Université catholique de Louvain
Faculté des Sciences
Département de Physique

CMS trigger strategies for the selection of MSSM Higgs bosons using electron tau-jet decay modes.

Olivier van der Aa
Promoteur: Professeur V. Lemaître

Dissertation présentée en vue de l'obtention
du titre de Docteur en Sciences

Louvain-la-Neuve, février 2004

CONTENTS

Introduction

The discovery of new physics phenomena is one of the important challenges facing particle physicists. Complex experiments are designed to reach the boundary of our knowledge in fundamental interactions. The technological complexity of those experiments is linked to the ever growing data analysis capacities of current software and computing systems. The expected data volume analysed at the Large Hadron Collider (LHC) will be several orders of magnitude above what has been recorded at LEP.

This complexity is for the sake of simplicity. All the efforts to build those complex detectors are given to uncover the foundation of fundamental forces engraved in simple and elegant principles. There exist twelve elementary matter particles, six leptons, and six quarks. There are four known forces in Nature that can act between them: the strong force, that holds atomic nuclei together; the weak force, is responsible for radioactive decay; electromagnetism, and gravity. The forces are transmitted between the matter particles by means of bosons. For instance the photon mediates the electromagnetic force. The simplicity lies in the equations describing the dynamics of those particles, in the form of the Standard Model. The elegance of the Standard Model is that above the enumeration of known elementary particles, the dynamics of the interactions sets out as a consequence of mathematical symmetries.

The Standard Model predictions have been tested to an astonishing level of precision, but it does not predict particle masses. The Higgs mechanism explains boson and fermion masses at the cost of introducing additional free parameters in the model. As we will see in the first chapter it is probably not the final word. The obvious glitch is that the particle predicted by the Higgs mechanism, the Higgs boson, as not been observed. In addition, new theories like supersymmetry, trying to solve the problems lying in the Standard Model Higgs mechanism, postulate the existence of a yet undiscovered family of particles.

The LHC is a new opportunity to further test the Standard Model and new ideas in understanding the nature of mass. One of the big challenges at the LHC is the very high event rate that detectors will have to cope with. In that context and since current technologies are not able to store all events, the data acquisition system will have to select them in real time. Rejected events will be lost forever, preventing any preprocessing. The trigger system selects a portion of the phase space as a

microscope focuses on a focal plane. Hence the observability of new signals depends on the flexibility of the trigger to "focus" on new phase space regions proposed by physics models.

The strategy of the Compact Muon Solenoid (CMS) detector is to analyse events in real time through reconstruction and selection algorithms running on a processing farm. The implementation of selection algorithms in a computer farm gives a flexibility in terms of reprogramming that is not possible with hardware based systems.

The development of the High Level Trigger algorithm was being pursued by the CMS collaboration during the beginning of my thesis. Their development where summarised in the Data Acquisition Technical Design Report (DAQ TDR) [1] which includes the work done on the electron-tau trigger briefly described in the Chapter 3. The electron tau trigger was introduced to improve by 10% the selection efficiency of the electron tau-jet final states in the decay of a MSSM Higgs with a small 2% in terms of trigger rate increase. During the development of the reconstruction algorithms for the measurement of electrons we have studied π_0 rejections using neural networks techniques as discussed in Section 2.1.3. Although it is a nice exercise, the rejection power was not important enough to motivate the use of such a sophisticated algorithm for π_0 rejection in the online environment.

In the development of the electron tau trigger we understood, the need to develop a software to ease the combination of selection algorithms. At that time, the definition of a trigger selection was the task of the physicist developing an analysis. Each analysis had his own trigger implementation where the selection was performed. In addition the modification of the trigger selection required to be more flexible and configurable.

From those observation we developed the High Level Trigger (HLT) steering software described in Section 3.4. It provides a set of trigger elements that can be combined at wish to form the HLT selection algorithm. A user can query the High Level Trigger to know its response and which trigger elements have been fired. The objects that have been selected ($e, \tau, \mu, jets, MET$) can be retrieved for further analysis. Each trigger element is configurable (e.g.: E_t thresholds, isolations, reconstruction algorithm used, etc). The specification of the selection logic only requires to write an XML that contains the elements that have to be connected. This is valuable when analysis have to be compared or to guarantee that they use the same trigger settings. At the time of writing this thesis, the HLT steering software is used in most of CMS analysis to perform selection of the Monte-Carlo data samples.

Using the developed High-Level Trigger steering software, we have studied the trigger strategy to be implemented in order to select MSSM Higgs decaying into electron tau-jet. With that particular benchmark channel, we show how the offline analysis drives the choice of triggering patterns. Hence a method to devise privileged trigger patterns for the search of a particular signal is presented. We show that the trigger patterns that have the highest efficiency after the High Level Trigger have not the highest signal over background ratio. For the present analysis the dominant trigger path is where one electron is identified. After performing a simple analysis of the selected data we show that the single electron efficiency is reduced in favour of the triggers combinations that include the tau-jet, the electron and the missing transverse energy.

This analysis has been possible with the deployment of a computing cluster in our laboratory in which we have contributed for the network architecture, CPU hardware, mass storage and software choices and installation. This is described in Appendix B. It currently provides the necessary tools for the users in the laboratory to produce and analyse events with the CMS simulation and reconstruction software.

In parallel with the CMS activities, we have contributed to the search of the decay of a Higgs boson into W boson pairs at LEP with the ALEPH detector. The analysis, described in Appendix A, has been performed in the context of a fermiophobic Higgs boson. In that context, the decay of the Higgs in W pairs becomes dominant for Higgs masses above 90 GeV. Our analysis concentrated on the channels with a least two leptons, which is included in the final limit in [2].

The physics units in this dissertation are the so-called natural units. The speed of light c is chosen as the unit for velocity, and Planck's constant $\hbar$ as the unit for angular momentum: $c = \hbar = 1$.

CHAPTER

ONE

The Standard Model and Beyond

1.1 The Higgs mechanism

The Higgs mechanism was introduced to explain the mass of the Z and W bosons [3]. Two "ingredients" form the basis of the Higgs mechanism. The first "ingredient" is the Goldstone theorem. Let us assume a Lagrangian invariant under a symmetry group with N generators. If the Lagrangian embeds a function of its fields that breaks the symmetry down to a group with g generators, then there will be $N-g$ massless fields appearing in the theory. These are called Goldstone bosons. The second "ingredient" is the particularity of gauge symmetry. In the case of a Lagrangian invariant under gauge symmetry, the massless Goldstone bosons are redefined as additional degrees of freedom of the gauge fields. Hence the gauge fields corresponding to the broken symmetry acquire mass.

The Goldstone theorem can be applied to the standard model $SU(2)_L \times U(1)_Y$ symmetry in order to give mass to the W and Z gauge bosons. One has to introduce a new field, called the Higgs field Φ, a complex doublet of scalar fields . The Lagrangian density associated with this field is chosen as,

$$\mathcal{L} = (D_\mu \Phi)^\dagger D^\mu \Phi + \mu^2 \Phi^\dagger \Phi - \lambda (\Phi^\dagger \Phi)^2. \tag{1.1}$$

with the covariant derivative given as

$$D_\mu \Phi = (\partial_\mu + igW_\mu^a T^a + ig' B_\mu)\Phi, \tag{1.2}$$

with T_a being the $SU(2)_L$ group generators and W_μ^a, B_μ the gauge fields. The g, g' couplings are expressed in terms of the Fermi coupling, the electron charge and the W mass as: $g^2 = \frac{M_W^2 G_F}{\sqrt{2}}$, $e = g \sin\theta_W$ and $g' = g \tan\theta_W$. The last two terms of Eq. 1.1 constitute the Higgs potential invariant under $SU(2)_L \times U(1)_Y$. Its shape is determined by the μ and λ parameters, and if $\mu^2/\lambda > 0$, its minimum is nonzero and equal to $\Phi^\dagger \Phi = \frac{\mu^2}{2\lambda}$.

The point is that the configurations of Φ minimising the potential are degenerated. In other words, the ground state is not invariant under $SU(2)_L$, while the whole Lagrangian is. The Goldstone theorem states that there will be three Goldstone bosons since the gauge symmetry has three of its four generators that do not leave the ground state invariant.

To see how the gauge boson acquires mass, let us choose a vacuum expectation value (v.e.v) for this field to be real with zero charge,

$$\langle \Phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}, \tag{1.3}$$

where $v = \mu/\sqrt{\lambda}$. One can then redefine the Φ field as

$$\Phi(x) = e^{iT^a \theta_a(x)} \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}. \tag{1.4}$$

It is always possible to apply an $SU(2)_L$ gauge transformation to rotate away the θ_a dependence of the Φ field. The three degrees of freedom embedded into the θ_a dependence form the massless Goldstone bosons. Then, expanding the kinetic term of Eq. 1.1 with the explicit value of Φ and expressing the physical fields as $Z_\mu = \cos\theta_W W^3_\mu - \sin\theta_W B_\mu$ and $W_\mu = 1/\sqrt{2}(W^1_\mu + iW^2_\mu)$, we have:

$$(D_\mu \Phi)^\dagger D^\mu \Phi = \frac{1}{2}\partial_\mu H \partial^\mu H + (v+H)^2 \left(\frac{g^2}{4} W^\dagger_\mu W^\mu + \frac{g^2}{8\cos^2\theta_W} Z_\mu Z^\mu \right). \tag{1.5}$$

We can see a mass term appearing for the Z and $W^\pm$ bosons as well as a coupling between the gauge bosons and the Higgs field. They are related to the Higgs v.e.v in the relation: $M_W = M_Z \cos\theta_W = gv/2$, while the photon remain massless.

At this stage, it has only been possible to explain the gauge boson mass with the Higgs mechanism, but nothing has been said about the origin of the fermion masses. In order to add a mass term in the standard model Lagrangian, one has to contract the left-handed leptons with a doublet in order to form an invariant quantity. As an example, we shall construct a term that mixes a first generation left-handed lepton doublet with the right-handed lepton and the Higgs field Φ as:

$$-G_e l^\dagger_L \Phi e_R = \frac{G_e v}{\sqrt{2}} (e^\dagger_L e_R) \quad m_F = \frac{G_e v}{\sqrt{2}}, \tag{1.6}$$

with $l_L = \begin{pmatrix} \nu_L \\ e_L \end{pmatrix}$. Then the electron has acquired mass that is proportional to the v.e.v of the Higgs field. Although the Higgs field can give mass to the fermions, it arises at the cost of a new coupling (G_e) for each new fermion to account for a complete mass spectrum. The Higgs mechanism is an elegant construction to give mass to the bosons mediating the weak force and to a lesser extent to the fermions.

One shortcoming is that the recently observed neutrino mass is not explained in the standard model framework. Another shortcoming of the standard model Higgs is the so-called fine-tuning problem. The measured Higgs mass is given by

$$m^2_H = (m^2_H)_0 + \Delta m^2_H, \tag{1.7}$$

where $(m_H^2)_0$ is the "bare" Higgs mass and Δm_H^2 are corrections from loop diagrams. The correction to m_H^2 from a loop containing a fermion is given by

$$\Delta m_H^2 = \frac{\lambda_F^2}{16\pi^2}(-2\Lambda_{UV}^2 + 6m_F^2 \ln(\Lambda_{UV}/m_F) + \cdots). \tag{1.8}$$

In the equation above, λ_F is the coupling constant of the Higgs to the considered fermion. Λ_{UV} is a cut-off that can be interpreted as the energy scale at which new physics enters that alters the behaviour of the theory. If no new physics occur between the Higgs mass scale and the Plank scale, then Λ_{UV} amounts to the Plank mass $m_p = 10^{19}$ GeV. The corrections to the bare Higgs mass evolve in quadrature with the Λ_{UV} cut-off with a negative sign. If the Higgs mass is around 100 GeV, then the Higgs bare mass should be at the Planck mass so that the correction term cancels the bare mass to obtain the expected mass around 100 GeV, this is what is referred to as the fine-tuning problem. One can also compute the effect of the boson loops that contribute to the Higgs mass. The expression is similar to the one of Eq. 1.8 but the correction comes with a positive sign. This means that, if one could construct a theory where each boson could be associated with each lepton, then the quantum corrections would cancel exactly.

1.2 Extensions to the Standard Model

In the previous section we found a motivation to introduce a new symmetry that relates fermions to bosons since it would avoid the fine-tuning problem. In this section we will discuss briefly the Minimal Supersymmetric Standard Model (MSSM).

The principle of supersymmetry is to introduce a link between matter particles (fermions) and force carriers (bosons). Thus one has to build an operator Q that generates such transformations

$$Q|\text{Boson}\rangle = |\text{Fermion}\rangle; \qquad Q|\text{Fermion}\rangle = |\text{Boson}'\rangle. \tag{1.9}$$

As derived in [4] the Q operators should follow the " "anti-commutation" rule:

$$\{Q_\alpha, Q_\beta^\dagger\} = 2\sigma_{\alpha\beta}^\mu P_\mu. \tag{1.10}$$

where P^μ is the generator of space-time translation and σ^μ are the Pauli matrices[1]. The particle states of the supersymmetry algebra are grouped in supermultiplets. A supermultiplet contains a fermion and the boson associated by the Q operator. Moreover, one can show that the mass operator P^2 commutes with Q, implying that the mass of the particles contained in a supermultiplet must be the same. The Q operators also commute with the generators of the gauge symmetries. Therefore particles in the same supermultiplet must have the same electric charges, weak isospin and colour degrees of freedom. For each existing fermion (boson) a new boson (fermion) should be introduced and it should match the previous requirement in order to fit in the supermultiplet representation.

[1] with $\sigma^0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$ $\sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

In the previous paragraph we stated that the masses of the particles contained in a supermultiplet must be the same. This means that the supersymmetry must be broken since no partner of the known fermions and bosons have been observed with the same mass.

In the MSSM the Higgs sector consists of two Higgs doublets, to give mass both to the up and down type fermions. These two Higgs doublets have positive and negative hyper charges values $Y = \pm 1$. They are $\Phi_d = (\phi_d^0, \phi_d^-)$ that couples to down type fermions and $\Phi_d = (\phi_u^+, \phi_u^0)$, that only couples to up type fermions. Minimising the Higgs potential yields vacuum expectation values for the Higgs field. They can always be chosen of the form:

$$\Phi_d^{vev} = \begin{pmatrix} 0 \\ v_d \end{pmatrix}, \Phi_u^{vev} = \begin{pmatrix} v_u \\ 0 \end{pmatrix}. \tag{1.11}$$

In the equation above, the vacuum expectation values can be related to the W boson mass and the electroweak gauge couplings by the following relation:

$$v_u^2 + v_d^2 = v^2 = \frac{2m_W^2}{g} \approx (174 \text{ GeV})^2. \tag{1.12}$$

Although the vev's are related to the W mass, the ratio of the vev's is not fixed by present experiments. We will define this ratio as $\tan\beta = v_u/v_d$.

After $SU(2)_L \times U(1)_Y$ has been broken, three degrees of freedom of the Higgs doublet become the longitudinal modes of the Z^0 and $W^\pm$ bosons. The remaining five degrees of freedom constitute the MSSM Higgs sector. These mass eigenstates are related to the gauge eigenstates Φ_u, Φ_d through a phase transformation that involves two angles α and β. They are given by:

$$A^0 = \sqrt{2}(\cos\beta \text{Im}[\phi_u^0] + \sin\beta \text{Im}[\phi_d^0]) \tag{1.13}$$

$$H^+ = (\cos\beta\phi_u^+ + \sin\beta\phi_d^{-*} \tag{1.14}$$

$$\begin{pmatrix} h^0 \\ H^0 \end{pmatrix} = \sqrt{2} \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} \text{Re}[\phi_u^0] - v_u \\ \text{Re}[\phi_d^0] - v_d \end{pmatrix}. \tag{1.15}$$

The angle α can be computed by diagonalising the CP-even Higgs squared-mass matrix.

$$\cos^2(\beta - \alpha) = \frac{m_h^2(m_Z^2 - m_h^2)}{m_A^2(m_H^2 - m_h^2)} \tag{1.16}$$

One can see that there are three neutral Higgses, A^0, H^0, h^0 and two charged Higgses $H^\pm$. At tree level, the Higgs masses depend only on two parameters, namely, m_{A^0} and $\tan\beta$. They are expressed as:

$$m_{H^\pm}^2 = m_{A^0}^2 + m_W^2 \tag{1.17}$$

$$m_{h^0} = \frac{1}{2}\Big(m_{A^0}^2 + m_Z^2 - \sqrt{(m_{A^0}^2 + m_Z^2)^2 - 4m_Z^2 m_{A^0}^2 \cos^2 2\beta}\Big) \tag{1.18}$$

$$m_{H^0} = \frac{1}{2}\Big(m_{A^0}^2 + m_Z^2 + \sqrt{(m_{A^0}^2 + m_Z^2)^2 - 4m_Z^2 m_{A^0}^2 \cos^2 2\beta}\Big) \tag{1.19}$$

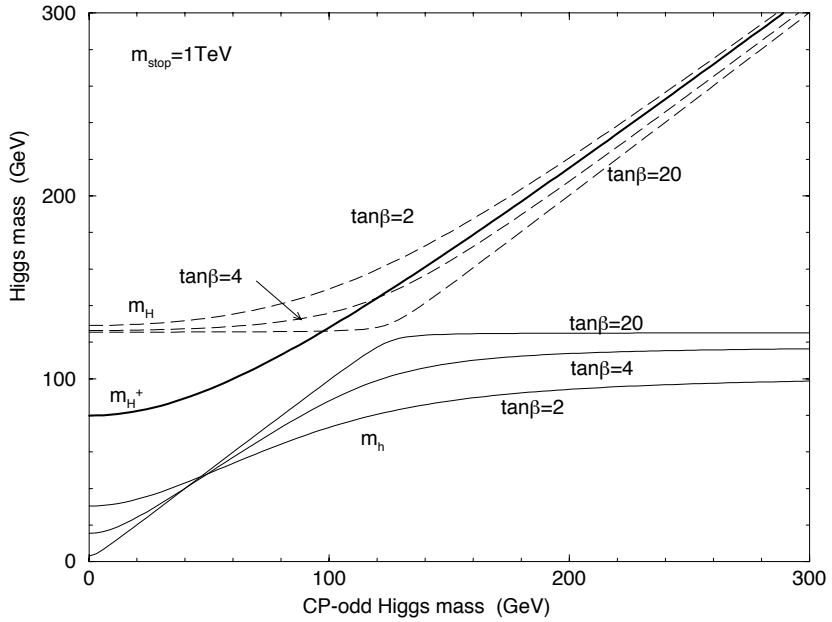


Figure 1.1: Masses of the MSSM Higgs bosons as a function of m_A for different values of $\tan\beta$ [5].

From expression 1.18 one can derive an upper bound on m_h at tree level,

$$m_h^2 \leq m_Z^2 \cos^2 2\beta, \tag{1.20}$$

but the radiative corrections can push the mass up to m_h^{max} around $110 - 140$ GeV depending of $\tan\beta$, the trilinear Higgs stop coupling (A_t), and the top mass. The mass behaviour of the Higgses is shown in Fig. 1.1. Two mass regimes can be identified:

- $m_A \gg m_h^{max}$: The three Higgses $m_A \sim m_H \sim m_H^{\pm}$ are degenerate in mass. $m_h \sim m_h^{max}$ and h couples to bosons and fermions as the Higgs of the standard model (SM).
- $m_A \ll m_h^{max}$: the H behaves like the SM Higgs while the h and A will be degenerate in mass.

The coupling behaviour of the MSSM Higgses to fermions and bosons can be compared to the SM coupling. Table 1.1 summarises the ratio of the MSSM coupling to fermions and bosons to the corresponding SM couplings. The couplings of h and H are generally suppressed with respect to the SM Higgs couplings. The table simplifies when considering the two regimes discussed previously. In the large m_A regime, one can derive from Eq. 1.16 that $\cos(\beta - \alpha) \ll 1$. This implies that the h will couple to bosons as the SM Higgs does and that the couplings of H to bosons will be strongly suppressed. In this scenario, the coupling to down-type fermions is enhanced by a

decay mode	h	H	A
$u\bar{u}$	$\cos\alpha/\sin\beta$	$\sin\alpha/\sin\beta$	$\cot\beta$
$d\bar{d}$	$-\sin\alpha/\cos\beta$	$\cos\alpha/\cos\beta$	$\tan\beta$
W^+W^-,ZZ	$\sin(\beta-\alpha)$	$\cos(\beta-\alpha)$	0

Table 1.1: Couplings of the MSSM neutral Higgses relative to SM couplings. d is one of the down-type fermions (d, s, b, e, μ, τ) and u stands for $(u, c, t, \nu_e, \nu_\mu, \nu_\tau)$

$\tan\beta$ factor both for the H and the A. Since the coupling is proportional to the fermion masses, the third generation will be the dominant in the H and A decay modes. This also explains why the associated production with $b\bar{b}$ pairs is important at LHC. The decay into tau pairs is also a important decay channel. In the region $m_A \ll m_h^{max}$ the roles of the H and the h are reversed, and the H behaves as in the standard model since $\sin^2(\beta-\alpha) \ll 1$ while the coupling of the h to gauge boson is suppressed.

In the next section we will briefly discuss the experimental constrains on the MSSM parameter space.

1.3 Limits on the MSSM Parameter Space

The LEP has pursued several searches for the supersymmetric Higgs boson. The search was carried out in $e^+e^- \to h^0A^0$ and in $e^+e^- \to h^0Z^0$ channels. The cross section for $e^+e^- \to h^0A^0$ goes as $\cos^2(\beta-\alpha)$ and as $\sin^2(\beta-\alpha)$ for $e^+e^- \to h^0Z^0$. The non observation of events in these channels leads to an exclusion limit in the $m_h, \sin^2(\beta-\alpha)$ region with $m_h > 89.8$ and $m_A > 90.1$.

Putting a limit on $\tan\beta$ is indirect and a choice of the MSSM free parameters is required to adduce a limit. The $m_{h^0}^{max}$ max scenario [6] is the framework used to derive the most conservative limit on the $\tan\beta$ parameter. In this framework which maximises the radiative correction to the CP even Higgs as a function of $\tan\beta$, the LEP combined results exclude the range $0.5 < \tan\beta < 2.4$ at 95% CL. The radiative corrections to m_h comes[2] from the $\tilde{t} - t$ and also from $\tilde{b} - b$ for large $\tan\beta$.

The difference between the maximal mixing and no mixing benchmark comes in the radiative corrections to the m_h. An upper bound on the one loop correction to m_h is given as:

$$m_h^2 \lesssim m_Z^2 + \frac{3g^2 m_t^4}{8\pi^2 m_W^2}\left[ln\left(\frac{M_S^2}{m_t^2}\right) + \frac{X_t^2}{M_S^2}\left(1 - \frac{X_t^2}{12M_S^2}\right)\right], \tag{1.21}$$

assuming $M_{\rm SUSY} = M_{\tilde{t}_L} = M_{\tilde{b}_L} = M_{\tilde{t}_R} = M_{\tilde{b}_R}$, and $M_S^2 = M_{\rm SUSY}^2 + m_t^2$ and $X_t = (A_t - \mu\cot\beta)$. The Higgs mixing parameter μ appears in the MSSM superpotential.

[2]The tilde above the particle means the supersymmetric partner of that particle, $\tilde{t}$ stands for the stop

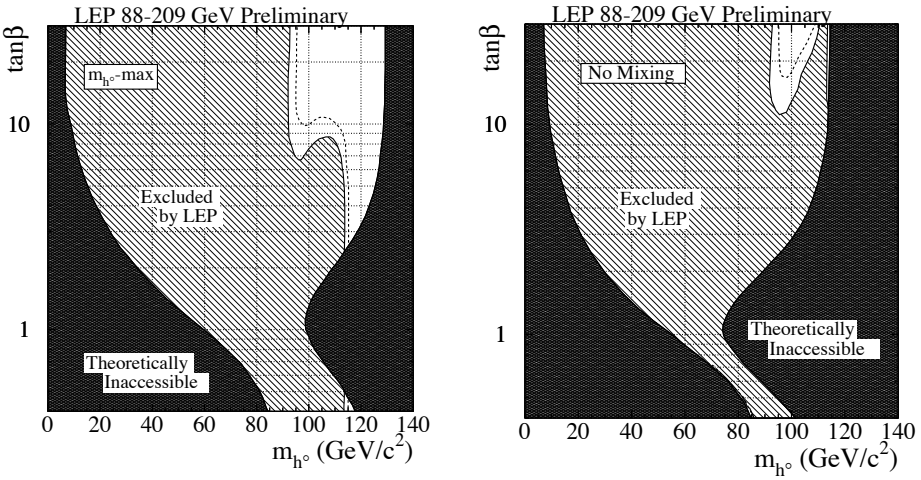


Figure 1.2: Exclusion region [7] in the $m_h, \tan\beta$ plane. The cases of the $m_{h^0} - \max$ (Left) and no mixing scenario (Right) are shown.

It comes as an off-diagonal term in the neutralino mass matrix that mixes the up and down higgsinos to form two neutralinos. A_t denotes the trilinear Higgs-stop coupling. The m_t^4 dependence on the h mass upper bound shows that the higgs mass can be significantly larger than its tree level value. The mass correction increases with X_t reaching its maximal value for $X_t = \sqrt{(}6)M_S$ which is called the maximum mixing case. $X_t = 0$ correspond to the minimal mixing.

The exclusion regions for the $\tan\beta$ parameter are much more stringent in the no mixing scenario. The combined LEP limits exclude the regions where $0.7 < \tan\beta < 10.5$ at 95% CL. As we have discussed, the limit on the $\tan\beta$ parameter is particularly sensitive to the top mass. The recent measurement of the top mass indicates a larger value for m_t than the one of Tevatron Run 1 [8] . This implies that the theoretically excluded $\tan\beta$ region is reduced. As a simple rule of thumb, one can show [9] that a variation of m_t by δ_{m_t} induces a shift in the m_h prediction by about δ_{m_t}.

As we have seen in the previous section, the τ lepton will play an important role in the search for the MSSM Higgses. In the next section we will describe the principal features of the τ lepton.

1.4 Decay of the tau lepton

As we have seen in Table. 1.1 the enhanced coupling of the MSSM heavy Higgs to tau leptons motivates the search for tau final states. The first hint of the existence [10] of the τ was observed in 1975 by M.L. Perl's[3] group, at SLAC-LBL in e^+e^- annihilation. They observed $e^\pm\mu^\mp$ events with undetected particles. The missing momenta spectra

[3]Nobel Prize winner in 1995 with F. Reines for the discovery of the tau lepton

suggested that at least two additional particles were produced in each event. Since then, the τ lepton has been intensively studied.

Decay channel	Branching Ratio ± Errors
$\tau^- \rightarrow \nu_\tau \pi^0 \pi^-$	25.42 ± 0.14%
$\tau^- \rightarrow e^- \widetilde{\nu_e} \nu_\tau$	17.84 ± 0.06%
$\tau^- \rightarrow \mu^- \widetilde{\nu_\mu} \nu_\tau$	17.36 ± 0.06%
$\tau^- \rightarrow \nu_\tau \pi^-$	11.06 ± 0.11%
$\tau^- \rightarrow \nu_\tau \pi^0 \pi^0 \pi^-$	9.17 ± 0.14%
$\tau^- \rightarrow \nu_\tau \pi^+ \pi^- \pi^-$	9.12 ± 0.10%
$\tau^- \rightarrow \nu_\tau \pi^0 \pi^+ \pi^- \pi^-$	4.37 ± 0.09%
$\tau^- \rightarrow \nu_\tau \pi^0 \pi^0 \pi^0 \pi^-$	1.08 ± 0.10%
$\tau^- \rightarrow \nu_\tau K^-$	0.6 ± 0.023%
$\tau^- \rightarrow \nu_\tau \pi^0 \pi^0 \pi^+ \pi^- \pi^-$	0.5 ± 0.04%
$\tau^- \rightarrow \nu_\tau \pi^+ \pi^- K^-$	0.48 ± 0.04%
$\tau^- \rightarrow \nu_\tau \pi^0 K^-$	0.45 ± 0.03%
$\tau^- \rightarrow \nu_\tau \pi^0 \bar{K}^0 \pi^-$	0.37 ± 0.04%
$\tau^- \rightarrow \nu_\tau \pi^- K^+ K^-$	0.15 ± 0.007%
$\tau^- \rightarrow \nu_\tau \pi^0 \pi^- \eta$	0.17 ± 0.02%
$\tau^- \rightarrow \nu_\tau \bar{K}^0 \pi^-$	0.8 ± 0.04%

Table 1.2: Dominant branching fractions of the tau lepton decays taken from [11].

