cross section for on shell particles and so decay into on shell particles are more probable, but not always energetically possible
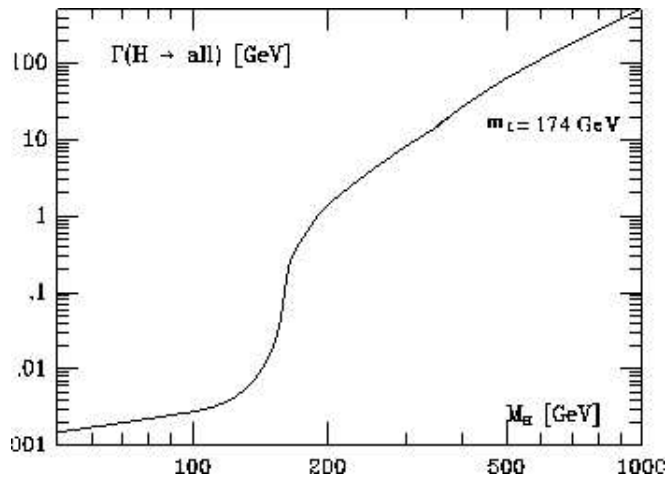
Figure 2.5: Total decay width Γ_H of the SM Higgs boson increases with the Higgs mass M_H. For ($M_H < 200$ GeV) the width is smaller than 2 GeV.
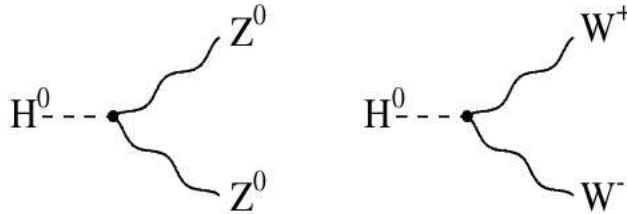

Figure 2.6: LO Higgs decays to vector bosons Z^0 et $W^{\pm}$

(as in this case). The decay to two Z bosons persists together with the decay to WW the dominant mode.

The gauge bosons are short living particles and cannot be detected directly, therefore final branching ratios are multiplied by the branching ratio corresponding to the decay to observed products (e.g. here $\sigma_{ZZ \rightarrow 4\mu}$). However, the Higgs coupling to the Z bosons is much stronger then the coupling to the other, much lighter particles (τ, b, etc.), as shown in fig. 2.10, even though one Z boson is off shell and therefore the production cross section σ is smaller. Therefore this branching channel is important.

The decay to top quarks is suppressed, for in this case, none of the quarks can be on shell. As soon as the Higgs is heavy enough to decay to two on-shell $W^{\pm}$ bosons, the decay to Z^0 bosons is suppressed until the decay to two on shell Z^0 bosons becomes possible as well. With the CMS experiment at the LHC it should be possible to discover the Standard Model Higgs boson in the whole mass region with a significance $\sigma > 5$, where $\sigma = N_S/\sqrt{N_B}$, N_S being the number of signal events and N_B being the number of background events in a considered interval. The significance is the quantity that describes whether a peak in a distribution is a real signal or rather a statistical fluctuation. If $\sigma > 5$ the peak (in our case) is statistically significant and a strong hint for new physics.

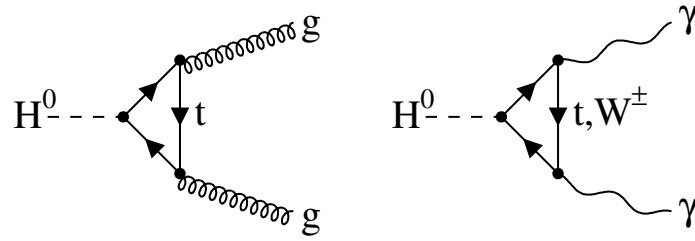A summary of the expected situation is given in fig. 2.11.

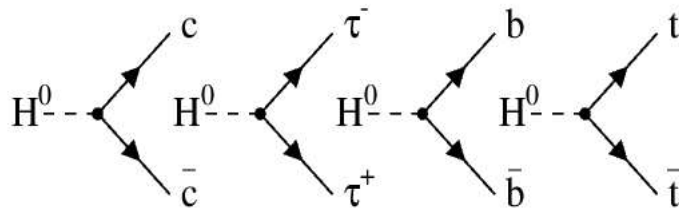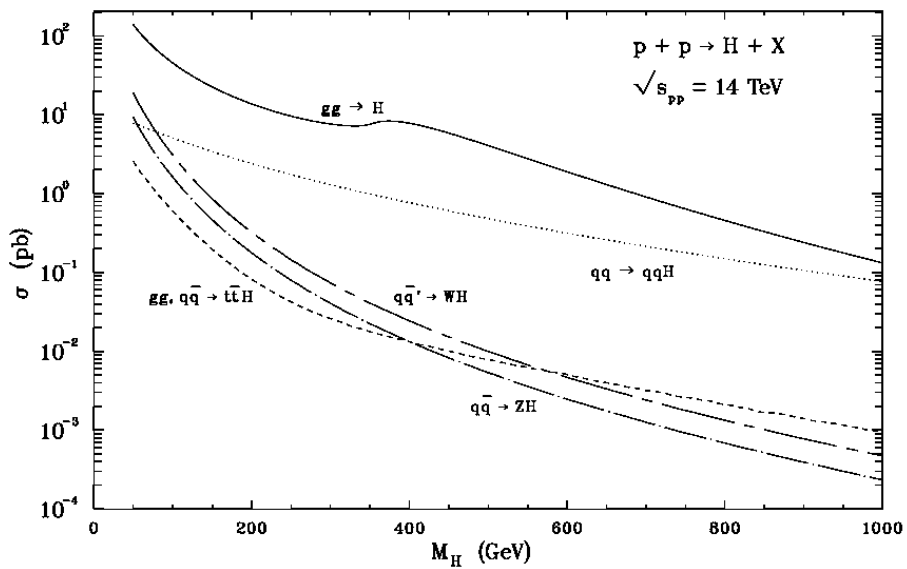Figure 2.7: Higgs decay to massless particles at LO via loops

Fig
ma

Fig. 5b

Figure 2.9: Higgs production (SM) cross sections σ at the LHC versus the Higgs mass M_H for the most important processes
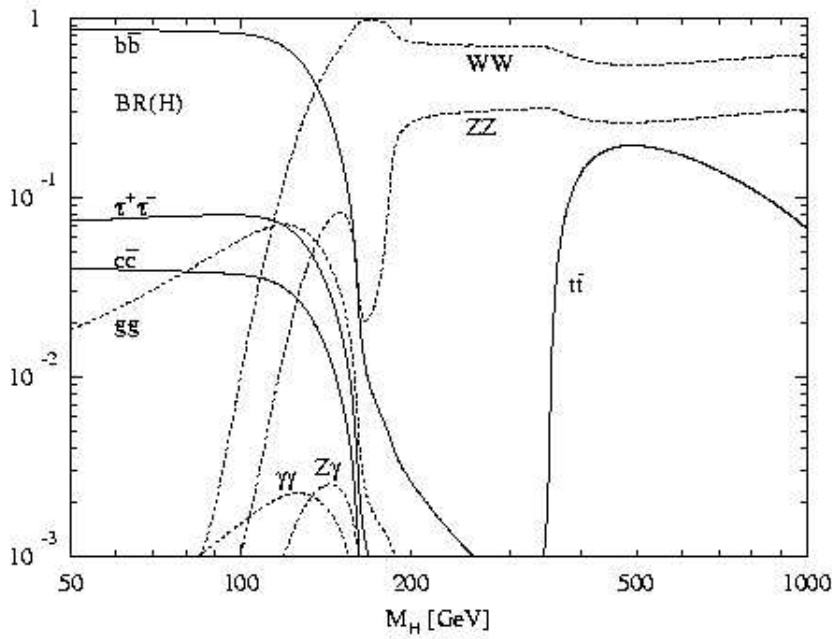
Figure 2.10: Main branching ratios BR(H) of the Higgs decay as a function of the Higgs mass in the SM
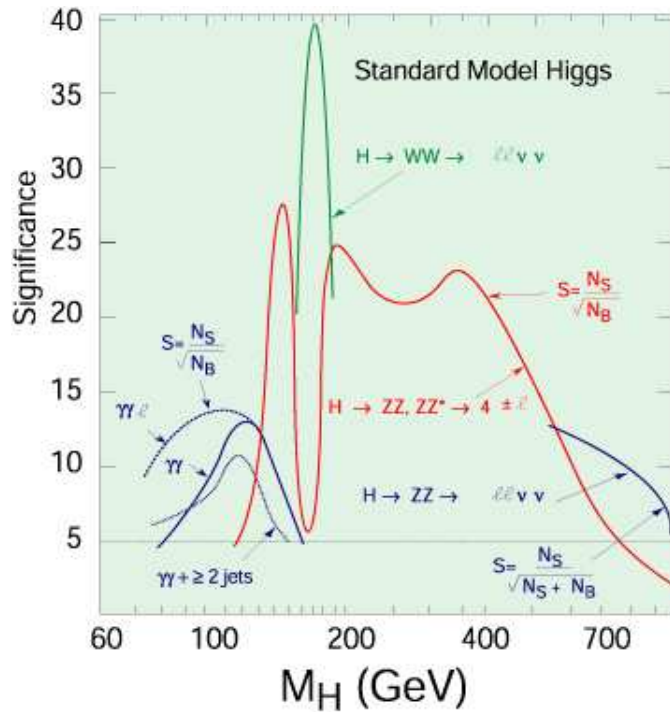

Figure 2.11: The significance $N_S/\sqrt{N_B}$, where N_S is the number of signal events and N_B the number of background events in the considered interval, as a function of the Higgs mass M_H in the SM, $L = 10^5 pb^{-1}$

Chapter 3

The Large Hadron Collider (LHC) and the CMS detector

3.1 The Large Hadron Collider
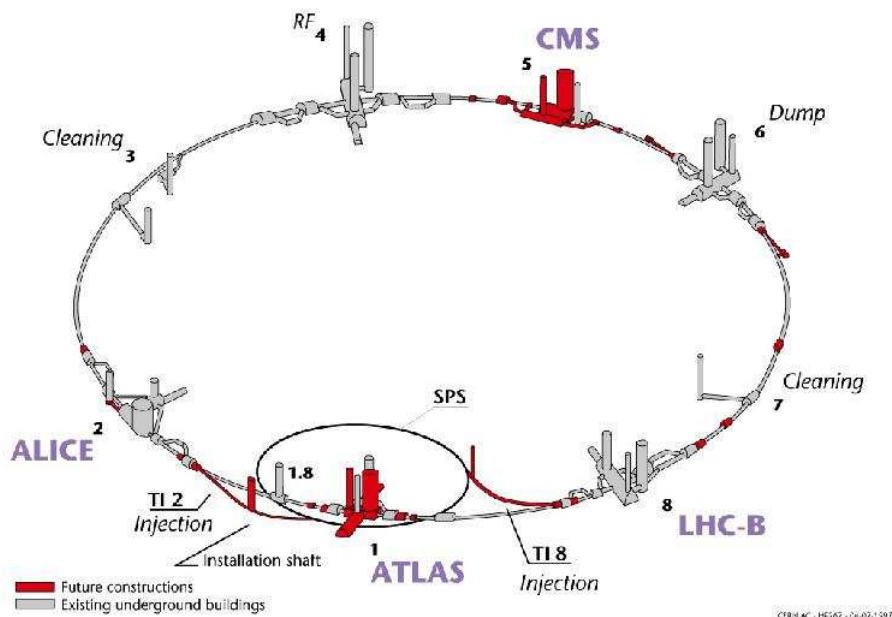
Figure 3.1: Layout of the LEP/LHC tunnel with the four detectors ALICE, ATLAS, LHCb and CMS

The Large Hadron Collider, which is under construction at CERN [1], is supposed to start the production of proton-proton collisions in 2007, with a center of mass energy of 14 TeV. Two proton beams each with energies of 7 TeV circle in opposite directions in a superconducting ring which is installed in the 26.7 km long LEP /LHC tunnel shown in fig. 3.1. The most important goal is the search for the electroweak

---

[1]Conseil Europeen pour la recherche Nucleaire

Collision Energy	$\sqrt{s} = 14$ TeV
Magnetic Fields	8.36 T
Total length	26.7 km
Diameter	8.6 km
High Luminosity	$10^{34} \text{cm}^{-2} s^{-1}$
Low Luminosity	$10^{33} \text{cm}^{-2} s^{-1}$
Intergrated Luminosity (one year, low luminosity) L_{int}	$20 fb^{-1}$
Bunch time separation	25 ns
Number of Bunches	2835
Number of particles per bunch	1.1×10^{11}

Table 3.1: Main LHC parameters

symmetry breaking, i.e. the search for the Higgs bosons. Among other interesting subjects there are the search for supersymmetric particles, Standard Model tests, CP violation etc. At four interaction points the detectors are foreseen : ALICE (A Large Ion Collider Experiment), ATLAS (A Toroidal LHC Apparatus), CMS and LHCb (Large Hadron Collider beauty experiment).

The design value of the luminosity at CMS and ATLAS interaction points is $L = 10^{34} cm^{-2} s^{-1}$ for high luminosity and $L = 10^{33} cm^{-2} s^{-1}$ for low luminosity. This lumnosity is achieved by filling each of the two rings with 2835 bunches of 10^{11} particles. Superconducting niobium-titanium magnets produce magnetic fields up to 8.36 T to focus and force the proton beams into the right trajectories - beam lifetimes of at least ten hours are expected. A collection of the most interesting LHC parameters is presented in tab. 3.1.

3.2   The CMS detector

The CMS detector [TDRcms] has been particularly designed to detect the Higgs boson, as well as possible signatures from new physics or particles at the LHC, even at low luminosity. The detector is built by an international collaboration consisting of over 1800 physicists from 150 institutions from 31 countries. A general view of the CMS detector is shown in fig. 3.2. The detector is divided into a barrel region and two endcap regions; Its central part is a 13 m long, superconducting solenoid of 6 m diameter - it is the biggest superconducting solenoid magnet with the highest field and stored energy ever. Overall it is nearly 22 m long with a width of 14.6 m, and a total weight of about 12500 tons. It will generate a field of 4 Tesla, which corresponds to a stored energy of 2.5 Giga Joules.

CMS is a large technologically advanced detector made up of many layers. Each of these is designed to perform a specific task and together they will allow CMS to identify and precisely measure the energies of all the particles produced in LHC
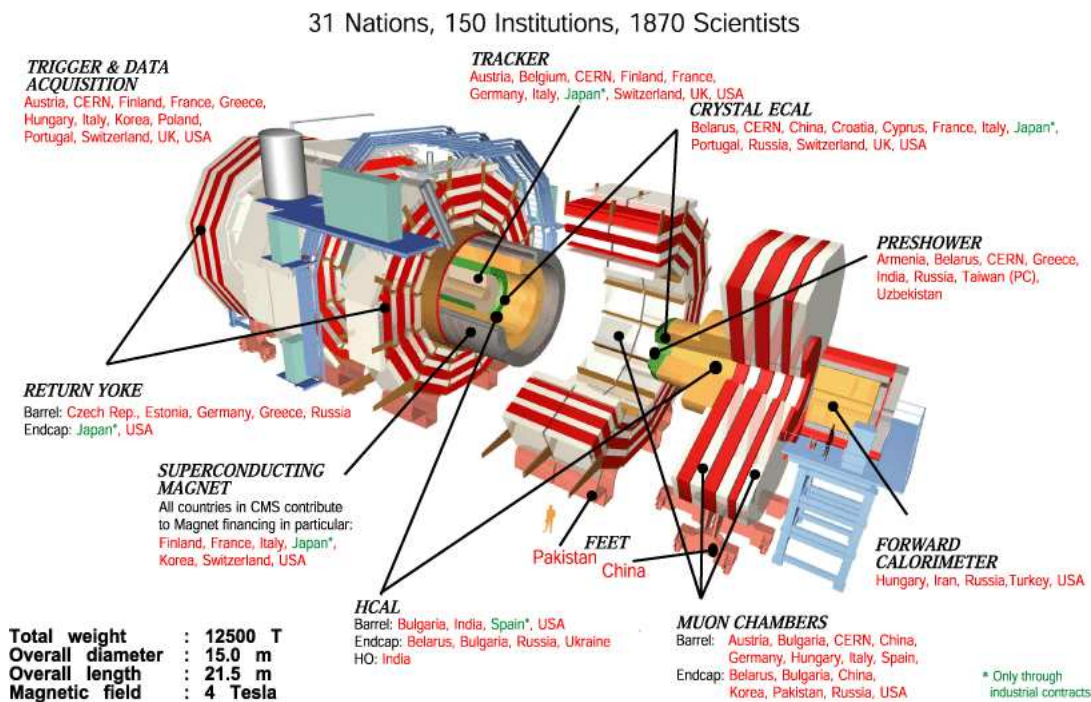
31 Nations, 150 Institutions, 1870 Scientists

TRIGGER & DATA ACQUISITION
Austria, CERN, Finland, France, Greece, Hungary, Italy, Korea, Poland, Portugal, Switzerland, UK, USA

TRACKER
Austria, Belgium, CERN, Finland, France, Germany, Italy, Japan*, Switzerland, UK, USA

CRYSTAL ECAL
Belarus, CERN, China, Croatia, Cyprus, France, Italy, Japan*, Portugal, Russia, Switzerland, UK, USA

PRESHOWER
Armenia, Belarus, CERN, Greece, India, Russia, Taiwan (PC), Uzbekistan

RETURN YOKE
Barrel: Czech Rep., Estonia, Germany, Greece, Russia
Endcap: Japan*, USA

SUPERCONDUCTING MAGNET
All countries in CMS contribute to Magnet financing in particular: Finland, France, Italy, Japan*, Korea, Switzerland, USA

FEET
Pakistan China

FORWARD CALORIMETER
Hungary, Iran, Russia, Turkey, USA

HCAL
Barrel: Bulgaria, India, Spain*, USA
Endcap: Belarus, Bulgaria, Russia, Ukraine
HO: India

MUON CHAMBERS
Barrel: Austria, Bulgaria, CERN, China, Germany, Hungary, Italy, Spain,
Endcap: Belarus, Bulgaria, China, Korea, Pakistan, Russia, USA

* Only through industrial contracts

Total weight	: 12500 T
Overall diameter	: 15.0 m
Overall length	: 21.5 m
Magnetic field	: 4 Tesla

Figure 3.2: CMS detector in 3D

proton-proton collisions. The layers of the CMS detector are arranged like a cylindrical onion around the collision point.

The basic detector layers are, from inside out :

- Inner tracking system to measure the momenta of charged particles in the magnetic field. It consists of pixel detectors and silicon micro strip detectors.

- Electromagnetic calorimeter(ECAL) to determine the energies and locations of photons and electrons. The ECAL consists of $PbWO_4$ crystals.

- Hadron calorimeter (HCAL), a copper scintillator sandwich, to detect jets and determine missing energy.

- Very forward calorimeter (VFCAL) to complete the measure of the missing energy and to tag forward jets; it has a similar structure as (HCAL).

- Muon system to identify and measure muons. It consists of drift tubes and cathode strip chambers.

Some further information can be found in tab. 3.2, a detailed description is avaliable in e.g. [TDRcms].

A particle emerging from the collision and traveling outwards will first encounter the tracking system - this will measure precisely the positions of passing charged particles allowing to reconstruct the tracks. Charged particles will follow spiraling

Layer	Geom. Accep.	Resolution
Inner tracking system	$|\eta| < 2.4$	$\Delta P_t/P_t = 1\%$ (5%)
ECAL	$|\eta| < 3$	Barrel $2.7\%/\sqrt{E}$; Endcap $5.7\%/\sqrt{E}$
HCAL	$|\eta| < 3$	$\Delta E/E = 110\%/\sqrt{E} + 5\%$
Muon system	$|\eta| < 2.4$	$\Delta P_t/P_t = 1\%$ (5%) for 100 GeV (1TeV)

Table 3.2: The resolution in E_t and P_t and the Geometrical Acceptances for the main CMS layers. Instead of θ often the pseudorapidyty η is used; Definition of Pseudorapidity: $\eta = 1/2 \; ln((P_{tot} + P_z)/(P_{tot} - P_z)) = 1/2 \; ln((1 + \cos\theta)/(1 - \cos\theta))$
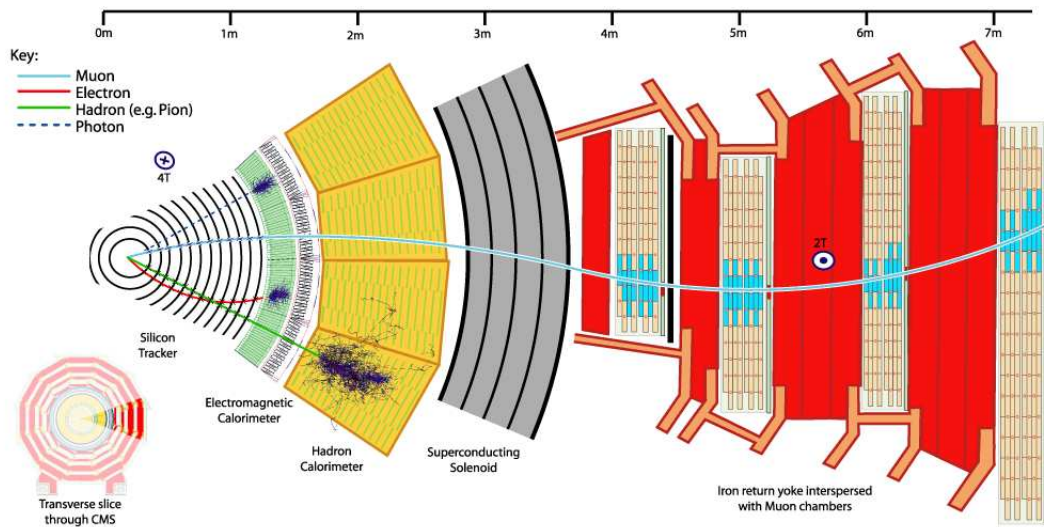
Figure 3.3: Overview over particles tracks in the CMS detector

paths in the CMS magnetic field and the curvature of their paths will reveal their momenta. The energies of the particles will be measured in the next layer, the so called calorimeters. Electrons (e^-, e^+), photons and particle jets will be stopped in the calorimeters, allowing their energy to be determined. The first calorimeter is designed to measure the energy of electrons and photons with high precision- since these particles interact electromagnetically, it is called an electromagnetic calorimeter. Particles which interact by the strong interaction, hadrons, deposit most of their energy in the next layer, the hadronic calorimeter. It is made of copper and plastic sandwiched together-the copper saps the energy of particles while the plastic produces light as they pass allowing to measure their energy. The only particles to penetrate beyond the HCAL are muons and neutrinos. The muons will be measured in the muon chambers - their momentum will be estimated from the bending of their paths in the CMS magnetic field. Neutrinos hardly interact and their presence will only be seen by adding up all the momenta of all the detected particles ("Missing E_t or P_t"). An illustration of the tracks of the particles mentioned above can be seen in fig. 3.3.