The large tau lepton mass, 1.7 GeV, provides the availability of hadronic decay channels (Table. 1.2), which constitute about 65 % of the tau branching ratios. Throughout this dissertation, the hadronic tau decays will be referred as tau-jet. One prong will refer to the decay of the tau in one pion or kaon while three prong will refer to all other hadronic decays.

The tau lepton is a short-lived particle, with a mean life time of 290 ps. This means that in the decay of a heavy Higgs boson one could see the secondary vertex of the τ.

CHAPTER

TWO

The CMS experiment

The Compact Muon Solenoid (CMS) detector is a multipurpose detector that is being built at CERN. Although its design is "compact", CMS is 21.6 m long, has a diameter of 14.6 m and weights 12500 tons. Before the start of the LHC, CMS will be lowered into one of the four caverns (Fig. 2.1) that host the four LHC experiments. By mid 2007 the LHC is expected to collide protons with a centre of mass energy of 14 TeV at a bunch crossing frequency of 40 MHz. The nominal design luminosity of the LHC is 10^{34} cm^{-2}s^{-1}. It implies a major challenge to analyse a data volume of the order of one peta-byte each year.

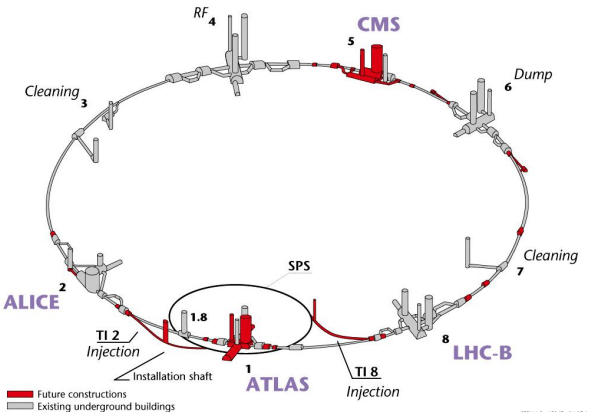


Figure 2.1: Schematic view of the LHC tunnel. Several access pits and caverns are sketched. CMS will be installed at the opposite of the CERN Meyrin site.

2.1 Particle detectors

A schematic view of a quarter of the CMS detector is sketched on Fig. 2.2. Therein the sub-detectors constituting the complete CMS machine are arrange like the layers of an onion. In the following sections we will discuss the different sub-detectors. Several sub-detectors are located inside the coil, which provides a magnetic field of 4T. When going from the interaction point outwards one finds the pixel detector, the silicon-tracker, the electromagnetic calorimeter and the hadronic calorimeter. The muon chambers are located outside the coil.

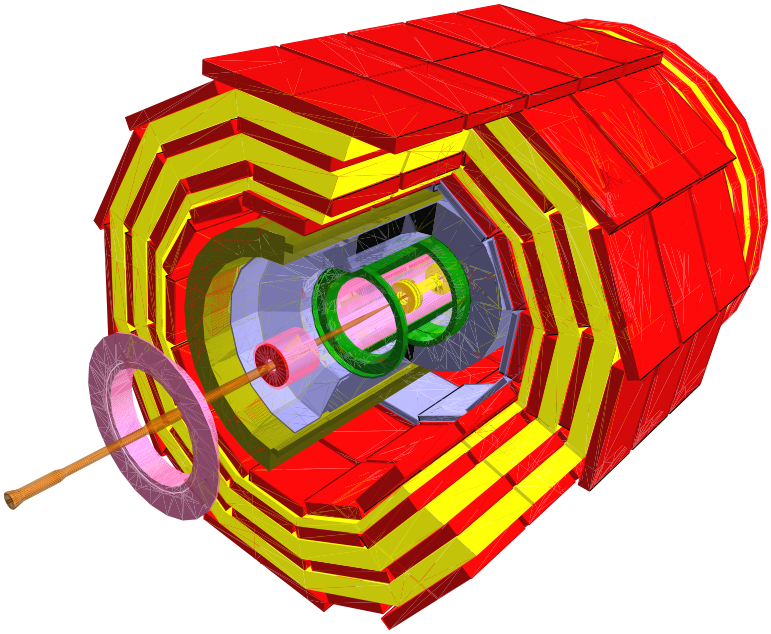


Figure 2.2: 3D view of the whole CMS detector

2.1.1 Pixel detector

The pixel detector is the first sensitive material that particles will cross after leaving the interaction point. It is used for vertex finding and its role is particularly important for the High Level Trigger selection since high QCD jet rejection can be achieved by achieved by analysing its data in a time that is affordable for the HLT. In addition, b quark tagging, is another motivation for the development of such a device.

The pixel detector is a small cylindrical device [12] shown on Fig. 2.3. It is made

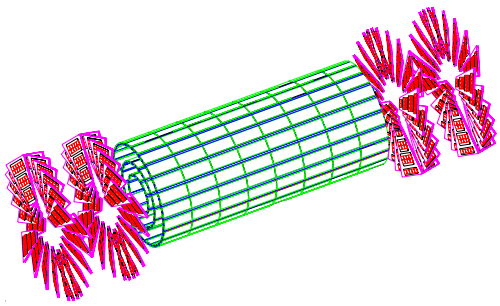


Figure 2.3: Schematic view of the pixel detector with its three layers and two endcaps. The typical pad size is $(150 \times 150)\ \mu\text{m}^2$.

of three barrel layers and two pairs of end-cap disks. The total pixel detector length is 120 cm. Its internal radius is 4 cm and external radius of 15 cm. The detector has a two-hit coverage that extends up to $\eta = 2.2$.

In order to achieve high resolution on the vertex position, analog signal interpolation is used. The 4T field induces a charge drift used to improve the position resolution. Hence turbine geometry in the end-caps disks has been chosen to have a non zero Lorentz angle and profit from this effect to improve the hit position measurement.

When running at the nominal LHC luminosity, about 1000 tracks cross the pixel barrel, hitting about 9000 pixels every 25 ns since one track hits three pixel on average. The high number of channels of the pixel detector $\sim 45 \times 10^6$ and a typical occupancy of 10^{-4} translates to a counting rate of about 10 kHz per pixel.

The pixel detector resolution is $\approx 15\ \mu$m in the z coordinate and about 10 μm in the $r\phi$ coordinate. Two methods are used for the finding of the primary vertex in the z coordinates; the histogramming method and the divisive method. Both methods use a clustering technique to combine tracks that have a similar longitudinal impact parameter. A χ^2 is minimised to define the primary vertex. A detailed description can be found in [13].

Both methods obtain a longitudinal resolution on the vertex position of $\sim 50\ \mu$m and $\sim 40\ \mu$m for the low and high luminosity regime, respectively.

2.1.2 Silicon Strip Tracker

The CMS tracker described in [14] and [15] has been designed to have a resolution $\sigma_{P_t}/P_t \sim 0.015 \times P_t\%$ (P_t is expressed in GeV) and a reconstruction efficiency of 96% for isolated high P_t tracks. If one does not take into consideration the multiple scattering, the tracker resolution is proportional to:

$$\frac{\sigma_{P_t}}{P_t} \propto \frac{\epsilon P_t}{0.3BL^2}, \tag{2.1}$$

where B is the magnetic field in Tesla, ϵ the position measurement error perpendicular to the trajectory, and L is the track length in the transverse plane. The diameter of the tracker being fixed, the magnetic field has to be maximised while ϵ must be as small as possible. Hence, the whole tracker is immersed in a 4T magnetic field and made of more than 15000 detector modules arranged in a barrel and end-caps with a radii of 1.1 m.

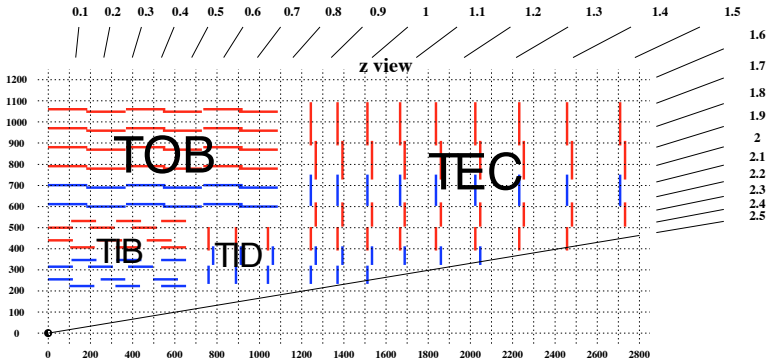


Figure 2.4: Longitudinal view of one quadrant of the Silicon Strip Tracker. TIB, TOB, TEC, TD acronyms are described in the text. Sensors labelled in red (blue) are single sided (stereo).

Fig. 2.4 shows a longitudinal view of the different elements involved in the whole tracker. They are four different detector "units", the Tracker Inner Barrel (TIB), the Tracker Inner Disks (TID), the Tracker Outer Barrel (TOB), and the Tracker End Caps (TEC) totalling 206 m^2 of silicon strip detector. Each track is made of up to 13 hits when taking into account the three layers of the pixel detector.

Figure 2.5: A TEC module and its reading electronics.

The module, Fig. 2.5, is the basic 'unit' of the tracker system. They are made of one, two or four silicon sensors supported by a carbon fibre frame. As shown on Fig.2.4, some layers are equipped with stereo sensors made of two silicon detectors assembled back to back with a tilt angle. This allows a 3D measurement of the track

hit.

The strip pitch is not the same for TIB and TOB since the track density is lower in the outer barrel than in the inner barrel. TOB has a pitch of $120 - 200$ μm while TIB,TEC,TID has $80 - 120$ μm. Using charge sharing between strips, it is possible to measure the hit position in the r, ϕ plane with a resolution $\sigma_{r,\phi} = 40 - 60\mu$m.

The reconstruction efficiency of the tracking algorithms is almost 100% up to $|\eta| = 2$. It uses Kalman filter techniques [16] that describe the track with five parameters evolving discretely with the number of collected hits. The transverse momentum resolution is smaller than 3% for a muon of less than 100 GeV up to $|\eta| = 1.75$. The resolution improves with decreasing the P_t as expressed in equation 2.1 but the multiple scattering effect grows with $1/P_t$. This explains why the resolution for muons of 10 GeV is almost the same as for 1 GeV muons.

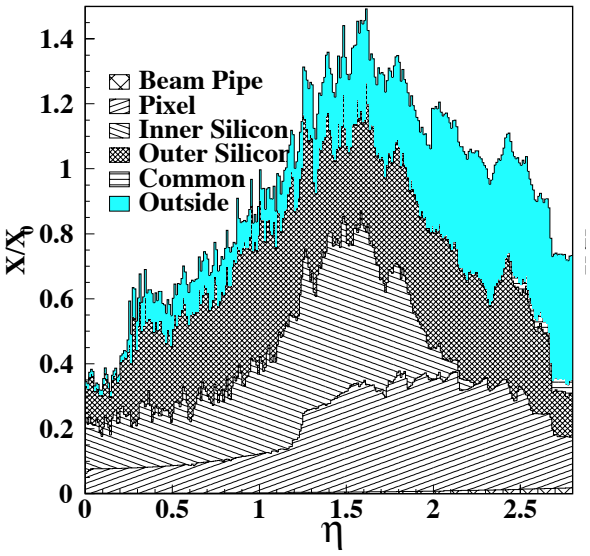


Figure 2.6: Material budget of the CMS tracker n terms of radiation lengths as function of η

The material budget induced by the installation of the tracker is important for the performance of the calorimeters. It implies that the tracker material should be minimised in order to avoid secondary interactions, excessive multiple scattering and bremsstrahlung. In the $|\eta|$ range of 1.2 to 2 the tracker material constitute more than one radiation length as shown on Fig. 2.6.

2.1.3 Calorimeters

ECAL

The electromagnetic calorimeter (ECAL) [17] has to measure electrons energy and position with high precision, in addition their identification should occur with high efficiency. The ECAL is made of a barrel and two end-caps covering the region $|\eta| < 1.5$ for the barrel and $2.5 < |\eta| < 1.5$ for the endcaps.

The observation of the Higgs decaying to two photons is a good benchmark channel to qualify the calorimeter quality. In that channel, the mass resolution is directly related to the energy and angular resolutions by,

$$\frac{\sigma_{m_H}}{m_H} = \frac{\sigma_{E_1}}{E_1} \oplus \frac{\sigma_{E_2}}{E_2} \oplus \frac{\sigma_\theta}{\tan\theta/2}, \tag{2.2}$$

where E_1, E_2 (in GeV) are the energies of the two photons and θ is their angular separation in radians.

In order not to degrade the intrinsic calorimeter resolution, a compact design has been chosen that can fit into the magnet. The calorimeter is made of 76000 Lead Tungstate ($PbWO_4$) scintillating crystals that have a small radiation length $X_0 = 0.89$ cm. Their quick scintillation response allows collection of 80% of the light in the approximately 25 ns that is required for the LHC conditions. One drawback of the ($PbWO_4$) crystals is their low light yield, that requires photo detectors with intrinsic amplification. Their light yield is also strongly dependent on their temperature, requiring that they operate in a stable temperature environment of 18° ±0.05°. This stability is achieved with water flow.

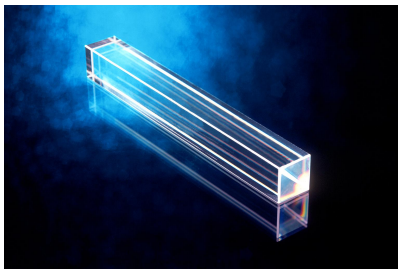


Figure 2.7: Sample of a $PbWO_4$ crystal. The crystal length is 23 cm and the front face has a surface of 22×22 mm^2.

The energy resolution terms in equation 2.2 can be parametrised as

$$\frac{\sigma_E}{E} = \frac{0.027}{\sqrt{E}\ (\text{GeV})} \oplus \frac{0.155}{E(\ \text{GeV})} \oplus 0.5. \tag{2.3}$$

The first term is related to the photo-statistics; the second term accounts for the electronic noise and pileup energy. Finally, the last term contains several contributions

including the longitudinal shower containment, non-uniformity of light collection and leakage currents in the photo detectors.

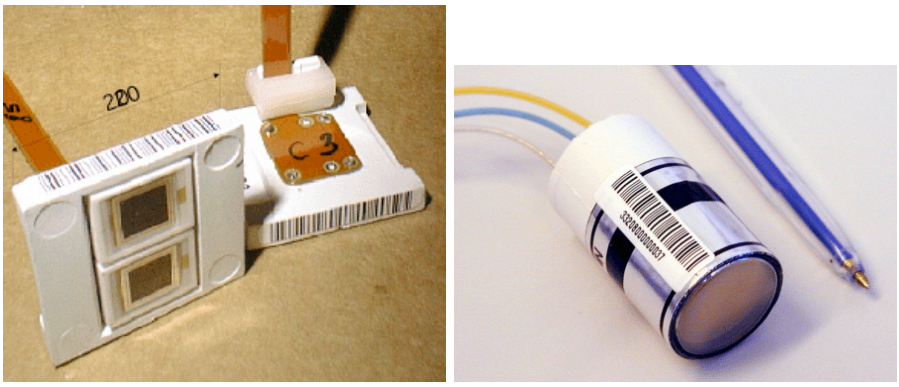


Figure 2.8: APD (left) and VPT (right) detectors used for the light collection.

Two technologies have been chosen to read the ECAL crystals, corresponding to the varying magnetic field orientations and radiation that the detectors have to hold. In the barrel, the crystals will have to withstand up to 0.3 Gy/h and up to 15 Gy/h in the end-caps. For what concerns the position resolution, 1 mrad is obtained in the barrel. For a 115 GeV Higgs, the contribution of the photon energy resolution dominates the mass resolution.

The normal orientation of the magnetic field to the crystals axis in the barrel forbids the use of vacuum photo-triodes (VPT) shown in Fig. 2.8). Avalanche photo diodes (APD) are used instead and mounted in pairs at the back of the crystals. The low yield of $PbWO_4$ crystals need the APD to operate with a gain of 50. In the end-caps VPT are used since they are far more radiation resistant than APD. Their operation gain is around 10.

The ECAL end-cap is equipped with a preshower [18] detector to distinguish between photon pairs originating from a neutral pion and single photons. On the end-cap region, a pion with $E_T = 60$ GeV will result in two photons only separated by only a few millimetres while they will be one centimetre apart in the barrel. This cannot be resolved by the crystal themselves and requires the use of a silicon detector called preshower. The preshower consists of two layers of silicon detector (with orthogonal strips), sandwiched between two layers of lead absorber (1st $\sim 2X_0$, 2nd $\sim 1X_0$). A pion rejection factor of 3 is obtained in the energy range of the $H \rightarrow \gamma\gamma$ channel.

The ECAL barrel is not equipped with a preshower. Other approaches are needed to reject π_0. As an example, we have estimated the rejection of π_0 with an artificial neural network that uses the cluster shapes to distinguish between the deposit of a π_0 and of electrons based on the method developed in L3 [19]. The neural network is constructed with two hidden layers of 15 perceptrons, fed by 10 input variables sensitive to the shape of the cluster and energy deposit ratios. For a given cluster we define four energy variables:

- $S1$: Energy of the most energetic crystal in the cluster
- $S4$: Energy of the max 2×2 crystal array containing $S1$
- $S9$: Energy of an array of 3×3 crystals centred on $S1$
- $S25$: Energy of an array of 5×5 crystals centred on $S1$

In addition, the cluster geometrical shape is accounted for with the following variables defined on the array of 5×5 crystals as:

$$\sigma_{ij} = \frac{\sum_{k=1}^{25} E_k (x_k^i - \langle x^i \rangle)(x_k^j - \langle x^j \rangle)}{\sum_{k=1}^{25} E_k} \tag{2.4}$$

$$\lambda_{\pm} = \frac{\sigma_{\eta\eta} + \sigma_{\phi\phi} \pm \sqrt{(\sigma_{\eta\eta} - \sigma_{\phi\phi})^2 + 4\sigma_{\eta\phi}^2}}{2}, \tag{2.5}$$

with $ij = \{\phi\eta, \phi\phi, \eta\eta\}$. The variables used as input to the neural network are combinations of the variables described above. They are summarised in Table 2.1

$v_1 = \frac{S1(2)}{S9(2)}$	$v_2 = \frac{\lambda_-(2)}{\lambda_+(2)}$	$v_3 = 10(\sigma_{\eta\phi}(2) + 0.05)$
$v_4 = \frac{S1(1)}{S9(2)}$	$v_5 = \frac{S9(1) - S1(1)}{S25(1) - S1(1)}$	$v_6 = \frac{S1(1)}{S4(1)}$
$v_7 = \lambda_-(1)/\lambda_+(1)$	$v_8 = 10(\sigma_{\eta\phi}(1) + 0.05)$	$v_9 = 10\sigma_{\eta\eta}(1)$
$v_{10} = 2\sigma_{\phi\phi}(1)$		

Table 2.1: Neural network input variables. For the case where there is only one cluster associated with the super-cluster, the variables $v_{1,2,3}$ are set to zero.

The neural network [20] has been trained on a sample of π_0 and electrons with a p_t of 30 GeV. Fig. 2.9 shows the distribution obtained both for electron and π_0 test samples.

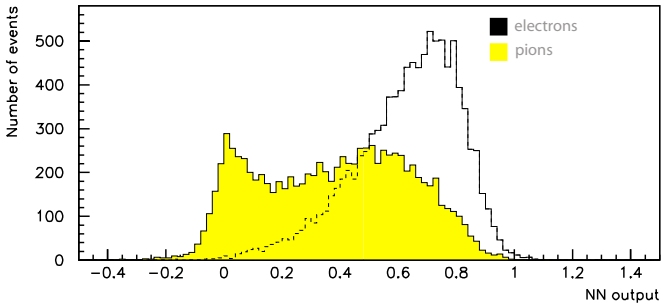


Figure 2.9: Neural net output for 30 GeV pions and electron on the test sample

HCAL

The Hadronic calorimeter [21],[22] is immediately downstream of the ECAL calorimeter. It is a sampling calorimeter made in a sandwich type of design as shown in Fig. 2.10. Copper absorber plates act as radiators, and are alternated with active elements made of plastic scintillator read out using wavelength-shifting plastic fibres. Copper alloy is a "relatively low Z" material which minimises the multiple scattering as muons go through the HCAL. The η and ϕ segmentation in the barrel corresponds to arrays of 5×5 ECAL crystals: 0.087×0.087 in η, ϕ.

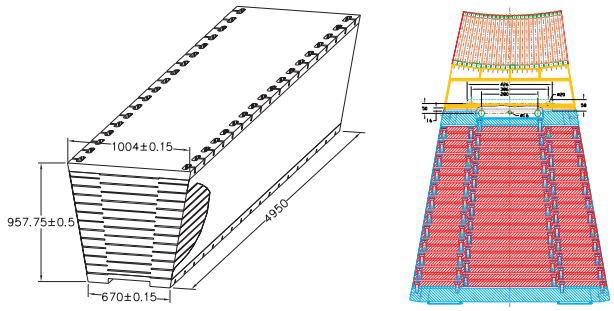


Figure 2.10: Isometric (left) and phi (right) view of a HCAL wedge. On the right, the ECAL is mounted on top of it. Units on the figures are in mm.

The barrel $|\eta|$ coverage extends up to 1.5, the end-caps up to 3. The total length of the HCAL barrel is 6.5 interaction length taking into account the ECAL. Hence the energy leakage from the barrel should be measured. Therefore, an outer barrel calorimeter is placed between the first muon chambers and the magnet cryostat. Finally the very forward calorimeter (VF) is 10 meters away from the interaction point and is read with quartz fibre that can sustain high radiation doses.

The typical resolution for a pion interacting in HCAL is given by

$$\sigma_E/E = 0.9/\sqrt{E} \oplus 0.048. \tag{2.6}$$

For a pion of 100 GeV the resolution is about 10 %.

2.1.4 Muon detectors

The passage of muons through matter is described by the Bethe-Bloch formula. The bremsstrahlung component is very low since it is proportional to m^{-2}, hence muons are highly penetrating particles. Therefore, the muons chambers are placed around the magnet cryostat. The muon chambers [24] are used to improve the P_t resolution by increasing the lever arm (L) that appears in equation 2.1.

Three technology were used to build the muon detector. Resistive plate chambers (RPC) are used up to $\eta = 2.1$, for the bunch crossing time identification. They are

made of two Bakelite plates separated by a few mm and coated with graphite on the outside. The graphite forms a cathode that is fed with a high voltage. An aluminium strip reads the avalanche generated by a particle crossing the detector. CMS has chosen to use RPCs in avalanche mode. This mode requires amplification at the front end since the collected charge is not as high as in streamer mode. The RPCs have a time resolution of about 1 ns, which makes them very good at identifying the bunch crossing time.

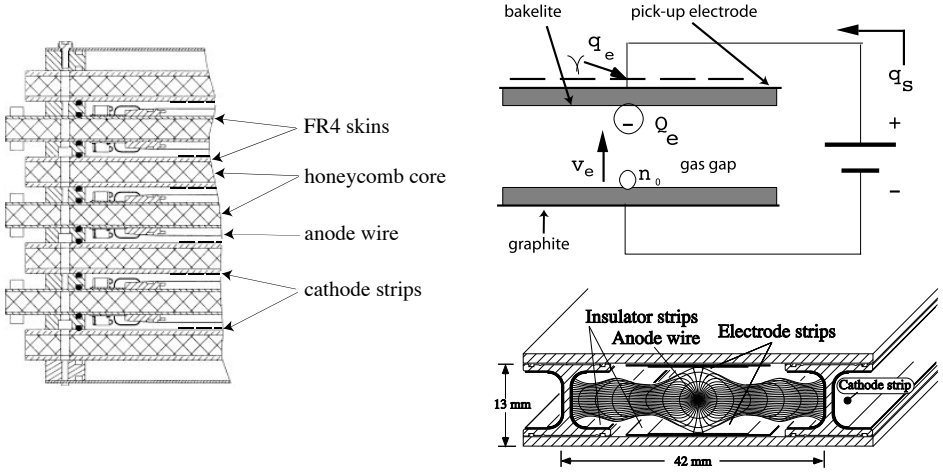


Figure 2.11: On the left, schematic view of cathode strip chambers. Top right: the principles of a resistive plate chamber are shown. A drift tube cell is illustrated bottom right.

In the barrel, the neutron induced background is sufficiently low to use Drift Tube (DT) chambers up-to $\eta = 1.2$ working in streamer mode. The basic element of a DT [25] is a drift cell with a cross section of 42×13 mm^2 and 2 to 3 meters long. The cell is filled with Ar/CO_2. The avalanche signal is captured by an anode wire in the centre of the cell. The DT chamber is made of 3 sets of drift layers. Each layers is made of several cells. The two outer layers measure the track hits in the bending plane. The middle layer measures the track position along the beam line. The arrival time of the track hits defines the distance between the track and the wire. A resolution of 100 μm is achieved in the r, ϕ plane.

In the end-cap, the neutron background is too high to use drift tubes. Therefore Cathode Strip Chambers (CSC) [26] are used to cover the detection of muons up to $|\eta| = 2.1$. A CSC layer is made of radially arranged cathode strips, perpendicular wires act as anode. The azimuthal track position is evaluated by measuring the charge sharing on the radial strips. The obtained azimuthal position resolution is lower than for the DT; $\sigma_{r\phi}$ ranges between 100 and 240 μm.

2.2 Trigger system

2.2.1 Overview

The CMS experiment must be capable of finding very rare events which may have cross sections smaller than the nb while the total inelastic cross section of the order of 50 mb. The wide range of physics processes together with their production cross sections is shown on Fig 2.12.

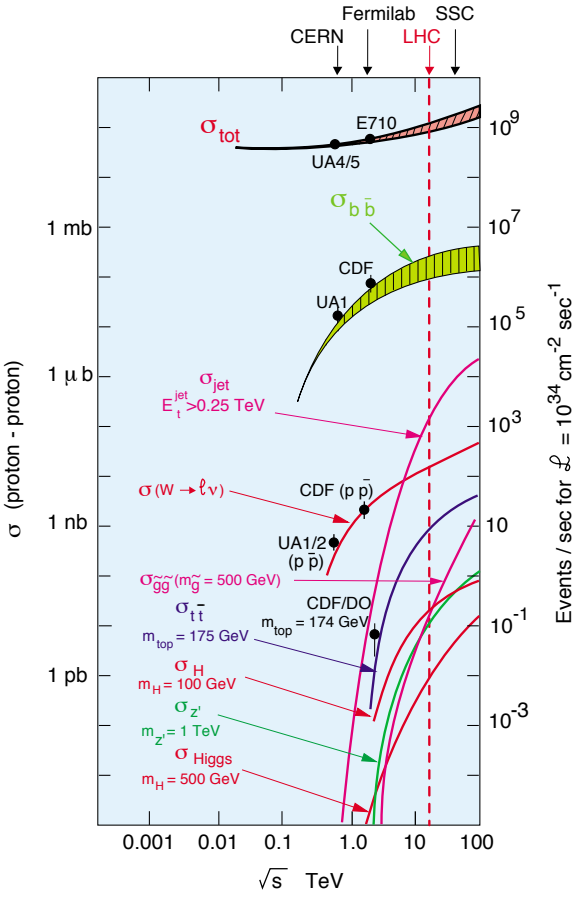


Figure 2.12: Inclusive proton-proton cross sections for basic physics processes. Interaction rates for a luminosity of 10^{34} cm^{-2}s^{-1} are shown on the right-hand scale.

The resulting total interaction rate reaches a giga Hertz. This rate accounts for the 20 collisions occurring at each bunch crossing in the high luminosity regime[1] with a bunch crossing frequency of 40 MHz. An online selection strategy must be put in

[1] $\mathcal{L} = 10^{34} cm^{-2} s^{-1}$

place to extract the rare signal events from this huge interaction rate. This selection must occur online since the storage of such an event rate is impossible with current technologies. CMS has chosen to reduce the event rate in two steps. In the first step, called the Level-1 trigger, the data rate is reduced by a factor 400 with a purpose built trigger processor. At this stage, not all detector information is available, and only rough transverse energy, jet location, electron identification and muon identification are used to select or reject the event. In the second stage, a data reduction of 1000 is accomplished by a computing cluster. The software fulfilling the data reduction is called the High-Level Trigger (HLT).

The following sections will detail the Level-1 trigger system and give an introduction to the High Level Trigger which will be discussed in more detail in the next chapter, where we shall discuss the data acquisition rate.

2.2.2 The Level-1 trigger

The requirement for the Level-1 trigger is to be able to operate at 40 MHz synchronously with the LHC clock. This is achieved using FPGA[2] technology, that enables using a purpose build logic while keeping the system flexible for reprogrammation.

The total latency of the system will be 3.2 μs of which 2 μs is due to signal propagation time. Each detector channel is pipelined in order to operate without dead time.

The Level-1 trigger has been designed to identify four of each of the following objects with the highest transverse energy: muons, isolated electrons or photons, non-isolated electrons or photons, central jets, forward jets and isolated hadrons or τ-jets. A logic called condition logic applies conditions on those particles candidates (objects). A particle condition is either an E_t or P_t threshold and the selection of η, ϕ region. In addition, the condition logic performs the computation of $\Delta\phi, \Delta\eta$ between particles. Each bit generated by particle conditions is combined with an AND-OR logic where combination (algorithms) are possible. Inside the Global Trigger, 128 different algorithms can be programmed to form a selection strategy. It can also take into account the total missing transverse energy as well as the jet counters, that count the number of jets above specified thresholds.

The output of the L1 trigger is a bit mask telling which of the algorithms have a positive outcome. In the case of a positive result at L1, the bit mask is used by the High Level Trigger to refine the selection for the particle candidates and take the final decision to keep or reject the event.

Electron identification

The Level-1 Electron/Photon Trigger algorithm [27] is schematically described in Fig. 2.13. The algorithm is based on towers made of 5×5 crystals in the barrel.

[2]Field Programmable Gate Array.

Defined as such, an ECAL tower matches an HCAL tower. A sliding window contains 3×3 towers spanning the entire (η, ϕ) plane with the E_t defined as the sum of the central tower and the most energetic of its four neighbours. An non-isolated electron

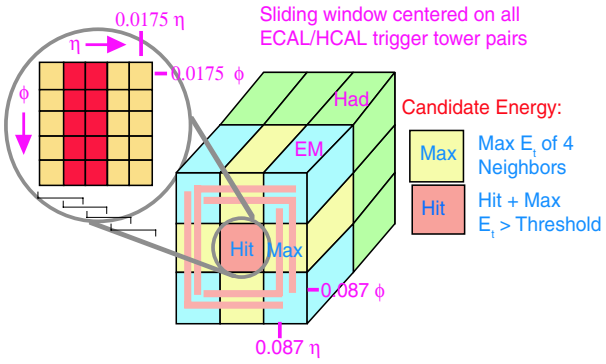


Figure 2.13: Schematic view of the electron/photon identification algorithm at L1.

candidate requires that the ratio of the HCAL tower to the ECAL central tower is lower than 0.05 in order to reject pions. In addition, the HCAL tower E_T summing the central tower and its most energetic neighbour has to be below a programmable threshold. Two additional criteria have to be met to define an isolated electron candidate. Firstly at least one of the L-shaped set of towers should be below a programmable threshold (~ 1.5 GeV). Secondly one requires that 90% of the energy deposited in the central tower localised in a row of 2×5 crystals. The first requirement acts as an isolation criteria avoiding the vetoing of the candidate if there is energy leakage in the adjacent tower. The second allows to lower the E_t threshold on the electron candidate while keeping the rate constant.

The four most energetic candidates of each electron type is send to the global trigger to apply the energy thresholds.

Jet and tau jet identification

The jet trigger uses a sliding window made of 12×12 trigger towers sliding by increments of four towers. The trigger towers are shown in Fig. 2.14. The Jet E_t is computed by summing the transverse energy of the 12×12 calorimetric towers while for the $\tau -$ jet it is only the central tower.

The identification of tau-jets, or pencil like jets requires the towers to pass an additional requirement. The nine regions should have a energy deposit that follows the pattern shown on the right of Fig 2.14. this identification is only computed in the region with $\eta < 1.95$

Finally eight jets (four central, four forward) and four tau-jets candidates are transmitted to the global trigger. The number of jets counted above a settable

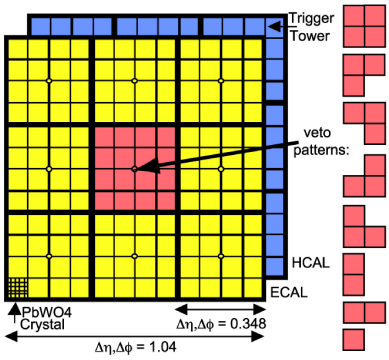


Figure 2.14: Schematic view of the sliding window used for jet/tau-jet identification.

threshold is also transmitted to the global trigger and can be used as a discriminating variable.

Muon identification

All three muon systems described in Sec. 2.1.4 participate in the decision taking. In the barrel, the electronics of each drift tube layer computes a track vector. A correlator combines the track vectors from the three layers of one chamber into one track vector. In addition the quality and crossing time of the chamber's track vector is provided to the global muon trigger. The CSC trigger primitives are formed the strips hits. They are combined to give 1/2 strip resolution and a track vector is computed from six layers of chambers

The regional triggers of the DT, CSC, and RPC deliver a list of muon candidates found in their respective parts of the detector. The Global Muon Trigger (GMT) combines the muon candidates and finds the four highest-rank muon candidates of the whole detector in every bunch crossing. The rank increase with P_t and quality of the muon candidate[3]. The typical reconstruction resolution at Level-1 is about 20% for the barrel and endcap region.

The Global Trigger

As said in the overview, the L1 Global Trigger (GT) receives (E_T, η, ϕ) of the four ranked object candidates of each type. Isolated and non Isolated electron/photons, central jets, forward jets, taus, muons as well as the sum E_t, missing E_T and jet count. The global trigger implements the selection on the basis of these conditions and their combination. The combination of object conditions form an algorithm and

[3]The quality is calculated based on the proximity of the CSC and RPC muon candidates in the barrel. In the endcap the comparison is done between DT and RPC muon candidates

128 such algorithms can be computed in parallel in the GT. A logical OR of those 128 algorithm defines the Level-1 response.

The design of the GT allows topological selection. As an example the search for the $H \rightarrow ZZ \rightarrow \mu\mu ee$ could, in principle, be implemented in two stages. In the condition logic, thresholds are defined for the two most energetic muons and electrons. Two other conditions require the electrons and muons to be back to back. The two bit that form the electron and muon conditions are combined with an "AND" logic to form one algorithm.

2.2.3 High Level Trigger

The High Level Trigger (HLT) processing starts as soon as a Level-1 trigger accept is issued by the Level-1 Global Trigger Processor every 10 μs. The task of the HLT is to further reduce the rate of data by a factor $\sim$ 1000 to fulfil the requirement of an output rate not exceeding $\sim$ 100Hz. This means that the HLT system has to make its decision each 10 ms while keeping its efficiency as high as possible for the interesting physics processes. In order to achieve the data reduction by a factor $\sim$ 1000 requires a massive computing power connected to the CMS data acquisition system (DAQ). The best estimates for the mean computing time of each HLT node is $O(10^{-2})$ s with a data input rate of $O(10^5)$ Hz. This means that the computing farm that will host the HLT system will be constituted of about 1000 CPU's. The role of the CMS DAQ is to provide the events to the computing elements of the HLT system. With an estimated event size of 1MB per event at the Level-1 output rate the DAQ system will require a total bandwidth of 100GB/s. This will be provided by a builder network described in the next chapter.

The CMS DAQ design is particular in the sense that it was chosen to minimise the use of custom technologies to benefit from the rapid growth of standard network and computing technologies. This choice implies that there are no special resources allocated to the design of a so called Level-2 trigger. The by-product of this decision is that the selection strategies will only be defined in the HLT software. It will certainly give more flexibility for the evolution of the selection strategies that physicists will require when developing new analyses.

Another specificity of the CMS DAQ is it's modularity. It is expected that the LHC instantaneous luminosity will build up with time. Therefore it is not necessary to have the full bandwidth at the first day of beam crossing. The DAQ design allows the adding of bandwidth by chunks of 12.5kHz as needed as the luminosity buildup.

CHAPTER

THREE

The High Level Trigger

3.1 Reconstruction Algorithms

The preparation for the analysis of events being stored during the data taking phase of CMS requires the development of a complete simulation and reconstruction chain. The simulation chain is summarised in Appendix B, it starts by simulating the physics occurring during parton collisions and ends with the simulation of the electronic response of the CMS detector. The output of the electronic response simulation is called the digis. After the simulation comes the reconstruction which aims at identifying and measuring electrons, muons, taus, jets and missing Et in the events. The reconstruction is decoupled from the simulation, since the same software should be used for the reconstruction of real events and of simulated events. In this section, we will describe the electron and tau reconstruction and selection algorithms that are used in the analysis presented in Chapter 4. We briefly mention b-jet reconstruction at the end of Chapter 4. The identification of other particles is described in [28],[29] and [1].

The High-Level Trigger needs reconstructed particles candidates to take its decision. Hence, its algorithms are called after reconstruction. The decision taking is based on logic rules that are defined by the physics program of the experiment. In Sec 3.4 we present the High-Level Trigger steering software that we have developed. It provides a simple way of combining selection algorithms to implement the trigger table and modify it as required.

3.1.1 Electron reconstruction

The electron reconstruction [30], in CMS, is complicated by the presence of the tracker material as presented in Sec 2.1.2. As an example, for electrons with $P_T = 35$ GeV in the barrel, the mean energy loss is 43.6% before exiting the tracker volume. The

energy loss carried by photons is mainly distributed in the ϕ direction such that clustering algorithms will combine sub-clusters to collect a maximum of the energy radiated by the electron.

The electron identification proceeds in three steps. First a clustering algorithm collects the energy deposit from the ECAL crystals. These clusters are then combined to form a super-cluster belonging to one electron to improve the energy and position resolution. In a second step, the the hit position in the first layer is estimated from the cluster position assuming the vertex to be at the nominal interaction point. If at least two hits are found, then the electron super-cluster is classified as an electron. In the third step, the pixel hits are used as seeds for the tracking algorithm to reconstruct the electron track.

L2 Clustering algorithms

Two different algorithms are used to construct the clusters and super-cluster forming an electron candidate.

In the barrel the Hybrid algorithm proceeds by finding the most energetic crystal above 1 GeV (seed), then it collects crystals around that seed to form a domino of 1×5 crystals as shown on Fig. 3.1. The algorithm then built dominoes of 1×3 or 1×5 crystals around the initial domino, depending on the energy of the central crystal. The collection of dominoes stops as soon as the more than 10 dominoes are built on the right and on the left of the initial domino. Dominoes with energy less than 0.1 GeV are discarded. Then a super-clustering algorithm collects sub-clusters with a seed crystal of more than 0.35 GeV in the ϕ direction to form the electron super-cluster.

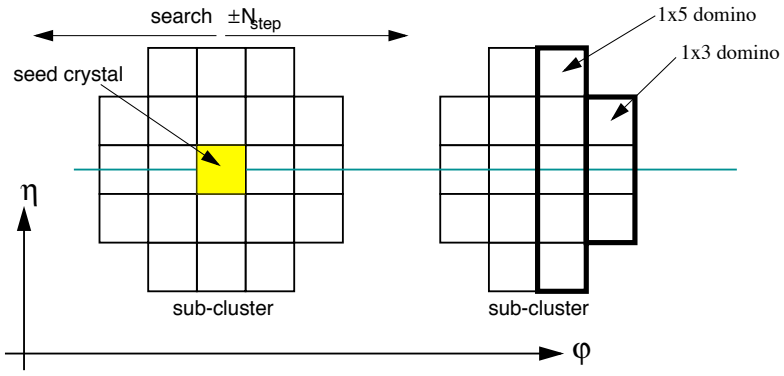


Figure 3.1: Schematic view of the Hybrid (super)clustering algorithm taken from [30].

In the end-caps it is the island algorithm is used. A list of seeds, i.e. crystals with transverse energy above threshold, is constructed. The algorithm starts with the most energetic seed and collects crystals in the sequence sketched in Fig. 3.2. The

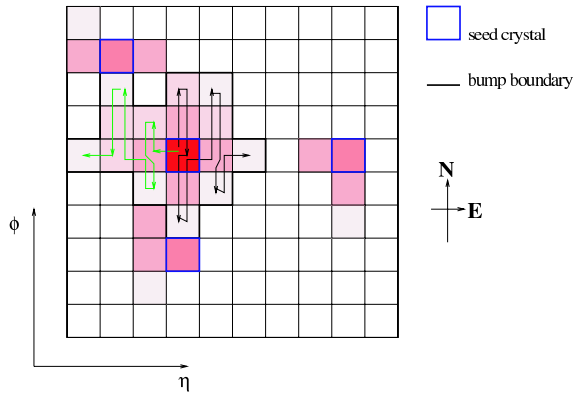


Figure 3.2: Schematic view of the Island clustering algorithm taken from [30].

seed collection of crystals is started by navigating around the most energetic seed in the ϕ direction until an energy rise or a hole is found. A hole is a crystal that has no energy because zero suppression has been applied. It then goes in one of the adjacent set of seeds in η and continues the ϕ collection until a energy rise or hole is found in the η direction. Both right and left η directions are harvested for the seed collection. Each aggregated seed is removed from the original seed list to avoid double counting of energy in different clusters. The Island algorithm design allows resolutions of the disintegration of a pion in two photons, and was used in a neural network for π_0/e discrimination (Sec 2.1.3). At this stage the clusters have to be collected to construct a super-cluster containing the photons emitted by the electron. The super-clustering builds a narrow η window and collects the clusters identified by the island algorithm to form the super-cluster of the electron candidate. The effect of the super-clustering on the energy resolution is shown on Fig. 3.3 for 30 GeV electrons. The reconstruction performance in terms of resolution in the barrel and in the end-cap is shown on Fig. 3.4

L25 Pixel matching

At this point we have no clue whether the super-cluster is the product of an electron or a photon that has converted in the tracker volume. Since the super-cluster position is calculated to match the real electron/photon position, the hit position of the non radiated electron can be estimated in the pixel detector. The presence of hits in the pixel detector can be used to qualify the super-cluster as being an electron or a photon since most of the tracker material lies beyond the pixel detector [31].

The expected hit position on the pixel layers is estimated by propagating the electron inward assuming a vertex at $(0, 0, 0)$. A window in the ϕ, z plane is opened and hits are sought for in the first pixel layer ($\Delta\phi \sim 40$ mrad, $\Delta z = \infty$) . If a hit is found, the vertex is recomputed using the super-cluster position and the hit position, if no hit is found, the search is repeated in the next layer. Using the newly estimated

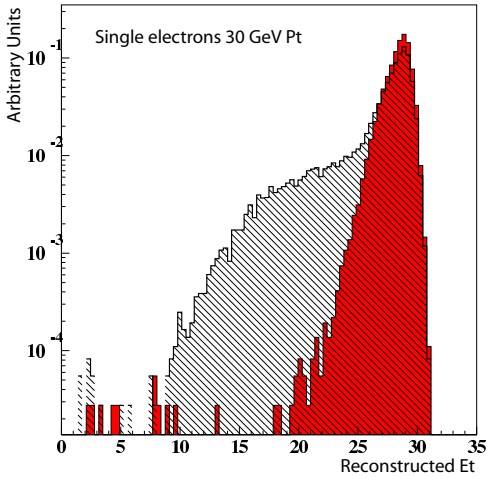


Figure 3.3: Reconstructed E_t of electrons with $P_t = 30$ GeV using the island cluster algorithm. The dark shaded histogram is computed with super-clustering while the light histogram is without super-clustering [30].

vertex position, compatible hits are searched in the second pixel layer. If no hit is found, then the search continues in the third pixel layer. A this stage the super-cluster will be marked as an electron if one of the two layers has a compatible hit.

The performance of the pixel matching algorithm to reject QCD jets is given as the efficiency to accept electrons for a given QCD jet rejection. Assuming three pixel layers, for a QCD rejection of 11, the electron selection efficiency with respect to the Level-2 is 97 %.

L3 electron track reconstruction

In order to further reduce the event rate, the electron track is reconstructed inside-out if compatible hits with the super-cluster were found in the pixel detector. With the evaluated momentum, the E/P is computed and the candidate is kept if the $E/P < 1.5$. The main background consists of charged pions that have passed the pixel/super-cluster matching requirement. Note that the E/P cut is applied for the HLT selection but it is not applied for the electrons candidates used in the analysis detailed in Chapter 4

In order to select electrons and positrons, an additional constraint is applied on the geometrical separation between the track and the supercluster position. The distance between the track direction extrapolated near the ECAL and the super-cluster ($\delta\eta = \eta_{L2} - \eta_{track}$) is required to be less than 0.0025.

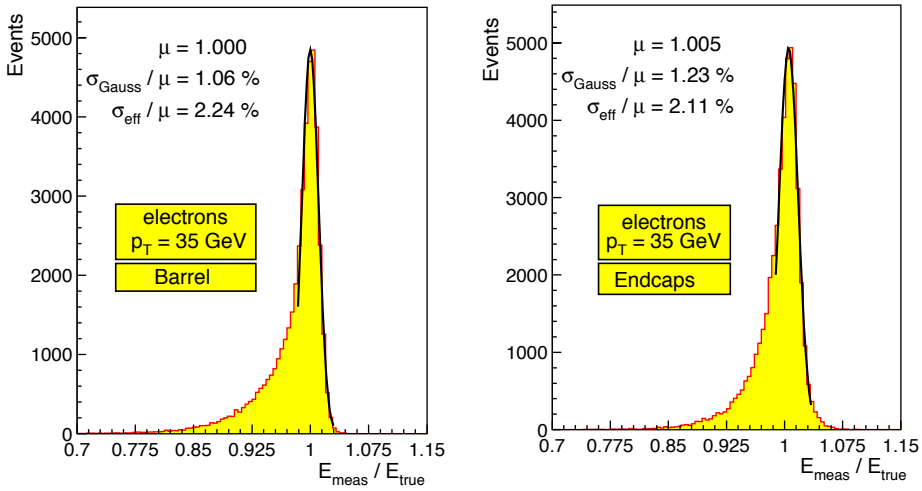


Figure 3.4: Distribution of the ratio of the measured and reconstructed energy for 35 GeV electron without pileup [1]. On the right for electrons in the end-cap and for electron in the barrel on the right.

3.1.2 Hadronic tau reconstruction

In this selection we describe the algorithms developed for the selection of hadronic tau decays called tau-jets. As described in Sec 1.2 MSSM Higgs couplings imply that a good tau identification will improve the discovery potential of MSSM Higgses in the high $\tan\beta$ regime. Tau jets have the property of being pencil like or narrow jets.

L2 Jet reconstruction

The tau jet reconstruction starts by applying an iterative jet cone algorithm [32] in the region of the four tau jets candidates identified by the L1 trigger. The iterative cone algorithm searches the maximum transverse energy object and defines a cone around its direction. The object is made of the ECAL and HCAL towers summed in E_T. Any object within that cone will be merged to form a proto-jet. The proto-jet direction is calculated from the energy weighted directions of the constituents, and a cone is defined around the new direction to form a new proto-jet. The procedure is repeated until the proto-jet does not change significantly. The constituents are removed from the list of objects, and the procedure is repeated until no objects are left in the list.

Once the most energetic jet is identified, the transverse energy is summed for all crystals that lie in a ring around the jet position. The ring has internal and external radii of 0.13 and 0.4 respectively The total transverse energy collected in the ring must be below an isolation threshold (5 GeV). This reduces the number of QCD dijet events by a factor 10, with an efficiency of 60% for the signal. Fig. 3.5 shows the signal efficiency for different values of the isolation threshold.

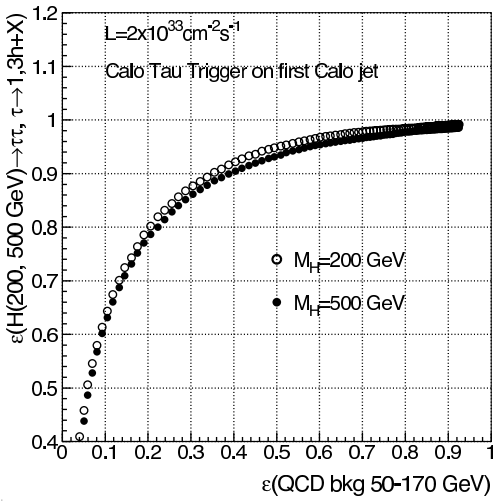


Figure 3.5: Selection efficiency of $H \rightarrow \tau\tau \rightarrow \tau_{jet}\tau_{jet}$ by the L2 tau calorimeter algorithm as a function of the efficiency to select qcd dijet events [1]. The assumed luminosity is $2 \times 10^{33}\text{cm}^{-2}\text{s}^{-1}$.

L25 tau-jet isolation with the tracker

In order to identify the tau hadronic final states, a jet rejection algorithm as been designed [33]. The algorithm proceeds by searching for tracks that match the taujet candidate as shown in Fig. 3.6. When a tau jet candidate is found in the calorimeter, a matching cone is opened in the tau direction. Tracks with a $P_t > 3$ GeV are sought in the matching cone and the most energetic one is identified. A new cone, the signal cone, is opened around the leading track, any track found inside the isolation cone that is not included in the signal cone will lead to the rejection of the tau-jet candidate.

The jet rejection achieved with this algorithm is shown on Fig. 3.7 for the QCD bin with $P_T = [50, 170]$ as a function of the selection efficiency for a 200 GeV Higgs.

3.2 Electron tau trigger settings

In this section, we present how we could enhance the trigger efficiency for the selection of electron tau-jet final states in the decay of MSSM Higgses to a pair of taus. The definition of a new trigger is guided by the event rate constraint. In CMS, the total L1 bandwidth is fixed to 1/3 of the maximum L1 rate of 50 kHz at low luminosity. Then those 16 kHz are equally divided into four class of triggers: electron/photon, muon, tau-jet, jets and combined channels. Our goal is therefore to add a combined trigger to increase the single electron efficiency on the Higgs signal with a rate increase smaller than 2 kHz at L1.

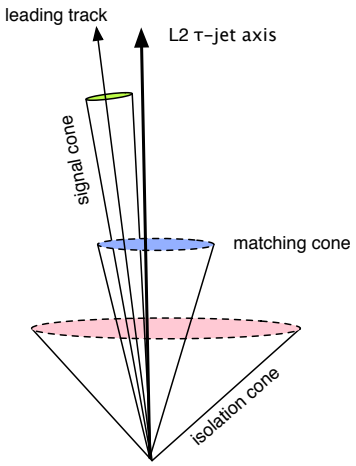


Figure 3.6: Schematic view of the cones used in the isolation of a tau-jet candidate. The cone sizes are: $R_{match} = 0.1$, $R_{signal} = 0.064$, $R_{isol} = 0.45$.

At low luminosity (2×10^{33} cm^{-2}s^{-1}) the study evaluates the efficiency of the following trigger equation:

$$E_T^e > c_1 \vee (E_T^e > c_2 \wedge E_T^{tau-jet} > c_3) = \mathsf{true}. \tag{3.1}$$

The single electron trigger threshold has been fixed by the maximum event rate allowed, i.e. 3.3 kHz at High Luminosity after the Level-1 Trigger. The corresponding E_t cut value for the single electron (c_1) is 23 GeV.

Remaining free parameters are the two thresholds c_2 and c_3 associated with the $e\tau$ trigger. The procedure is to fix the maximum additional rate due to the $e\tau$ trigger and find which values of c_2 and c_3 maximise the selection efficiency. The event rate has been evaluated with a QCD sample enriched with electrons [34]. Fig. 3.8 shows curves of iso-rate with the selection efficiency as a function of c_2 and c_3.

If one assumes an additional L1 rate of 823 Hz, the efficiency increase due to the $e\tau$ trigger is 4% at L25 with the following cuts: $c_2 = 16$ and $c_3 = 53$. The gain is small at low luminosity, but grows to 10% at high luminosity as shown in [1]. We observe on Fig. 3.8 that for a given trigger rate, the highest efficiency is achieved by lowering the electron threshold and rising the tau-jet one. This is due to the P_T spectrum differences in the hadronic and leptonic decays. As shown on Fig. 3.9 the presence of two neutrinos in the leptonic decays implies that the electron spectrum as a higher fraction of events at low P_T, while it is not the case for the tau-jet. It as to be noted that the lower electron E_t bound is fixed by the data sample that prevented to further lower the electron E_t cut without biasing the event rate estimation.

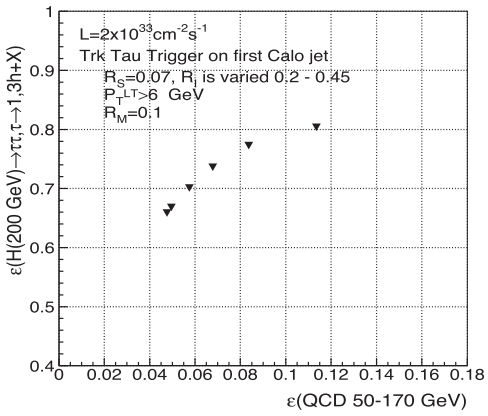


Figure 3.7: Selection efficiency of the tracker isolation algorithm for a 200 GeV Higgs as a function of the QCD jet efficiency. Each point corresponds to a different size of the isolation cone (R_i).

3.3 Event Building

In order for an event to be analysed, it has to be provided to the filter unit (FU) that will process the event. After a Level-1 accept, the event data are scattered in multiple memories. The event builder will collect all these event fragments in one place, the builder unit (BU). Finally the FU obtains its events from the BU, to be further analysed.

Fig. 3.10 shows the different elements involved in DAQ system. The lower part of the figure shows the filter sub-farms that analyses the events stored in the Builder Units. The filter task is included in one of the processors connected to the filter farm network. It can request the event manager to provide an event that has been stored in the Builder Unit (BU) after a Level-1 accept. In the upper part of the Fig. 3.10 one sees the builder network, the readout units (RU) and the detector frontends. Each RU collects the information of several events from a collection of subdetectors frontends. This means that at the RU level, one event data is split in about ≈ 500 readout units. The data frames of each event have to be concatenated into the BU to be read by the filter farm task. In order to achieve this, a readout builder network acts as a switch fabric to collect each event fragment into one data structure that contains the whole event data. Assuming an average event size of 1 MB the switch fabric will have to sustain data throughputs of 800 Gb/s.

The event building proceeds in two phases. In a first step, the data from groups of eight front-end drivers (FED) is concatenated via a FED builder into the readout unit (RU) to form what is called a super-fragment. At this point the data from one event is stored in 64 RU. In a second concatenation step, the super-fragments that constitute one event are sent to a Builder Unit (BU) through the Readout Builder network (central part in Fig. 3.10).