All in all, CMS will have 15 000 000 individual channels, all of which will be controlled by powerful computers. They will synchronize the detector with the LHC

accelerator, making sure that CMS is ready to record interesting collisions. It is expected to have on average 800 millions collisions per second. Not all of theses will produce interesting results - most of the time protons will just graze past each other. Head-on collisions will be rare, and the processes which produce new particles rarer still. The Higgs boson, for example, is expected to appear about once every day. The Main Trigger concepts to cope with these data amount will be described in the next section.

3.3 Data acquisition with the CMS detector

For a bunch crossing frequency of 40 MHz and a design luminosity of $10^{34} cm^{-2} s^{-1}$ approximately 25 inelastic collisions occur every 25 ns corresponding to an interaction rate of the order of 1 GHz. This input rate of 10^9 interactions every second has to be reduced by a factor of at least 10^7 to 100 Hz, in order to match the capabilities of the mass storage and offline computing systems. For this purpose, CMS plans to have a multi-level trigger system. The first stage of this rate reduction is performed by the Level-1 (L1) trigger which decides the acceptance or rejection of an event within a few microseconds after a collision.

During that period the full detector information is kept in the detector front-end pipelines. The maximum output rate of the L1 trigger is 100 kHz, which is determined by the speed of the detector electronics readout and the input of the data acquisition system. To account for the limited reliability of rate predictions a safety factor of three is taken into account, therefore a maximum L1 rate of 30 kHz is foreseen. This rate is shared equally between muon and calorimeter triggers. The L1 trigger system is organized into three subsystems: the L1 calorimeter trigger, the L1 muon trigger, and the L1 global trigger. The second stage is performed in software by the High Level Trigger (HLT) system (normally L2 and L3).

The HLT stages have much longer processing times and are therefore based on computer farms. Further offline analysis is planned to be also performed by global networks like e.g the GRID.

The quality of the selection algorithms is of high importance. To guarantee that, the HLT algorithms are designed using modern object oriented software techniques and implemented in C++, using well defined interfaces to a common framework. The L2 Trigger rate is designed to be about 5 KHz, the L3 Trigger rate about 100 Hz.

How the multi level Trigger system works in case of the muons is described more detailed below.

3.4 Data acquisition with the muon system

The muon system consists of 3 level trigger [Neu]. The L1 muon trigger is organized into subsystems representing the different muon detector systems, the Drift Tube (DT) trigger in the barrel, the Cathode Strip Chamber(CSC) trigger in the endcaps up to $|\eta < |2 : 4$ and the Resistive Plate Chamber (RPC) trigger covering both the
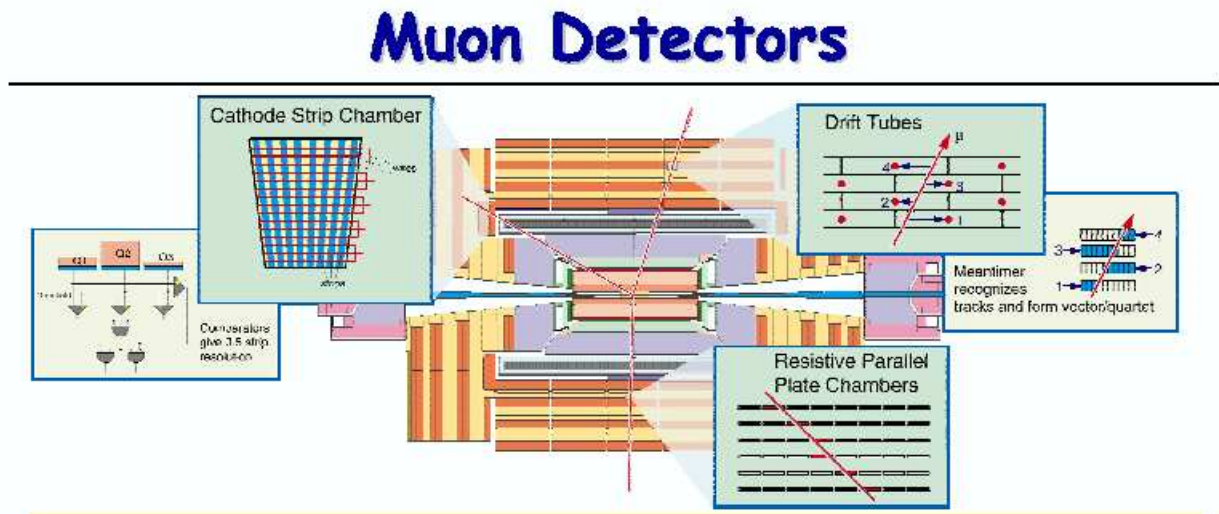
Figure 3.4: Level 1 Muon Trigger at CMS consisting of the Drift Tube (DT) trigger , the Cathode Strip Chamber(CSC) and the Resistive Plate Chamber (RPC) trigger

barrel and the endcap regions up to $|\eta| < 2.1$. The L1 muon trigger has a Global Muon Trigger (GMT) that combines the trigger information from the DT, CSC and RPC trigger systems and sends this to the L1 Global Trigger. Centrally produced muons are measured three times: in the inner tracker, after the coil and in the return yoke. There are two Muon Triggers: A Single Muon Trigger with a threshold of 20 GeV and a di-muon trigger with a threshhold of 7 GeV each - if one of this Trigger thresholds is reached, the event is stored. These thresholds are chosen to get mostly so called "prompt" muons triggered, that means muons from decays of W, Z ,Higgs, top, b and c quark decays - these are the events that want to be kept. Non prompt muons (for example ϕ, K $\pm$ decays, punchthrough of hydronic showers, cosmic muons etc.) have mostly lower P_t values.

At L2 Trigger Level only muon and calorimeter data are used, but in contrast to the Level-1 trigger it has access to event data with the full detector granularity and resolution. This allows to reject fake Level-1 muon candidates and reduces the contribution of non-prompt muons. The resulting P_t resolution is of the order of $\Delta P_t/P_t \approx 12\%$ and the L1-rate is reduced by a factor of 10.

The Level-3 muon trigger performs a matching of muon and tracker data. Using the full tracker information significantly improves the muon P_t measurement to $(\Delta P_t/P_t \approx 1.5\%)$. A more detailed description of the algorithms used by he High Level Trigger is given in chapter 5.

Chapter 4

The local computing infrastructure

The full detector simulation including all the physical processes that are expected to happen at LHC/CMS requires a great amount of CPU power. In 2002 a computing cluster with 18 computing nodes was established at the university Karlsruhe (not only for the CMS group) to achieve this aim.

Figure 4.1: A visual impression of the cluster with the file server, a portal machine and the computing nodes

The computing infrastructure is complex and it was challenging to configure this environment. The first general problem is, that clusters like this need certain working conditions, for example a constant room temperature. In case of a failure of the cooling system in the cluster room, the temperature rises rapidly (more than 20 CPUs!) and a damage of the expensive hardware is risked. Therefore it is very important to have a monitoring tool that watches the important parameters and can automatically shut down the cluster in case of emergency.

Our strategy was twofold. On the one hand, we installed temperature sensors in the room and, on the other hand, we read out the mainboard temperature sensors. In more detail, we wanted to use the CPU and the mainboard temperature sensor and, to prevent a fan failure, also to watch the fan speeds. The actual data and its average temperature, summaries of the last days and weeks etc. should be displayed on a web page.
In the following a brief overview of the network layout is given and the main concepts of the monitoring tool are described.

4.1 Hardware setup and the network layout

Figure 4.2 shows a schematic view of our network structure. There are 18 computing nodes (5 x Athlon XP 1600+ and 13 Athlon T-bird 1300 MHz, with 512 MB SD-RAM, 40 GB hard disk space, 100 Mbit ethernet cards) connected to a file server and to the portal machines via a private IP link (each node has a static IP address in the range of 192.168.101.0 - 192.168.101.18). The portal machines have a public IP address, i.e they are accessible from other domains, while the computing nodes are only visible from the portals inside the private network. A detailed description of the setup and configuration of the cluster can be found in [Pat].
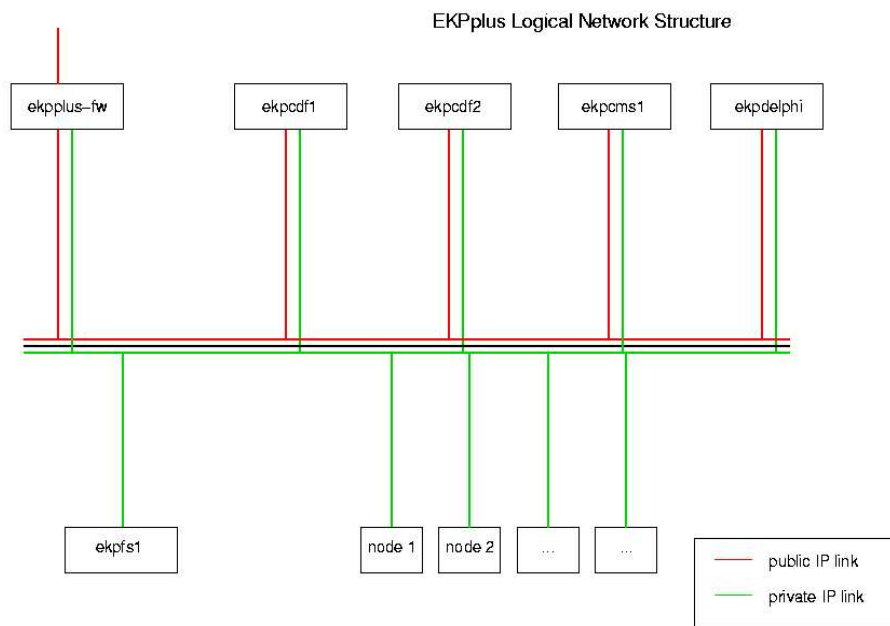
Figure 4.2: Layout of our network topology: Above the firewall ekpplus-fw and the 4 portal machines ekpcdf1 and ekpcdf2, ekpcms1 and ekpdelphi, below the file server ekpfs1 and the nodes are shown

The monitoring tool should be able to connect to all the computing nodes and the portal machines, i.e. it has to run on a machine inside the private network.

4.2 Development of the temperature monitoring tool

In order to access the sensors of the ASUS A7S-VM mainboards of each node from the employed Linux operating system, we decided to use the package lm-sensors. The sensor package is included in many Linux distributions and can also be downloaded [Lm].

Although our mainboard type was not officially supported by lm-sensors, we managed to find a module that works, with the helpful support of the developers. This tool returns the temperatures, fan speeds and a collection of voltages, an example of the output is given below:

```
w83782d-i2c-0-2d
Adapter:  SMBus AMD756 adapter at 50e0
Algorithm:  Non-I2C SMBus adapter
VCore 1:  +1.61 V (min = +1.40 V, max = +1.69 V)
VCore 2:  +2.51 V (min = +1.40 V, max = +1.69 V)
+3.3V: +3.32 V (min = +2.97 V, max = +3.63 V)
+5V: +4.94 V (min = +4.50 V, max = +5.48 V)
+12V: +12.01 V (min = +10.79 V, max = +13.11 V)
-12V: -11.47 V (min = -13.21 V, max = -10.90 V)
-5V: -5.07 V (min = -5.51 V, max = -4.51 V)
V5SB: +5.07 V (min = +4.50 V, max = +5.48 V)
VBat:  +3.15 V (min = +2.70 V, max = +3.29 V)
fan1:  0 RPM (min = 3000 RPM, div = 2)
fan2:  4856 RPM (min = 3000 RPM, div = 2)
fan3:  0 RPM (min = 750 RPM, div = 8)
temp1:  +42.0C (limit = +60C, hysteresis = +50C) sensor = thermistor
temp2:  +56.0C (limit = +60C, hysteresis = +50C) sensor = thermistor
temp3:  +211.0C (limit = +60C, hysteresis = +50C) sensor =
PII/Celeron diode
```

In a configuration file it is possible to customize the output of lm_sensors. Our first approach was to write a daemon that runs on each node and that is able to send via TCP/IP various informations about the status of the node, e.g. load and temperature. Technically, the daemon is a main program with a plug-in for each monitored quantity. Therefore new features can be added easily with new plug-ins and the interface to the main program is always the same. In case of the temperature, the lm_sensors program is executed, its output is parsed and the CPU temperature, mainboard temperature and fan speeds are send to a central client. This client provides a logging file for each node and sends every 6 minutes a temperature data request to the daemon, which returns a confirm that it is still running and sends the current data. The 18 logging files are supposed to be visualized and made available via a web page. For this task a graphical user interface using ROOT [Bru] was written in C++.
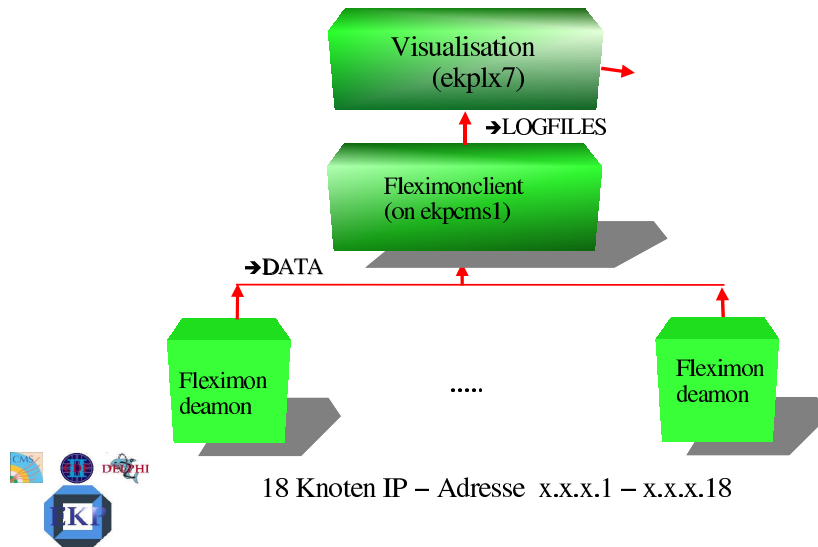
Figure 4.3: The logical structure of the temperature monitoring tool: A daemon runs on each node and sends its data to a central client. This client provides logging files and performs the visualization

ROOT is not only a powerful and comfortable histogram and analyzing framework written in C++, it offers also the possibility to create graphical user interfaces (this feature is less comfortable, however). Here, a simple graphical interface was written that offers the possibilities to start and stop the monitoring and displays the current histogram of the monitored value for each node. All histograms are automatically displayed on a HTML web page.

In summary, the whole program consists of a daemon, written in C and running on each node, a client writing logging files for each node and a graphical interface visualizing the data and computing mean values and standard deviations. This concept worked well. However, then we decided the client to run on a machine without an X-Server, therefore in the present version the graphical user interface cannot be used. Instead, the client creates directly postscript objects. The other parts of the temperature monitoring tool are running at all times with a good performance. In case of emergency, i.e. when certain thresholds (of temperature, for example) are exceeded, the system is automatically shutting down.

4.3 Computing setup for the simulation

The setup of the CMS Software packages in this computing environment was cumbersome. A simplified overview of the infrastructure is given in fig. 4.5. The whole software surrounding including all tool packages is installed on the portal machine. Via NFS (Network File System) mounts the necessary paths are visible on the computing nodes (Black thin arrow). NFS protocol provides transparent remote access to shared file systems across networks. The nodes are supposed to access the necessary executables from the portal and additional data (Red arrow) from either the
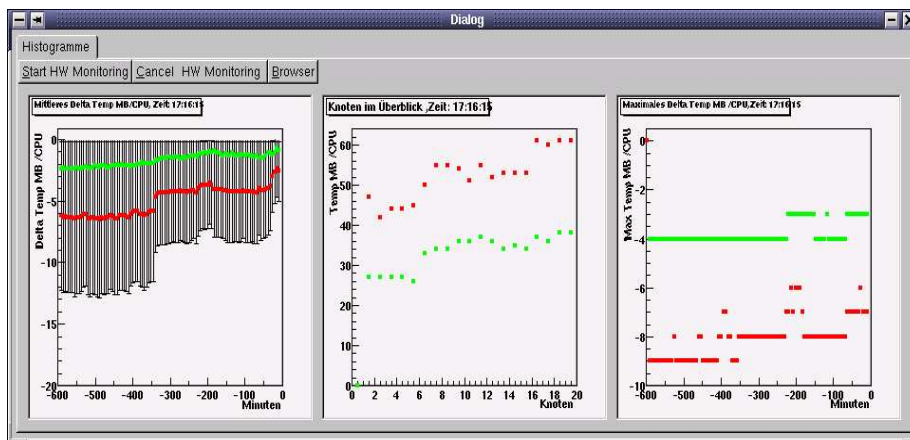
Figure 4.4: Visual impression of the graphical interface for the temperature monitoring tool with start and stop buttons and 3 different histograms

portal machine or the two file servers. The file servers are again mounted on the portal and computing nodes via NFS. After the job is terminated the result is written on the file server or the portal machine (Green arrow).

Job submission is managed by the open source batch system Open PBS, which offers a short, medium and long queue to the user and looks after a fair allocation of the resources. For the submission of jobs to the queues one has to provide a shell script, that sets the environment and defines input and output parameters. Open PBS offers a graphical user interface (XPBS) which is intuitive to use and helps keeping an overview over the jobs and their status. A visual impression of XPBS is given in fig. 4.6.

The main points leading to additional efforts necessary when installing the CMS software in this environment were in summary:

1. Our new hardware was not fully supported by the operating system used at CERN. Therefore several patches had to be performed.

2. CMS jobs had to be submitted via PBS

3. To keep things manageable, paths with a unique name convention (i.e portal paths begin with /portal/portal name/....) were used for our cluster and cloning of CERN paths like "/cern" was impossible. Also there was not a central afs server that is used at CERN for distributed file access. The CMS distribution of the software packages is based on shell scripts, environment variable and the management tool SCRAM. Unfortunately, several shell scripts contained hard coded path names like "/cms", "/cern" or "/afs", as well as automatically set environment variables.

4. Some files were simply missing and had to be added.

.. 18 Nodes ..

→ NFS Mount

→ Software
 + Data

→ Computing
 output

Portal : ekpcms1

Fileserver 1
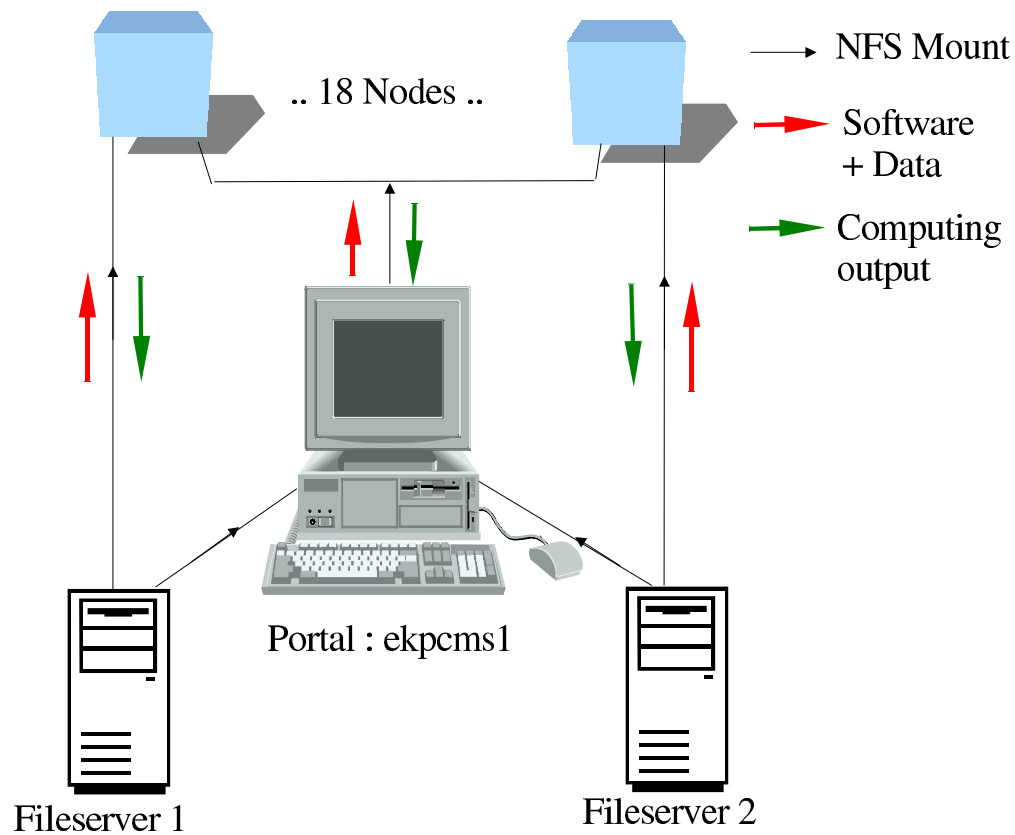
Fileserver 2

Figure 4.5: The structure of the computing environment

5. Instead of a central AMS server, responsible for the access of objectivity databases, we accessed our databases e.g. on the fileserver via NFS. Objectivity via NFS causes the main problems, when accessing the same dataset from different nodes at the same time.