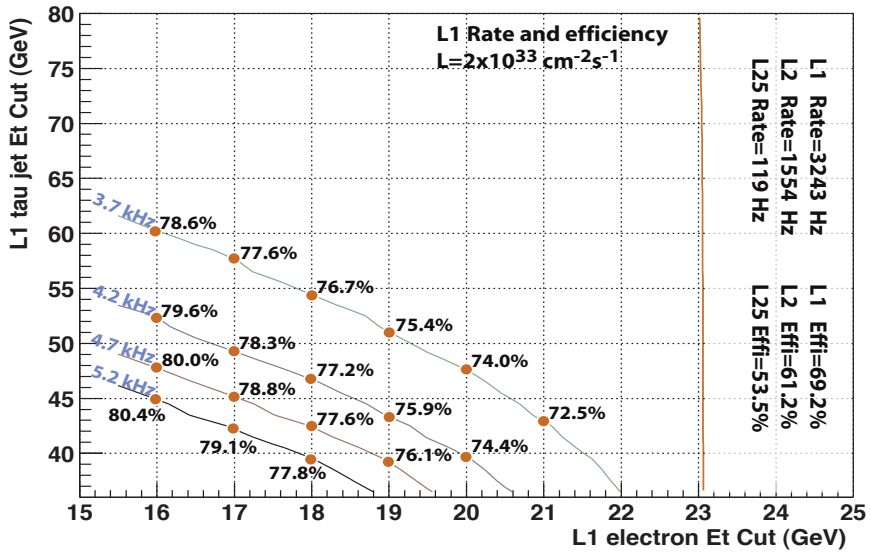


Figure 3.8: Event rate and efficiency for the selection of electron tau-jet final states in the Higgs decay (200 GeV).

The bandwidth of the builder network links is specified by the RU throughput. The DAQ system is organised in eight slices that sustain a total throughput rate of 100 kHz. Each DAQ slice has to cope with 12.5 kHz. The super-fragments are constituted of eight fragments coming from the FED, each 2 kB long. This means that each RU has to sustain a throughput of 200MB/s.

Each time the Level-1 trigger issues a Level-1 accept, the event fragments are pushed to the Readout Units. If the RU starts to be overloaded by incoming data, a throttling mechanism allows to reduce the Level-1 trigger rate. The final concatenation is an on demand mechanism. If one BU has freed an event from memory, it can request the event manager to allocate a new event that will be pulled from the corresponding RU through the RU Builder network. Once the BU is populated with the content of one event, the Filter Task can request it for analysis.

3.4 High Level Trigger software prototype

The filter task that we introduced in the previous section has to implement the selection required by the physics programme of the CMS experiment. It is clear that the selection done *in situ* should be carefully carried out, since it will have a direct impact on data quality and the possibility of observing new signals. One has to remember that an event lost at the filtering stage is lost forever.

One of the objectives of the ORCA reconstruction software is to provide tools to identify and measure leptons, jets and missing momentum. The decision to select an

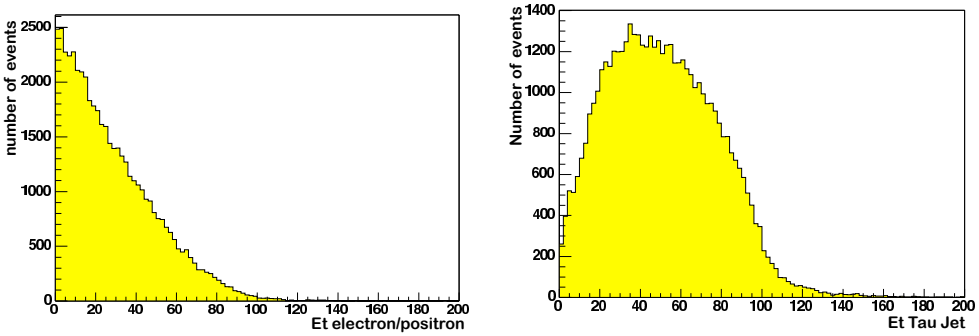


Figure 3.9: P_T spectrum of the tau-jet on the right and for the electron on the left at generator level for a 200 GeV MSSM Higgs decaying into a pair of taus.

event can only be done on the output of these reconstruction algorithms by combining their output in a logic that should match the physics objectives of the experiment.

The argument above was the reason for the development of a software package that allows to specify easily the selection logic of the High-Level Trigger (HLT). This software provides a set of elements that define trigger-primitives (e.g. selection of a tau-jet, calorimetric cluster, muon track, etc). It also provides a way to combine these trigger-primitives to form a logical equation that is evaluated to verify whether the analysed event matches the selection criteria. In the next sections, both the trigger-primitives and the logic that combine them will be referred to as elements.

As an example, the search for an isolated electron with a P_t higher than a chosen value is performed in an element. Several elements can then be combined to form the trigger equation logic, for example, searching for a muon with a P_t higher than 14 GeV and an isolated electron. The HLT software also provides the option of selecting events in a mode that optimises the time used for the selection. This "optimisation" mode proceeds by evaluating the elements that have a quickest outcome. When two elements are combined with a logical "or", there is no need to evaluate the outcome of both elements if there is one that is true. It is preferable to first evaluate the element that gives the quickest response. This is implemented in the prototype and one can devise a trigger mode where the distribution of evaluation time per element is gathered. After this learning mode, the trigger would run in optimisation mode using the mean computing time to evaluate the trigger logic in the most clever path.

The following section will describe the main features of the HLT prototype and the requirements that have driven its development.

3.4.1 Requirements

As far as the HLT is concerned, requirement from several sources have to be taken into account. The most important one comes from the Level-1 interface and the

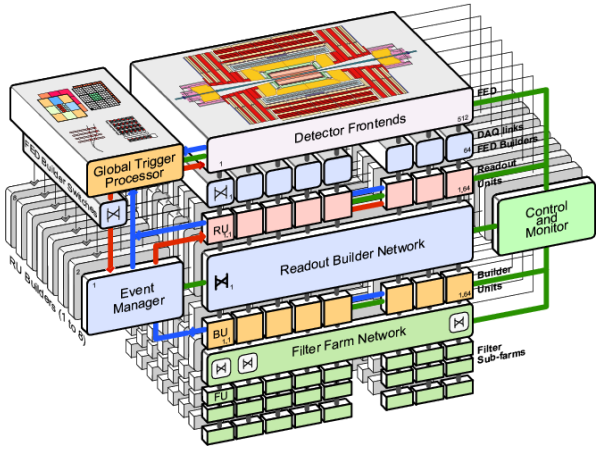


Figure 3.10: 3D view of the data acquisition system. The lower part of the sketch contains the filter farm where the HLT software is executed. The top part show the frontend drivers and the middle part represent the event builder network.

way selection parameters are specified. The last requirement is addressed by the core frameworks [35] [36] used by the HLT. The next two sections detail these requirements.

The L1 interface

The HLT has to be seen as a consumer of both the Level-1 decision and of the reconstruction algorithms that are provided by the ORCA framework [37]. The HLT has to access the Level-1 decision word and the L1 candidates. The decision word can be used as a veto for HLT elements in order to evaluates the elements that have a corresponding true decision at Level-1. For example, the HLT could work in a mode where the muon selection is not performed if there is no muon candidate at Level-1. It has not yet been determined if the veto mode will be the standard working mode, but a least it should be possible to switch it on or off. The muon and electron reconstruction starts from the Level-1 candidates, but it is not necessarily the case for the jet selection at HLT.

Selection parameters setting

The specification of a trigger table and the settings of the reconstruction parameters have to be handled coherently. This means that any changes of the parameters that could change the HLT selection behaviour must be linked to the stored events. The ORCA framework has been modified in this sense. It provides a mechanism to configure the event selection and reconstruction by specifying all the parameters needed in the computation of quantities used by the HLT. The configuration that was used to select an event is then attached to it.

It should be noted that the trigger configuration must be able to change on a run by run basis. The pre-scale settings as well as the selection settings could be modified on request from the monitoring system. For example, if a noisy channel starts to increase the trigger rate, one could apply a pre-scale. Another example is the drop in luminosity that would require lowering the trigger thresholds. All these requirements have been taken into account in the HLT software and its core framework.

The HLT steering code

The HLT steering code should provide:

- a single place for HLT elements (single electron, single muon, ...) implementation
- a dynamical HLT definition, that can be redefined without the need to even restart the program,
- a possible optimisation of the HLT evaluation sequence,
- a Boolean HLT decision,
- a full HLT response as a bit pattern (as is the case for the L1),
- off-line checking facilities to monitor the behaviour of the elements. For example, one could request the trigger rate of a chosen element.

The steering code has to be based on the L1 output and also on the existing reconstruction code. In addition, the induced overhead has to be minimised (which implies a code that is as simple as possible). Another design constraint of a more organisational nature is the dependency of the HLT on other ORCA code. In this context, any HLT code should be designed to depend only on reconstruction packages, to still allow the sequential release/build procedure used in ORCA. If this is achieved, analysis sub-packages should be able to use HLT in the early steps of their tasks.

3.4.2 Core HLT steering code

The core of the HLT is the steering code, which implements the required functionalities as described in section (3.4.1).

The trigger system can be seen as a logical equation. Its value should be true for an event to be selected. Usually, the trigger logic is a "OR" function of several trigger elements that computes reconstructed quantities to which a selection criteria is applied. The trigger logical equation can therefore be represented as a binary tree. Each node or element of the tree is either an operand or a logical operator. An operand is an trigger element that applies a selection algorithm on reconstructed quantities, while a logical operator only combines the outcome of two operands.

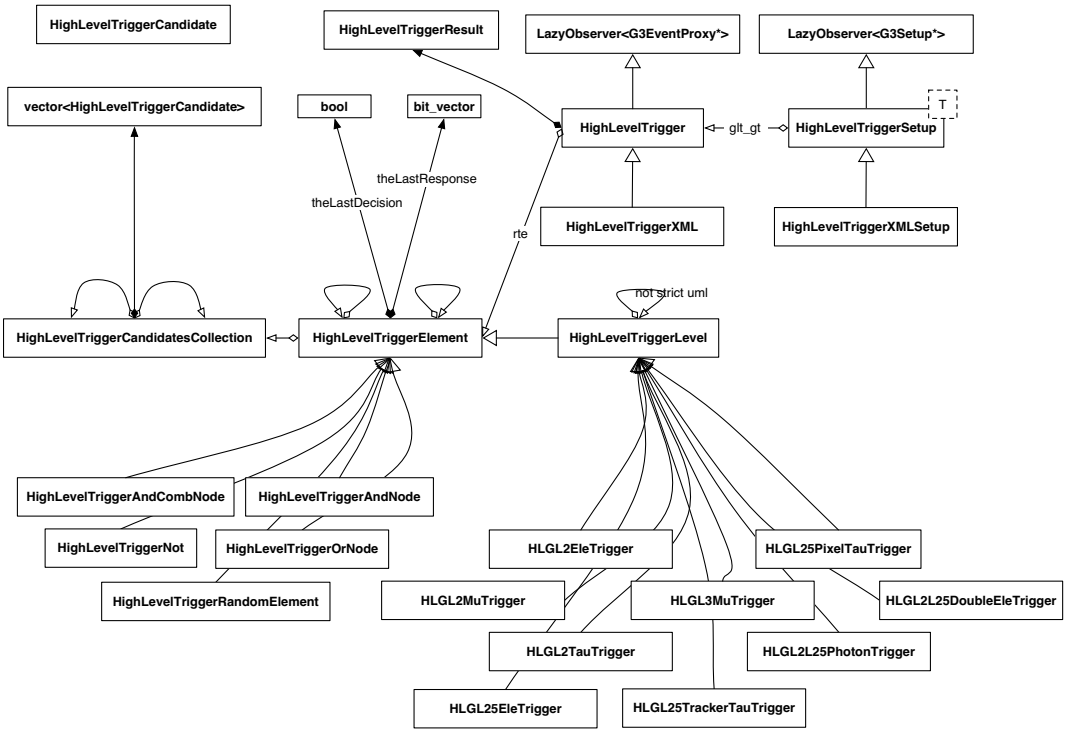


Figure 3.11: UML diagram of the High Level Trigger software prototype

Such a tree design is valuable in order to simplify the evaluation of the trigger outcome by using recursive techniques. Navigation in the tree is natural and eases the configuration of each node.

The core of the steering code consists of several classes as shown in Fig. 3.11. Five classes are central in the HLT steering code.

1. **HighLevelTriggerElement**: The base class of any operand or logical operator. Since it acts as the representation of the tree nodes, it has two references to other HighLevelTriggerElements. Its behaviour is to observe physics events and produce three quantities. A Boolean value (true/false) states the outcome of the specific implementation of this element. To detail the internal state of the decision, the HighLevelTriggerElement constructs an array of bits expressing the detail of the selection. Eventually, a list of HighLevelTriggerCandidates is produced and corresponds to the particles that have triggered this Element.

2. **HighLevelTriggerLevel**: This class is a specialisation of the HighLevelTriggerElement that has only one parent. This is useful in the situation where a trigger element must only deal with the validation of a particle candidate without combining several elements. For example, the search for muons is implemented by deriving from a HighLevelTriggerLevel to implement the differents

selection levels of the muons that will be described later. Logical operators usually derive from the HighLevelTriggerElement. Examples are given in the table 3.1

3. **HighLevelTriggerCandidate**: This is what elements produce when they have found particles that match chosen selection criteria. They are Lorentz Vectors with a vertex, charge and particle identification.

4. **HighLevelTrigger**: This class owns the root of the trigger tree. It is therefore responsible for initiating the construction of the trigger trees. This class can be derived to implement specific trees. For example the **HighLevelTriggerXML** is a HighLevelTrigger implementation that read an XML[1] file and builds the trigger tree. One can imagine that, in the future, the trigger tree will be built from the information stored in a condition database. The structure of the High Level Trigger involves several sub triggers for the selection of electrons, muons, taus, jets, photons, missing E_T and combinations of those. This implies that the HighLevelTrigger can hold several trigger trees that are evaluated in sequence until a positive outcome if found.

5. **HLGL1Veto**[2]: This class is a specialisation of the trigger level that accesses the Level-1 trigger decision bits. This is used in the case of a veto applied to the trigger. One can choose to veto the High Level Trigger with the output of the Level-1 Trigger. For example, if one assumes that the HLT selection of tau-jets should not be performed if there is no corresponding Level-1 candidate, then the HLGL1Veto can be attached on top of the corresponding element.

Most of the other classes found in the Fig. 3.11 are implementations of the HighLevelTriggerElement that does the concrete work of selecting events on the base of reconstructed quantities. These are detailed in the Table 3.1. The naming convention for the trigger elements is not strict. Level-1 is defined as all algorithms that are implemented in the FPGA architecture of the first level trigger. The selection algorithms that follow the Level-1 are named with a number that grows with the selection sequence. In general, Level-2 algorithms uses only the calorimeter and muon chamber data to compute a trigger decision. Level-2.5 involves the pixel detector, and Level-3 the tracker data.

3.4.3 Trigger evaluation sequence

As already mentioned, the HLT response is evaluated recursively, and the tree leaves correspond to the lowest HLT trigger level. More precisely, the evaluation request is propagated recursively from the root element up to all the leaves. The actual evaluation then starts from the leaves down to the root, going through all the HLT steps and stopping as soon as one element in the chain fails. This result in an evaluation sequence that can be optimised at each node, as we will discuss in the next section.

[1] eXtensible Markup Language, `http://www.w3.org/XML/`
[2] Not shown in Fig. 3.11

HLGL2TauTrigger	Level 2 single tau trigger
HLGL25PixelTauTrigger	Level 2.5 tau validation with pixel
HLGL25TrackerTauTrigger	Level 2.5 tau validation with tracker
HLGL2EleTrigger	Level 2 single electron trigger
HLGL25Eletrigger	Level 2.5 single electron trigger
HLGL2L25DoubleEleTrigger	Level 2 or Level 2.5 double electron trigger
HLGL2L25PhotonTrigger	Level 2 or Level 2.5 single photon trigger
HLGL2MuTrigger	Level 2 single muon trigger
HLGL3MuTrigger	Level 3 single muon trigger
HLGL2JetTrigger	1,2,3,4 Jet trigger (calorimetric)
HLGL2MetTrigger	Missing Et Trigger
HLGL3EleTrigger	Level 3 single electron trigger
HLGL3PhotonTrigger	Level 3 photon trigger
HLGL3BJetTrigger	Level 3 TrackCounting b tagger
HLGL2JPsiTrigger	Specific J/ψ tagger
ttHJetTagging	$t\bar{t}H$ trigger
HighLevelTriggerAndCombNode	Logical and between 2 triggers. Candidates have to be separated.
HighLevelTriggerAndNode	Logical and between 2 triggers
HighLevelTriggerNot	Logical not
HighLevelTriggerOrNode	Logical or between 2 triggers
HighLevelTriggerRandomElement	random trigger

Table 3.1: Elements implemented in the HLT prototype. "CombAnd" is a logical "AND" with the additional restriction that the candidates are not in the same regions. More generally, the "CombAnd" can be used to build combined triggers of non exclusive objects (like an electron and a tau). In the future, more combination requirements can be added on request.

In order to clarify how the HLT behaves, let's assume that one wants to select events with the trigger corresponding to the following logical equation:

$$E_t(e) > 23 \vee (E_t(e) > 14 \wedge E_t(\tau) > 48). \tag{3.2}$$

It will be translated in the tree shown in Fig. 3.12. The two nodes in grey[3] are the logical operators of the equation while each white columns are the operands.

The evaluation sequence of the previous trigger equation will be achieved as follows:

1. The request will go up to element 3 , via the "OR" node and element 2.

2. Element 3 will be evaluated and return false.

[3] green for the colour print

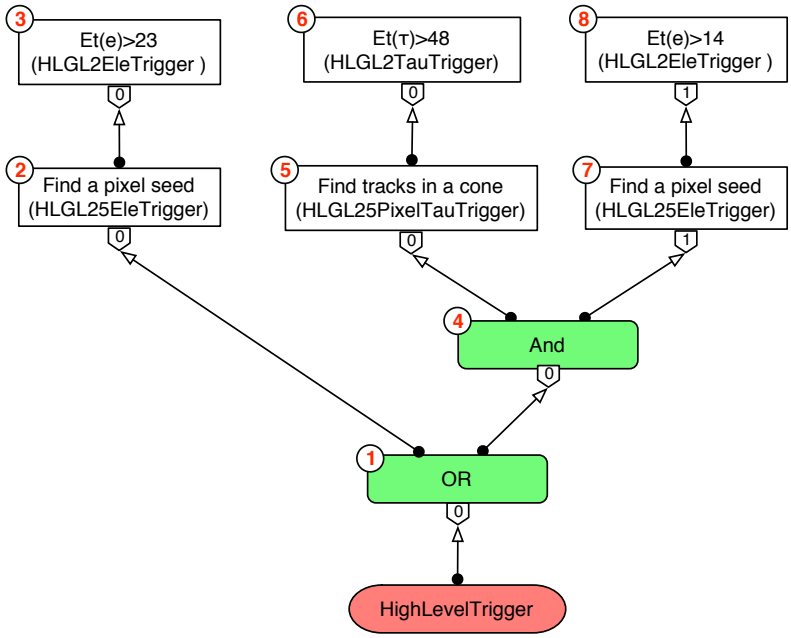


Figure 3.12: Tree view of the logical equation (3.2). Below each box the bit corresponding to the outcome of each element is shown. The bit a the bottom of the tree states the result of the whole trigger, in this case the event is rejected.

3. Since element 3 is false, the attached element 2 (HLGL25EleTrigger) will not be evaluated and will return false.

4. The request will then go up to element 6, via the "OR" node, the "AND" element and element 5.

5. Element 6 will be evaluated and return false.

6. Element 5 (HLGL25PixelTauTrigger) must now be evaluated, but since the previous level returned false it will be skipped and return false.

7. The request will then go up to element 8, via the "OR" node and the "AND" element and the attached element 7.

8. Element 8 will be evaluated and return true as well as element 7.

9. The "AND" element will be evaluated and return false.

10. The "OR" element (the root node) will be evaluated and return false.

If the High-Level Trigger has several trees (sub-triggers), it will request the response of each sub-trigger in sequence until a positive response is found, in which

case the event is accepted. If all sub-triggers give a negative output, then the event is rejected.

In the next section, we will describe how the evaluation sequence can be optimised to reduce computing time.

3.4.4 Optimisation of the evaluation sequence

As described in Sec 3.4.3 , the evaluation sequence starts at the leaves of the tree to end at the root element of all sub-trigger that constitute the whole trigger. In order to optimise the evaluation time, each "OR" element can be configured to optimise the computing time needed to get the quickest output.

As an example, in the case of an "OR" element there are two daughters to evaluate. If the first daughter gives a positive response, then the second daughter does not need to be evaluated. If the response is negative, then the second daughter should be evaluated. Intuitively, the daughter that gives the quickest output should be evaluated first. This is true only if the daughter trigger probabilities are the same. If they are not, one has to compute what arrangement will give the smallest mean computation time.

Let $p_{(1,2)}$ be the probability for the first and second daughter to give a positive response , $ta_{(1,2)}$, $tr_{(1,2)}$ the mean time for a daughter to accept or reject an event. The mean time to accept an event if one evaluates daughter 1 first is given by:

$$\langle Ta\rangle_{12} = p_1 ta_1 + (1-p_1)p_2(tr_1 + ta_2) \tag{3.3}$$

Permuting 1 and 2 gives the mean time,$\langle Ta\rangle_{21}$, for the 21 sequence. One has then to choose the order that minimises the mean accept time. In the case of the "OR" element the mean reject time is fixed since both daughter have to be evaluated.

In the case of an "AND" element, it is the mean reject time that can be optimised in a similar procedure, while the mean accept time is fixed. The mean reject time is given by:

$$\langle Tr\rangle_{12} = (1-p_1)tr_1 + p_1(1-p_2)(ta_1 + tr_2). \tag{3.4}$$

In both the "OR" and "AND" situations, one can find the order which will give the smallest mean computing time for the acceptance or rejection of an event. In the previous section, we have seen that each sub-trigger is evaluated in sequence until a positive response was found. The optimisation in the case of an "OR" with several sub-triggers is an extension of the "OR" with two elements. We have to find an order of the sub-triggers that minimises the mean computing time to accept an event. If the order of the sub-triggers is defined by the sequence $\{k_1 \cdots k_n\}$ then the mean accept time is given by:

$$\langle Ta\rangle_{\{k_1 \cdots k_n\}} = p_{k_1} ta_{k_1} + \sum_{i=2}^{n} \left[p_{k_i} \prod_{j=1}^{i-1} (1-p_{k_j}) \right] \left[ta_{k_i} + \sum_{l=1}^{i-1} tr_{k_l} \right], \tag{3.5}$$

sub-trigger	p_i	ta_i	tr_i
1	0.5	60	12
2	0.3	60	50
3	0.1	20	10
4	0.6	90	40

Table 3.2: Acceptation probabilities for each subtrigger with rejection and acceptation times for each subtrigger selection time as a function of the trigger bits permutations.

with $ta_i(tr_i)$ the mean accept (reject) time for the sub-trigger i and p_i being the probability of trigger i to give a positive response. The expression for the mean reject time is given as:

$$\langle Tr\rangle_{\{k_1\cdots k_n\}} = \prod_{i=1}^{n}(1-p_{k_i})\sum_{i=1}^{n} tr_{k_i} = \langle Tr\rangle, \tag{3.6}$$

the latter is clearly invariant under any permutation of the k_i; all sub-triggers have to be evaluated to state on the rejection of an event. The total mean computation time is given as:

$$\langle T\rangle = \langle Ta\rangle_{\{k_1\cdots k_n\}} + \langle Tr\rangle \tag{3.7}$$

The minimisation of expression 3.5 requires the evaluation of $n!$ arrangements, which is not realistic since the trigger table can be constituted of ~ 20 sub-triggers. The optimal arrangement is obtained by sorting the sub-trigger with the order defined as:

$$||kl|| < ||lk|| \Leftrightarrow p_k ta_k + (1-p_k)p_l(tr_k + ta_l) < p_l ta_l + (1-p_l)p_k(tr_l + ta_k). \tag{3.8}$$

Although not yet proved, we have empirically tested this assertion by comparing the mean accept time of the arrangement sorted as prescribed with all possibles permutations. The defined order was the one used in the "OR" with two elements, and by using the introsort algorithm [38] the computing needed to order the trigger bits is guaranteed to be less than $\mathcal{O}(N\log(N))$.

The effect on sorting the sub-trigger to reduce the mean time used to compute the trigger response can be dramatic for events that are selected with high efficiency. In this situation the mean computing time given in Eq. 3.7 is dominated by the mean accept time $\langle Ta\rangle$.

For instance, assume a simple trigger made of four sub-triggers with the probabilities, accept and reject mean time[4] given in Table 3.2. The mean accept time for each sub-trigger arrangement is shown in Fig. 3.13 compared to the optimal arrangement 1342. The best arrangement gives a mean time of 68.4, and the worse gives 100.4, about 46% difference !

In the case of the HLT running condition most events must be rejected. Only one event on $\mathcal{O}(1000)$ is kept for further analysis. This means that the $p_i \sim \mathcal{O}(10^{-3})$, and

[4] In arbitrary units

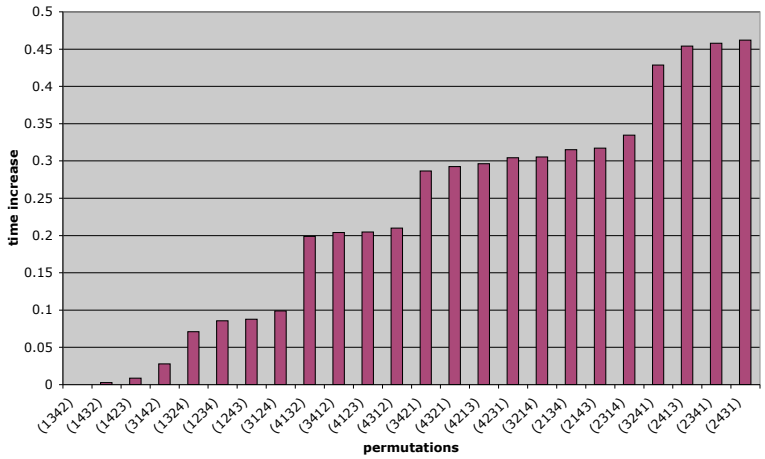


Figure 3.13: Time increase for the mean accept time $\langle Ta \rangle_{\{k_1 \cdots k_4\}}$ in each possible arrangement of the sub-trigger.

the optimisation does not change significantly the mean computing time. This can intuitively be understood, since in most cases all triggers bits have to be computed to be sure that no trigger has given a positive response. Although the mean accept time will not be significantly modified by the optimisation procedure, the tail of the time distribution will be reduced. As far as the DAQ is concerned it means that the risk of having buffer overflows that would increase the dead time, will be reduced.

The estimate of the time and probabilities that enter the optimisation of the evaluation sequence should be done online. The HLT includes a mechanism to gather histograms for the selection and rejection time of each HLT element. The trigger probabilities and timing can then be extracted from these algorithms. One could imagine that the trigger table would contain the trigger probabilities and timing gathered during a test run. The sub-trigger and elements order would then be statically defined at the start of the run when the trigger table is loaded to build the trigger tree. A more elaborated configuration would allow a dynamic definition of the evaluation sequence based on the real time learning of the time and probabilities of each element. It has to be discussed whether the optimisation mode does not add to much complexity in the trigger response monitoring. A change of the trigger settings would imply a change of the trigger evaluation sequence. In addition the partial evaluation of the trigger response imply further processing on the offline farm, and this additional CPU cost has to be evaluated.

3.4.5 HLT timing

The timing of the HLT was performed on Drell-Yan[5] (DY) events without pileup. The time computation is done while evaluating each sub-trigger. The computation

[5] $q\bar{q} \rightarrow Z, \gamma*$

is stopped when all sub-triggers are evaluated, so that the quoted timing should be taken as an upper bound on the computation time. The optimisation described in the previous section was not enabled. The trigger table used is given in appendix C.1.