After several months, however, we completed the installation of the CMS software at Karlsruhe.
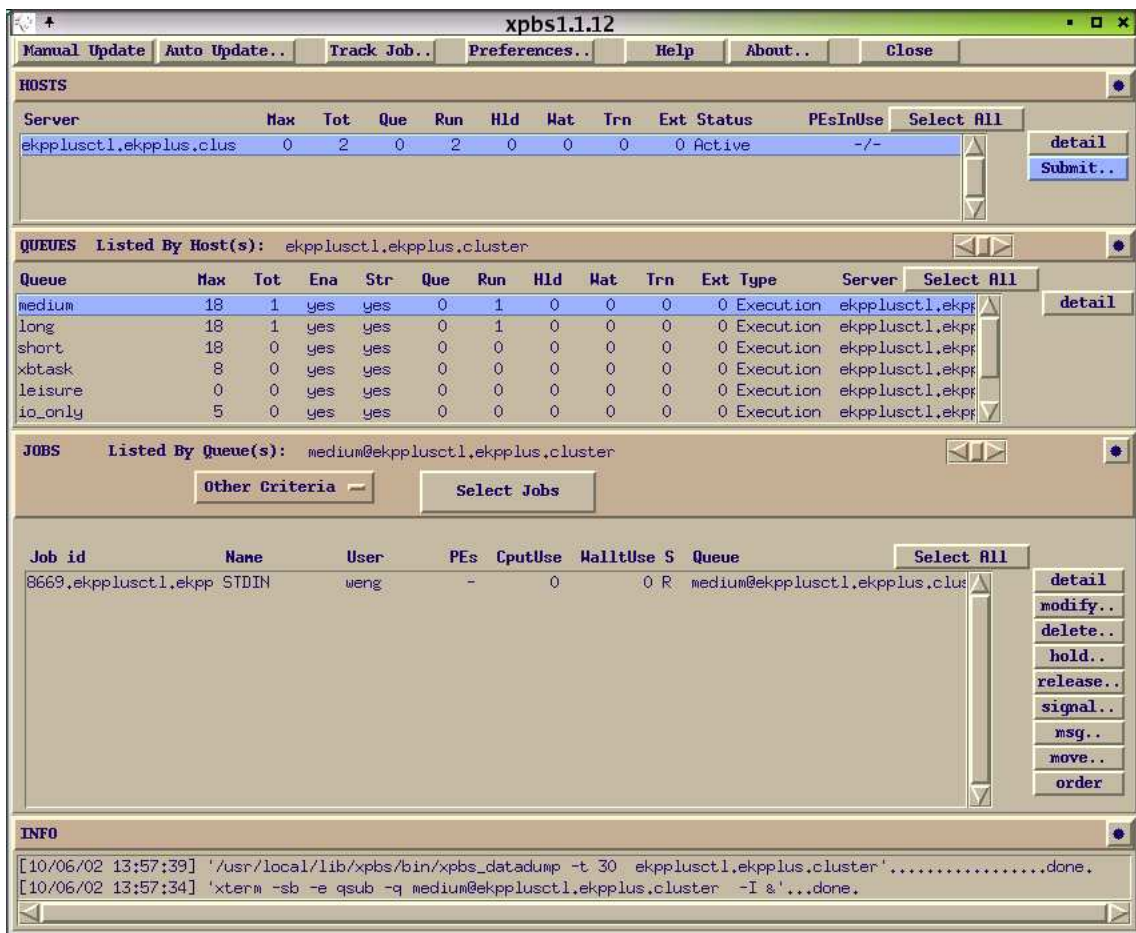
Figure 4.6: Xpbs - the graphical front end of the batch system OPEN PBS

Chapter 5

The CMS detector simulation and the analysis framework

5.1 Event generation and full detector simulation

The CMS Software to perform a full detector simulation is a complex collection of standalone programs and tool kits. The main software packages are (not complete):

1. *PYTHIA* [PHY] and *COMPHEP* [COMPH] : These are the two event generators that were used (PYTHIA 6.158 and COMPHEP 41.10).

2. *CMSIM* [CMSIM] : The full detector simulation based on GEANT3, written in Fortran. Its successor is Oscar, based on GEANT4 , the object oriented version of GEANT. We used CMSIM 125.

3. *OBJECTIVITY* : A commercial package to create and administrate databases.

4. *COBRA* [COBRA] (Coherent Object-oriented Base for simulation, Reconstruction and Analysis): A general framework with many necessary tools which is supposed to provide the software common to ORCA, IGUANA and (in the future) OSCAR. We used COBRA 6.2.1.

5. *ORCA* [ORCA] (Object-oriented Reconstruction for CMS Analysis) : A toolkit in C++ containing the classes and methods to access the reconstructed data from different detector parts. We used ORCA 6.2.3.

6. *IGUANA* [IGUANA] (Interactive Graphics For User Analysis)) : A graphical tool for the inspection of databases and viewing the CMS detector. An visual impression of IGUANA is shown in fig. 5.2. We used IGUANA 3.1.0.

7. *SCRAM* (Software Configuration, Release and Management): This tool performs version management and sets up the environment for the user to create and run executables. The basic management of source code is done with CVS (Concurrent Versioning System). We used SCRAM V0_19_6.
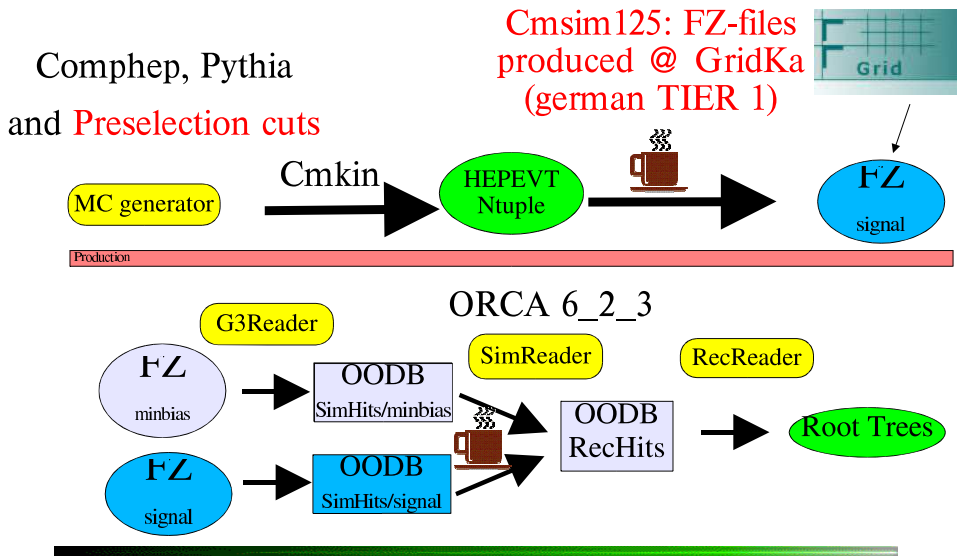
Figure 5.1: CMS simulation chain

In the following it will be described how these single programs form together the full simulation and analysis chain.

First one needs an event generator to create the signal and background events. The included tool CMKIN is used to call the correct Pythia routines and set LHC specific parameters automatically. There are example programs distributed, that have to be adapted to the particular local environment. They are controlled by data cards, where the physical process has to be specified. Also preselection cuts can be set here in event-selection subroutines- The output of all these FORTRAN based routines is an HEPEVENT Standard n-tuple. Because Pythia does not contain all processes needed for the analysis, also COMPHEP had to be used as event generator with an interface to cmkin for the $Zb\bar{b}$ background. The HEPEVENT Ntuple serves as input for the detector simulation, which reads in the geometry of the CMS detector and simulates the interaction with the detector. This is the longest part of the simulation, it takes about one minute per event, depending on the number of particles and the topology of the event. The output are ZEBRA files, so called FZ-files, that are typically very big for one n-tuple of e.g. 10000 events which therefore has to be divided because otherwise the maximum file size is exceeded. For this reason we wrote several FZ-files (normally 500 events per FZ-file) and gave each a different run number. These FZ-files can now be stored in the objectivity database specified by the environment variable OO_FD_BOOT by running the two ORCA jobs "writeHits" and subsequently "writeAllDigis". This database can than be accessed by analysis programs, like e.g. the example programs in ORCA or the analysis program developed during this thesis, KaMuonAnalysis. The latter is described in the next chapter.

All in all the full simulation takes at least 2 minutes / event. For this reason it is very important to simulate fully only relevant events, i.e. to perform some efficient preselection cuts.
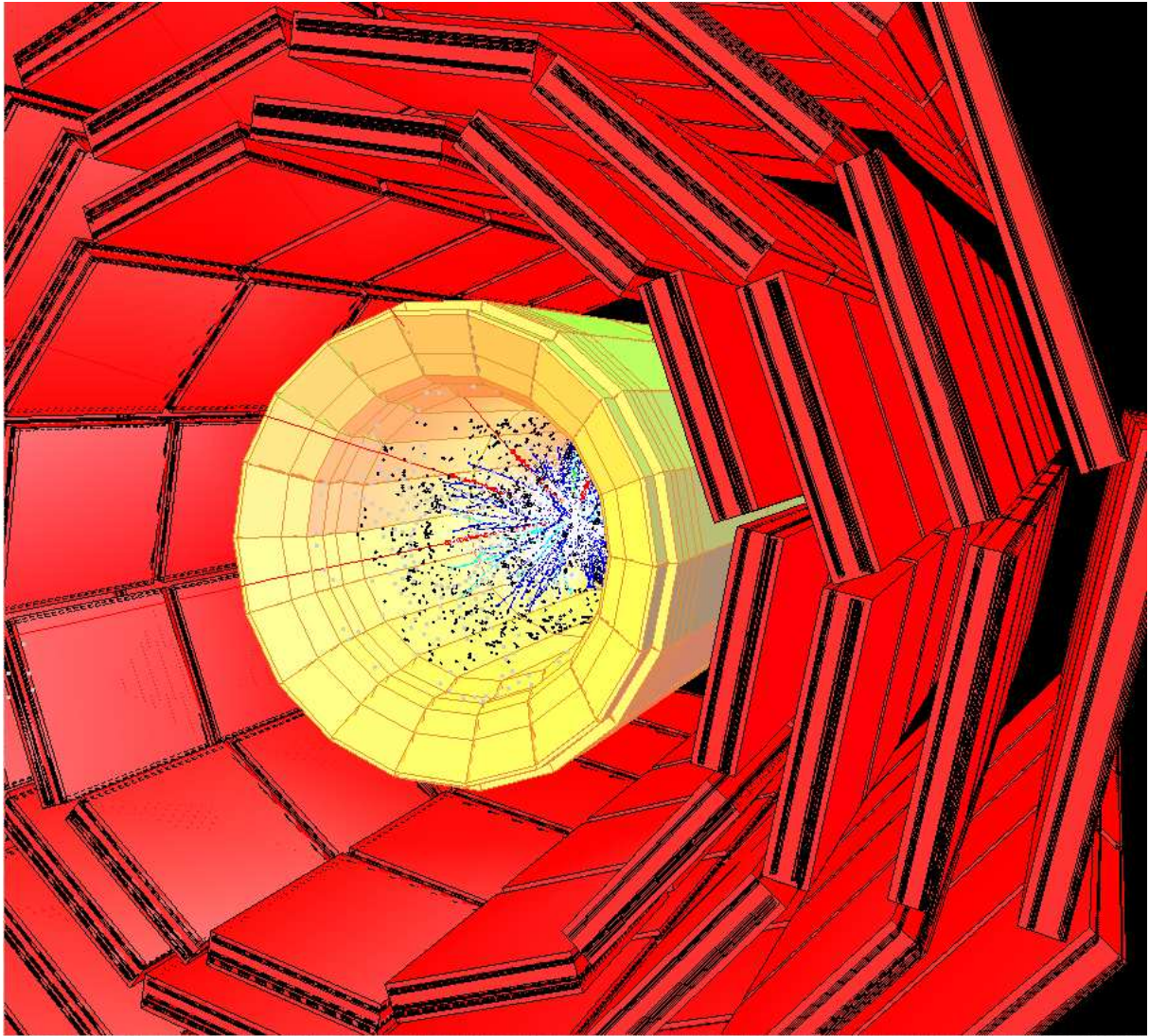
Figure 5.2: A visual impression of IGUANA. Shown is a 4 μ event from a database produced at Karlsruhe. One can see the calorimeter and the muon chambers.

Once the objectivity databases with the Monte Carlo data and the data of the full detector simulation are produced, the aim is to get the desired muon information. During this work over 3000 lines of code were developed to perform the analysis in the CMS framework.

5.2 The main program - the class KaMuonAnalysis

In ORCA the user has to define an own analysis class that performs the desired task
- in this case the class KaMuonAnalysis. The class has a constructor, a destructor
and (among other methods) the public method `analysis`, which will actually do
the analysis. Furthermore, there is a private method `update`. This method is
mandatory and will be called if a new event is processed by ORCA. Here it will
then just call the analysis method of the class, provided the pointer to the event is
not the NULL pointer. All methods and data members are defined in the header
file KaMuonAnalysis.h, of which a documentation sheet created with DOXYGEN
is included in the appendix; (The whole code is documented with doxygen and
a reference manual, either in html or as postscript, is available.) Since the data
members are private they can only be modified by methods of our analysis class.
The whole software structure within this framework is complex and it is useful to
have a closer look at the implementation of the main methods. The constructor
is automatically called when a new instance of the class is created, necessary
initializations are done there:

```
KaMuonAnalysis::KaMuonAnalysis() {
e.g.:  file=new TFile( ''test.root '', ''recreate ''); // new ROOT
file
int bla = 0; // bla = 0 :-)
cout << ''========================================== '' << endl;
cout << '' Constructor called ''<<endl;
cout << ''========================================== '' << endl;
}
```

Similarly, the destructor is automatically called whenever an instance of the class
is deleted. Consequently, this is the place to do some final calculations and print
a summary. One should also delete class pointers and close the ntuples / rootfiles
used.

```
   KaMuonAnalysis::~KaMuonAnalysis() {

e.g.  close (file); // (closing the root file)

e.g.  delete class pointer; // (deleting class pointer)
cout << '' ========================================= '' << endl;
cout << '' Number of events analyzed:  ''<< eventsAnalysed << endl;
cout << '' Number of runs analyzed:  ''<< runsAnalysed << endl;
cout << '' ========================================= '' << endl;
}
```

Finally the method where the analysis is done:

```
   void KaMuonAnalysis::analysis(G3EventProxy* ev) {
cout << ''========================================== '' << endl;
```

```
cout << '' Muon Analysis processing event (run:event) ''
<< ev->simSignal()->id().runNumber() << '':  ''
<< ev->simSignal()->id().eventInRun() << endl;
cout << '' --- Events analyzed :  '' << eventsAnalysed << endl;
cout << ''=============================================== '' << endl;
(Get Monte Carlo and Reco Data) ...

}
```

Since this method is called for each event, it needs `G3EventProxy* ev` as parameter. In our analysis we get first of all the Monte Carlo data as generated by Pythia and access then, step by step, L1, L2 and L3 Muon trigger data. Only L3 Trigger data are saved in the rootfile and used for our analysis. To run the code, one first has to check the Buildfile, in which the name of the executable, the source file and the needed libraries are specified as a XML file. Linking the wrong libraries means that some reconstruction steps cannot be performed and no particles are found. This can happen, too, if detector components are switched off, which is specified in the (hidden) .orcarc file - both Buildfile and .orcarc files are documented in the ORCA manual. To compile and link the executable one has to type "scram b bin". In the next sections the implementation of the analysis methods is described more detailed and a short overview of the used ORCA classes is given.

Accessing the muon data in ORCA

We want to loop over the L1 muons, for instance. This can be done in ORCA by using an iterator. This iterator needs to know about the current event; therefore it must be called in a method where `G3EventProxy* ev` is accessible. The code then looks the following way:

```
L1MuGMTSetup* setup = Singleton<L1MuGMTSetup>::instance();
L1MuGlobalMuonTrigger* gmt = setup->GlobalMuonTrigger();
::GMTcands_const_iter gmt_iter;
for ( gmt_iter = gmt->begin(); gmt_iter != gmt->end(); gmt_iter++ )
{
(*gmt_iter).ptValue() // get the desired quantity
...
}
```

The important point here is the use of an iterator `gmt_iter`. It is a template - here it is used for objects of the class `L1MuGlobalMuonTrigger`. It reduces loops over all muon candidates to a simple `for loop`. The member functions of the iterated class are used to access information about the current muon candidate like its transverse momentum p_t, its energy E, its pseudo-rapidity η etc. If we want to access L2-Muon data, the classes and iterator names are different, but the principle of accessing data of a certain detector component is always the same.

The L1 Trigger

The L1 Global Muon Trigger (GMT) combines the trigger information from three different detector components, see chapter 2 : the Drift Tube (DT), the Cathode Strip Chamber (CSC) and the Resistive Plate Chamber (RPC). The DT and CSC
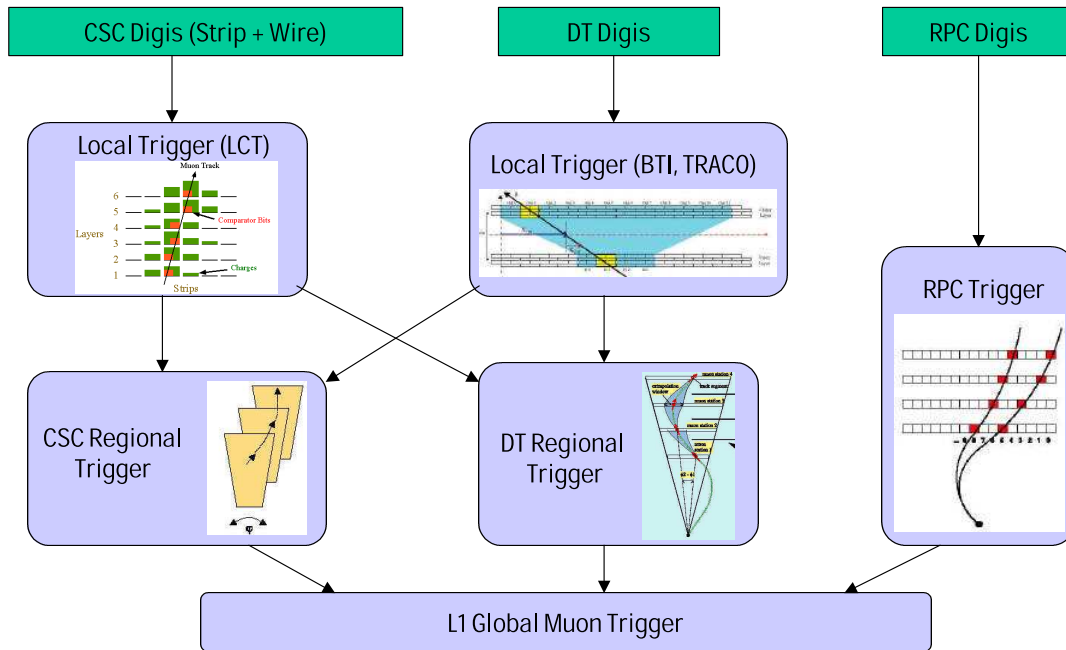
Figure 5.3: Principle of the L1 Global Muon Trigger algorithm: The best muon candidates are determined from Drift Tube (DT), the Cathode Strip Chamber(CSC) and the Resistive Plate Chamber (RPC) data

Track Finders are running in parallel. The four best DT and the four best CSC muons are sent to the GlobalMuonTrigger. It performs a DT/CSC-RPC matching based on distance in (η, ϕ) -space and determines the four best muon candidates in the entire CMS detector which are finally sent to the Global Trigger(GT). In addition to finding the four best muons the GMT appends two bits set by the Calorimeter Regional Trigger, a MIP bit and an isolation bit, also called quiet bit, to each of the muon data. The MIP bit is set if the calorimeter energy is consistent with the passage of a minimum ionizing particle, the isolation or quiet bit is set if a certain energy threshold in the trigger towers surrounding the muon is not exceeded. Both bits are used in the GT to suppress background and to improve selectivity. A schematic overview of the organization of the L1 Trigger can be seen in fig. 5.3

The L2 Trigger

The first step is here to reconstruct track segments in the DT and CSC chambers by clustering hits in the regions of interest defined by the Level-1 muon trigger. Then the reconstructed tracks segments are combined using a Kalman-Filter algorithm
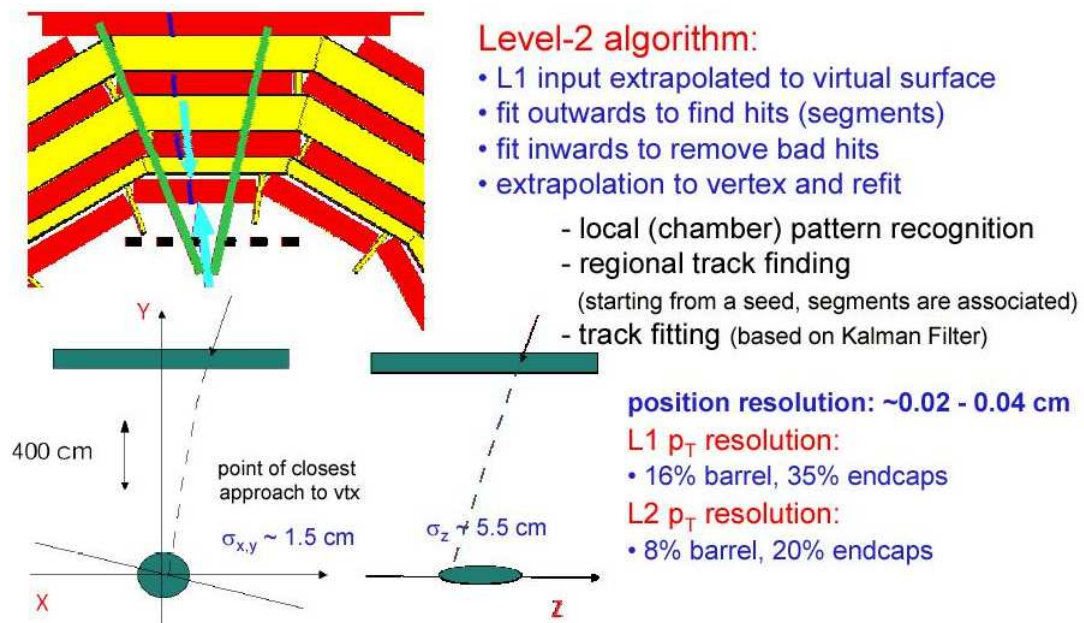
Figure 5.4: Principle of the L2 Muon Trigger algorithm. The Muon chambers and calorimetry data are used

in order to reconstruct a local muon trajectory. Bad measurements are rejected by applying a χ^2 cut. After that, tracks are projected through the calorimeters to the nominal vertex position in order to assign a p_T value at the interaction point. This allows to reject fake Level-1 muon candidates. From the vertex the trajectory is refitted and then the next DT /CSC chamber station is taken. The projection/refit procedure is repeated until one arrives at the outside of the detector.

The L3 Trigger

Starting from a muon reconstructed at Level 2, the muon trajectory is extrapolated from the innermost muon station to the outer tracker surface, taking into account the muon energy loss in the material and multiple scattering. Then the algorithm searches for silicon layers compatible with the muon trajectory and defines a region of interest to start regional track reconstruction. Tracks are reconstructed starting from existing hits in the outer silicon layers going inward by adding one detector at a time. To resolve ambiguities all reconstructed tracks are then refitted including the reconstructed hits in the muon chambers from the original Level-2 muon. Finally, they are selected on the basis of a χ^2 cut.
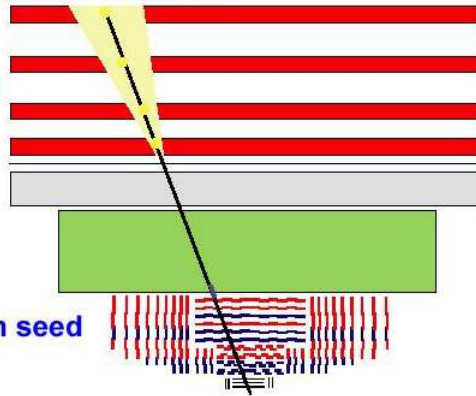
Figure 5.5: Principle of the L3 Muon Trigger algorithm. The results of the L2 Trigger are improved by matching with tracker data.

5.3   Saving data using root trees

The reconstruction data obtained from ORCA must be saved to perform further analysis without rerunning the long reconstruction job. We had the choice between Hbook Ntuples and ROOT Trees. The Hbook Ntuples are used in Examples delivered by the CMS software; a C++ wrapper, that allows the creation of HBook Ntuples in the object oriented framework does exist and works. However, we discovered some disadvantages:

- Variable names in the Hbook Ntuple have limited size (max 10 character).

- C++ ideas are not supported by a FORTRAN code wrapper.

- Errors in parameters passed to HBook are not discovered by the compiler but causes segmentation faults at runtime instead, when the fortran routines are called. This makes the development process rather slow.

Because of this points, we decided to use root trees. The main advantages are :

- There are no limitations on variable names. Sub folder structures like branch or leaf allow further structuring.

- ROOT is written in C++.

- Errors are quickly discovered by the compiler/interpreter.

- Nice visualization and viewing tools (e.g. Tree Viewer see fig. 5.6) exist

- Analysis code used in KaMuonAnalysis can be directly used and optimized in root macros, which makes code development very efficient.

- ROOT trees are very flexible and easy to handle.

- Dynamic allocation of memory is available per event, i.e. only as much memory as necessary is allocated.
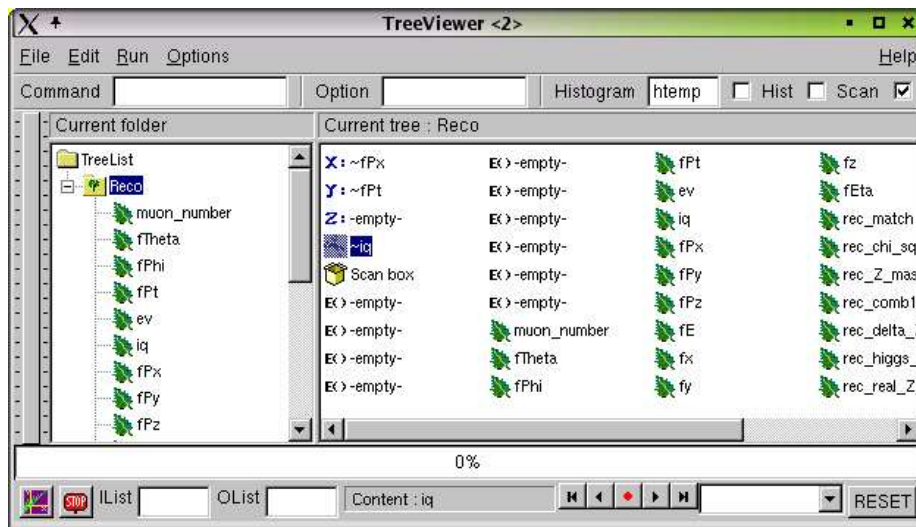
Figure 5.6: ROOT viewing tool TreeViewer

We defined the class KaMuonTree, where for the reconstructed and the generated data there is one ROOT-Tree (ROOT TTree class) containing the different muon parameters , i.e. Px , E etc. The muon data are saved in arrays with dynamically allocated size, depending on the number of muons. Reading the data back can be done fast and comfortably by looping over the arrays. Correlation between the different quantities are kept, while, for example, by storing in histograms, these correlations are lost. We developed, however, also the class KaMuonRefHistos(containing ROOT histograms of the TH1F class), where we save the same data in histograms - this is done for practical needs, namely to compare the data of both classes and to ensure that the code works and the data are saved correctly. Both classes were created using the root interpreter `cint`. As interface to the KaMuonAnalysis there are fill methods, where the data obtained from the ORCA classes can be copied to the tree / histogram. Details can be viewed in the reference manual. For further analysis standalone executables linked against root libraries were used, that read the root files, plotted the important quantities and performed final calculations.

5.4 The muon isolation class

A characteristic of our signal is a final state with 4 isolated muons. The isolation is exploited to distinguish the signal muons from muons that come from background.

We have used the muon isolation package "MuonIsolation" in ORCA to perform this task. The general principle of isolation algorithms is the analysis of the detector response in a region around the direction of the object under study, here the muon. Since non-isolated muons are accompanied by jets, while isolated ones have only uncorrelated soft particles from pile-up in their proximity, a muon can be defined "isolated" by comparing some kind of detector measurement in a cone around the muon direction (hereafter called algorithm direction) with a predefined threshold. The geometrical definition of the cone is given by the condition being the distances in pseudo rapidity and azimuthal angle between the deposit and the algorithm direction. The muon itself contributes to the detector measurement inside the cone. This contribution can be subtracted to avoid biasing the algorithm.

The threshold on the detector measurement in the cone should be η-dependent in order to guarantee a flat efficiency in η, and must be higher than the average pile-up contribution but low enough to efficiently reject jet originated deposits. Several quantities can be used as detector measurement in the cone. The algorithms used here are based on the transverse energy deposit in calorimeters: the transverse energies is measured in the towers of the hadronic calorimeter (HCAL) and is combined with the reconstructed transverse energy deposit in the electromagnetic calorimeter (ECAL). At L2, data from calorimeters are already available and the calorimeter isolation algorithm can be applied.
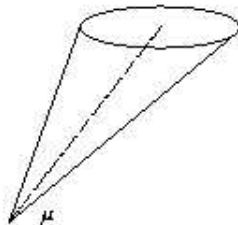
Figure 5.7: The cone around the muon. If $\sum E_T$ in the cone is less than a threshold, the muon is considered isolated

The isolation algorithms consist of two logical steps: the extraction of the signal deposited in a cone around the muon, that is specific to the detector used, and the actual isolation cut, i.e. the comparison of the cone content with the threshold. Both, the size of the cone and the threshold can be optimized in order to maximize the rejection on the background while keeping the efficiency above a given nominal value. The calorimeter isolation algorithm uses as input the parameters of the L2 reconstructed muon at the impact point (i.e. the point of closest approach to the beam line in the plane transverse to the beam). The muon direction at the impact point is the best approximation of the direction of the hypothetic jet and is used in the definition of the algorithm direction.

The extraction of the energy deposits is done independently in the ECAL and the HCAL. In the case of ECAL the measured quantity is the $\sum E_T$ in the crystals around the muon direction at the impact point. In the case of HCAL, whose

segmentation is much coarser than that of ECAL, the algorithm direction is defined instead as the center of the tower pointed to by the muon direction at the vertex; the measured quantity is the $\sum E_T$ of the towers whose center belongs to the cone. This guarantees that the same number of towers contributes to all cones of a given size.

In order to reject pile-up deposits, the HCAL towers with reconstructed E_T below 0.5 GeV and the ECAL crystals with reconstructed E_T below 0.2 GeV are neglected. To avoid electronic and detector noise, an additional energy threshold of 0.12 GeV in the barrel ECAL, 0.45 GeV in the endcap ECAL and 0.6 GeV in HCAL is applied. These values correspond to 3 standard deviations of the nominal noise level. To subtract the energy deposited in the cone by the muon itself, the impact point of the muon on the calorimeters is calculated by propagating the muon to the boundary between ECAL and HCAL. The actual isolation variable is constructed from both, HCAL and ECAL deposits in the cones. Since the sensitivities of ECAL and HCAL calorimeters are different, a weighting parameter has been introduced and the isolation variable is defined as:

$$E_T^{WEIGHT} = \alpha \sum E_T^{ECAL} + \sum E_T^{HCAL} \qquad (5.1)$$

In the MuonIsolation package optimization studies have been performed, depending on luminosity and pseudo rapidity. As expected, the thresholds are higher for high luminosity and increase with pseudo rapidity. In the packages automatically the best thresholds and cone sizes (as determined by the optimization studies) are used. More detailed information can be found in a CMS Note [Iso].

Chapter 6

Studies of the detector performance

In order to interpret the "measured" (or, in this case, simulated) data correctly, it is important to understand the detector properly. The important quantities are the efficiency and the resolution of the detector.

In the following the resolution will be defined as $\Delta X := X_{reconstructed} - X_{generated}$ i.e. the difference between the detector measured value and the true generated value, with $X = \phi$, θ, P_t etc.

The efficiency is defined in our particular case as the number of events with 4 reconstructed muons (with the L3 Muon Trigger) divided by the number of the generated events with 4 "reconstructable" muons. "Reconstructable" means in this case that there were 4 muons at generator level within the acceptance (preselection cuts) and, additionally, with $|\eta| < 2.4$, because the detector is only sensitive in this interval.

In our case, such detector studies were of essential importance, because they did not only check the applied reconstruction algorithms in ORCA, but also the produced data bases with fully simulated events.

6.1 Muon reconstruction efficiency

The results of the efficiency (as defined above) studies are summarized in tab. 6.1.

The efficiencies are not very convincing, taking into account that the reconstruction efficiency of the L3 trigger is specified to be about 97 % for a single muon, depending on the transverse momentum and η [Neu]. The efficiency to find 4 muons should be $(0.97)^4 \approx 88.5$ %. Unfortunately, the L1 trigger has an η coverage limited to 2.1 (because of the limit of the RPC), while L2 and L3 cover an η region of 2.4. However, the L2 and L3 triggers search muons only in the regions of interest, defined by the L1 trigger, so that only muons with $|\eta| < 2.1$ can be found by the L3 trigger. This problem has already been pointed out by [Zion]. The reconstruction efficiency can be significantly improved using the offline Muon Reconstructor. Our effort to use the offline Muon Reconstructor, which searches independently of the L1 Trigger within a range of $|\eta| < 2.4$, was not successful, due to several technical problems:

Sample	Efficency to find 4 muons
$M_H = 125$ GeV	56.0 %
$M_H = 130$ GeV	57.0 %
$M_H = 140$ GeV	60.0 %
$Zb\bar{b}$	35.0 %
$t\bar{t}$	42.0 %
ZZ	47.0 %

Table 6.1: The L3 Muon Trigger efficiency to find 4 muons for signal and background samples. Remarkable is the fact that there are differences in efficiency between the signal and background samples of over 10 %.
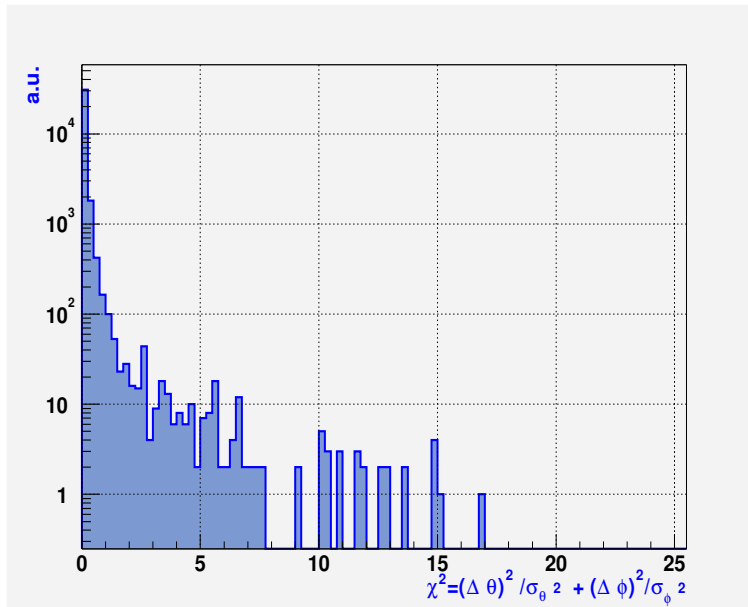
Figure 6.1: The $\chi^2 = \frac{\Delta_\phi^2}{\sigma_\phi^2} + \frac{\Delta_\theta^2}{\sigma_\theta^2}$. It was possible to match 99 % of the reconstructed muons within a 2.6 χ^2 cut.

the Muon Reconstructor often crashes and the resolution was very bad, but, in a first estimation, the efficiency has increased by about 20%. After our feedback to the developer at CERN, this reconstruction package was updated in ORCA_6_3_0.

In summary, there is a good chance to improve the efficiency and thus the discovery potential of this channel. It is also promising here, that the efficiency is different for signal and background in our preliminary estimate and consequently that indicates another potential mean for background rejection. Here sureley room for improvements is left.

In case the L3 trigger has found 4 muons, this does not guarantee, however, that they really correspond to the generated muons and are not of other origin like minimum bias muons, fakes etc. This subjet will be discussed below.
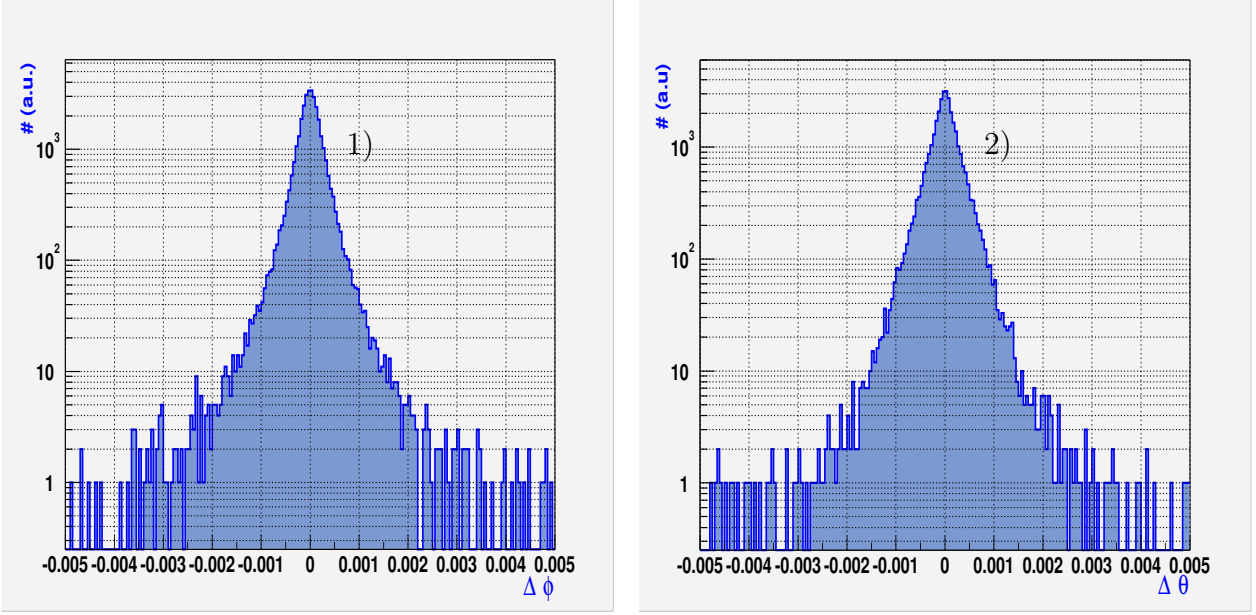
Figure 6.2: Resolution in ϕ (1) and θ (2) for muons found by the L3 Muon Trigger

6.2 Muon matching and Z mass computation

If one considers the case of n generated and reconstructed muons, it is obvious that the sorting of the muons on generator and detector level does not persist unaltered. Consequently a criterion to find for each generated muon its reconstructed partner - or, if more than 4 muons could be reconstructed, to find the combination corresponding to the Monte Carlo muons, is necessary. We decided to perform a matching in θ and ϕ and took as matching criterion the χ^2 function :

$$\chi^2 = \frac{(\phi_{rec} - \phi_{gen})^2}{\sigma_\phi^2} + \frac{(\theta_{rec} - \theta_{gen})^2}{\sigma_\theta^2} = \frac{(\Delta_\phi)^2}{\sigma_\phi^2} + \frac{(\Delta_\theta)^2}{\sigma_\theta^2} \tag{6.1}$$

By applying a χ^2 cut of 2.6, corresponding to 99 % of the total events, it was possible to determine whether a muon could be matched to the Monte Carlo truth or not. The result can be seen in fig. 6.1.

A similar problem is which pair of the muons forms a real Z and which came from the decay of an off shell Z . We used here the same matching algorithm with the different criterion $\Delta_{M_Z} = M_{2\mu} - M_Z$ instead of Δ_ϕ or Δ_θ, that has to be minimized. All possible combinations of oppositively charged muon pairs were tried and the pair closest to the real Z mass was defined as the on-shell Z. The other pair was considered as the virtual Z^*.

Once the matching was performed, it was possible to study the detector resolution of other quantities like the transverse momentum P_t or the resolution in the reconstructed invariant mass of the 4 muons. The results can be seen in tab. 6.2 and in the figures 6.4, 6.6, 6.7.

Since the resolution is dominated by the tracker resolution, we compared our values with the numbers specified in [TDRtracker]. (Since the tracker delivers the

quantity	σ	percentage within 3 σ cut
$1/P_t$	0.0016 1/GeV	98.9 %
θ	0.4 mrad	98.4 %
ϕ	0.4 mrad	98.8 %
M_H	1.25 GeV	99.0 %
M_Z	1.1 GeV	96.3 %

Table 6.2: The L3 Muon Trigger Resolution ΔX for some important quantities

most accurate measurements in P_t or ϕ, it is sufficient to compare the results with the tracker design values.) For instance, muons with a P_t smaller than 10 GeV and bigger than 100 GeV are supposed to have a $\Delta P_t/P_t$ at the order of $0.007 - -0.02$, depending on the pseudorapidity - the value determined in fig. 6.6 is around 0.02. Or, considering the resolution in ϕ, the design value of $\Delta\phi$ for muons with a P_t in the same range as above should be $0.2 - 3$ mrad, depending on η, again, the result here is 0.4 mrad. We have used the resolutions ΔX and the reconstruction efficiencies to perform crosschecks whether our local installation and production chain is compatible with the environment and the databases available at CERN. For this reason, the database for $M_H = 130$ GeV was reproduced and the quantities summed up in tab. 6.2 compared - the results for the two databases were compatible.