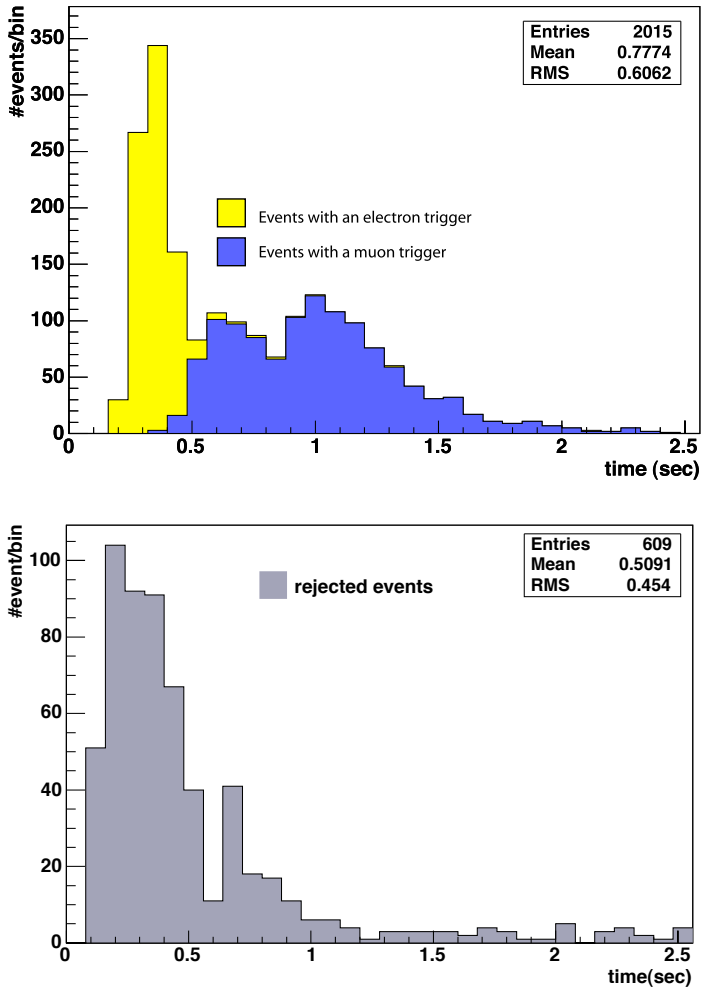


Figure 3.14: Time distribution for $q\bar{q} \to l^+l^-$ events without pileup. The upper plot shows the events accepted by the HLT, the lower one shows the rejected events.

The timing shown in Fig. 3.14 was computed on a 2.4 GHz Xeon processor with 512 kB cache and 2 GB RAM. The obtained mean time is 727 ms. The considered Drell-Yan background was preselected for events with an invariant mass above 70 GeV and the obtained $\sigma \times BR$ is 435 pb. This translates to a rate of 0.8 Hz at low luminosity which means that the processing of these events is not a problem even with one CPU.

For $t\bar{t}$ events the mean time is much longer since there is more activity with high

P_t tracks. If we play the same game as for the DY events we find that the mean time to compute the HLT response is 1810 ms for a cross section of 500 pb and a Level-1 efficiency of 70% it translates to 1400 ms to process the 0.7 Hz of $t\bar{t}$ events coming from the Level-1 at low luminosity.

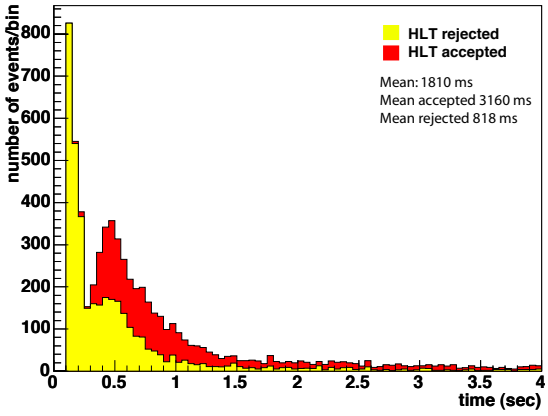


Figure 3.15: Distribution of the computing time required by the HLT to compute its response. This has been done on $t\bar{t}$ events. Each entry in the histogram corresponds to an event accepted by the level 1 trigger.

We have seen that the online processing CPU power to analyse DY and $t\bar{t}$ events is rather small since two or tree machines could be enough to cope with these events at low luminosity. Those backgrounds will probably be an important background in Higgs searches as we will see in the next sections, but it will not be a problem for the online selection. The point is that the event rate at Level-1 is dominated by QCD dijet events.

In order to evaluate the HLT performance, timing measurements have been performed for a given category of events. We have chosen QCD events using the P_T bin $30-50$ GeV. The obtained mean time per Level-1 accept is 410 ms for a cross section of 0.195 mb, and a Level-1 efficiency of 2%. At low luminosity (2×10^{33} cm^{-2}s^{-1}) it translates to a Level-1 accept rate of 7.8 kHz. With one CPU, the mean computing time should then be 0.128 ms. Knowing that the mean computing time for QCD events is 410 ms this means that 3200 CPUs are required to analyse events in real time. Taking into account the doubling of computing power each 18 months foreseen by Moore's[6] law, the number of required CPUs will be reduced to 1000 CPU at the start of the LHC.

[6]The original paper can be found on `ftp://download.intel.com/research/silicon/moorespaper.pdf`

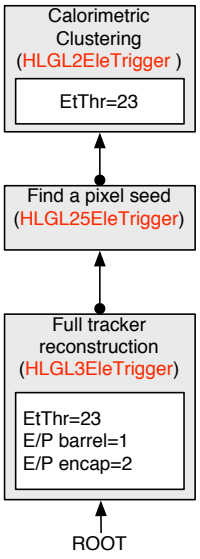
3.4.6 Configuration of the HLT

The coherent handling of the parameters that define the selection is an important point for the High Level Trigger. The COBRA framework provides a mechanism for specifying what are the parameters that define the behaviour of an algorithm. Each algorithm has to announce to the framework what are the names and type of the parameters that can be configured to alter the algorithm behaviour. In addition a class implementing an algorithm should give a default value to the parameters. The reconstruction of the objects produced by an algorithm requires the specification of a value for the parameters that define the algorithm behaviour, otherwise the default values are used. The configuration is implemented in the RecConfig class and querying of an algorithm is done with an instance of the RecQuery. The behaviour of the RecQuery/RecConfig mechanism is used in the on demand reconstruction. For a given event, if the same algorithm is queried twice, then the computation is redone only if the RecQuery is different from the first query.

In the High Level Trigger the algorithms are the trigger elements and their product are the HighLevelTriggerCandidates and boolean response. The HLT definition is performed by implementing a HighLevelTrigger derived class where the complete trigger is fully specified by a set of RecQuery. Each RecQuery is associated with an element and specifies the parameters that define the selection criteria of this element. In addition the RecQuery also specifies the names of the two daughter that the element is connected to and their associated RecQuery. If no parameters are specified when building a RecQuery for an element, then the default configuration of the element is used for the selection. The construction of the HLT RecQuery's follows a factory design pattern [39] that can be extended to meet the online HLT configuration requirements. One could imagine for instance that the trigger is built from the data stored in the condition database and its implementation would not require the modification of the HLT core.

Fig. 3.16 shows an example of the definition of simple selection criteria using the mechanism described in the previous paragraph. The example shows how to construct a single electron trigger with three levels. At Level-2 the search for a calorimetric cluster with $E_t > 23$ GeV is performed; then at Level-2.5 a hit in the pixel detector is required for the full tracking to be performed. Finally at Level-3 an E/P cut is required when the full track reconstruction is performed. The detail of the algorithm is not important at this point and will be given in the next sections. The HighLevelTriggerTest is a HighLevelTrigger derived class that implements the setup method to define the trigger tree.

In the HLT prototype a HighLevelTrigger that read its configuration from an XML file has been implemented. This is particularly interesting in a situation where the selection criteria need to be changed without recompiling the High Level Trigger library that specifies the selection. Changing the selection logic or parameters is just a question of changing an xml file that could even be generated via an adapted graphical user interface. Fig. 3.17 shows the xml files that correspond to the selection specified in Fig. 3.16.




```
virtual  void HighLevelTriggerTest::setup() {
    // Define the Calorimetric Clustering parameters
    RecQuery l2("HLGL2EleTrigger");
      l2.setParameter("EtThr",23);
    // Connect the L2 with the Pixel seed search
    RecQuery l25("HLGL25EleTrigger");
      l25.setComponent("mother1",l2);
    // Connect the L3 with the L25 pixel seed search
    RecQuery l3("HLGL3EleTrigger");
      l3.setParameters("EtThr",23);
      l3.setParameter("EovP_barrel",1);
      l3.setParameter("EovP_endcap",2);
      l3.setComponent("mother1",l3);
   // finally add the tree to the list of root's
   addRootTriggerElement(l3);
};
```

Figure 3.16: Example of the configuration for a single electron trigger with three levels. The required code to construct the trigger tree is shown on the right

```
<?xml version="1.0" encoding="Latin-1" standalone="no"?>
<!DOCTYPE GlobalTrigger SYSTEM "./HighLevelTrigger.dtd">
<GlobalTrigger>
 <L3EleTrigger>
   <L25EleTrigger>
     <L2EleTrigger>
       <Parameter value="23" name="EtThr"/>
     </L2EleTrigger>
   </L25EleTrigger>
   <Parameter value="23" name="EtThr"/>
   <Parameter value="1" name="EovP_barrel"/>
   <Parameter value="2" name="EovP_endcap"/>
 </L3EleTrigger>
</GlobalTrigger>
```

Figure 3.17: xml specification of the trigger tree for the selection of an electron with $E_t > 23$ GeV.

CHAPTER

FOUR

High Level Trigger Analysis

In this chapter we will discuss how the high level trigger strategies are important for the observability of the MSSM higgs sector. A benchmark channel has been chosen, and for this channel we evaluate which trigger patterns dominate in the final analysis. At analysis time, physicists will have to choose from trigger streams to search for new signal. Hence a method to define those trigger streams is developed in this chapter. In the following sections, we set out the signal and background samples used for the exercise. The trigger efficiencies (L1 and HLT) are shown. The L1 and HLT selected sample is then analysed as it will be done offline, with selection cuts to extract the signal from the background. This analysis does not pretend to be as complete as it could be when real data becomes accessible. It is rather an exercise to show how the HLT trigger pattern dominate the signal to background ratio and the signal significance that can be extracted.

4.1 Generated events

This analysis is developed for the benchmark channel involving the decay of a neutral heavy Higgs decays to pair of tau leptons, one tau decays leptonically giving an electron, and the other decays hadronically resulting in a tau-jet.

The dominant production channel for high values of $\tan\beta$ is the fusion of b-quarks. Three masses have been generated for $\tan\beta = 20$, respectively $m_A = 200$, 300 and 400 GeV. The cross section at leading order was computed using the HQQ program [40]. We used HDECAY [41] for the branching ratio computation. Although NLO calculations have been published [42], which enhance the total cross section, we used the LO cross section in this analysis. The diagram giving the most important contribution is shown on Fig. 4.1.

The generated backgrounds are shown with their diagram examples in Table 4.1. The W+jets background has been generated with ALPGEN [43], with the decay of

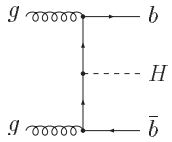


Figure 4.1: Leading Order diagram dominating the cross section for the signal

the W forced to electron, and the requirement to have two jets. It has been generated since one of the jets could fake a tau-jet. Drell-Yan background was generated for $M_{Z/\gamma^*} > 70$ GeV to reduce the cross section. The decay was forced into electron, muon and tau pairs. We used TOPREX v4.02 [44] for the generation of the $t\bar{t}$ and Wtb background. All other backgrounds where generated with PYTHIA [45] v6.215. The $t\bar{t}$ and Wtb background was generated without forcing into specific decay channels. The irreducible background $Zb\bar{b}$ was generated with COMPHEP [46].

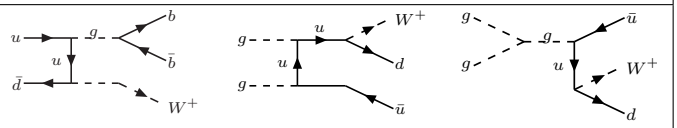
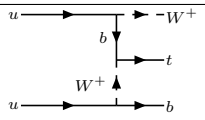
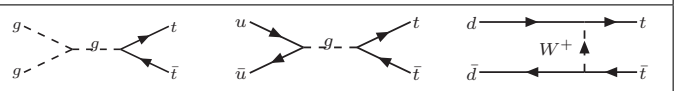
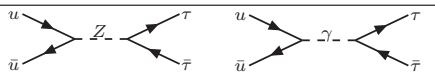
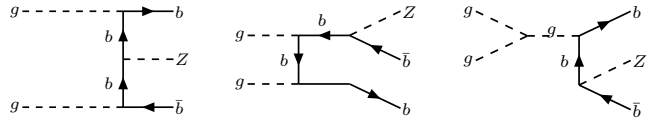
Background	Typical diagram
W+Jets	
Wtb	
$t\bar{t}$	
Drell Yan	
Zbb	

Table 4.1: Sample of the diagram involved in the backgrounds for this analysis.

Computed cross sections multiplied by branching ratios are shown on Table 4.2. With the available computing power in our laboratory (Appendix B), we have generated a total of 1.5E5 events without pileup. Events have been processed through the simulation and reconstruction chain described in B.2. The studied benchmark assumes 100 fb^{-1} of integrated luminosity. The event weights are given by $w_i = (\sigma \times R_i \times L_{int})/n_i$, with R_i is the branching fraction, L_{int} the integrated luminosity,

Process	Selection	$\sigma \times$ BR (pb)	n events	weight
$Z/\gamma \to l^+l^-$, $l = (e, \mu, \tau)$	$M_{Z/\gamma^*} > 70$ GeV	435	74204	587
$t\bar{t}$	no	499	9637	5100
W+jets ($W \to e\nu$)	$min(Jet_{P_t}) > 20$ GeV, $\|\eta\|(Jet) < 2.5$, $\Delta R_{min}(j_1, j_2) > 0.4$	1270	14230	8800
Wtb	no	52	9900	532
$Zb\bar{b} \to b\bar{b}\tau^+\tau^-$	no	5	5900	91
$Hb\bar{b} \to b\bar{b}\tau^+\tau^-$(etau)	no ($m_A = 200$)	0.55	9860	5.5
$Hb\bar{b} \to b\bar{b}\tau^+\tau^-$(etau)	no ($m_A = 300$)	0.10	9397	1.1
$Hb\bar{b} \to b\bar{b}\tau^+\tau^-$(etau)	no ($m_A = 400$)	0.031	9241	0.3

Table 4.2: Cross section for each generated signal and background. Below in the table etau means that the tau pair was forced to decay into electron tau-jet final state. The selection column specifies what type of kinematic selection has been applied to the generated sample.

σ the cross section and n_i the number of events. The event weights are summarised on the right of Table 4.2

Clearly one would require much more background statistics to produce an analysis quoting the expected signal over background ratio for 100 fb^{-1}.

4.2 Trigger efficiencies

At low luminosity regime, i.e. 2×10^{33} cm^{-2}s^{-1}, the summarised trigger tables are given in Table 4.3. A detailed table is given in C.2

sub-trigger	L1 threshold	HLT threshold
electron	$E_t > 29$	$E_t > 26$
double electron	$E_t^1 > 17 \wedge E_t^2 > 17$	$E_t^1 > 14.5 \wedge E_t^2 > 14.5$
single tau	$E_t > 86$	$E_t > 80$
double tau	$E_t^1 > 59 \wedge E_t^2 > 59$	$E_t^1 > 50 \wedge E_t^2 > 50$
etau	$E_t^e > 18 \wedge E_t^\tau > 35$	$E_t^e > 18 \wedge E_t^\tau > 35$
τMET	$E_t^{miss} > 40 \wedge E_t^\tau > 35$	$E_t^{miss} > 40 \wedge E_t^\tau > 35$

Table 4.3: Level-1 and HLT summarised trigger table. The superscript index indicate which candidate have been used. They are ordered by decreasing transverse momentum. The electron threshold at L1 is higher than at HLT due to a bug in the ORCA version 8.2.2 used.

The selection efficiency for each background and signal is shown in Table 4.4. It is clear that the selection efficiency should grow with the Higgs mass as shown in

Process	$\mathcal{E}$ L1	$\mathcal{E}$ HLT	$\mathcal{E}_{\mathcal{R}}$ HLT
$Z/\gamma \rightarrow l^+l^-$	0.49	0.39	0.81
$t\bar{t}$	0.70	0.39	0.56
$W + jets$	0.57	0.42	0.72
wtb	0.61	0.36	0.59
$Zb\bar{b} \rightarrow b\bar{b}\tau^+\tau^-$	0.44	0.19	0.43
signal $m_A = 200, \tan\beta = 20$	0.60	0.42	0.70
signal $m_A = 300, \tan\beta = 20$	0.78	0.63	0.81
signal $m_A = 400, \tan\beta = 20$	0.86	0.75	0.86

Table 4.4: Selection efficiency at L1 and HLT. The last column contains the HLT trigger efficiency relative to the L1 accepted events. $\mathcal{E}_{\mathcal{R}} = N_{HLT}/N_{L1}$ where N_{HLT} is the number of events passing the HLT and N_{L1} the number of events selected at Level-1.

Table 4.4. The $W+jets$ process is the background selected with the highest efficiency. This is because of the high P_t electron, as shown on Fig. 4.2. The efficiency difference between the $Zb\bar{b}$ and the Drell-Yan events is only due to the different final states.

The repartition of the events in the different trigger algorithms as been analysed. For each background and signal events the relative contribution of the high level trigger algorithms are calculated. In the series of figures in Table 4.3, the single electron trigger is dominating the selection efficiency.

In the case of the signal (Table 4.4) there is a clear increase of trigger efficiencies with increasing Higgs masses. Again the single electron is dominating the trigger efficiency, but we have to see the overlap between the triggers to state on the dominance of the single electron trigger. On Fig. 4.5 the efficiencies are shown for each trigger pattern that are by definition exclusive. Most of the trigger efficiency is attributed to two trigger patterns where there is only a single electron trigger or a single electron trigger with an etau trigger. The remaining efficiency consists of patterns with a τMET trigger.

$$\left(\begin{array}{c|cccccc} \text{bit} & e & ee & \tau & \tau\tau & e\tau & \tau\text{MET} \\ \hline e & 1 & 0.0174 & 0.194 & 0.103 & 0.682 & 0.173 \\ ee & 0.806 & 1 & 0.194 & 0.0484 & 0.629 & 0.21 \\ \tau & 0.472 & 0.0102 & 1 & 0.179 & 0.537 & 0.603 \\ \tau\tau & 0.742 & 0.00752 & 0.529 & 1 & 0.749 & 0.173 \\ e\tau & 0.892 & 0.0178 & 0.289 & 0.136 & 1 & 0.285 \\ \tau\text{MET} & 0.343 & 0.00896 & 0.491 & 0.0476 & 0.431 & 1 \end{array}\right) \quad (4.1)$$

At this stage we have seen that the e $\wedge$ eτ trigger and the e trigger alone account for 44% of the selection efficiency for the signal. Followed by the trigger that combines τMET, yielding 35% efficiency. The "correlation" between the sub-triggers is seen in the matrix (4.1). Each entry $M(i,j)$ is read as the probability for the trigger in the column j to be fired if the trigger in the line i is fired. $M(i,j) = P(T_j = \text{true}|T_i = \text{true})$ with $T_{(i,j)}$ the trigger response of trigger i and j. The e and eτ triggers are

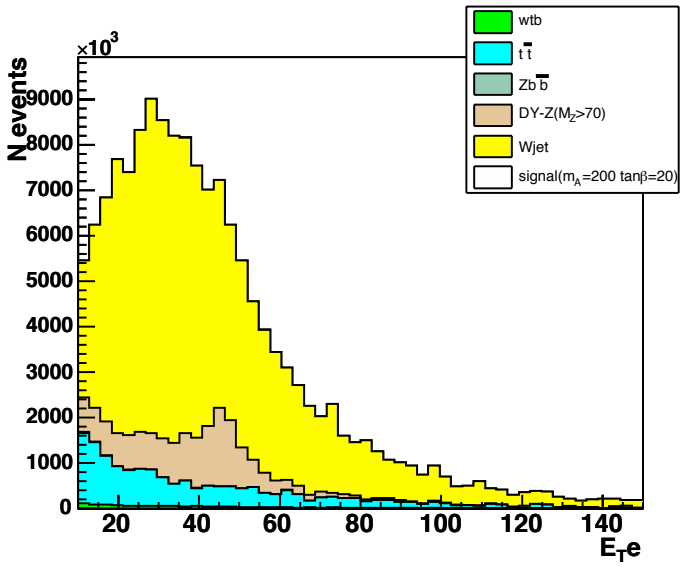


Figure 4.2: Electron E_t spectrum for all backgrounds before L1 and HLT selection.

highly correlated indicating that the e and eτ is not the best choice to enhance the signal efficiency. In the "correlation" matrix we can search for the trigger with the minimal maximum "correlation". It will define the trigger least correlated to the others. For the present matrix, the τMET fits this definition. Although this trigger does not dominate the selection efficiency it will become important after applying additional cuts to extract the signal from the background.

4.3 Methodology for event selection

In this section we present the statistical method used for the offline analysis presented in Sec 4.4. We also introduce the mass reconstruction technique used for this channel.

4.3.1 Statistical approach

The CL_s method has been used to evaluate the search potential of the analysis. In this method one defines a statistical estimator Q that allows to estimate the probability that an observed value is more background-like or more signal-like. For a simple case of a counting experiment, Q can be the number of observed events. In the original Zech's [47] frequentist derivation of upper limits, the CL_s is defined as

$$CL_s = P(n_{s+b} \leq n_0 | n_b \leq n_0). \tag{4.2}$$

Let us assume that n_0 is the number of counts observed in an experiment. n_{s+b}

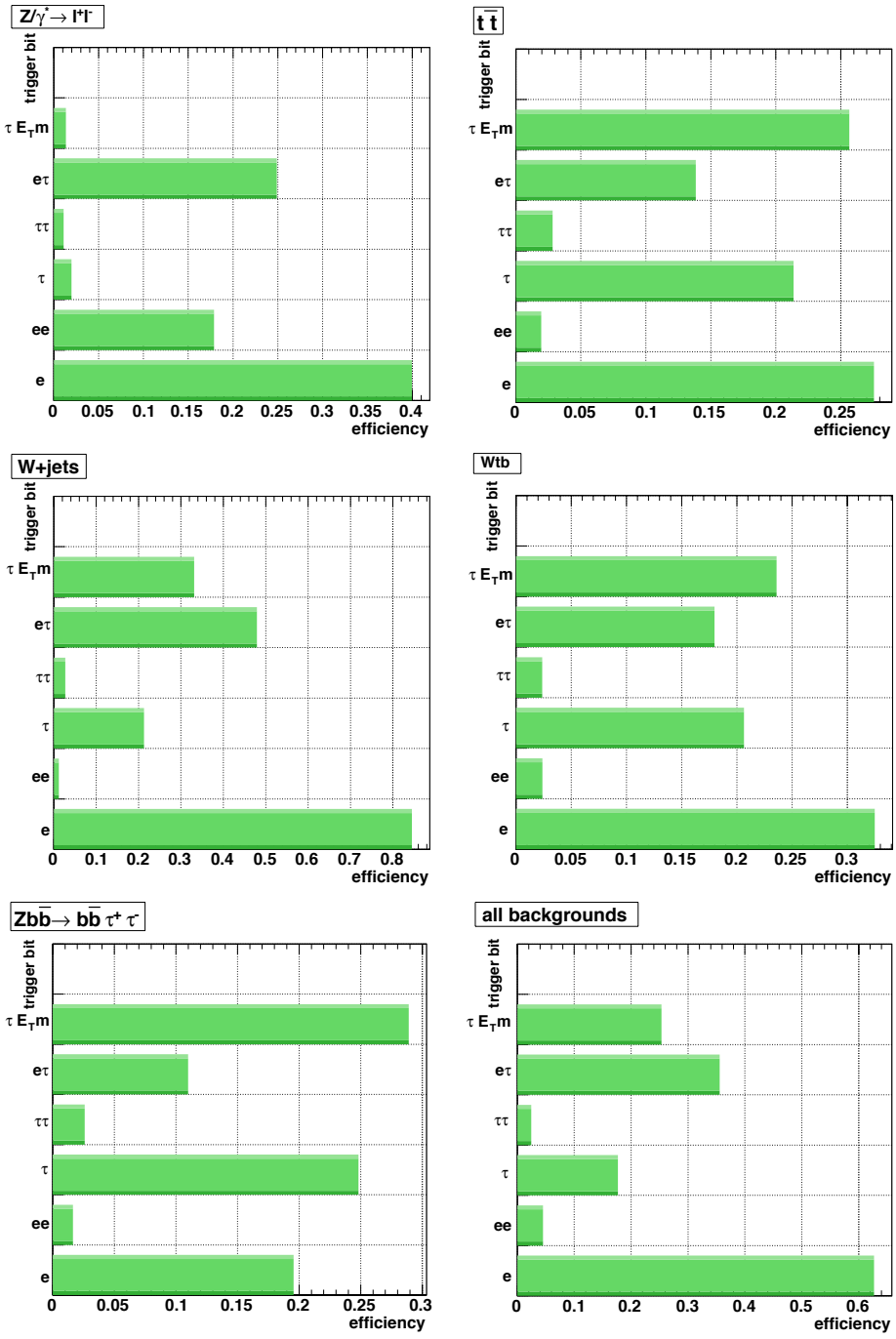


Figure 4.3: Repartition of the trigger decision in trigger selection classes for each background. The trigger class is shown on the y axis the x axis shows the efficiency of the trigger class calculated as N_x/N_{hlt}. N_x is the number of events that as triggered the trigger algorithm x. N_{hlt} is the number of events that where triggered by the high level Trigger.

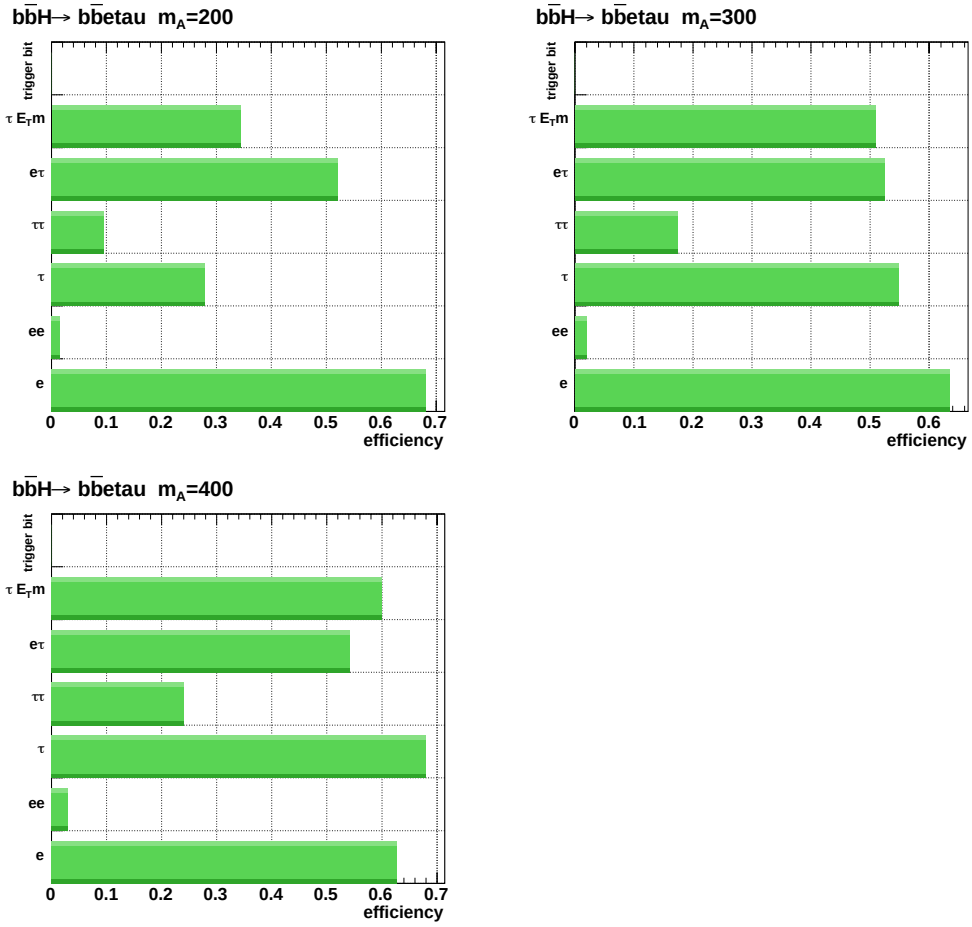


Figure 4.4: Repartition of the trigger decision in trigger selection classes for the three signal samples. The trigger class is shown on the y axis the x axis shows the efficiency of the trigger class calculated as N_x/N_{hlt}. N_x is the number of events that as triggered the trigger algorithm x. N_{hlt} is the number of events that where triggered by the high level Trigger.