In summary, one can say the resolution and efficiency results are compatible with the design values defined in the tracker specification [TDRtracker] and can be regarded as a benchmark that our software setup produces reliable results.
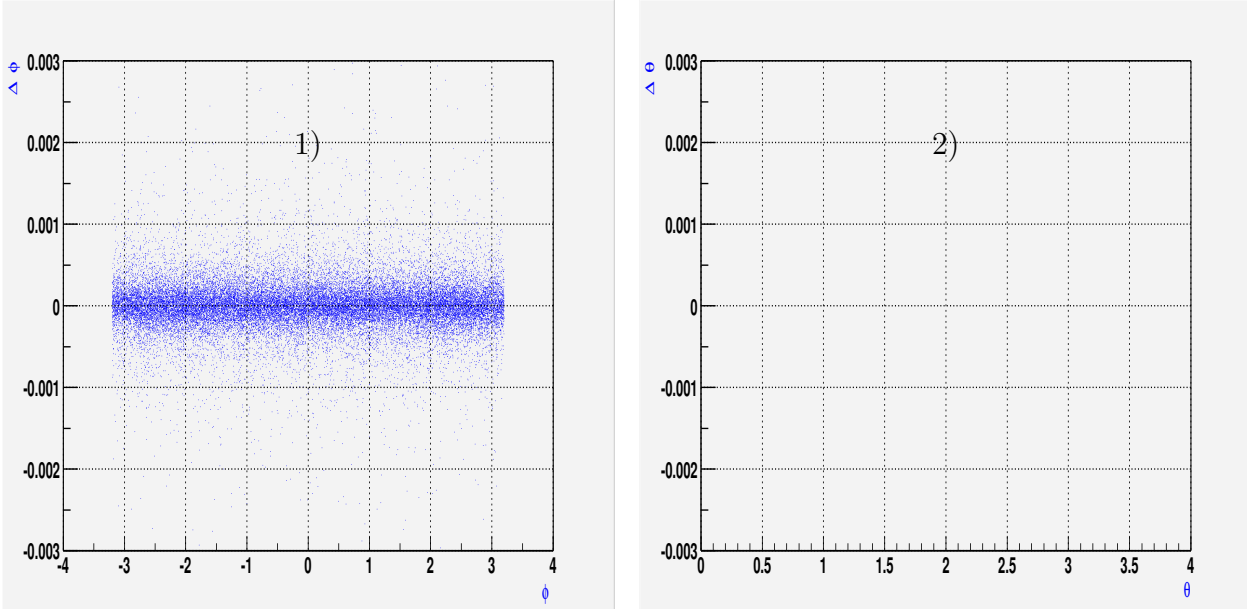
Figure 6.3: Difference $\Delta\theta$ and $\Delta\phi$ between generated and reconstructed values of ϕ and θ as a function of ϕ (1) and θ (2), as a 2D Plot
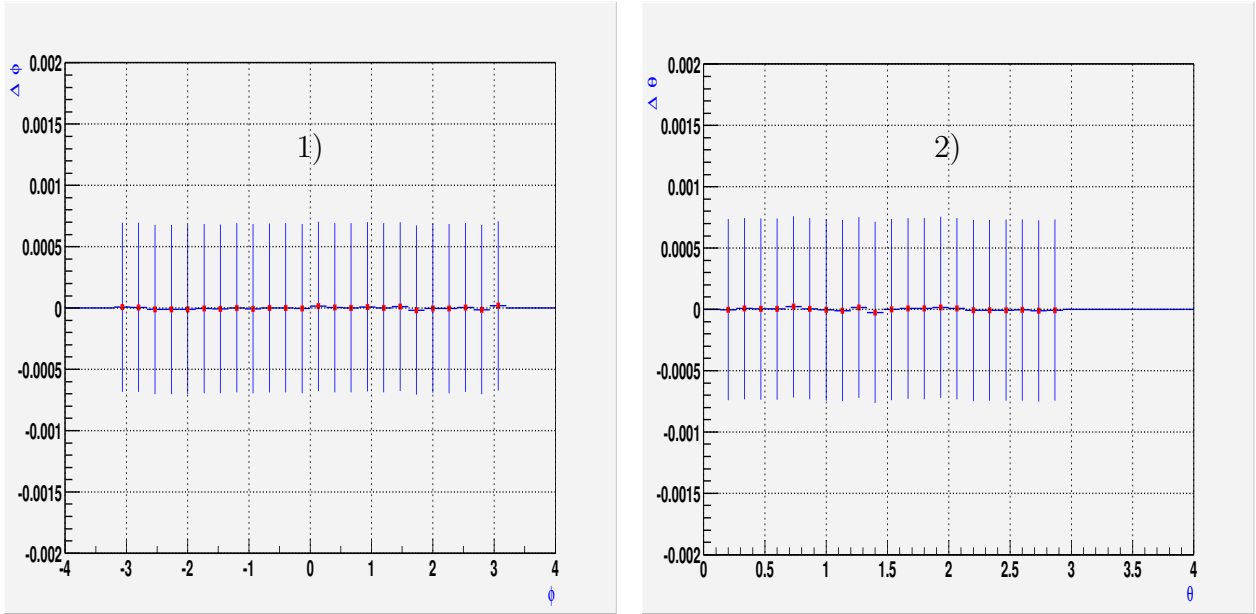
Figure 6.4: Difference $\Delta\theta$ and $\Delta\phi$ between generated and reconstructed values of ϕ and θ as Profile Plot : The red point is the mean of the projected y values on the x axis, the error bars correspond to the spread(RMS). Plot (1) shows $\Delta\phi$ versus ϕ, plot (2) shows $\Delta\theta$ versus θ.
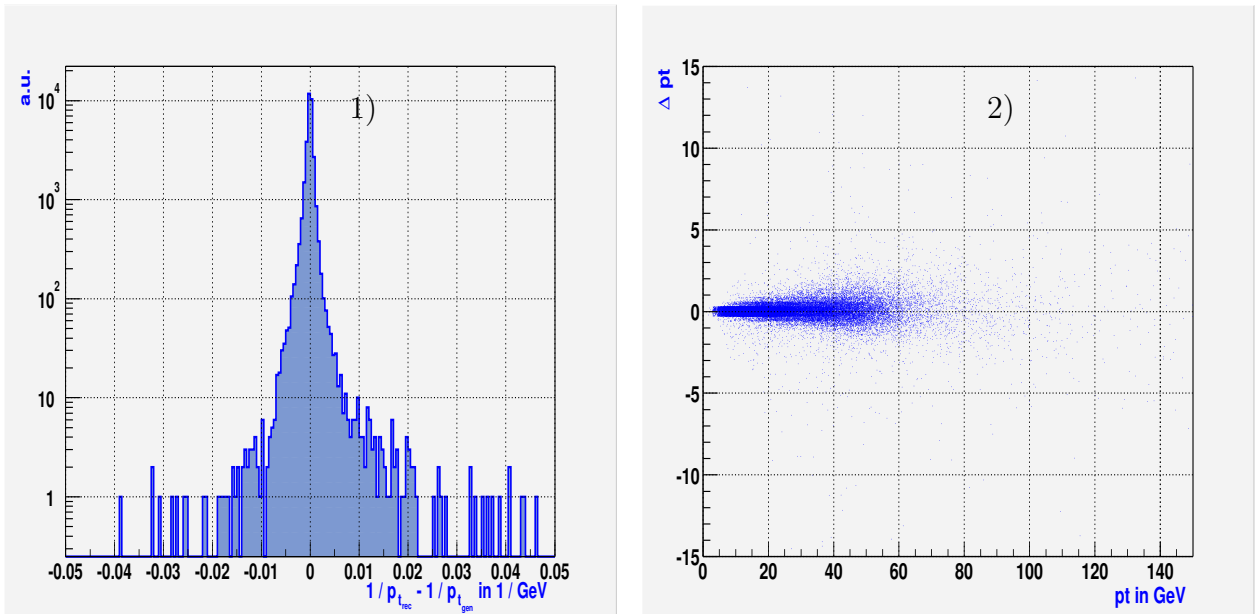
Figure 6.5: Resolution $1/P_{t_{rec}} - 1/P_{t_{gen}}$ (1) and ΔP_t versus P_t as 2D Plot (2)
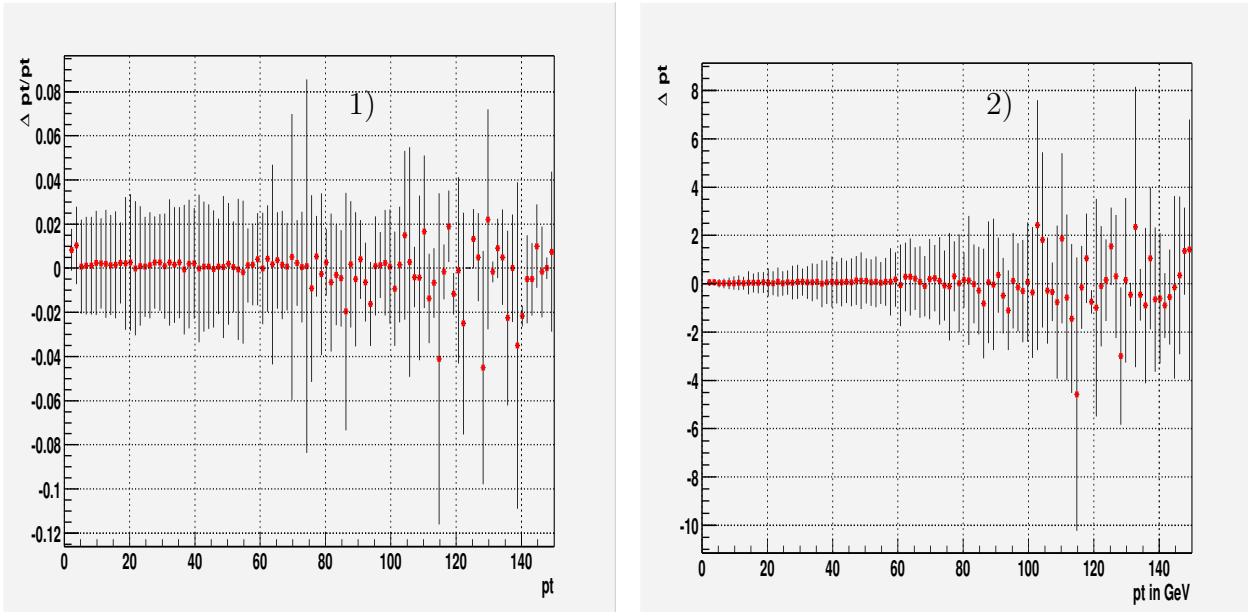
Figure 6.6: Scatter plots of the resolution in P_t. The red point is the mean of the projected y values from a 2D plot like (2) on the x axis, the error bars correspond to the spread (RMS). Plot (3) shows $\Delta P_t/P_t$ versus P_t (relative error) and plot (4) shows ΔP_t versus P_t (absolute error, projection of plot (2)). The relative error does not change very much over a large P_t region < 100 GeV.
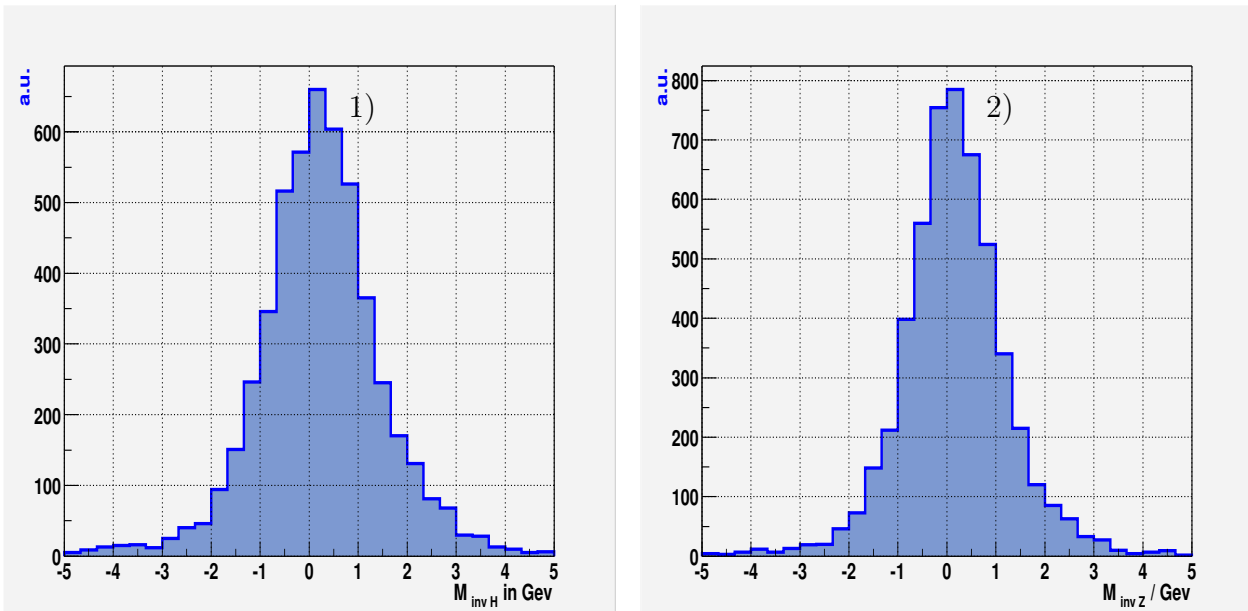
Figure 6.7: Resolution in $M_{inv4\mu}$, $\sigma = 1.25$ GeV and $M_{inv2\mu}$, $\sigma = 1.1$ GeV (real Z mass).

Chapter 7

Analysis of the channel $H \rightarrow 4\ \mu$

The decay of the Higgs boson to 2 Z bosons, each one decaying to 2 μ, lead to a final state with 4 muons, which can be relatively well detected with the muon system of CMS. Background processes with 4 muon final states are relatively rare - therefore this decay channel is called the "golden channel". However, the branching ratio $BR(H \rightarrow ZZ^*)$ varies as a function of the Higgs mass, therefore only some mass regions can be considered. In this work the mass region $125 < M_H < 140$ GeV of LHC for the low luminosity phase is considered. This range is, as mentioned in the introduction, well compatible with the theoretical predictions.

7.1   Cross sections for $\mathbf{H} \rightarrow \mathbf{ZZ^*} \rightarrow 4\mu$
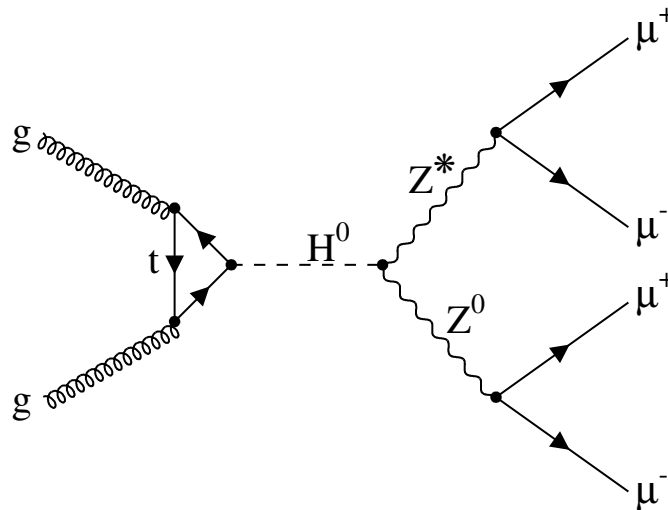
Figure 7.1: The Feynman graph of the signal: The Higgs boson is produced via gluon fusion to tops (main production contribution) and decays in one on-shell and one off-shell Z boson (Z and Z^*). The Z boson pair decays to two muon pairs.

The Higgs boson can be produced with any of the mechanisms mentioned in Chapter 2. However, gluon fusion is the dominant process and the most probable

production of the signal is the one shown in fig. 7.1. As mentioned before, the cross section σ for the decay of the Higgs boson into four muons depends on the Higgs mass. The production cross section for the Higgs boson decreases with the Higgs mass, while the branching ratio of the Higgs decays into two Z bosons increases with Higgs mass. For the final state of our channel the two Z bosons must decay into $\mu^+\mu^-$, with a branching ratio of $\mathrm{BR}(Z \rightarrow \mu^+\mu^-) = 3.36\%$ [PDG]. The total cross section of Higgs into four muons is summarised in tab. 7.1 and visualised in fig. 7.2. It can be seen that the k factor for gluon fusion is around 2. In this work the NLO cross section for Higgs production is used.

σ in pb at LHC	M_H=125 GeV	M_H=130 GeV	M_H1 =140 GeV
LO $\sigma_{gg \rightarrow H}$	13.8	13.0	11.5
NLO $\sigma_{gg \rightarrow H}$	33.2	30.9	27.0
LO $\sigma_{qq \rightarrow Hqq}$	4.6	4.5	4.2
NLO $\sigma_{qq \rightarrow Hqq}$	4.5	4.3	4.1
LO $\sigma_{qq \rightarrow HV}$	1.4	1.2	0.9
NLO $\sigma_{qq \rightarrow HV}$	1.6	1.4	1.1
LO $\sigma_{qq,gg \rightarrow Htt}$	0.3	0.3	0.2
NLO $\sigma_{qq,gg \rightarrow Htt}$	0.2	0.2	0.2
LO σ_{tot}	20.2	19.0	16.8
NLO σ_{tot}	39.8	37.0	32.5

Table 7.1: Production cross sections for the Higgs boson at the LHC. There are four different processes considered at LO and NLO. The calculations were done with HIGLU, HQQ, V2HV, and VV2H [Spira].

7.2 Background processes

Background processes are all processes which have a signature similar to that of our signal, i.e. all processes with 4 muons in the final state. The relevant backgrounds at the LHC for this channel are:

- The process where two quarks produce two Z boson, decaying leptonically into muons with a probability of 3.36%. We also considered here the case where one Z boson or both decay into a pair of taus, which decay with a probability of 17.37% into muons. The contribution of the $ZZ \rightarrow 2\,\mu\,2\,\tau$ background was roughly estimated in this analysis to be about 14%. This is reasonable, because in the case of the direct decay of the two Z bosons into 4 muons, only the events coming from a real and virtual Z* pass the preselection cuts, events with muons coming from two real (on-shell) Z bosons are rejected. If the decay of one real Z into 2 taus and one real Z into 2 muons is taken into account, the muon pair coming from the tau decay has a smaller invariant mass and the probability to pass the Z* preselection cut is higher. Therefore
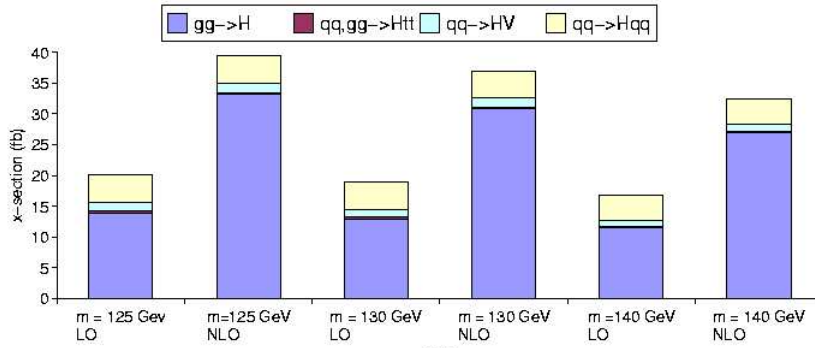
Figure 7.2: Production cross sections for the Higgs boson at the LHC. There are four different processes considered at LO and NLO. The exact values can be seen in tab. 7.1.
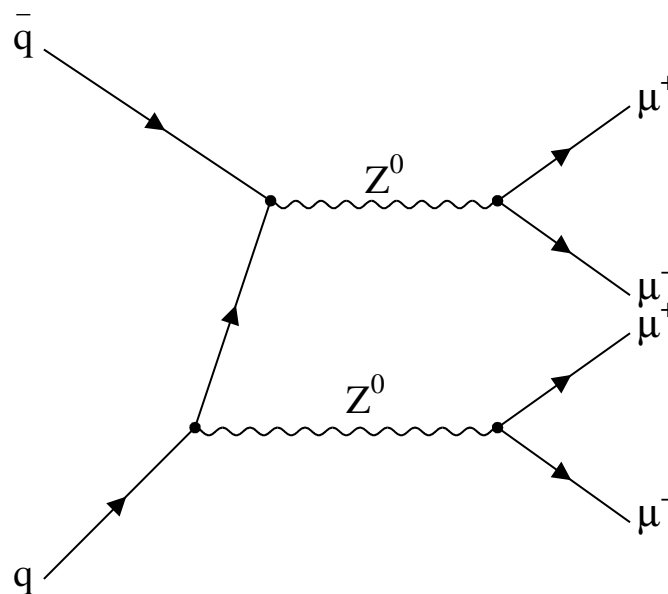
Figure 7.3: The ZZ background: two Z bosons are produced and decay into two muons each. This leads to a 4 muon final state.
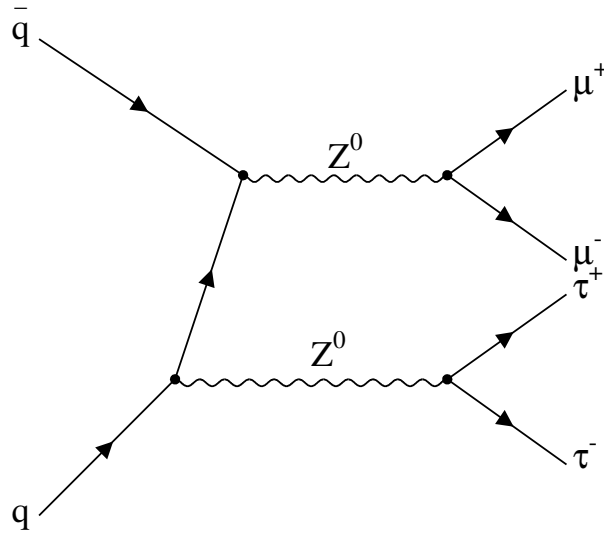
Figure 7.4: The ZZ 2 τ background: Two Z bosons are produced, one decaying into two muons, the other decaying into two taus. The two taus may decay into muons which leads again to a 4 muon final state.

events with two on-shell Z bosons, that have a larger production cross section than ZZ*, are accepted more often and the contribution of $ZZ \rightarrow 2\,\mu\,2\,\tau$ is not negligible. This process is shortly referred to as ZZ and is shown in fig. 7.3 and in fig. 7.4.

- Two gluons decay into two $b\bar{b}$ pairs. One pair forms a Z boson decaying leptonically into muons, the other forms b jets that can contain muons. This process is shortly referred to as $Zb\bar{b}$ and is shown in fig. 7.5.

- the QCD production of a top anti-top pair where both the top and the anti-top decay into a W boson and a bottom quark. The W boson decays with a probability of 10.4% into muons while the b-jets can also contain muons. Therefore a 4 muon final state is possible. This process is shortly referred to as $t\bar{t}$ and can be seen in fig. 7.6.

In the analysis 3 signal samples with the masses 125 GeV, 130 GeV and 140 GeV with 10000 events each and 3 samples for the main backgrounds were used: 7500 $Zb\bar{b}$ events, 10000 $t\bar{t}$ events and 10000 ZZ events. The number of events considered above are the accepted events after the preselection processed through the full detector simulation.
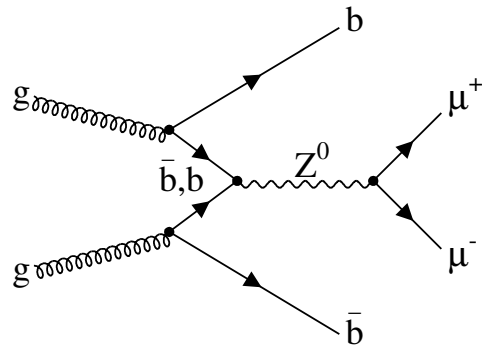
Figure 7.5: The $Zb\bar{b}$ background: The Z bosons decay into two muons and the b form jets, that might contain muons, so a 4 muon final state can be achieved.
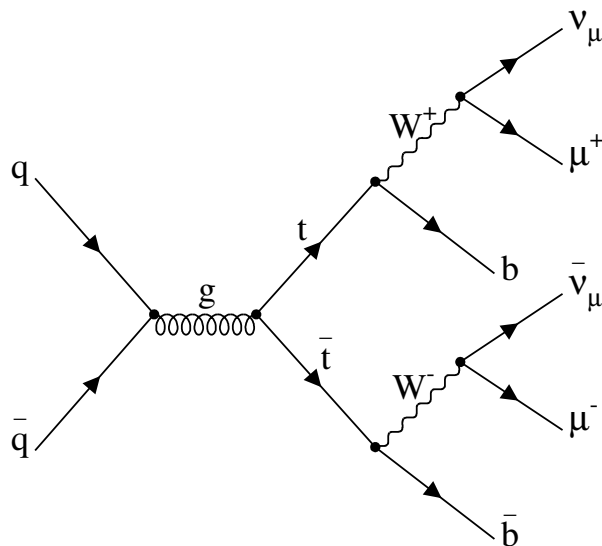
Figure 7.6: The $t\bar{t}$ background: The top quarks decays nearly at 100% to a W boson and a b quark - W bosons may decay into a muon and neutrino which gives 2 muons, while two additional muons can come from b jets.

7.3 The analysis and the applied cuts

As mentioned before, weak preselection cuts were necessary at generator level, in order to simulate only relevant background events that would remain even after imposing our final selection criteria. Our detector resolution studies made it possible to decide which events could be cut on generator level without biasing the background on reconstruction level.

We applied the following cuts at generator level :

1. $|\eta| < 2.5$ for all four μ

2. $P_t > 3.0$ GeV of all four μ

3. 10 GeV $< m_{inv}$ 2 μ < 60 GeV ("Z*")

4. 60 GeV $< m_{inv}$ 2 μ < 110 GeV ("Z")

5. 100 GeV $< m_{inv}$ (4 μ) < 150 GeV

Table 7.2 and tab. 7.3 give an overview of how many signal and background events are expected for one year of running LHC at low luminosity ($L_{int} = 20\,fb^{-1}$). Herefore the total production cross sections σ_{tot} at generator level for the Higgs boson have to be multiplied by the branching ratio to the four muon final state and the acceptance ξ defined by the preselection cuts. For the background the branching ratios and the acceptance has to be considered, too. In table 7.2 the total Higgs production cross sections at NLO from tab. 7.1 are taken and they are multiplied by the branching ratio BR($H \rightarrow 4\mu$). The branching ratio BR($H \rightarrow ZZ$) was computed using the FORTRAN code HDECAY [Spira]. The acceptance ξ was determined as the ratio of the accepted events within our preselection cuts and the total number of generated events. The studies were done with NLO cross sections for Higgs production. Since the k -factor for Higgs production by gluon fusion is close to 2, this is the biggest (theoretical) uncertainty in the whole analysis and the newest theoretical calculations may change the results. PYTHIA performs LO calculations only, so that we had to take the LO event topology from PYTHIA and rescaled the cross sections to NLO. This is an optimistic choice, because the topology of the NLO Higgs production is different from the LO topology used in PYTHIA, which can influence e.g. the reconstruction efficiency in a negative way.

The final cuts applied to the reconstructed variables were the same as presented in an earlier diploma thesis performed with the fast simulation CMSJET. The aim was to estimate how the results change using:

1. The Monte Carlo data directly from PYTHIA or COMPHEP (generator level)

2. The result of the fast simulation (CMSJET)

3. The results achieved here with fully simulated signal and background samples

M_H $inGeV$	σ_{tot} in pb	BR($H \rightarrow 4\mu$)	Accep. ξ	$\sigma_{tot} * BR * \xi$	Exp. Events
125	39.76	$2.89 * 10^{-5}$	56.0 %	0.64	12.8
130	37.07	$4.39 * 10^{-5}$	57.3 %	0.93	18.6
140	32.52	$7.75 * 10^{-5}$	60.0 %	1.51	30.3

Table 7.2: Expected events for $L_{int} = 20 fb^{-1}$. They are determined from the production cross section σ for the Higgs boson at the LHC multiplied by branching ratios BR to 4 muon final states and the acceptance ξ taken from PYTHIA. The production cross section for the Higgs boson decreases with the Higgs energy, while the branching ratio of the Higgs into two Z bosons increases with the Higgs mass. The calculations were done with HIGLU, HQQ, V2HV,VV2H and HDECAY[Spira].

background	σ_{tot} in pb	$\sigma_{tot} * BR$	Accep. ξ	$\sigma_{tot} * BR * \xi$(fb)	Exp. Events
$Zb\bar{b}$	746 (LO)	25.6 pb	0.24. %	61.4	1228
$t\bar{t}$	886 (NLO)	8.86 pb	0.74 %	65.5	1310
ZZ	15.8 (NLO)	73 fb	1.37 %	0.95	19

Table 7.3: Production cross section for the backgrounds at the LHC. For the $Zb\bar{b}$ and the $t\bar{t}$ background the branching ratio $BR(Zb\bar{b} \rightarrow 2\mu)$ is considered and only events with a four muon final state are accepted. The ZZ background contains decays to muons and taus (four times $\sigma_{ZZ \rightarrow 4mu}$), only events with a four muon final state are accepted. The theoretical cross section are taken from [YEL].

In a first approach we studied the significance, namely the ratio $\sigma = N_S/\sqrt{N_B}$ at generator level. The final cuts at generator level were applied (except the isolation cut) in order to compare the results with the results after the final cuts on the reconstructed variables. The final cuts were:

1. $|\eta| < 2.4$ for all four μ

2. $P_t > 20.0$ GeV of the first μ, $P_t > 10.0$ GeV of the second μ and $P_t > 7.0$ GeV for the last two muons.

3. 12 GeV $< m_{inv}$ 2 $\mu < 60$ GeV ("Z*")

4. 84.7 GeV $< m_{inv}$ 2 $\mu < 96.7$ GeV ("Z")

5. All 4 muons should be isolated

In the plot there are the distributions of the sum of the 4 muon invariant masses for the three considered Higgs masses $M_H = 125$ GeV, $M_H = 130$ GeV and $M_H = 140$ GeV and the three main background samples normalised for an integrated luminosity $L_{int} = 20 fb^{-1}$ which corresponds to one 1 year of LHC running (low luminosity). The three peaks have tails to the lower mass range, because in the decay process the two Z bosons can radiate (a photon) and so the invariant mass of the four muons can be lower than the initial Higgs mass. The results are shown
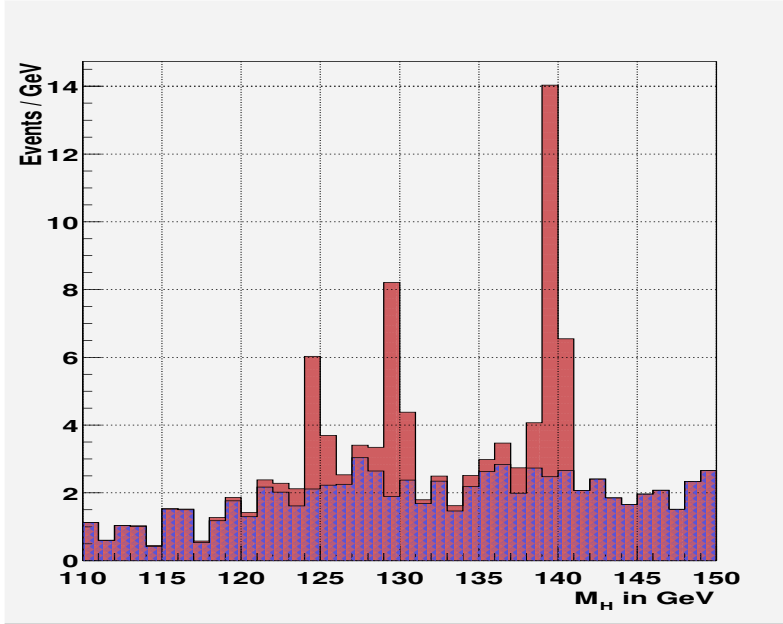
Figure 7.7: The distribution of the sum of the 4 muon invariant masses for the three Higgs mass samples $M_H = 125$ GeV, $M_H = 130$ GeV and $M_H = 140$ GeV plotted at generator level together with the main backgrounds Zbb, tt, ZZ normalised to $L_{int} = 20fb^{-1}$.

Sample	η Cut	P_t Cut	Z_1 Cut	Z_2 Cut	Expected Events
$M_H = 125$ GeV	94.9 %	70.8 %	47.9 %	47.8 %	6.1
$M_H = 130$ GeV	94.7 %	74.6 %	54.0 %	53.9 %	10.1
$M_H = 140$ GeV	94.7 %	81.4 %	63.2 %	63.1 %	19.1
$t\bar{t}$	96.1 %	31.5 %	6.9 %	2.6 %	33.4
Zbb	95.4 %	14.5 %	12.2 %	3.7 %	45.4
ZZ	92.8 %	49.2 %	38.7 %	38.2 %	7.3

Table 7.4: Cut efficiencies on signal and background at generator level.

in fig. 7.7 and tab. 7.5. To compute the significances $\sigma = N_S/\sqrt{N_B}$ an interval of 2 GeV was considered ($M_H \pm 1$ GeV).

The cuts are the same for all three samples. By optimising the cuts for each sample the significance can still be improved, the signal for $M_H = 140$ GeV is already significant at more than 5 σ. The question is now what happens after the full simulation chain, i.e. how the significances change and whether the signal still can be seen. For this purpose the full simulation of the three signal and the the three background samples was performed to study the results with the final cuts now on the reconstructed variables.

New was here the isolation cut: all 4 muons should be isolated from other particles e.g. jets, as described in chapter 5. This cut has not been applied at generator level.

Significances for $L_{int} = 20fb^{-1}$ on generator level					
M_H [GeV]	S	$B_{Zb\bar{b}}$	$B_{t\bar{t}}$	B_{ZZ}	$S/\sqrt{B_{tot}}$
125	5.4	4.0	1.3	0.6	2.3
130	8.9	3.6	2.7	0.6	3.4
140	16.8	4.5	2.7	0.6	6.0

Table 7.5: Number of expected events within the final cuts on generator level and the significances σ.

Sample	P_t Cut	Z_1 Cut	Z_2 Cut	Iso Cut	Expected Events
$M_H = 125$ GeV	81.3 %	53.3 %	52.8 %	51.3 %	3.7
$M_H = 130$ GeV	84.6 %	60.6 %	59.0 %	58.4 %	6.3
$M_H = 140$ GeV	89.6 %	69.7 %	68.4 %	66.2 %	11.8
Zbb	31.8 %	26.8 %	8.5 %	2.6 %	11.3
$t\bar{t}$	54.0 %	11.6 %	4.5 %	4.5 %	23.9
ZZ	64.1 %	50.6 %	49.3 %	47.1 %	4.2

Table 7.6: Cut efficiencies on signal and background for reconstructed variables.

7.4 The achieved results

The results for the significances after the full detector simulation can be seen in fig. 7.10 and in tab. 7.7. We scaled the results of [Drol] for $L_{int} = 300fb^{-1}$ to the luminosity considered here to compare the results for the sample $M_H = 130$: For the luminosity $L_{int} = 300fb^{-1}$ a significance of $S/\sqrt{B_{tot}} = 10.9$ was obtained in [Drol] that corresponds to $S/\sqrt{B_{tot}} = 2.81$ for $L_{int} = 20fb^{-1}$, while our result was $S/\sqrt{B_{tot}} = 2.3$ so that the results are compatible and the significances, as expected, smaller. However, this is only a rough estimate, because for this luminosity the adequate pile up has to be considered and we only worked with databases for low luminosity pile up.

The background consists of 60.3% $t\bar{t}$ events, 28.6% $Zb\bar{b}$ events and 11.1% ZZ events and has an error of 6.5 % according to $\Delta f = \sqrt{\sum_i c_i^2 n_i}$, where n_i is the number of events after all cuts for each background and c_i a normalisation factor, so that $f_i = n_i c_i$ is equal to the number of expected events for one year LHC running for the background component f_i. The total number of background events for this luminosity is consequently $f = \sum_i c_i n_i$. For each background component the error was calculated according to the formula

$$\Delta f_j^2 = \sum_i \frac{\partial f}{\partial n_i}^2 \Delta^2 n_i \tag{7.1}$$

where Δf_j is the error of the background component to be computed and Δn_i the error of the component i. The error Δc_i was not considered. Since the number of events n_i in the region $M_H \pm 1$ GeV that was used for the computation of the

significances, is small, the interval of 40 GeV was used to estimate the background error. The results were $\Delta_{t\bar{t}} = 3.2\%$, $\Delta_{Zb\bar{b}} = 3.25\%$ and $\Delta_{ZZ} = 0.7\%$, so the background is known with a rather small error. In fig. 7.11 we plotted once again the signal for $M_H = 140$ GeV and the background and added "toy data" according to the number of expected events for one year LHC within our acceptance after the final cuts. The red line is a Gaussian fit for the signal peak and the blue line is a linear fit of the background. This plot shows how the situation may look like after one year LHC.

Significances for $L_{int} = 20\,fb^{-1}$ after full simulation					
$M_H\,[GeV]$	S	$B_{Zb\bar{b}}$	$B_{t\bar{t}}$	B_{ZZ}	$S/\sqrt{B_{tot}}$
125	2.38	1.2	0.99	0.32	1.5
130	3.96	1.2	1.65	0.33	2.3
140	7.34	0.9	2.1	0.36	4.1

Table 7.7: Number of expected events within the final cuts after full simulations and the significances σ.

7.5 Optimisation of the cuts

The best results can be obtained by optimising the cuts according to the Higgs mass. To find the best possible cuts we studied the shapes of the distribution of some variables for signal and background. The normalised P_t distribution, for instance, can be seen in fig. 7.8. The muons are sorted that way, that the muon with the highest P_t is numbered μ_1, the one with the second highest P_t is numbered μ_2 and so forth. In this plot the red line corresponds to the Higgs sample ($M_H = 125$ GeV here) and the other colours are the three background samples. It can be seen, that cuts on the last two muons are reasonable, because a great fraction of the $Zb\bar{b}$ and $t\bar{t}$ background has two muons (μ_3 and μ_4) with rather small P_t values ($P_t < 10$ GeV).

We also analyzed the shapes of the invariant mass distributions of the two muon pairs that form the best real Z boson mass and the virtual $Z*$ mass (taking into account the correct charge combination). Since we have already applied preselection cuts at generator level, the cuts on the Z masses appear not to be so efficient any more. Their distributions can be seen in fig. 7.9.

We optimised the cuts by varying one cut with the other cuts fixed in a region of interest defined by the distributions in fig. 7.8 and fig. 7.9, and determined an interval, where the significance reaches a maximum. It was possible to slightly improve the significance σ. The P_t cuts on the last two muons were changed to $\mu_3 = 10$ GeV and $\mu_4 = 7$ GeV for $M_H = 125$ GeV and to $\mu_3 = 11$ GeV and $\mu_4 = 9$ GeV for $M_H = 130$ GeV and $M_H = 140$ GeV. This can be seen in tab. 7.8 and fig. 7.12. The error of the total background Δf is 8.4%, computed , as above, in an

interval of 40 GeV. However, the number of expected background events is now very small -in tab. 7.8 two times there are no background events at all left. Therefore more integrated luminosity is necessary to get realible results.

Significances for $L_{int} = 20 fb^{-1}$ after full simulation					
$M_H \ [GeV]$	S	$B_{Zb\bar{b}}$	$B_{t\bar{t}}$	B_{ZZ}	$S/\sqrt{B_{tot}}$
125	1.4	0.5	0.1	0.2	1.7
130	2.3	0.0	0.2	0.2	3.8
140	5.4	0.2	0.0	0.2	5.3

Table 7.8: Number of expected events within the optimised cuts and the significances σ. The statistics of the background is very low.

7.6 Possible improvements and further studies

Several improvements are still possible, listed in the following :

1. An important point for this channel is the rejection of the $t\bar{t}$ and $Zb\bar{b}$ background, that form together nearly 90% of the whole background considered in this analysis. The criteria to determine isolated muons can still be improved using the Muon Isolation package or another jet finder algorithm. The approach here in this work is very simple and fast simulation studies indicate, that a better suppression of muons in jets should be possible.

2. The cuts can be optimised using neural networks, like e.g. NEUROBAYES. This FORTRAN based package is developed at the IEKP institute at Karlsruhe [Fei] and a C++ interface to read in RootTrees is available. For this reason an analysis starting from the developed Root trees class KaMuonTree can be performed in straightforward manner.

3. The whole analysis has to be re-accomplished for the high luminosity phase. Here new data bases with more underlying pile up events have to be created.

4. It is also possible to consider backgrounds with a three muon final state and to estimate the probability to get a 4 muon event because of measuring one additional muon from a minimum bias event.