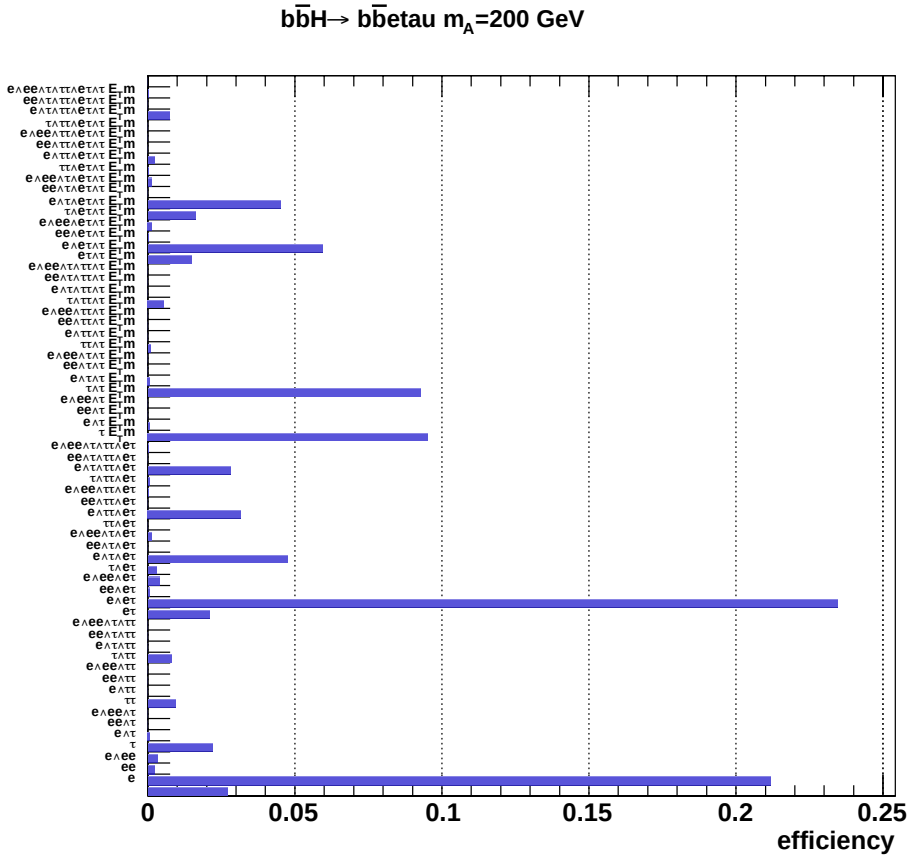


Figure 4.5: Efficiencies for each trigger pattern. The efficiency is computed as N_x/N_{HLT} where x is one of 2^6 exclusive trigger class. The legend on the left indicates what pattern has been fired. As an example the e pattern means the single electron trigger has been fired but no other.

is the number of counts one would see in the hypothesis of signal+background and n_b is the number of counts in the background only hypothesis. Then CL_s is equal to the probability of stating that there is a signal while there is only background. One can translate the equation (4.2) into

$$CL_s = \frac{P(n_{s+b} \leq n_0)}{P(n_b \leq n_0)}. \tag{4.3}$$

CL_s must be minimised when developing an analysis. n_0 is taken as the value for which the background distribution is maximum. Generally the test statistics is more complicated than a simple event count. In most real examples one defines a statistical test as a likelihood ratio. If n_i is the observed number of events in the bin i and s_i, b_i the expected number of signal and background then, Q is defined as:

$$Q = \frac{\prod_{i=1}^{N_{bins}} \frac{e^{-(s_i+b_i)}(s_i+b_i)^{n_i}}{n_i!}}{\prod_{i=1}^{N_{bins}} \frac{e^{-(b_i)}(b_i)^{n_i}}{n_i!}} \tag{4.4}$$

In equation (4.4), Q is a probability ratio. The numerator evaluates the probability that the observed values for n_i are in agreement with the signal+background hypothesis while the denominator is the same probability but in the background only hypothesis. One then defines CL_s as

$$CL_s = \frac{P_{s+b}(Q \leq Q_0)}{P_b(Q \leq Q_0)}, \tag{4.5}$$

where Q_0 is the observed value.

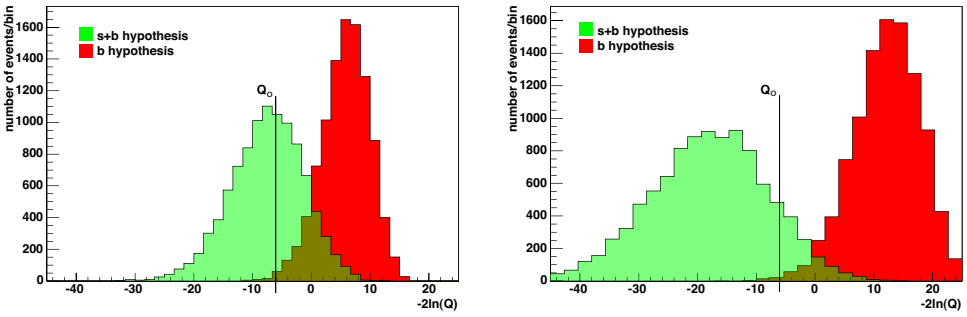


Figure 4.6: Distribution of the statistical variable Q in the background and signal+background hypothesis. The left plot assumes $s + b = 8$ events and $b = 6$ events. The right plot assume the same background hypothesis and $s+b = 13$. Q_0 is the observed value shown with a vertical line.

The (Probability Density Function) PDF of Q is different in the signal+background and background only hypothesis. The generation of the PDF for integration is done assuming the two hypotheses and generating events with a Poisson distribution. Fig. 4.6

illustrates this in the case of a counting experiment. The plots illustrate two situations with different signal and backgrounds. Clearly the CL_s is lower on the right-hand plot than on the left-hand one since the sensitivity of counting events for a high signal/background ratio is better than for lower signal/background. In order to optimise an analysis, one as to develop a selection strategy that better separates the Q distributions in order to obtain the lowest CL_s. In order to fix the selection cuts one fixes Q_0 to be at the maximum of the background statistics. Thus it assumes that there is only background and then computes the expected CL_s (noted $\langle CL_s \rangle$). The selection cuts have to minimise the CL_s in order to minimise the risk to claim a signal discovery if the background-only hypothesis is true. For the analysis developed in the following sections, we will use the hypothesis that there is signal and background and evaluate the probability that what we observe is due to the background only. Therefore, Q_0 is chosen at the peak of the Q distribution for the s+b hypothesis and we will develop a set of cut to minimise $P_b(Q \leq Q_0)$ or maximise $\langle CL_s \rangle$.

4.3.2 Mass reconstruction

The Higgs mass reconstruction must take into account the missing energy taken by the three neutrinos. The missing transverse momentum is used in conjunction with the hypothesis that the neutrinos are emitted in the direction of the original tau lepton. Let $\vec{v_1}$ and $\vec{v_2}$ the momentum vectors of the visible tau decay products. We can write,

$$\vec{p}_{\tau_1} = (1+\alpha)\vec{v}_1, \tag{4.6}$$

$$\vec{p}_{\tau_2} = (1+\beta)\vec{v}_2. \tag{4.7}$$

α and β account for the momentum of the neutrinos. The energy fraction taken by the visible tau decay products is given by $x_1 = 1/(1+\alpha)$ for τ_1 and similarly $x_2 = 1/(1+\beta)$ for τ_2. The missing transverse momentum $\vec{P_T^m}$ is written as

$$\vec{P}_T^m = \alpha\vec{v}_{1T} + \beta\vec{v}_{2T}. \tag{4.8}$$

Solving α and β proceeds by computing the cross product of $\vec{P}_T^m$ with $\vec{v}_{1T}$ and $\vec{v}_{2T}$ ($\vec{v}_{1T}$, $\vec{v}_{2T}$ being the transverse component of $\vec{v}_1$ and $\vec{v}_2$)

$$\vec{P}_T^m \times \vec{v}_{1T} = \beta\vec{v}_{2T} \times \vec{v}_{1T}, \tag{4.9}$$

$$\vec{P}_T^m \times \vec{v}_{2T} = \alpha\vec{v}_{1T} \times \vec{v}_{2T}. \tag{4.10}$$

α and β solutions are given by:

$$\alpha = \frac{P_T^m \sin\delta\phi_2}{v_1 \sin\delta\phi}, \beta = \frac{P_T^m \sin\delta\phi_1}{v_2 \sin\delta\phi}, \tag{4.11}$$

where $\delta\phi_{1,2}$ are the angles in the transverse plane between the missing momentum and the two tau directions given by $\vec{v}_{1T}$ and $\vec{v}_{2T}$. In addition $\delta\phi$ is the angle between two taus in the transverse plane. Finally, if $m_{\vec{v}_1\vec{v}_2}$ is the invariant mass constructed with the visible decay products then, the mass of the tau pair is obtained by rescaling $m_{\vec{v}_1\vec{v}_2}$ with x_1 and x_2 as:

$$m_{\vec{\tau}_1\vec{\tau}_2} = \frac{m_{\vec{v}_1\vec{v}_2}}{\sqrt{x_1 x_2}} \tag{4.12}$$

The effect of rescaling the invariant mass is show on Fig. 4.7, for signal events with $m_A = 200$ GeV. $\vec{v}_1$ is taken as the total momentum of the tau visible decay products and $\vec{v}_2$ the electron momentum. In the following sections, the rescaled reconstructed mass is used as the discriminant variable for the computation of the signal over background expectation.

4.4 Offline selection for $m_A = 200$ GeV

By "Offline" we mean a selection made after the events are written to tape, to further extract the signal out of the background. Two exclusive selections strategies are defined based on the number of reconstructed electron candidate in the sample. It has to be mentioned that the offline electrons are not required to pass the E/P cut on the track. This means that some of the one-prong tau decays are seen as offline electrons. After the HLT we observe 50% of the events have one offline electron and a tau-jet, but in the remaining 50% they is no identified tau-jet but instead two offline electrons are seen. Therefore we define two exclusive streams (or classes), the single electron and double electron stream.

In the "single electron" stream, the reconstruction algorithms sees only one electron candidate and one tau-jet candidate not colinear with the electron. In the "double electron" stream we do not see a tau-jet not colinear with an electron candidate and there are more than one electron candidate.

The identification quality is shown in Fig. 4.8 for the one electron stream, and in Fig. 4.9 for the di-electron one. The distance is defined as

$$Dr = \sqrt{(\eta_{mc} - \eta_{off})^2 + (\phi_{mc} - \phi_{off})^2}, \tag{4.13}$$

with off the reconstructed particle and mc is the corresponding Monte Carlo particle. As seen on the Fig. 4.8 the identification quality of the tau-jet is lower than for the electron. Nevertheless in 99% of the events in the one electron stream, the particles are correctly identified.

As expected, it is not the case in the double electron stream. In that case, the most energetic reconstructed electron does not always match the real electron as seen on Fig. 4.9. In a non negligible fraction of events the most energetic electron does not match the real electron.

4.4.1 One electron stream

The dominating background for the situation with one electron is the W+jets background. Reconstructing the transverse mass of the electron and missing E_t as $m_T^2 = E_T^2 - P_T^2$, where $E_T = (E_T^m + E_T^e)$ and $\vec{P}_T = \vec{E}_T + \vec{E}_T^e$ shows a peak at 80 GeV as on Fig.4.10.

Cutting at $M_T < 18$ GeV rejects 93% of the W+jets background. The next background to overcome is the $t\bar{t}$ background. Since the electron and tau are coming

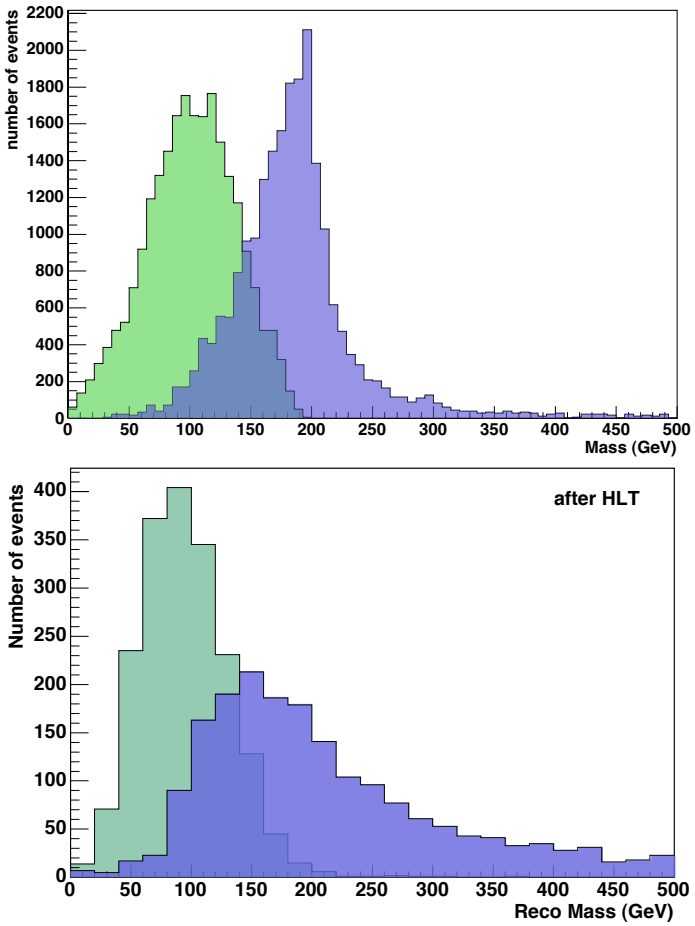


Figure 4.7: On the top, mass computed with the tau Monte-Carlo decay products including (blue) and not including (green) the missing transverse momentum, for a generated mass of 200 GeV. On the bottom, the same quantities are shown using the reconstructed information after HLT selection.

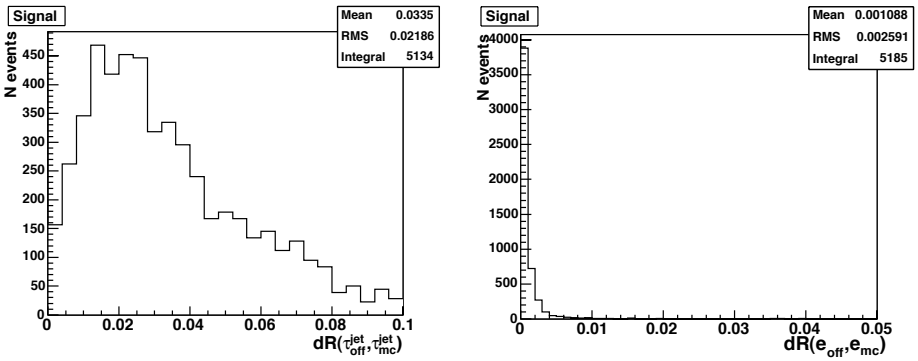


Figure 4.8: Reconstruction quality in the in the single electron class. On the left, distance in the (η, ϕ) plane, between the reconstructed tau-jet with the highest E_T and the true tau-jet. On the right, the same distribution is shown but for the electron reconstruction quality.

from the decay of different W bosons, the angular separation between the electron and tau-jet candidate can be used to further reduce the $t\bar{t}$ background. Fig. 4.11 illustrate this for events passing the transverse mass cut.

After applying $\Delta\phi > 2.5$ the $t\bar{t}$ background still dominates. Given the jet multiplicity in $t\bar{t}$ events, the electron is expected to be better isolated in the signal sample than in the $t\bar{t}$ background. Therefore, we define the isolation (iso_e) as the distance in the (η, ϕ) plane to the closest track with a P_t above 3 GeV. The 3 GeV cut is the highest possible cut for which a decent signal efficiency is obtained. Fig. 4.12 shows that a cut with $\text{iso}_e > 0.25$ rejects 70% of the $t\bar{t}$ background.

After this cut is applied, the Drell-Yan remains an important background. To reduce it, we use the fact that the missing transverse energy in the signal is relatively higher than in the Drell-Yan background. The missing transverse energy (MET) distribution depends on the Higgs mass hypothesis and for small Higgs masses the MET cut becomes less discriminant. For the 200 GeV hypothesis, it is particularly useful to cut on MET > 25 GeV as shown on Fig. 4.13.

Table 4.5 summarises the selection cuts described in the previous paragraphs. The first column ($s/\sqrt{b}$) does not take into account the Monte Carlo statistical error, and is an indication of the expected $s/\sqrt{b}$ one could achieve if the present selection does not select a statistical fluctuation of the background enhancing the signal. The last two columns are computed with the statistical error taken into account. The first estimator s/σ_b uses the weights to give the variance estimator for the background, given as $\sigma_b = \sqrt{\sum_{i=1}^{n} \alpha_i^2 n_i}$. α_i are the background weights of Table 4.2 and n_i are the number of MC events. The second estimator computes the significance using the CL_b with the s+b hypothesis. The expected significance is ~ 0.1 after selection, its low value is due to the small Monte Carlo statistics. The significance that does not include the statistical error, is an indication of what it could be when the MC statistical error is small. It assumes that the present selection does not introduce a

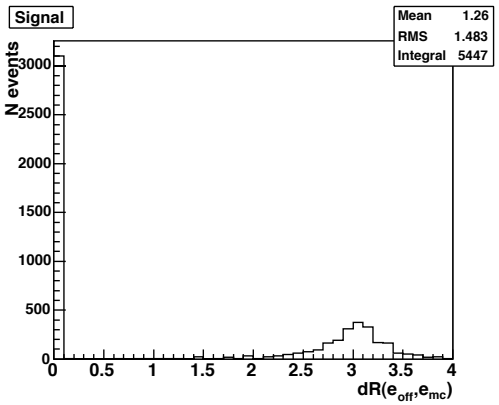


Figure 4.9: For the two electron class, distance in the (η, ϕ) plane, between the reconstructed electron with the highest E_T and the true electron.

strong bias.

Description	Cut	N_s	N_b	$s/\sqrt{b}$	s/σ_b	$S(CLb_{s+b})$
trigger	L1∧HLT	23557	9.2E7	2.4	0.027	0.032
	$1e \wedge 1\tau$	5441	3.4E6	2.9	0.032	0.040
Anti W+jets	$M_T < 18$ GeV	3114	415327	4.8	0.060	0.072
Anti $t\bar{t}$	$\delta\phi > 2.5$	2957	290545	5.4	0.068	0.084
	$iso_e > 0.25$	2583	131003	7.1	0.082	0.102
Anti Z	MET > 25	1758	58628	7.3	0.072	0.083

Table 4.5: Summary of the selection cuts applied for the single electron class.

Trigger patterns after selection

We can now have an insight on the trigger patterns with highest efficiency after offline selection. Before selection the dominating trigger was the single electron but it is not anymore the case after selection as seen on Fig. 4.14. This means that it is not always the trigger with highest efficiency at HLT that contributes to the highest signal over background. The enhancement in the trigger efficiency in the exclusive representation is shown on Fig. 4.15.

The single electron trigger has almost disappeared, since we required to have one electron and one tau. The double electron trigger also vanishes for the same reason. On the contrary, all triggers that are in combination with a τMET trigger are enhanced. Another way to quantify the importance of the τMET trigger is to see how the $s/\sqrt{b}$ decreases when this trigger is disabled. The expected significance after selection is 7.3 if the τMET is enabled (Table 4.4). Disabling the trigger reduces

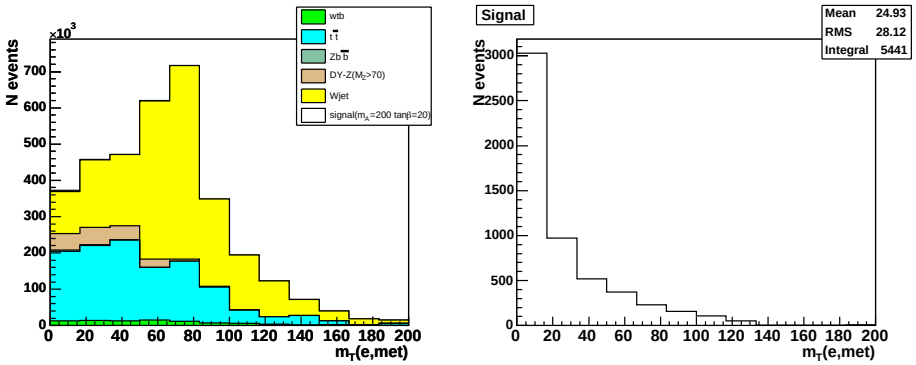


Figure 4.10: Transverse mass constructed with the electron and missing E_T, for the signal on the right and background on the left after L1 and HLT selection.

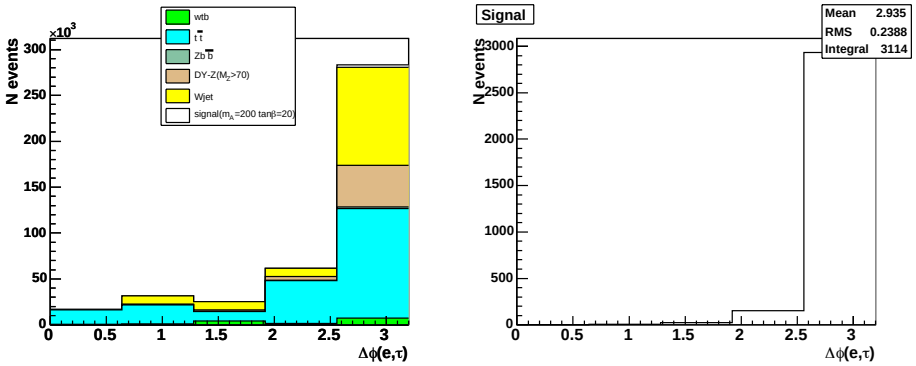


Figure 4.11: $\Delta\phi$ angle between the electron and tau-jet after the transverse mass cut is applied.

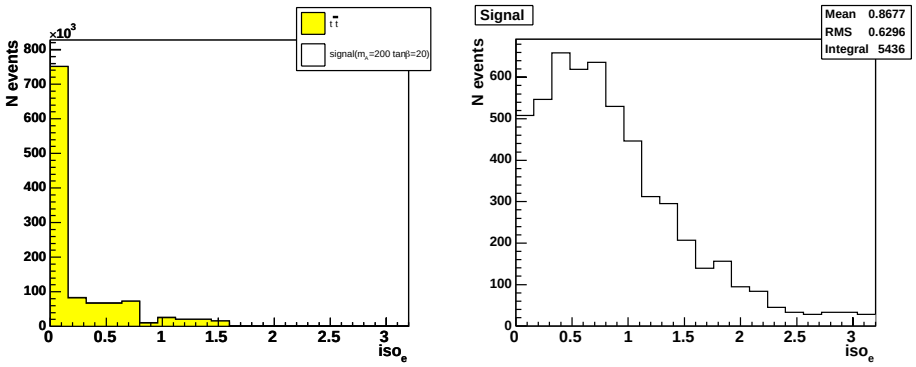


Figure 4.12: Isolation of the electron before the transverse mass cut. On the left for the $t\bar{t}$ background and for the right on the signal.

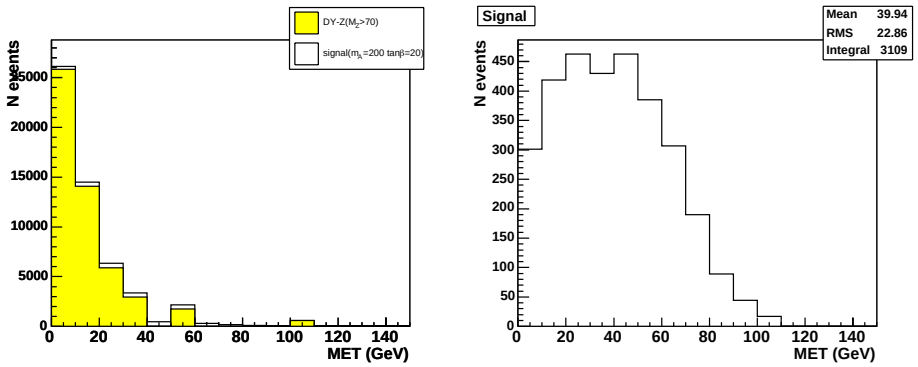


Figure 4.13: MET for the signal (right) and Drell-Yan background (left) after the transverse mass cut.

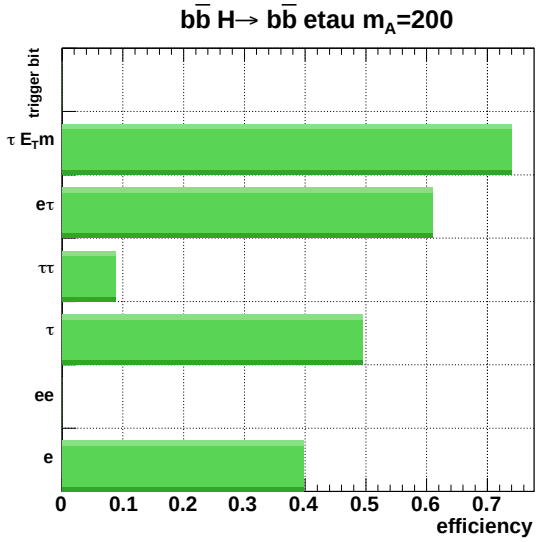


Figure 4.14: Trigger efficiency after 'offline' selection

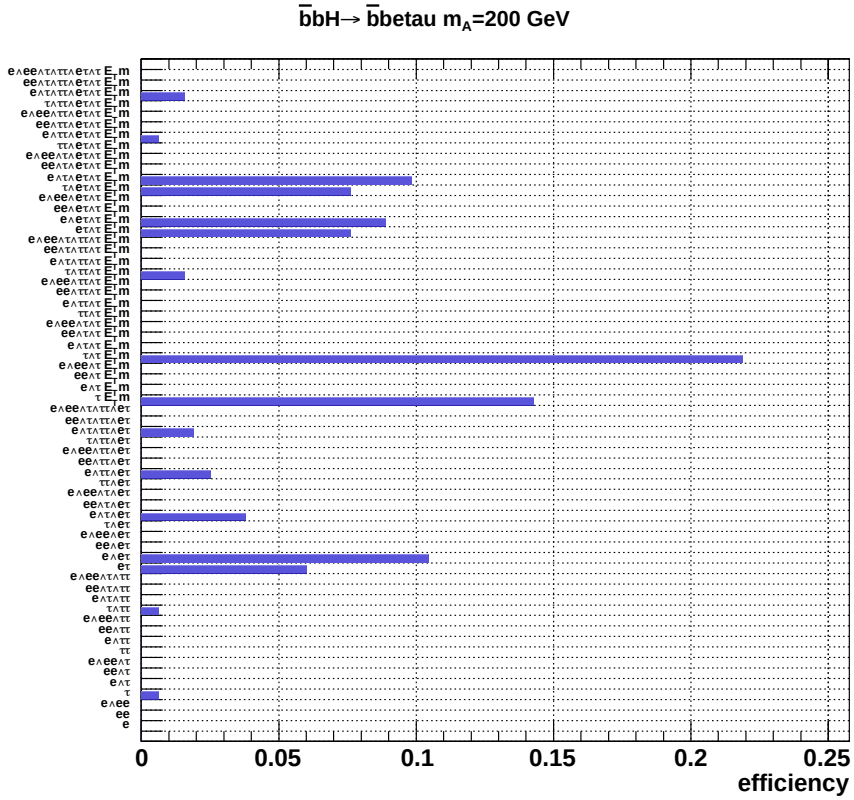


Figure 4.15: Efficiencies for each trigger pattern. The efficiency is computed as N_x/N_{HLT} where x is one of 2^6 exclusive trigger class. The legend on the left indicate what is the pattern that has been fired.

the significance to 3.2. In addition we give the "correlation" matrix after selection is shown in Eq. 4.14.

$$\left(\begin{array}{c|cccccc} bit & e & ee & \tau & \tau\tau & e\tau & \tau\text{MET} \\ \hline e & 1 & 0 & 0.432 & 0.168 & 1 & 0.528 \\ ee & 0 & 0 & 0 & 0 & 0 & 0 \\ \tau & 0.346 & 0 & 1 & 0.115 & 0.5 & 0.859 \\ \tau\tau & 0.75 & 0 & 0.643 & 1 & 0.75 & 0.429 \\ e\tau & 0.651 & 0 & 0.406 & 0.109 & 1 & 0.594 \\ \tau\text{MET} & 0.283 & 0 & 0.575 & 0.0515 & 0.489 & 1 \end{array}\right) \quad (4.14)$$

One important point is how the trigger rate is enhanced when the τMET trigger is added. We have estimated the rate on a di-jet QCD sample with pile-up a low luminosity. We have used P_T bins going from 20 GeV to 80 GeV totalling 58000 events. The total L1 rate with this sample is 5.4 kHz after L1. This rate reduces the total rate by 3%. The evaluation of the HLT rate would require much more statistics.

4.4.2 Two electron stream

In the case where there is no HLT tau candidate, we use the second most energetic offline electron as the tau candidate. As shown on Fig. 4.9 the events around $Dr = 0.3$ are events where the charged pion is taken as most energetic offline electron and the real electron sits in the second offline electron. The so-called two electron class has a much lower startup signal over background ratio since the $Z \rightarrow ee$ events contribute in this stream while it was not in the single electron stream.

Description	Cut	N_s	N_b	$s/\sqrt{b}$	s/σ_b	$S(CLb_{s+b})$
trigger	L1∧HLT	23557	9.2E7	2.4	0.027	0.032
	$2e$	5525	$1.7E7$	1.3	0.016	0.021
Anti W+jets	$M_T^{iso} < 20$ GeV	2494	2.6E6	1.54	0.025	0.03
Anti Z	$\lvert m_{ee} - 90\rvert > 10$	1919	1.4E6	1.62	0.021	0.02
Anti $t\bar{t}$	$\max(i_{e_1}, i_{e_2}) > 0.2$	1847	7.7E5	2.1	0.027	0.017
Anti Z	MET > 25	1205	2.9E5	2.21	0.026	0.016

Table 4.6: Summary of the selection cuts applied for the double electron class.

The complete selection criteria are given in Table 4.6. As in the single electron stream the W+jets background is reduced by 93% by a cut on the transverse mass constructed with the most isolated electron and missing E_t (M_T^{iso} in Table 4.6).

The single electron stream did not include the double electron decay of the Z, but it is present in the double electron stream. The Z peak appears clearly in the di-electron invariant mass m_{ee} as shown on Fig. 4.16 for events passing the L1 and HLT selection in the two electron stream. Cutting on $|m_{ee} - 90| > 10$ rejects 85% of the Drell-Yan background. The remaining background is attributed to an off-shell Z/γ and to decays in tau leptons and then to electron pairs that smear the invariant

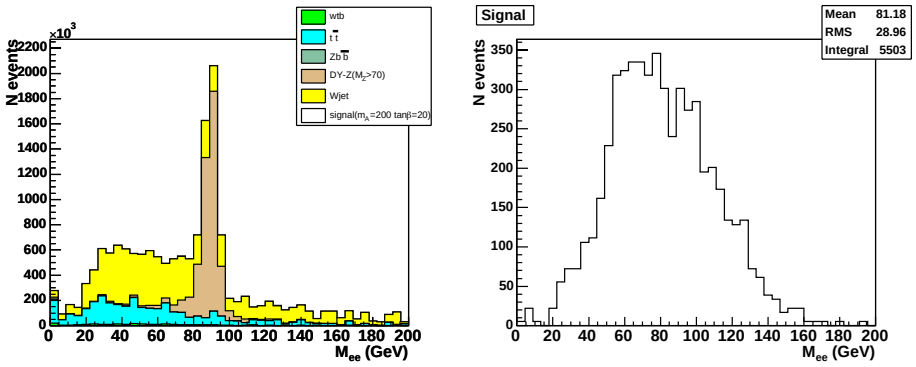


Figure 4.16: Reconstructed mass for the two electron candidates after L1 and HLT selection for the double electron stream.

mass. After the application of the two former cuts (on m_{ee} and M_T^{iso}), the rejection of the $t\bar{t}$ background is performed by requiring the most isolated electron to have an isolation of more than 0.2 in the (η, ϕ) plane. 80% of this particular background is rejected with such a cut. Fig. 4.17 shows the isolation after the W+jets the di-electron invariant mass cuts. The dominating backgrounds, are the W+jets and Drell-Yan,

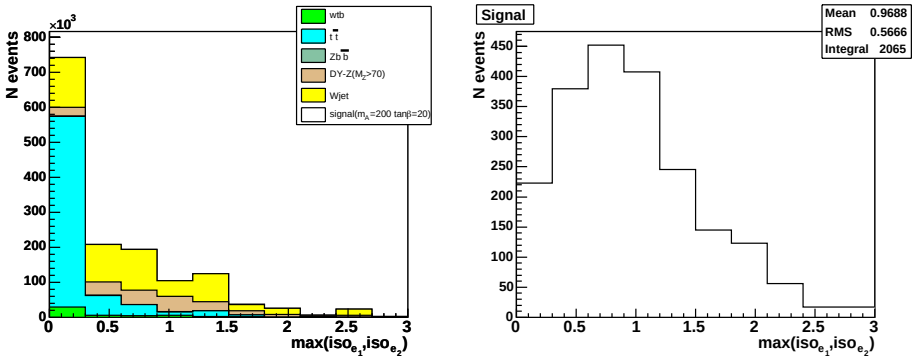


Figure 4.17: Maximum of the isolation computed for the two electron candidates.

even after the invariance mass cut. To further reduce the Drell-Yan, we require the missing momentum to be higher than 25 GeV (Fig. 4.18). Less mass implies less missing momentum in the decay to two tau pairs.

As shown in Table 4.6, the signal over background gain after selection is not significant. The signal over $s/\sqrt{b}$ is even higher after the L1 and HLT selection than after the complete selection. This is due to the higher significance of the single electron stream. The double electron stream is an attempt to select a subset of the signal events were no tau-jet has been identified but one track of the hadronic tau decay has been taken as an offline electron candidate.

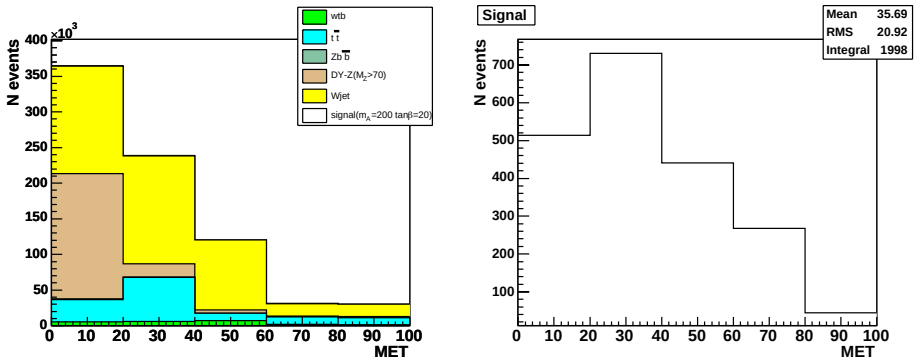


Figure 4.18: Total missing transverse energy.

4.5 Summary

Both in the "one electron" and "two electron" stream, the dominant background is W+jets. For the "two electron" the next dominant background is $t\bar{t}$, while it is the Drell-Yan for the "one electron" stream. In the "two electron" stream the rejection of the Drell-Yan background is efficiently achieved with a mass cut on the di-electron mass. The same cut could be applied on the di-tau pair mass computed using Eq. 4.12 but its resolution is degraded by the missing E_t resolution.

By combining the two analysis, the signal efficiency is 5.4% and it is 0.16% for the background. We start with 54230 signal events and 223×10^6 background events for 100 fb^{-1}. The "one electron" stream gives the best signal significance of 7.3 without including the statistical error. Using the statistical error the significance drops to 0.07. The significance including the statistical error is increasing as more cuts are applied, as shown on Table 4.5, indicating the pertinence of the selection up to the MET cut.

With the selection in place, we show that the trigger patterns with highest efficiency at HLT are not always the one with the highest significance for the signal. In particular, for the "one electron" stream, the single electron trigger efficiency is strongly reduced by the application of analysis selection cuts. The best trigger patterns seems to be a combination of $e\tau$ and τMET triggers.

The systematic error on the level of background has not been estimated since the significance is dominated by the statistical error. For the background, σ_b/b is at the level of 36% and $\sigma_s/s = 5\%$ for the "one electron stream" after selection. Clearly more statistics would be required to perform an analysis quoting a signal significance.

Finally, to illustrate the rejection power of the present analysis, the selection efficiency of each cuts in the "one electron" stream are shown on Fig. 4.19.

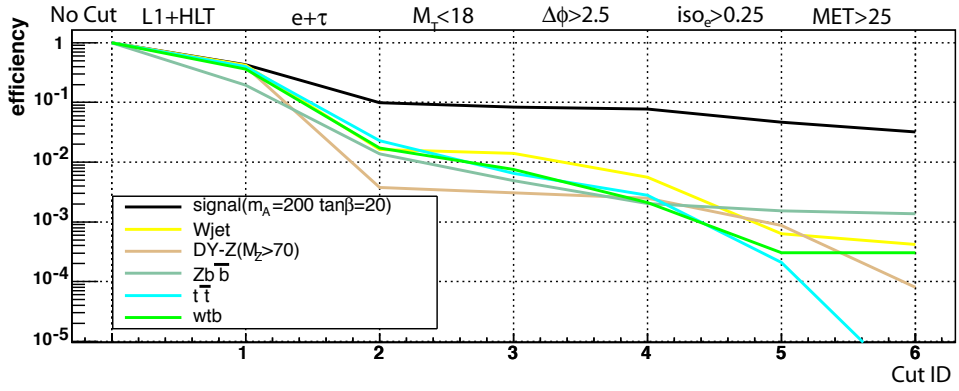


Figure 4.19: Selection efficiency after each cuts presented on Table 4.5. The efficiency is computed with the number of events before L1 and HLT selection.

4.5.1 Background estimations

The list of backgrounds considered in the present analysis is not exhaustive. For example, the W+jets background should also include the decay of W into tau which will contribute in higher missing momentum region. In addition one could include the WW background since no b-tagging is used in this analysis. However the cross section for the WW background is at the level of 0.14 pb.

The present analysis is therefore a test of a possible procedure to define trigger streams by looking at the trigger patterns surviving the off-line selection. This work should be done for every major channel, with higher statistics. Two independent Monte Carlo event samples should be used, one for the optimisation of the selection and another one to evaluate the significance of each trigger pattern after off-line selection to avoid any bias in the off-line selection. Another point is that the analysis should evaluate the effect of pile-up, to check how the selection is affected.

4.5.2 About the b tagging

We have not used the b tagging in the selection procedure presented in the previous section. The main reason for this is the soft spectrum of the b quarks in the signal compared to that of the backgrounds. Fig. 4.20 shows the P_T spectrum of the most energetic b in the signal and with the $t\bar{t}$ sample. In order to preverve a "good" efficiency for tagging the soft b-jet, we used the track-counting b tagging algorithm [49] with a minimum significance of 2.

The efficiency of the b tagging decreases with the P_T of the b quark, the efficiency for the selection of b quarks with $|\eta| < 2.5$ on the signal is 30% while it is 75% on the $t\bar{t}$ background. The b tagging efficiency is shown on Fig. 4.21 as a function of the b P_T assuming a sample of signal and $t\bar{t}$ background.

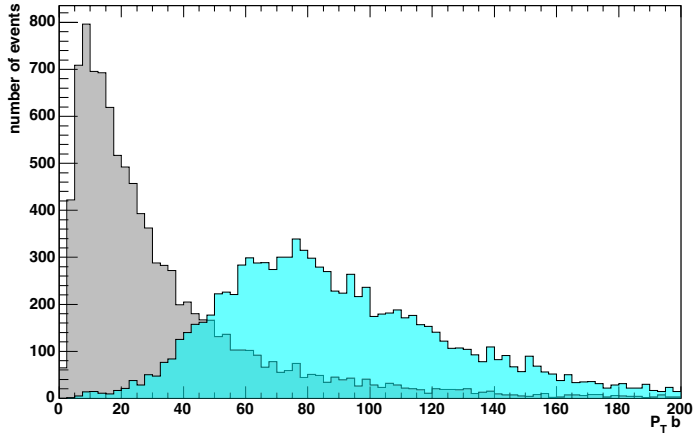


Figure 4.20: P_T spectrum (in GeV) at the generator level for the most energetic b for $t\bar{t}$ background in light blue and for the signal in grey (m_A = 200 GeV). The number of events is normalised such that the ratio of integrals is equal to one.

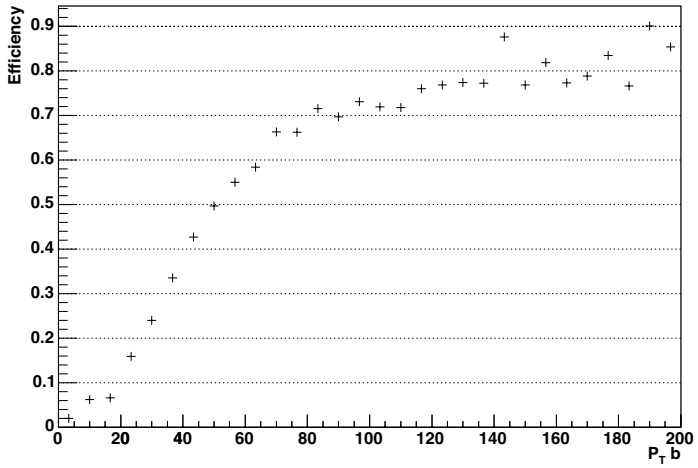


Figure 4.21: b tagging efficiency as a function of the b quark P_T.

If we use the selection of the single electron stream and we add the requirement to have at least one identified b jet, the significance not including the statistical error drops from 7.3 to 2. It seems therefore not optimal to apply such a tagging. However, the statistical error is large and a final conclusion will only be possible with a higher statistics. Conclusion about the trigger mix could also be dramatically affected.

Conclusion

Throughout this dissertation, we have shown the importance of the High-Level Trigger in the CMS detector one of the future LHC experiments. The development of new analyses in the forthcoming years will have to consider the trigger system as an integrand part of the analysis processes, since the online selection will be almost as elaborate as the offline analysis.

The prospects for the possible observation of new physics is conditioned by the quality of the simulation and reconstruction software. To prove the non observability of new physics signals, fast simulation tools are useful. But to prove the observability of a signal, we need to use detailed simulation and reconstruction software since fast tools provide a rather idealised simulation.

For what concerns the simulation and reconstruction, we have designed the High-Level Trigger steering software to set up the definition of trigger patterns by use of combinations of simple trigger primitives. This software is usable on simulated event samples and can also be used in the Event Filter of the CMS experiment. In that context we have devised a method for the optimisation of the mean computing time based on the real time gathering of timing statistics. The trigger tree can then be built based on the gathered information such that the mean accept time is minimised. In order to reduce the tails in the time to process events, the use of the optimisation mode could be valuable if the complete trigger response is not computed in the filter farm.

Using this tool, we have devised a method to identify the trigger patterns dominant in an analysis. The observability of MSSM heavy Higgses is used as a benchmark channel with $m_A = 200$ GeV and $\tan\beta = 20$. In this particular channel, the trigger pattern that is the most efficient after the High-Level Trigger is not the dominating pattern after the offline selection. It is expressed by the fact that the single electron trigger which is the one with highest efficiency after HLT is strongly reduced after the offline selection and the dominating trigger pattern becomes a mix of the $e\tau$ and τMET triggers. A general method for selecting trigger patterns contributing to an analysis could use this technique to compute the contribution from each trigger pattern after the offline analysis.

APPENDIX

A

Search for WW decay of a Higgs boson at LEP

The search for a fermiophobic Higgs boson has been primarily carried out in the $H \to \gamma\gamma$ channel, in which the Higgs boson couples to photons via a W loop. For fermiophobic Higgs boson heavier than 90 GeV, the predicted $H \to \gamma\gamma$ branching fraction becomes small relative to the predicted $h \to WW$ branching fraction (Fig. A.1) motivating a search in this new channel.

The main production processes at e^+e^- colliders for a fermiophobic Higgs boson are $e^+e^- \to Z^* \to HZ$ (Higgs-strahlung), WW and ZZ fusion. The cross section for the boson fusion production processes are considerably smaller than the Higgs-strahlung process for LEP center-of-mass energies. In the Higgs boson mass range kinematically accessible for Higgs-strahlung at LEP ($m_H \leq 120$ GeV/c^2), at most one on-shell W can be produced. Even if the partial differential width of the Higgs boson to WW is dominated by $H \to W^*W^*$ (Fig. A.1) rather than $H \to WW^*$, one of the virtual W's is expected to be near on-shell, and the other to have a much smaller mass and energy.

The full spectrum of $HZ \to WW^*ff$ contains a total of 96 different channels depending on the decays of the W's and the Z. These final states were grouped in four exclusive classes depending on the number of hard leptons in the final state. The analysis is performed on the data taken in the years 1999 and 2000 at center-of-mass energies ranging from 191 to 209 GeV. The luminosities and center-of-mass energies are shown in Table A.1.

A.1 Simulated samples

Sample of signal events were generated using `HZHA` [52] for Higgs boson mass between 90 and 117 GeV and for the 7 different centre-of-mass energies: 192, 196, 200, 202,

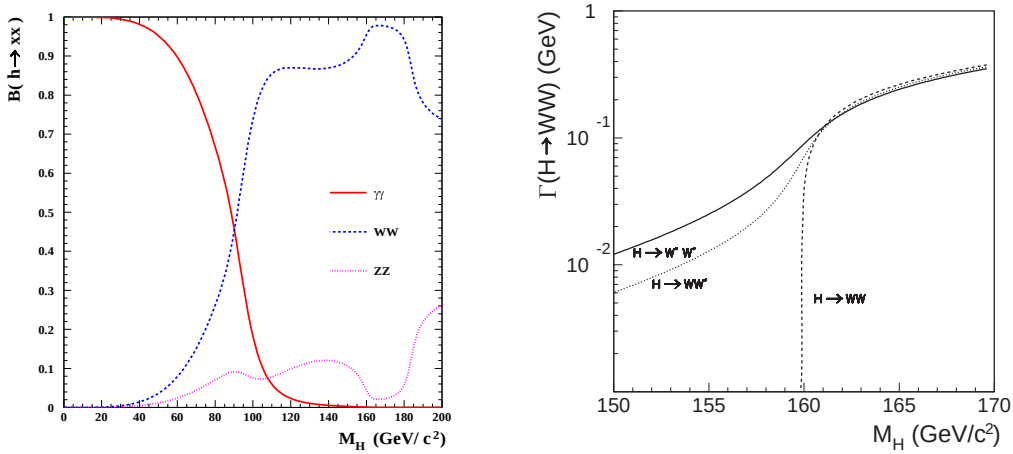


Figure A.1: On the left, branching fraction of the benchmark fermiophobic Higgs boson into boson pairs as calculated by HDECAY2 [50]. On the right, Comparison of the off-shell and on-shell contributions for $H \to W^+W^-$. The dashed line corresponds to the on-shell decay, the dotted line to the case where only one W is off-shell, and the solid line corresponds to the case where both W's are off-shell, from Ref. [51].

205, 207 and 208 GeV. In the `HZHA` code, there is no spin correlation between the W's coming from the Higgs boson decay. The signal events are therefore re-weighted to take into account this spin correlation. The event weight is computed as the ratio between the full four-fermions matrix element and the `HZHA` one:

$$w \approx \frac{|M_{H\to 4f}|^2}{(|M_{H\to WW^*}| \cdot |M_{W\to 2f}| \cdot |M_{W^*\to 2f}|)^2} \ , \tag{A.1}$$

with the formula from [53].

Event samples of all SM background processes relevant for the Higgs search were also generated: the Bhabha process was simulated with `BHWIDE` [54], muon and tau pair production with `KORALZ` [55], the $\gamma\gamma$ processes with `PHOTO2` [56], WW production with `KORALW` [57], the $q\bar{q}$ events with `KK2F` [58], and the remaining four-fermion processes with `PYTHIA` [45].

A detailed GEANT [59] simulation of the detector response was applied to both background and signal events.

The ALPHA++[60] package has been used to analyse data and Monte Carlo. ALPHA++ is the C++ version of the FORTRAN analysis package ALPHA. It has been cross-checked in the context of previous analysis [61] and is now fully validated.

A.2 Description of the selections

Four exclusive classes are defined according to the different final state topologies. The first class contains events with no lepton, it is the fully hadronic class. Signatures with more than one hard lepton, including all the leptonic Z decays, is set in the second class. The third and the fourth classes contain events with a single reconstructed hard lepton, and at least one soft lepton, respectively. For simplicity and conciseness, the term "lepton" is here used in the final state for electrons and muons only. It therefore includes taus decaying leptonically, but not one- and three-prong hadronic tau decays. Those last are only specifically studied for class two events, where it can be efficiently separated from the hadronic W boson decays. The topology classification is given in Table A.2. The notation convention is as follows: the decay of the Z is given in parenthesis, followed by the decays of the quasi on-shell and the off-shell W's; Q and q stand for a hadronic decay (hard or soft); L and l stand for a leptonic decay (hard or soft). The table gives as well the WW decay branching ratio corresponding to each class, with and without taking into account channels where at least one W decays into a tau and a neutrino.

After preselection, all the selection cuts are optimised in order to minimise the expected confidence level $\langle \mathrm{CL_s} \rangle$, defined in Sec. 4.3.1. For the signal, a Higgs boson mass hypothesis of 110 GeV/c^2 is chosen for the optimisation process. A blind analysis approach has been chosen over the use of an independent dataset for optimisation since a too large statistic would have been needed for the low branching ratio channels. To do this, data was not looked at until the end of the optimisation and selection process. Only the data taken in year 2000 and the corresponding simulated events are used during the optimisation process. The data taken in 1999 are therefore not included in numbers and plots presented in this section. They are included in final results presented in Sec A.4.

Table A.1: Integrated luminosities, center-of-mass energy ranges and mean center-of-mass energy values for data collected by the ALEPH detector for years 1999 and 2000.

Year	Luminosity (pb^{-1})	Energy range (GeV)	$\langle\sqrt{s}\rangle$ (GeV)
2000	7.3	$207 - 209$	208.0
	125.9	$206 - 207$	206.6
	79.7	$204 - 206$	205.2
1999	44.4	$201 - 203$	201.7
	87.2	$199 - 201$	199.5
	79.9	$195 - 197$	195.5
	28.9	$191 - 193$	191.6

A.2.1 The class separation and preselection

To separate the classes, the energies of the two most energetic leptons in the final state are used. The first lepton is labelled as "hard" if the energy is higher than 25 GeV. Events with a hard lepton are set in class 2 if the second lepton has more than 20 GeV, and in class 3 otherwise. To separate events in class 1 and 4, a linear discriminant is build with the energy of the most energetic lepton E_{L1}, the total missing three-momentum p_{miss}, and the most energetic lepton isolation I_{L1}:

$$2.3E_{L1} + p_{\mathrm{miss}} + 4.8\log(|I_{L1}|). \tag{A.2}$$

Events are classified as purely hadronic (class 1) if the discriminant is lower than 13, and are set in the fourth class otherwise. The cut values for the class separation are chosen in order to minimise the class cross-contamination, being 27%, 21%, 13% and 46% for the four classes, respectively. The corresponding numbers of events are presented in Table A.3.

Common precuts are applied to those classes in order to strongly reduce the $\gamma\gamma$ and l^+l^- backgrounds. The energy fraction in twelve degrees around the beam pipe E_{12} must be lower than 40% of the center-of-mass energy. There must be more than 3 charged particle tracks ($N_{ch} > 3$) or the acolinearity θ_{aco} must be lower than 170°. The event total invariant mass M_{tot} and transverse momentum P_t must satisfy

Table A.2: Final state topologies and branching ratios corresponding to the four classes. The decay of the Z is given in parenthesis, followed by the decays of the hard and the soft W. Branching ratios are given with and without considering the taus.

Class	Channels		BR without (with) taus	
1: Fully hadronic	(Q)Qq	(1a)	0.328	
	(ν)Qq	(1b)	0.094	
			0.422	(0.422)
2: >1 Hard lepton	(L)Qq	(2a)	0.032	
	(L)Ql	(2b)	0.010	
	(L)Lq,(L)τq	(2c),(2T)	0.010	
	(L)Ll	(2d)	0.003	
			0.054	(0.100)
3: 1 Hard lepton	(Q)Lq	(3a)	0.101	
	(Q)Ll	(3b)	0.031	
	(ν)Ll	(3c)	0.029	
	(ν)Lq	(3d)	0.008	
			0.171	(0.284)
4: 1 Soft lepton	(Q)Ql	(4a)	0.101	
	(ν)Ql	(4b)	0.029	
			0.130	(0.195)
			Total: 0.777	(1.000)

Table A.3: Number of candidates in each class for each signal topologies. "Classes" result of the presented separation while "signals" correspond to the ideal separation based on the true (generator level) leptons. Numbers are computed for a Higgs mass of 110 GeV/c^2 and BR=1 produced in year 2000.

	class 1	class 2	class 3	class 4
signal 1	**12.6**	0	0.1	1.3
signal 2	0.3	**1.6**	0.5	0.4
signal 3	1.8	0.4	**5.9**	2.0
signal 4	2.4	0.1	0.2	**4.1**

$M_{tot}/\sqrt{s} > 0.2 - 6P_t/\sqrt{s}$. Since it accounts for the largest branching fraction, events in the purely hadronic class are required to have a more than twelve charged particle tracks and $M_{tot}/\sqrt{s} > 0.35$ while the normalised longitudinal momentum $P_l/\sqrt{s}$ must be lower than 0.2.

A.3 Selection of events with two hard leptons

In this dissertation, we will discuss the case of class 2 which is characterised by final state with at least two "hard" leptons. Events in this class are further classified in the five topologies shown in Table A.2. Requiring the third most energetic lepton (e,μ) to have an energy lower than 8 GeV separates the two lepton channel from the topologies involving more than two leptons. The separation between HZ→(L)Ql and HZ→(L)Lq, (L)Ll channels is done by requiring the total hadronic energy (E_{had}) to be higher than 60 GeV. HZ→(L)Ll is then uniquely distinguished from HZ→(L)Lq since it contains strictly four tracks of charged particles in the final state. Within the two leptons channel, one finds events with taus decaying hadronically. The separation of these events from those where all W's decay hadronically is achieved with a cut on the total hadronic event energy.