5. After a first approximation, that muons coming from τ decays give a 14% contribution, it would be interesting to study the channel $H \to 2\,\mu\,2\,\tau \to 4\,\mu$ in more detail. Besides, the channel $H \to 4\tau$ is as well not entirely analyzed.
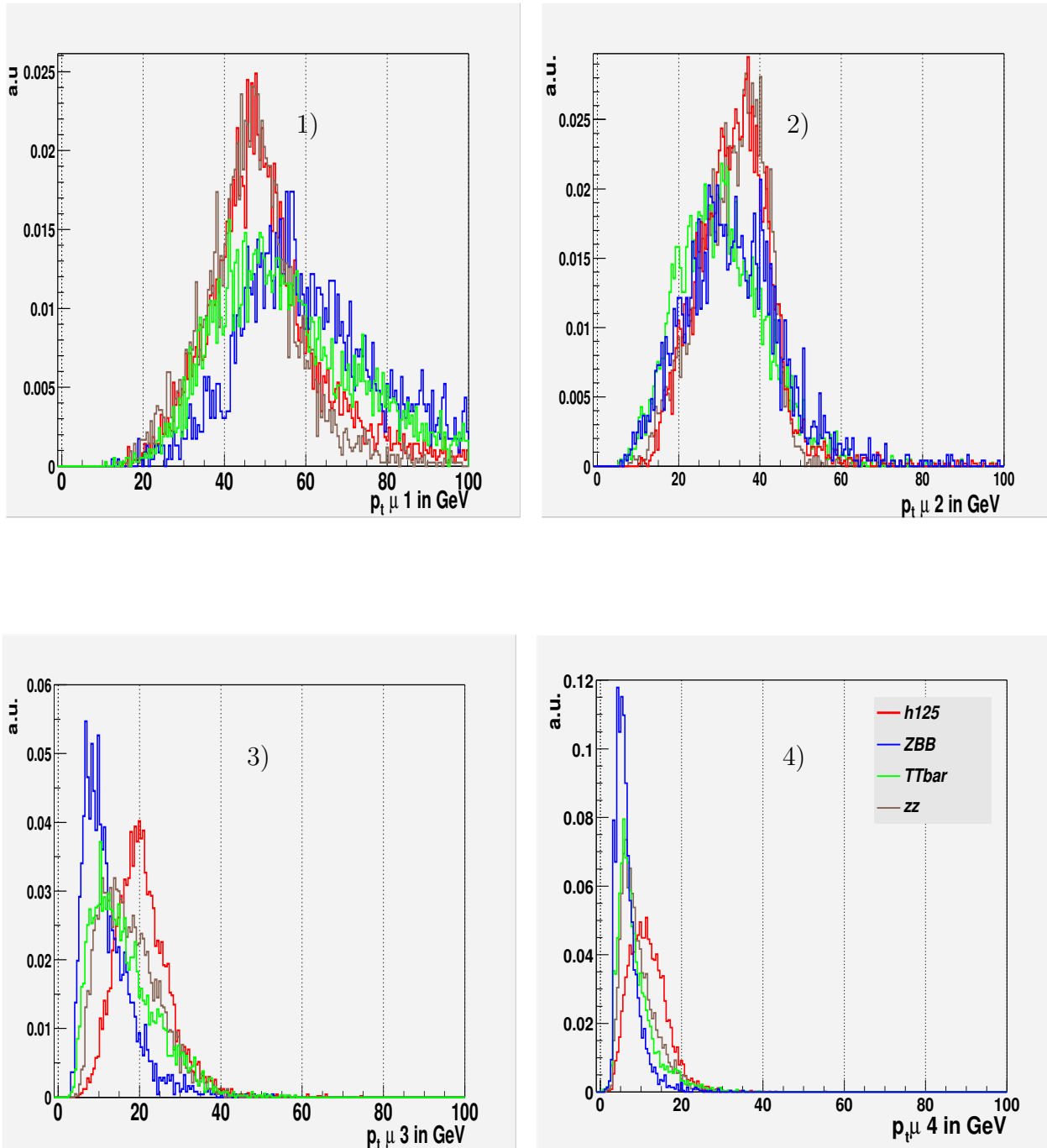
Figure 7.8: Normalised distribution of transverse momenta (P_t) for the four muons of signal ($M_H = 125$ GeV) and background events. The muons are sorted by their Pt values starting with the muon with the highest P_t. It can be seen that cuts on the last two muons could lead to a better separation.
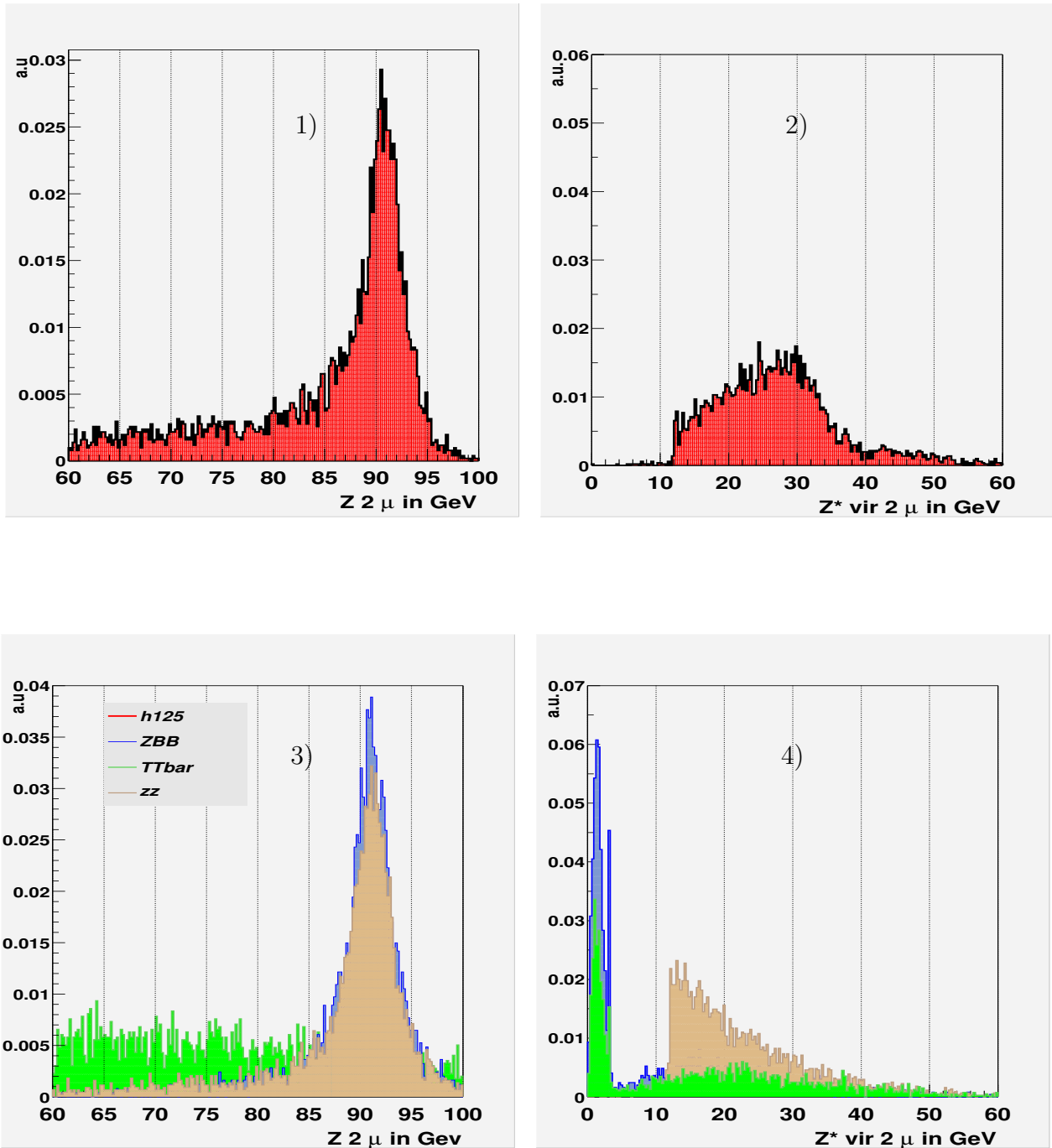
Figure 7.9: Normalised distribution of the invariant mass of the 2 muons where the invariant mass is closer to the Z mass (1), (3), and the other two muons that form the virtual Z* (2), (4), for signal $(1,2)M_H = 125$ GeV and background (3,4). Cuts on the two Z masses can significantly reduce the background contribution. The two peaks in (4) between 0 and 10 GeV are a consequence of the fact, that the more simple preselection pairing algorithm is different from the final pairing algorithm and invariant masses < 10 GeV may occur.
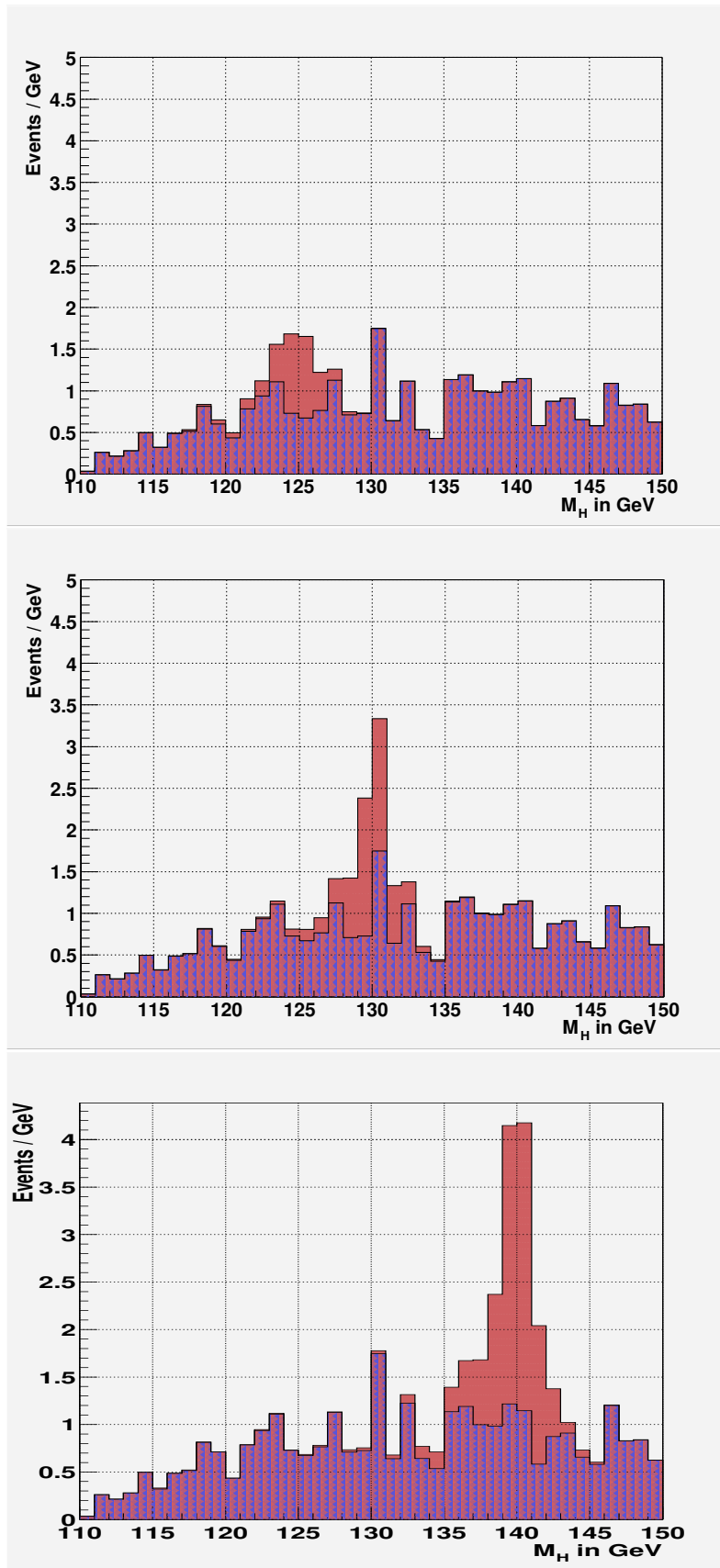
Figure 7.10: The distribution of the sum of the invariant masses of 4 muons in case of the three Higgs masses $M_H = 125$ GeV, $M_H = 130$ GeV a nd $M_H = 140$ GeV plotted after full detector simulation of the signal and the main backgrounds $Zb\bar{b}$, $t\bar{t}$, ZZ, normed for $L_{int} = 20 fb^{-1}$.
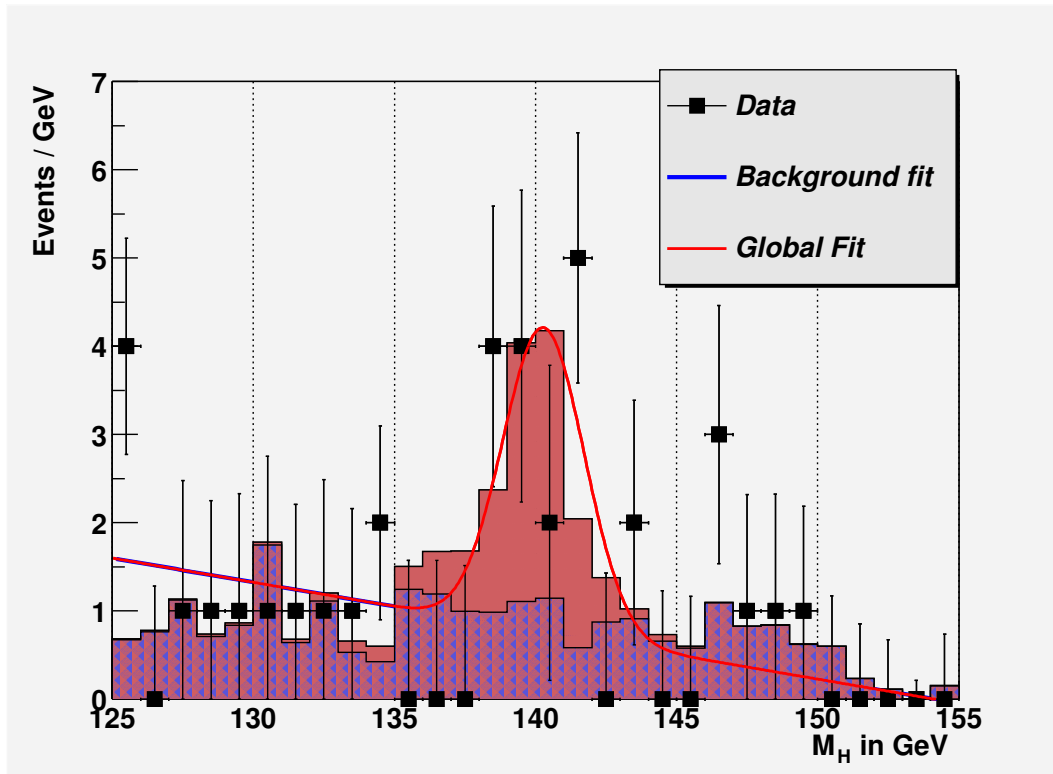
Figure 7.11: Signal for $M_H = 140$ GeV and the backgrounds after full simulation, with added "toy data" according to the number of expected events for one year LHC within our acceptance after the final cuts. The red line is a Gaussian fit for the signal peak and the blue line is a linear fit of the background. This plot shows how the situation may look like after one year LHC.
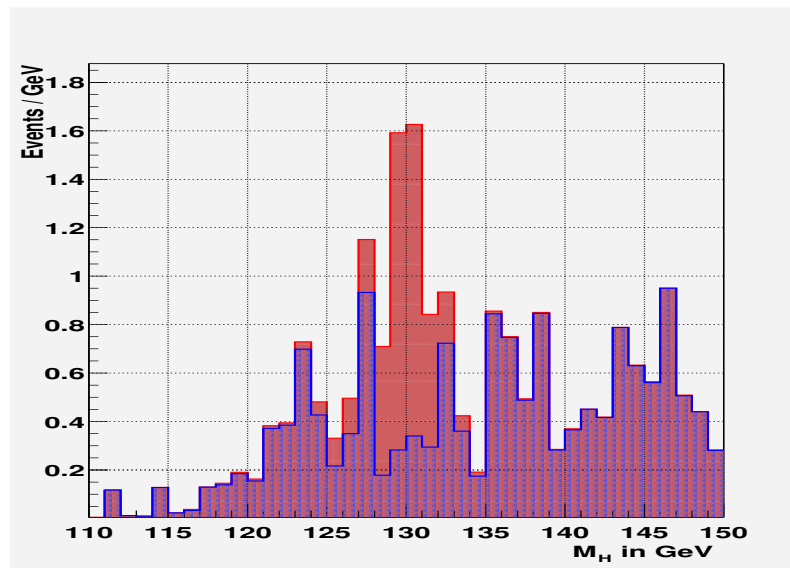
Figure 7.12: The Higgs mass $M_H = 130$ GeV with optimised cuts after full simulation together with the main backgrounds $Zb\bar{b}$, $t\bar{t}$, ZZ normed for $L_{int} = 20 fb^{-1}$.

Chapter 8

Conclusion and outlook

In this thesis for the first time a realistic study for the channel $H \to 4\mu$ is presented, in which simulated signal and background events were processed through the full detector simulation and reconstruction of the CMS detector.

For this purpose, the whole simulation and reconstruction chain, consisting of event generation, fragmentation, detector simulation, pile-up and overlay of minimum bias events and the object-oriented reconstruction, was set up, configured and operated successfully on a Linux PC cluster at the University of Karlsruhe.

Since the Standard Modell Higgs boson is expected to be light, but above 114 GeV, the Higgs discovery potential after the LHC start-up was investigated for Higgs masses of $M_H = 125$ GeV, $M_H = 130$ GeV and $M_H = 140$ GeV. The considered background channels are $Zb\bar{b} \to 4\mu$, $t\bar{t} \to 4\mu$, and $ZZ \to 4\mu$, including taus decaying to muons (e.g. $ZZ \to 2\mu 2\tau \to 4\mu$), where the latter constitutes a contribution of roughly 14 % to the the $ZZ \to 4\ \mu$ background. Reconstruction efficiencies were found to be different for signal and background, showing the importance of the realistic simulation of signal and background. Results from the full simulation were found to be largely compatible with predictions of earlier studies performed with a fast detector simulation, but the obtained signal significances are smaller. The significances $\sigma = N_S/\sqrt{N_B}$ found for 20 fb^{-1} (one year LHC running) were $\sigma_{M_H=125} = 1.5$, $\sigma_{M_H=130} = 2.3$ and $\sigma_{M_H=140} = 4.1$.

This study has also shown that improvements of the CMS reconstruction algorithmens are still possible and necessary. With higher statistics, i.e. during the high luminosity phase, this channel may also allow studies of the $H - ZZ$ coupling at lower Higgs masses.

Zusammenfassung

Das Standardmodell der Teilchenphysik gehört zu den größten intellektuellen Errungenschaften der modernen Physik. Alle Elementarteilchen, die im Standardmodell vorhergesagt werden, konnten nachgewiesen werden, außer eines einzigen Teilchens, des Higgs Bosons. Dieses Teilchen spielt im Standardmodell aber eine zentrale Rolle, da es den anderen Teilchen ihre Masse verleiht. Wenn es nicht gefunden wird, muß das ganze Standardmodell als korrekte Beschreibung der Natur in Frage gestellt werden.

Bisherige Suchen nach dem Higgs-Boson verliefen ergebnislos, jedoch ist es gelungen, große Bereiche in der Higgs-Massenskala auszuschließen. Eine untere Grenze wurde durch direkte Suche beim den LEP[1] Experimenten am CERN[2] etabliert, eine obere Grenze konnte durch Vorhersagen der Theorie (elektroschwache Präzisionsmessung) abgeschätzt werden, so daß das Higgs Boson in einem verhältnismäßig kleinen Massenfenster zwischen 114.4 GeV [3] und 193 GeV erwartet wird.

Ziel dieser Arbeit ist es, das Entdeckungspotential des Kanals $H \rightarrow 4\mu$ mit dem CMS[4]-Detektor am LHC[5] in dem von theoretischer Seite als wahrscheinlich angesehenen Massenbereich 125 - 140 GeV in der Phase der niedrigen Luminosität zu untersuchen. Um möglichst realitätsnahe Ergebnisse zu erhalten, wurde hierfür die neuste verfügbare, softwaretechnisch aufwändige, vollständige Detektorsimulation für Signal und Untergrund Ereignisse verwendet.

Die Studie besteht aus acht Kapiteln. In Kapitel 1 wird das Standardmodell der Teilchenphysik kurz skizziert. Großer Wert wird dabei auf die zentrale Bedeutung der Eichinvarianz gelegt; diese wird an einem kurzem Beispiel im Lagrange-Formalismus verdeutlicht und dann der Zusammenhang zu der fundamentalen Rolle des Higgs-Bosons hergestellt. Anschließend wird detailliert auf die spontane Symmetriebrechung und den Higgsmechanismus eingegangen.

Kapitel 2 stellt die möglichen Produktions- und Zerfallskanäle für das Higgs-Boson am LHC vor. Hierbei werden quantitative Angaben zu den Wirkungsquerschnitten, Verzweigungsverhältnissen etc. der erwarteten Prozesse gemacht und grundlegende Größen wie die Luminosität eingeführt.

Kapitel 3 widmet sich den technischen Einzelheiten des LHC Beschleunigers und des CMS Detektors. Der LHC soll 2007 in Betrieb genommen werden und dann für

---

[1]Large Electron-Positron collider
[2]Conseil Europeen pour la Rechérche Nucléaire
[3]in der gesamten Studie wurden $\hbar$ und c gleich 1 gesetzt
[4]Compact Myon Solenoid
[5]Large Hadron Collider

drei Jahre mit der niedrigen integrierten Luminosität von 20 fb^{-1} pro Jahr laufen, wonach die Phase der hohen Luminosität von 100 fb^{-1} pro Jahr vorgesehen ist. Der Aufbau und die Funktionsweise des CMS-Detektors wird kurz beschrieben, wobei auf das Trigger-System und das Myon-System genauer eingegangen wird, da sie für den hier betrachteten Kanal besonders wichtig sind.

In Kapitel 4 wird die Rechner-Infrastruktur am Institut beschrieben, die in diesem Jahr in Form eines Computing Clusters mit 18 Rechnerknoten an der Universität Karlsruhe aufgebaut wurde. Es wird zuerst die Netzwerkarchitektur erläutert, dann das Temperatur-Monitoring-System, das im Rahmen dieser Arbeit entwickelt wurde. Anschließend wird die Installation und Konfiguration der nicht trivialen, über das Netzwerk arbeitenden, CMS-Simulationssoftware und die hierfür benötigte Werkzeuge wie das Batch-System PBS erläutert.

Kapitel 5 umreißt die Software-Komponenten der CMS-Distribution und beschreibt die aufwändige Produktion der vollständig simulierten Signal- und Untergrund Datenbanken, anhand derer die eigentliche Analyse ausgeführt wird. Zu dieser wird im Anschluß das C++ Programm vorgestellt, das während dieser Arbeit entwickelt wurde, wobei auch die Art und Weise des Zugriffs auf rekonstruierte Daten im objektorientierten Rahmen erläutert wird. Danach werden die verwendeten Klassen und Algorithmen kurz beschreiben und die entwickelte Schnittstelle zu dem Analysepaket ROOT [Bru], mit der die Daten als RootTrees oder Histogramme abspeichert und so für weiter Analysen verfügbar gemacht werden.

In Kapitel 6 werden Studien zur Effizienz und Auflösung der CMS-Detektorsimulation vorgestellt. Zuerst wurde die Effizienz bestimmt, 4 Myonen zu rekonstruieren. Diese kann der folgenden Tabelle 1 entnommen werden.

Sample	Effizienz 4 Myonen zu finden
$M_H = 125$ GeV	56.0 %
$M_H = 130$ GeV	57.0 %
$M_H = 140$ GeV	60.0 %
Zbb	35.0 %
$t\bar{t}$	42.0 %
ZZ	47.0 %

Tabelle 1: Die L3 Myon Trigger-Effizienz 4 Myonen zu finden, für Signal und Untergrund Samples unter der Voraussetzung, daß 4 Myonen auf Generator-Ebene vorhanden waren mit $|\eta| < 2.4$. Bemerkenswert ist, daß sich die Effizienz für Signal und Untergrund um mehr als 10 % unterscheiden.

Anschließend wird das Zuordnungsproblem diskutiert: Betrachtet man den Fall von n generierten und ebenso vielen rekonstruierten Myonen, so kann sich die Nummerierung der Myonen in Pythia durch die Rekonstruktion ändern und ein Kriterium ist notwendig, das jedem generierten Myon einen eindeutigen, rekonstruierten Partner zuordnet. Hier wurde die Größe $\chi^2 = \frac{(\phi_{rec}-\phi_{gen})^2}{\sigma_\phi^2} + \frac{(\theta_{rec}-\theta_{gen})^2}{\sigma_\theta^2}$ verwendet und es wurden immer das generierte und rekonstruierte Myon mit dem kleinsten χ^2 einander zugeordnet. Nun konnten wir die Auflösung des Detektors untersuchen, indem wir die Differenz zwischen generiertem und rekonstruiertem Wert der einander
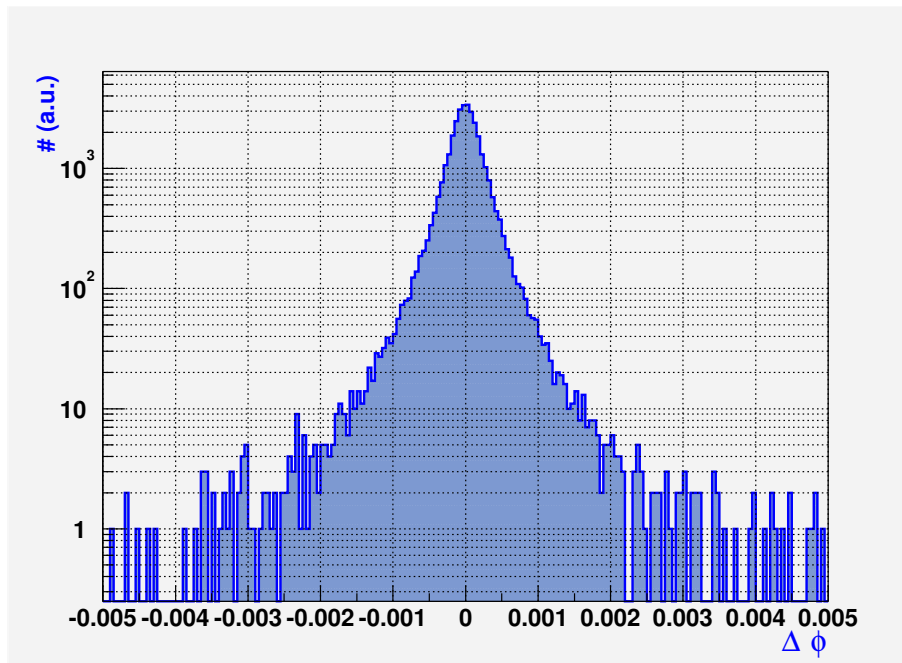
Abbildung 1: Auflösung $\Delta\phi$ in in rad.

durch das obere Kriterium zugeordneten Myonen $\Delta X := X_{rekonstruiert} - X_{generiert}$ betrachten, zum Beispiel die Auflösung in ϕ, die in Abbildung 1 zu sehen ist.

Die Ergebnisse der Auflösungsstudie in den betrachteten Größen sind in der folgenden Tabelle 2 nochmals zusammengefasst.

Größe	σ	Prozentsatz im $3\,\sigma$ Schnitt
$1/P_t$	0.0016 1/GeV	98.9 %
θ	0.0004	98.4 %
ϕ	0.0004	98.8 %
M_H	1.25 GeV	99.0 %
M_Z	1.1 GeV	96.3 %

Tabelle 2: Auflösung ΔX für verschiedene Grössen

Kapitel 8 stellt die Ergebnisse der eigentlichen Studie vor, in der die Signifikanz des Kanals für die drei Higgsmassen M_H=125 GeV, 130 GeV und 140 GeV untersucht wird. Der Feynmangraph des Signals ist in Abbildung 2 zu sehen.

Zuerst wird der Produktionswirkungsquerschnitt des Higgs Bosons für diese Massen diskutiert und die wesentlichen Untergrundprozesse werden benannt. Die in dieser Studie betrachteten Untergrundprozesse sind:

- Zwei Quarks produzieren zwei Z Bosonen, die beide leptonisch in zwei Myonen zerfallen. In dieser Studie wurde auch der Zerfall über Taus berücksichtigt z.B. $ZZ \rightarrow 2\,\mu\,2\,\tau$, der wieder zu einem 4 Myonen Endzustand führen kann, wenn die Taus wieder in Myonen zerfallen.
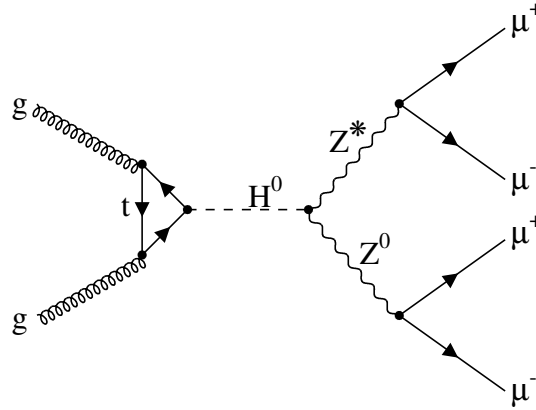
Abbildung 2: Der Feynmangraph des Signal: Das Higgs Boson wird durch Gluonfusion produziert und zerfällt in ein Z-Boson, ein on-und ein off-shell Z Boson(Z und Z^*). Das Z-Bosonenpaar zerfällt in zwei Myonenpaare.

- Zwei Gluonen zerfallen in zwei $b\bar{b}$ Quark Paare. ein Paar zerfällt leptonisch in zwei Myonen, das andere Paar bildet b-Jets, die auch Myonen enthalten können.

- Die QCD Produktion eines Top und Anti-Top Paares, die beide in ein W Boson und ein b Quark zerfallen. Das W Boson zerfällt jeweils in ein Myon und ein Myon Neutrino, während die entstehenden b-Jets auch wieder Myonen enthalten können.

Danach wird diskutiert, warum es bei der vollständigen Detektorsimulation notwendig ist, schwache Vorselektionschnitte auf Generatorebene durchzuführen: Die gesamte Simulationskette braucht ca. 2 Minuten pro Ereignis, so daß es essentiel ist nur relevanten Untergrund voll zu simulieren und nicht solchen, der ohnehin nicht in unserem Massenfenster liegt. Um vergleichen zu können wie sich die Ergebnisse bei vollständiger und der schnellen Simulation ändern werden die in [Drol] eingeführten Schnitte verwendet. Die Preselektionskriterien sind im einzelnen:

1. $|\eta| < 2.5$ für alle vier Myonen

2. $P_t > 3.0$ GeV für alle vier Myonen

3. 10 GeV $< m_{inv}$ 2 μ < 60 GeV („Z*")

4. 60 GeV $< m_{inv}$ 2 μ < 110 GeV („Z")

5. 100 GeV $< m_{inv}$ 4 μ < 150 GeV

Anschließend wird dann bestimmt, wieviele Ereignisse für diese Vorselektionschnitte für Signal und Untergrund bei einer integrierte Luminosität von 20 fb^{-1} (entspricht ein Jahr LHC bei niedriger Luminosität) erwartet werden, wenn man zusätzich auf Generatorebene die folgenden finalen Analyseschnitte durchührt :

1. $|\eta| < 2.4$ für alle vier Myonen

2. $P_t > 20.0$ GeV für das erste μ, $P_t > 10.0$ GeV für das zweite μ und $P_t > 7.0$ GeV für die letzten beiden Myonen.

3. 12 GeV $< m_{inv}$ 2 μ < 60 GeV („Z*")

4. 84.7 GeV $< m_{inv}$ 2 μ < 96.7 GeV („Z")

Das Ergebnis ist der folgenden Tabelle 3 zu entnehmen:

Signifikanz für $L_{int} = 20 fb^{-1}$ auf Generatorebene					
M_H [GeV]	S	$B_{Zb\bar{b}}$	$B_{t\bar{t}}$	B_{ZZ}	$S/\sqrt{B_{tot}}$
125	5.42	4.05	1.32	0.58	2.3
130	8.86	3.60	2.75	0.56	3.4
140	16.78	4.5	2.75	0.62	6.0

Tabelle 3: Anzahl der erwarteten Ereignisse nach den finalen Schnitten auf Generatorebene und die Signifikanz σ.

Danach werden die gleichen Schnitte auf die vollständig simulierten Daten angewendet und dann die Signifikanzen betrachtet, zu sehen in Tabelle 4.

Signifikanz für $L_{int} = 20 fb^{-1}$ nach vollständiger Simulation					
M_H [GeV]	S	$B_{Zb\bar{b}}$	$B_{t\bar{t}}$	B_{ZZ}	$S/\sqrt{B_{tot}}$
125	2.38	1.2	0.99	0.32	1.5
130	3.96	1.2	1.65	0.33	2.3
140	7.34	0.9	2.1	0.36	4.1

Tabelle 4: Anzahl der erwarteten Ereignisse nach den finalen Schnitten und nach der vollständigen Rekonstruktion, sowie die Signifikanz σ.

Das wesentliche Ergebnis ist, daß das Signal nach der Simulationskette noch zu sehen ist, auch wenn ein Jahr noch nicht ausreicht, um eine signifikante Aussage machen zu können, keine unerwarteten Effekte bei der vollständigen Detektorsimulation aufgetreten sind. Eregebnishistogramme zu der Tabelle 4 können in der Abbildung 7.10 eingesehen werden. Es werden noch Optimierungstudien zu den Schnitten durchgeführt und so die Signifikanz etwas verbessert. Anschließend werden Probleme und mögliche weitere Verbesserungen diskutiert. Die wichtigsten Punkte sind, daß mit der ORCA Klasse MuonRekonstructor die Rekonstruktionseffizienzen um schätzungsweise 20 % und somit die Ergebinisse verbessert werden können (dies ist im Rahmen dieser Arbeit wegen technischer Probleme noch nicht gelungen) Eine weitere Unterdrückung der wichtigen Untergründe $Zb\bar{b}$ und $t\bar{t}$ kann durch Verbesserung des Isolationskriterium erreicht werden. Die erhaltenen Signifikanzen sind im

großen kompartibel mit den Vorhersagen aus früheren Studien, die mit der schnellen Detektorsimulation durchgeführt wurden, aber die Signifikanzen sind kleiner. Ein weiteres wesentliches Ergebnisse ist, daß hier zum ersten Mal die realitätsnahe, vollständige CMS Simulations-und Rekonstruktionskette, mit den wesentlichen Bestandteilen Ereignisgenerator, vollständigen Detektorsimulation mit Pile-Up Generator und Analyse in dem objekt-orientierten Rahmen von CMS, auf einem Linux PC Cluster in der Universität Karlsruhe installiert und erfolgreich eingesetzt wurde.

Appendix A

KaMuonAnalysis Class Reference

`#include <KaMuonAnalysis.h>`

Inheritance diagram for KaMuonAnalysis:

```
┌──────────────────────────┐  ┌──────────────────────────┐
│ Observer< G3EventProxy *> │  │  Observer< G3SetUp *>    │
└──────────────────────────┘  └──────────────────────────┘
              ┌──────────────────────────┐
              │      KaMuonAnalysis       │
              └──────────────────────────┘
```
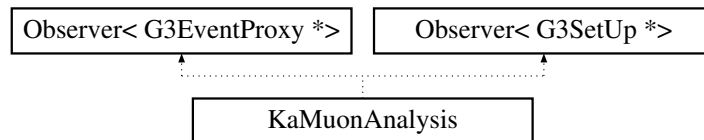
Public Methods

- **KaMuonAnalysis** ()

  A Analysis with Tracking through Muon Chamber and Tracker that reads out Muon - events from an Objectivity database and compares the generated and the reconstructed (L3 Reconstruction) Muons including Matching the muons and Z Mass reconstruction All defines are in **./interface/globals.h**.

- **∼KaMuonAnalysis** ()

  Destructor here we also save the root histo&tree in the file.

- void **initSetUp** ()

  Create Track Reconstructor.

- void **analysis** (G3EventProxy∗ ev)

  Here we get the Generated & Reco Muon Data as Lorentz Vector.

- void **getTrackerTracks** (G3EventProxy ∗ ev)
- void **FillRootTree** (int i,int block)

  Root fill Methods.

- void **FillRootHisto** (int i,int block)

Saving info in Root Histo class:(KaMuonRefHisto()); we first collect al info in this->xxx Variables (i.e this->isize) for each Muon; then , here , we copy it to the root histo ; This method is called i-times for i Muons.

- void **FindMinofMatrix** (int (∗col match),float (∗col quality),int (∗row match),float (∗row quality),int col dimension,int row dimension,int func)

    Finds minimum of a matrix ; to fill the matrix give func as parameter float func(col index,row index); This will be used to fill the col index,row index position of the matrix; WARNING: It would be nicer to have func as method pointers, but I had problems to implement this and pass a method pointer to another method of the same class (though it worked with func pointers in a test) Thats why i pass func as a define and switch it.

- float **delta theta phi** (int i rec,int i gen)

    Computes the diferenz of delta theta (rec,gen)squared and phi (rec,gen) squared.

- float **delta Z Zreco** (int col,int row)

    Computes the delta of Z mass - reco Z mass.

- float **delta Z Zgen** (int col,int row)

    Computes the delta of Z mass - gen Z mass.

- void **fillL1** ()

    fill level 1 trigger information.

- void **fillL2** (G3EventProxy ∗ ev)

    fill local muon reconstruction information.

- void **fillL3** (G3EventProxy∗ ev)

    fill global muon reconstruction information.

- void **fillL2** (int idx, const RecTrack& **muon**)

    fill L2 muon information.

- void **fillL3** (int idx, const RecTrack& **muon**, TrackAssociatorByHits∗ assoc)

    fill L3 muon Reconstruction , v. Neumeister.

- double **kink** (const RecTrack& **muon**) const

Public Attributes

- int **gen negative muon** [MAX MUON]

Methods.

- int **rec_negative_muon** [MAX_MUON]
- **KaMuonTree∗ myroottree**
- **KaMuonRefHisto∗ myroothisto**
- TFile∗ **file**
- TDirectory∗ **TrueH**
- TDirectory∗ **RecoH**
- float **theta_gen** [MAX_MUON]
- float **theta_rec** [MAX_MUON]
- float **phi_gen** [MAX_MUON]
- float **phi_rec** [MAX_MUON]
- float **rec_higgs_mass**
- float **gen_higgs_mass**
- int **comb1_Zmass** [MAX_MUON]
- int **comb2** [MAX_MUON]
- float **delta_Z_mass_reco** [MAX_MUON]
- float **Z_mass_reco** [MAX_MUON]
- float **Z_mass2** [MAX_MUON]
- float **Z_mass_gen** [MAX_MUON]
- float **delta_Z_mass_gen** [MAX_MUON]
- vector<RecTrack> **L3Tracks**

    L3 reconstruction.

- vector<const RecTrack∗> **GoodRecTracks**
- vector<RecTrack>::iterator **iRec**
- TrajectoryStateOnSurface **recL3State**
- vector<RawRecParticle> **MyMuons**

    generated particles.

- **RawParticle muon**

Private Methods

- void **upDate** (G3EventProxy ∗ ev)
- void **upDate** (G3SetUp ∗ su)
- void **getGeneratorParticles** (G3EventProxy ∗ ev)

    Getting Generator Data , only looking at Muons (pid =13).

Private Attributes

- TrackReconstructor∗ **myTrackFinder**
- RawGenLevelDetector **myRglDet**
- RawStableParticleFilter **myStableFilter**
- RawParticleTypeFilter **myMuonFilter**
- RawParticleRecon **myParticleRecon**
- RawParticleRecon **myMuonRecon**
- unsigned int **eventsAnalysed**

     num of events.

- unsigned int **runsAnalysed**
- unsigned int **lastrun**
- unsigned int **charge**
- bool **theInitialised**
- int **fbest_match** [MAX_MUON]
- float **fchi_squared_gen** [MAX_MUON]

     best fit: fbest_match[gen]=rec.

- int **fbest_matchrec** [MAX_MUON]

     chi_squared of best fit: fchi_squared_gen[gen]=rec.

- float **fchi_squared_rec** [MAX_MUON]

     best fit: fbest_match[rec]=gen.

- int **isize**

     temp variable for number of Muons generated.

- float **fPhi**

     temp variable for storing Phi in the Root tree /histo; the data is stored here , copied to the tree/histo invoid **FillRootTree***(int i,int block)* (p. **??**)*; etc ∗ and overwritten immediatly ! f in variablename stands for float , i for int.*

- float **fP**

     temp variable for storing Muons Phi in the Root tree /histo.

- float **fPx**

     temp variable for storing Muons P in the Root tree /histo.

- float **fPy**

     temp variable for storing Muons Px in the Root tree /histo.

- float **fPz**

  temp variable for storing Muons Py in the Root tree /histo.

- float **fE**

  temp variable for storing Muons Py in the Root tree /histo.

- float **fTheta**
- float **fEta**
- float **fPseudoEta**

  temp variable for storing Muons Rapidity in the Root tree /histo.

- float **fPt**

  temp variable for storing Muons pseudo Rapidity in the Root tree /histo.

- int **icharge**
- vector<Hep3Vector> **GoodGenMuon**

  variable for this-methods.

- vector<HepLorentzVector> **GoodjoMuon**
- vector<HepLorentzVector> **GoodRecMuon**
- vector<Hep3Vector> **GoodCluster**
- vector<Hep3Vector> **GoodTracksAtVertex**
- vector<Hep3Vector> **GoodTracksAtEcal**
- int **ncluster**
- int **ntracks**
- int **nmuons**
- int **matchmuon**
- int **matchcluster**
- int **matchtrack**
- float **EMuRes**
- float **ETkRes**

Static Private Attributes

- int **dataset**
- bool **printFlag** = SimpleConfigurable<bool>(true,"MuonReconstruction-Ntuple:Print")
- string **localRecAlgo** = SimpleConfigurable<string>("L2Muon-Reconstructor","KaMuonAnalysis:LocalMuonReconstructionAlgo")
- string **globalRecAlgo** = SimpleConfigurable<string>("L3Muon-Reconstructor","KaMuonAnalysis:GlobalMuonReconstructionAlgo")

- const int **MAXGEN** = 50
- const int **MAXGMT** = 20
- const int **MAXREC** = 30

List of Figures

List of Tables

Bibliography

[Berg] C. Berger, *Teilchenphysik - eine Einführung*, Springer, 1992

[Bru] R. Brun and F. Rademakers, ROOT - An Object Oriented Data Analysis Framework 3.02b, see also `http://root.cern.ch`

[CMSIM] *CMS Simulation Package CMSIM* User Guide and Reference Manual see also `http://cmsdoc.cern.ch/cmsim/cmsim.html`

[COBRA] Cobra homepage `http://cobra.web.cern.ch/cobra/`

[COMPH] *CompHEP - a package for evaluation of Feynman diagrams and integration over multi-particle phase space* **Feb 2000, hep-ph/9908288 v2**

[Drol] V. Drollinger, Die Suche nach dem Higgs-Boson-Studien für das CMS Experiment (diploma thesis), **IEKP-KA/1997-11**

[Eweak] The LEP Electroweak Working Group `http://lepewwg.web.cern.ch/LEPEWWG/`

[Fei] M. Feindt, `http://www-ekp.physik.uni-karlsruhe.de/~phit/neurobayes.html`

[HaMa] Francis Halzen, Alan Martin, *Quarks and Leptons*, John Wiley & Sons 1984

[Har] William Kilgore in collaboration with Robert Harlander *Inclusive Higgs Production at Hadron Colliders* 14th Topical Conference on Hadron Collider Physics, Karlsruhe see also `http://hcp2002.physik.uni-karlsruhe.de/`**Phys. Rev. Lett. 88, 201801 (2002)**

[Hig] P. W. Higgs, Broken Symmetries and the Masses of Gauge Bosons, *Phys. Rev. Lett.* **13** (1964) 508

[IGUANA] the IGUANA Homepage`http://iguana.web.cern.ch/iguana/`

[Iso] N. Amapane, M.Fierro and M.Konecki *High Level Trigger Algorithmus for Muon isolation* CMS Note in preparation, private comunication

[Kem] Y. Kemp, Studien zur Suche nach Higgsbosonen im WH-Kanal bei CDF für eine Higgsmasse von 120 GeV (diploma thesis), **IEKP-KA/2001-26**

[Lm] Linux System Hardware Monitoring lm sensors,
`http://www2.lm-sensors.nu/~lm78/`

[Neu] N.Neumeister *Muon Reconstruction and Online Event Selection in CMS*, LHC
Days in Split, Croatia, 8 - 12 October, 2002.
see also `http://neumeist.home.cern.ch/neumeist/talks/talks.html`
CMS: Electrons, Photons and Muons, Proceedings of the III International Symposium on LHC Physics and Detectors, 2001. *CMS high-level triggering*, Nuclear
Inst. and Methods in Physics Research, 2001

[Muon] *Muon Project Technical Design Report*, 15 December 1997, **CERN/LHCC
97-32**

[ORCA] *ORCA* 6_2_3 *Userguide* see also `http://cmsdoc.cern.ch/orca`
August 21, 2002

[Pat] Patrick Schemitz *Aufbau und Konfiguration eines Hochleistungscluster*,in
preparation **December 2002**, phd thesis

[PDG] Particle Data Group *Particle Physics Booklet*, July 2002

[PHY] *Pythia 6.2 Physics and Manual*, April 2002, **hep-ph/0108264**

[Spira] M.Spira `http://www.desy.de/~spira/proglist.html`

[TDRcms] *The Compact Muon Solenoid - Technical Proposal*, 15 December 1994,
CERN/LHCC 94-38 LHCC/P1 CMS TDR 5

[TDRtracker] *Tracker Technical Design Report* **15 April 1998**, CERN/LHCC 98-6
CMS TDR 5

[Wei] S. Weinberg, A Model of Leptons, *Phys. Rev. Lett.* **19** (1967) 1264

[Wil] Scott Willenbrock *Hadron Colliders, the Standard Model and Beyond*,
14th Topical Conference on Hadron Collider Physics, Karlsruhe see also
`http://hcp2002.physik.uni-karlsruhe.de/` **October 2002**

[YEL] *Proceedings of the workshop on Standard Model Physics (and more) at the
LHC* **CERN 2000 - 004, May 2000** ()

[Zion] Daniel Zer-Zion, Stefano Villa, *A new look at $H \to 4\mu$*, CMS week, CERN,
June 14, 2002

About me:

Figure 1: *Joanna Weng*

- 1977: Born in Warsaw.

- 1990-1997: Paulsen Gymnasium in Berlin.
  June 1997: "Abitur".

- 1997 -1999: Studies of Physics at the Technical University Berlin (TU).
  May 1997: "Vordiplom"

- 1999-2000: Studies of Physics with ERASMUS scholarship at the University Louis Pasteur (ULP), Strassbourg.

- 2000-2003: Studies of Physics at Karlsruhe University(TH).
  February 2003: "Diplom".

- Since February 2003: Doctoral student at IEKP[1], University of Karlsruhe and member of the HIK/GES[2] group at the FZK[3].

- Research Topics: Data analysis, simulation of the CMS detector, studies of the physical potential of the LHC, GRID computing.

---

[1]Institut für Experimentelle Kernphysik
[2]Grid Computing and e-Science
[3]Forschungszentrum der Helmholtz Gemeinschaft Karlsruhe

Danksagung

Hiermit versichere ich, die vorliegende Arbeit selbständig verfasst und nur die angegebenen Hilfsmittel verwendet zu haben.

Joanna Weng

Karlsruhe, den 7. Januar 2003