The mass reconstruction of the Higgs boson is based on the identification of the leptons coming from the Z boson. Therefore out of the three most energetic leptons, the pair of leptons with the same flavour and with a good estimation of the Z boson mass is selected. The Higgs mass is then computed using the formula:

$$M_H = \sqrt{(E_{tot} - E_{L1} - E_{L2})^2 - (\vec{P}_{tot} - \vec{P}_{L1} - \vec{P}_{L2})^2} \ , \tag{A.3}$$

where E_{L1} ($\vec{P}_{L1}$) and E_{L2} ($\vec{P}_{L2}$) are the energies (momentum) of the leptons associated with the selected pair. The variable M_H is used as a discriminant variable for the channel HZ→(L)Qq (class2a).

For the other subclasses involving more than two hard leptons, the Higgs mass resolution is degraded by the presence of one or more neutrinos. The following equation

is therefore used to evaluate the Higgs boson mass:

$$M_H^2 = (P_{tot}^\mu - P_Z^\mu)(P_{tot\mu} - P_{Z\mu}) + 2(P_{tot}^\mu - P_Z^\mu)P'_\mu \ , \tag{A.4}$$

where P_{tot}^μ is the total energy-momentum four-vector. P_Z^μ and P'_μ are the four-momentum of the Z boson and the neutrino. The first term of equation (A.4) corresponds to the equation (A.3), and the additional term introduces the correction needed to take the undetected neutrino into account and assumed to be emitted in the opposite direction of the electron. As before, P_Z^μ is reconstructed with the selected pair and $P'_\mu = (E_{L3}, -\vec{P}_{L3})$.

No mass estimator is used for the decay channel involving more than one neutrino. In the channel involving more than 3 leptons, the confidence level is computed in one bin, so it reduces to a counting experiment.

A.3.1 Class 2a: Events with hadronic W's decay

To select the HZ→(L)Qq channel - that shares 58 % of class 2 branching's ratio - one requires that the energy of the third lepton is less than 8 GeV and that the total hadronic energy (E_{had}) is more than 60 GeV. In addition to the preselection described in section A.2.1, the number of charged tracks is required to be greater than 8 in order to reject the $\gamma\gamma$ background.

The dominating background event after preselection cuts is $q\bar{q}$, semi-leptonic WW as well as ZZ events where one of the gauge boson decays into hadrons and the other in leptons. The rejection of $q\bar{q}$ and WW events is achieved by applying a cut on the variable y_{45}[1] and requiring the reconstructed mass of the two leptons to be in a 14 GeV window around the Z mass.

The remaining ZZ background is reduced by requiring the two leptons to be back to back and the Z energy to be less than 95 GeV. The details of the preselection and selection are shown in Table A.4.

After the final selection 0.67 background event and 0.54 signal event are expected. Two candidates are observed in the data. The expected confidence level $\langle CL_s \rangle$ is 0.57, making this sub-channel the one that will contribute the most to the expected limit for class 2. The reconstructed Higgs mass M_H (of equation A.3) is shown in Fig. A.2a and b before and after the full selection, respectively.

A.3.2 Class 2T: Leptonic Z and decay of the W in tau leptons

Events with two leptons from the Z boson and a hard tau decaying hadronically into one- or three-prong are grouped in the HZ → (L)τq channel. This is the only case where the one- or three-prong hadronic tau decay can be distinguished efficiently from the hadronic W boson decay. Events are required to have a third lepton (e or μ) with less than 8 GeV and a total hadronic energy of less than 60 GeV. This selection is

[1]The variable $y_{n(n+1)}$ is the transition value of the y_{cut} at which the event changes from a clustering with $n+1$ jets to a clustering with n jets.

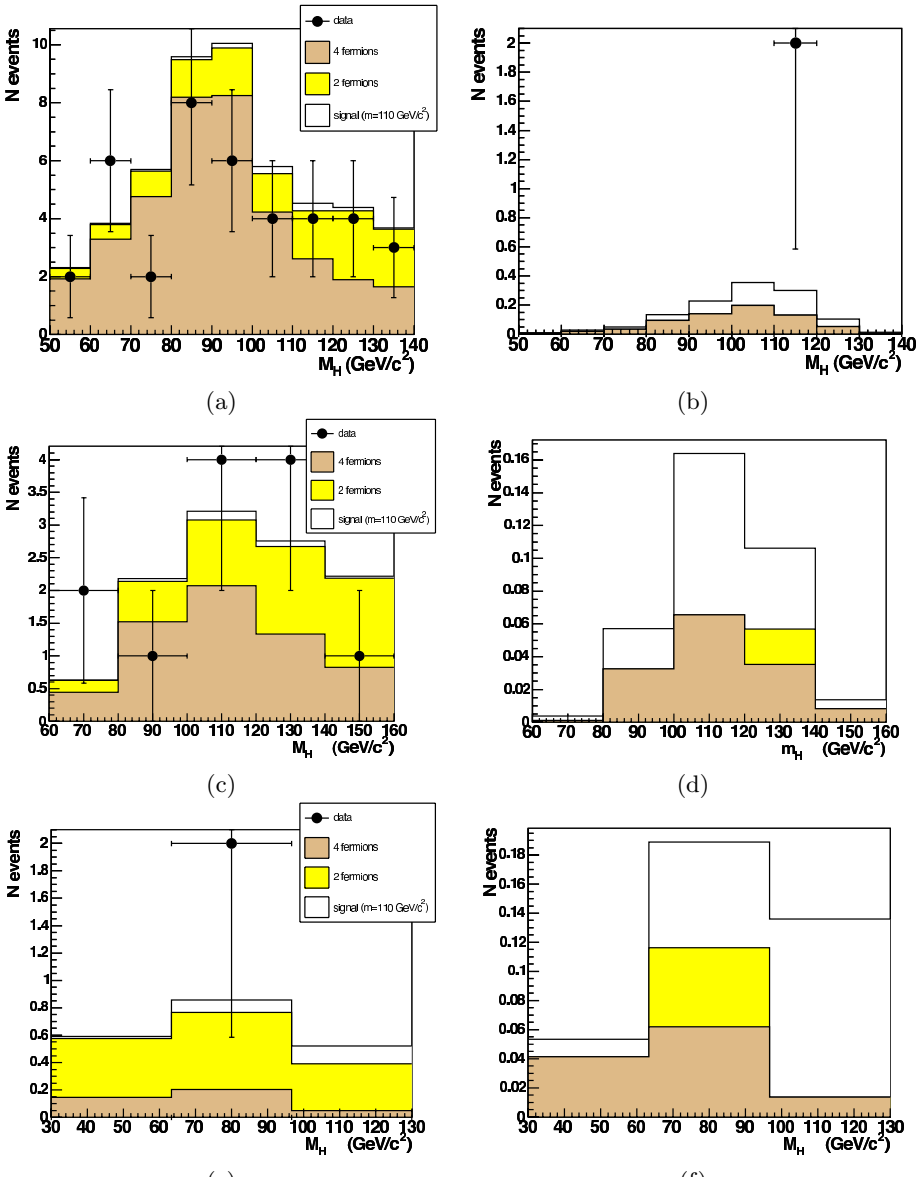


Figure A.2: Distributions of the discriminant variable M_H (equation A.3) for the decay HZ→(L)Qq plotted for the class 2a events after the preselection (a) and after the final selection cuts (b). In the plots (c) and (d), the distributions of the discriminant variable M_H (equation A.4) are shown before and after the final selection cuts for events in class 2b (HZ→(L)Ql). The two lower plots (e) and (f) show the same variable M_H, before and after the final selection cuts respectively, for events in class 2c (HZ → (L)Lq).

Table A.4: Selection criteria for the two leptons channel HZ → (L)Qq (class 2a). The numbers of signal (N_s), background (N_b) and data (N_d) events, as well as the value of the expected confidence level $\langle CL_s \rangle$ are given in the table. Energies, momenta and masses are expressed in GeV, GeV/c and GeV/c^2, respectively. Angles are in degrees.

	Cuts	N_s	N_b	N_d	$\langle CL_s \rangle$
Class 2	$E_{L1} > 25$ $E_{L2} > 20$				
Class 2a	$E_{L3} < 8$ $E_{had} > 60$				
Preselections	$E_{12}/\sqrt{s} < 0.4$ $M_{tot}/\sqrt{s} > (0.2 - 6 * P_t/\sqrt{s})$ $N_{ch} > 8$	1.03	48.8	39	0.85
Group 1 (anti WW,$q\bar{q}$)	$\lvert M_{Z_{rec}} - M_{Z_0} \rvert < 14$ $\ln(y_{45}) \geq -7$	0.62	6.29	2	0.69
Group 2 (anti ZZ)	$E_{L1} + E_{L2} < 95$ $\theta_{L_1 L_2} > 135$	0.54	0.67	2	0.57

introduced in order to differentiate the HZ → (L)Qq channel from the one involving W tau decays where the tau lepton decays hadronically. This is done by requiring that the total hadronic energy is below 60 GeV in order to select hadronic tau decays. The complete set of selection cuts are shown in Table A.5, the transverse momentum of the Z boson (P_{t_Z}) allows to further reject ZZ background events.

For the class 2T after all cuts, the expected value of the signal confidence level is 0.94. The observed value for the signal confidence level is 0.97.

A.3.3 Class 2b: Leptonic off-shell W and hadronic on-shell W

The HZ → (L)Ql channel is required to have a third lepton with more than 8 GeV and a total hadronic energy of more than 60 GeV. The selection proceeds in a similar way as in the HZ → (L)Qq channel by reconstructing the Z peak. The logarithm of the isolation for the lepton that is the most antiparallel to the missing momentum is used in the selection. That variable, $\ln(I_{L_A})$, is sensitive to the isolation of the soft lepton emitted by the off shell W. The lepton tends to be more isolated for the signal than for the background.

After the final selection the dominating background is ZZ events going into $l^+l^-\bar{b}b$. We expect 0.18 signal event and 0.16 background event. No event remains in the data after the final selection. Distributions of the reconstructed Higgs mass M_H (equation A.4) are shown in Fig. A.2c and d before and after the full selection, respectively. For the class 2b after all cuts, the expected value of the signal confidence level is 0.83. The observed value for the signal confidence level is 0.83.

Table A.5: Selection criteria for channel HZ → (L)τq (class 2T). The numbers of signal (N_s), background (N_b) and data (N_d) events, as well as the value of the expected confidence level $\langle CL_s \rangle$ are given in the table. Energies, momenta and masses are expressed in GeV, GeV/c and GeV/c^2, respectively. Angles are in degrees.

	Cuts	N_s	N_b	N_d	$\langle CL_s \rangle$
Class 2	$E_{L1} > 25$, $E_{L2} > 20$				
Class 2T	$E_{L3} < 8$ $E_{had} < 60$				
Preselection	$E_{12}/\sqrt{s} < 0.4$ $M_{tot}/\sqrt{s} > (0.2 - 6 * P_t/\sqrt{s})$ $P_t/\sqrt{s} > 0.002$	0.08	2.59	2	0.96
Selection	$P_{t_Z} < 60$ $\lvert M_{Z_{rec}} - M_{Z_0} \rvert < 23$	0.06	0.46	1	0.94

Table A.6: Selection criteria for the three lepton channel HZ → (L)Ql (class 2b). The numbers of signal (N_s), background (N_b) and data (N_d) events, as well as the value of the expected confidence level $\langle CL_s \rangle$ are given in the table. Energies, momenta and masses are expressed in GeV, GeV/c and GeV/c^2, respectively. Angles are in degrees.

	Cuts	N_s	N_b	N_d	$\langle CL_s \rangle$
Class 2	$E_{L1} > 25$ $E_{L2} > 20$				
Class 2b	$E_{L3} > 8$ $E_{had} > 60$				
Preselections	$E_{12}/\sqrt{s} < 0.4$ $M_{tot}/\sqrt{s} > (0.2 - 6 * P_t/\sqrt{s})$ $N_{ch} > 7$	0.29	10.7	12	0.92
anti $q\bar{q}$, WW	$\lvert M_{Z_{rec}} - M_{Z_0} \rvert < 20$ $\ln(y_{45}) > -8$ $\ln(I_{L_A}) > -7$	0.19	1.08	0	0.84
anti ZZ	$E_{L1} + E_{L2} < 98$ $\theta_{L_1 L_2} > 142$	0.18	0.16	0	0.83

Table A.7: Selection criteria for the tree lepton channel HZ → (L)Lq (class 2c). The numbers of signal (N_s), background (N_b) and data (N_d) events, as well as the value of the expected confidence level $\langle CL_s \rangle$ are given in the table. Energies, momenta and masses are expressed in GeV, GeV/c and GeV/c^2, respectively. Angles are in degrees.

	Cuts	N_s	N_b	N_d	$\langle CL_s \rangle$
Class 2	$E_{L1} > 25$ $E_{L2} > 20$				
Class 2c	$E_{L3} > 8$ $E_{had} < 60$				
Preselection	$E_{12}/\sqrt{s} < 0.4$ $M_{tot}/\sqrt{s} > (0.2 - 6 * P_t/\sqrt{s})$ $P_t/\sqrt{s} > 0.01$ $N_{ch} > 4$	0.23	1.73	2	0.82
Group1 (anti $q\bar{q}$, WW)	$\lvert M_{Z_{rec}} - M_{Z_0} \rvert < 23$ $\ln(I_{L_A}) > -11$ $\ln(y_{45}) > -11$	0.20	0.18	0	0.81
Group 2 (anti ZZ)	$P_{t_Z} < 60$	0.20	0.17	0	0.81

A.3.4 Class 2c: hadronic off-shell W and leptonic on-shell W

The HZ → (L)Lq channel is required to have a third lepton with more than 8 GeV and a total hadronic energy of less than 60 GeV .

Compared to the channel HZ → (L)Ql, the missing transverse momentum is higher. The preselection shown in Table A.7 rejects most of the background by requiring a total transverse momentum greater than 1% of the total center-of-mass energy. For this reason, the cuts that typically reduce the semi-leptonic WW background (group 1 in Table A.7) do not improve much the background rejection. The mass window for the reconstructed Z mass is also broader than in the HZ → (L)Ql case since the third lepton is sometime misidentified as one of the leptons coming for the Z. Finally a cut on the P_t of the Z boson is applied to further remove the ZZ background.

After the final selection, the expected confidence level $\langle CL_s \rangle$ is 0.81. No event is observed in the data while 0.20 signal event and 0.17 background event are expected. The reconstructed Higgs mass M_H (of equation A.4) is shown in Fig. A.2e and f before and after the full selection, respectively. The dominating background is due to ZZ events.

A.3.5 Class 2d: Fully leptonic

The HZ → (L)Ll channel is characterised by a small branching fraction (5% of the events in class 2) but with a clear topology: 4 leptons in the final state, one of them being soft. This channel is then selected by cutting on the energy of the third most energetic lepton ($E_{L3} > 8$ GeV) and on the total hadronic energy. One also requires strictly 4 charged tracks. To further reduce the l^+l^- a cut on the missing transverse momentum and on the acoplanarity, and on the thrust (T) is applied. Finally the

Table A.8: Selection criteria for the four lepton channel HZ → (L)Ll (class 2d). The numbers of signal (N_s), background (N_b) and data (N_d) events, as well as the value of the expected confidence level $\langle \mathrm{CL_s} \rangle$ are given in the table. Energies, momenta and masses are expressed in GeV, GeV/c and GeV/c^2, respectively. Angles are in degrees.

	Cuts	N_s	N_b	N_d	$\langle \mathrm{CL_s} \rangle$
Class 2	$E_{L1} > 25$ $E_{L2} > 20$				
Class 2b	$E_{L3} > 8$ $E_{had} < 60$ $N_{ch} = 4$				
Preselection	$E_{12}/\sqrt{s} < 0.4$ $M_{tot}/\sqrt{s} > (0.2 - 6 * P_t/\sqrt{s})$ $P_t/\sqrt{s} > 0.11$	0.10	3.13	3	0.96
Group 1 (anti ZZ)	$T < 0.98$ $\Phi_{\mathrm{aco}} < 176$	0.09	0.93	1	0.95
Group 2 (anti WW)	$\ln(I_{L_A}) > -9$	0.09	0.58	0	0.90

remaining WW events are rejected by requiring the most anti-parallel lepton with respect to the missing momentum to be well isolated. The selection criteria are detailed in Table A.8.

After the final selection one expects 0.09 signal event and 0.58 background event. No candidate is observed in the data.

A.4 Conclusions

Finally the selections are applied on the data taken in 1999. Table A.9 gives The numbers of signal, background and data events for the years 1999 and 2000,are given for all five channels. The expected number of signal events is about 1.24 while we expect 5.78 background events. In the data, we observe 9 candidates corresponding to a 1.2σ excess over the background only expectation.

A.4.1 Cross section upper limits

An upper limit on the cross section production at a given invariant mass is derived. Fig. A.3 presents the resulting 95% C.L. upper limit on $\xi^2 = B(H \to WW)\sigma(e^+e^- \to Hff)/\sigma^{SM}(e^+e^- \to Hff)$, as a function of the Higgs mass. The sensitivity of this channel is too small compared to the three other classes to give a limit lower than one for ξ^2. It is only by combining the four other channels that one can put a limit on the ξ^2 as described in [2].

Table A.9: Numbers of signal (N_s), background (N_b) and data (N_d) events, as well as the value of the expected and observed confidence level for class 2 selections.

Year	Class	N_s	N_b	N_d	$\langle \mathrm{CL_s} \rangle$	$\mathrm{CL_s}$
2000	2a	0.54	0.67	2	0.57	0.90
2000	2T	0.06	0.46	1	0.94	0.97
2000	2b	0.18	0.16	0	0.83	0.83
2000	2c	0.20	0.17	0	0.81	0.81
2000	2d	0.09	0.58	0	0.90	0.90
1999	2a	0.084	1.73	3	0.93	0.97
1999	2T	0.009	0.74	0	0.99	0.99
1999	2b	0.03	0.40	2	0.97	0.99
1999	2c	0.03	0.13	0	0.96	0.96
1999	2d	0.02	0.74	1	0.99	0.99
All combined		1.24	5.78	9	0.43	0.73

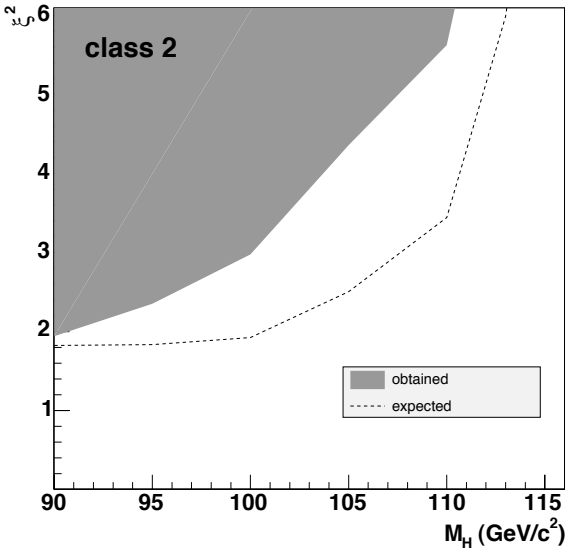


Figure A.3: Limit on ξ^2 (defined in the text) as a function of the Higgs boson mass. The dotted line corresponds to the expected limit while the 95% C.L. excluded region is shown by a grey area.

APPENDIX

B

Computing Farm

The analysis presented in Chap 4 required both hardware and software tools. On the hardware side a computing cluster of 24 nodes has been deployed in Louvain-la-Neuve. It's purpose is the generation of the events for the present analysis as well as the simulation of the CMS detector response. The cluster and its infrastructure will be outlined in the next sections.

B.1 Hardware setup

The chosen architecture as been based on existing cluster like the ones found at CERN and is separated in four elements:

- 24 Computing nodes: 12 dual Intel PIII-S 1.4GHz, 512 KB of L2 cache, 512MB RAM; 12 dual Xeon 2.4GHz, 512 KB of L2 cache, 1GB ECC RAM.
- One Ethernet Switch: 24 ports 10/100Base-TX and 2 copper Gigabit uplinks. Switch fabric speed of 9.6 Gb.
- 1 Workstation for software repository: Dual Intel Xeon 2GHz, 512 KB L2 cache, 512MB RAM.
- 1 Mass Storage: 2 RAID10 arrays of 630GB each and 2 RAID1 arrays of 157 GB each.

The 24 computing nodes are connected to the mass storage and the workstation using gigabit links. Gigabit links are necessary to sustain data throughput above 100 Mb/s during Monte-Carlo generation and digitisation with pileup. These high data rates required the use of a private network to guarantee that the commodity network of the institute is not disturbed during the event generation. In the private

network, the workstation is used as an IP router to guarantee network connectivity between the cluster nodes and the client machines submitting production jobs. In addition all necessary software's for the simulation of CMS are stored on this workstation. It is centrally managed and any user machine in the laboratory that need to run CMS software has only to mount one directory to start working. The whole cluster architecture is summarized on Fig. B.1 and a picture of the cluster can be seen on Fig. B.2.

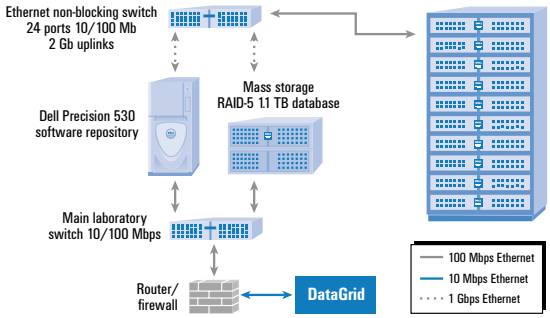


Figure B.1: Overview of the cluster architecture, from [62]. The 24 machines cluster rack is shown on the right. On the top we can see the ethernet switch with its two links to the workstation and the Mass storage.

B.2 Software setup

Several softwares are necessary to build the complete chain going from the event generation to the event simulation, reconstruction and, finally, the analysis. The Table B.1 shows a list of those software separated in two categories. The first category concerns all necessary services while the second lists the actual software developed by the CMS collaboration in order to simulate and reconstruct events.

In the reconstruction and simulation category ORCA plays a central role with two other actors, OSCAR/CMSIM[1] and CMKIN. CMKIN is a Pythia application for the Monte Carlo tool used for the generation of analysed events. The Pythia generated events are then fed to OSCAR/CMSIM that simulates the energy deposited in the detector matter by the particles involved in the reaction. For each sensitive volume the energy is computed and stored. The stored information is called Hits. Their computation is by far the most time consuming operation per event, as detailed in the next section. Once the Hits containing the information about the energy deposited in the sensitive volumes are stored, the ORCA software will simulate the response of whole electronic read-out chain. The end product is the electronic footprint of what each detector would have recorded in real data taking condition.

[1]OSCAR is the GEANT 4 C++ version of the CMSIM GEANT 3 detector simulation.

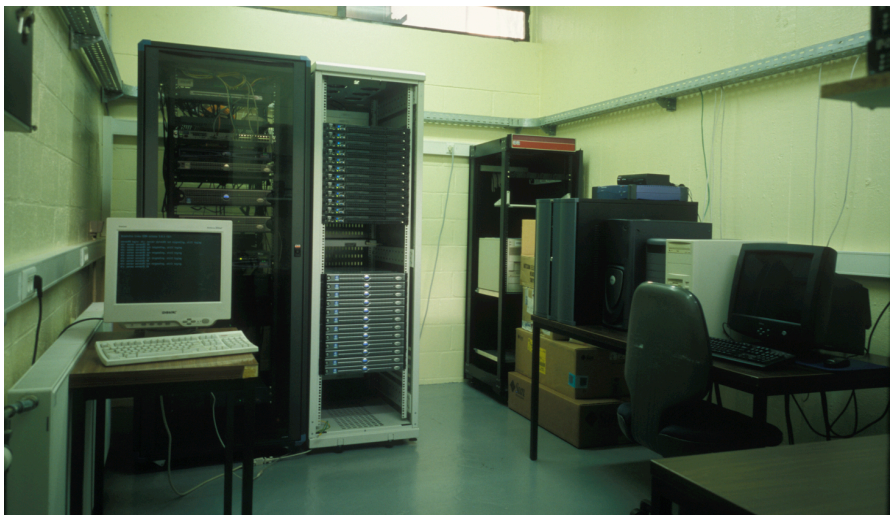


Figure B.2: Picture of the computing farm room. The central rack contains the 24 machines, with the twelve PIII on the top and the twelve dual Xeon on the bottom. On the right we can see the mass storage as well as the software repository.

From this point on, all software involved in further steps will be called reconstruction software. The goal of the reconstruction software is to deduce what event occurred in the collision, it proceeds by computing the energy deposits from the electronic response to obtain a list of particles and their momenta. that where present in the collision products. An important point is that the reconstruction software is designed in such a way that it is not aware of the real or simulated nature of the analysed events. In other words while the event reconstruction is currently only involved with simulated events this will not be the case at the LHC startup. Ideally one would plug an exact copy of the reconstruction software in the Filter Farm connected to the CMS central DAQ to analyse real data.

On the software service side, CONDOR [63] is used as the scheduler for the execution of CMS simulation and reconstruction jobs. Both machines in the cluster and desktop machines of the laboratory are used for CMS simulation. More than 216E3 CPU hours have been accumulated by the cluster to this date (2004), it includes computations of vortex dynamics in supra-conductors [64] and numerical integration of one-loop Feynman diagrams. The monitoring of the farm is achieved by SNMP and the gathered data can be visualised on the web with CACTI. CACTI allows to see the status of each machine through a web browser to ease the tracking of problems, even the machine temperature is monitored. For the software developments in the laboratory, a central CVS server has been installed. It is used as a storage for all analysis software,scripts, documentation, paper writings etc. In addition the content of the repository can be viewed on the web. The list on Table B.1 is not exhaustive since several script suites have been developed to ease the event generation and simulation in the laboratory, and the details of the analysis software are not given.

Service/Software	Description
RedHat 7.3	Linux Operating system.
Condor v6.6.1	Job Submission system implementing fair share scheduling.
SNMP	Simple Network Management Protocol for the monitoring of cluster nodes.
CVS	Concurrent Version System for software developments in the laboratory.
VIEWCVS	Web viewing tool of the CVS repository.
CACTI 0.8.2a	Web monitoring tool of the cluster and network parameters (cpu usage, network bandwidth, temperature, disk usage).
GCC 3.2.3	C/C++ ISO Compliant Compiler.
SCRAM	Build and configuration tool used for all CMS projects.
COBRA	Core software for object persistency and event dispatching framework.
ORCA	Object Oriented Program for CMS event reconstruction and simulation.
IGUANA	Tool to visualize the simulated and reconstructed events.
CMSIM	GEANT 3 simulation with the CMS detector geometry.
OSCAR	Object oriented Simulation for CMS Analysis and Reconstruction (GEANT 4).
CMKIN	Pythia application for Monte-Carlo event generation of SUSY and Standard model Physics.

Table B.1: Non exhaustive list of the softwares that are part of the cluster architecture.

B.3 Cluster performance

The performance of the cluster for the simulation of CMS is detailed in the Table B.2. The timing was done with CMSIM and ORCA v7.2.2, events where produced without pile-up. Clearly the most time consuming step is the GEANT simulation of the detector, the Monte-Carlo event generation is not included in the table since it is negligible compared to hit and digis computation time.

Simulation Step	timing $\left[\frac{\text{sec}\times\text{KMips}}{\text{evts}}\right]$	Event size $\left[\frac{\text{kB}}{\text{evts}}\right]$
Geant 3 simulation (hits)	120	370
Detector simulation (digis)	20	360
Total	140	730

Table B.2: Timing of the GEANT and ORCA simulation steps.

The power of the computing cluster is shown on Fig. B.3. The installation of the cluster was done in two steps. In 2002 we installed twelve PIII machines named cluster02. End of 2003, twelve Xeon machines where added, they are named cluster03. The CPU power of each cluster is detailed in the figure. In addition we see the computing power available from the users desktop machines (FYNU). The cluster totals 140 kMips which translates to $\sim$ 3800 events produced in an hour. The event rate when digitising with pileup at high luminosity($10^{34}\text{cm}^{-2}\text{s}^{-1}$) is reduced by a factor three.

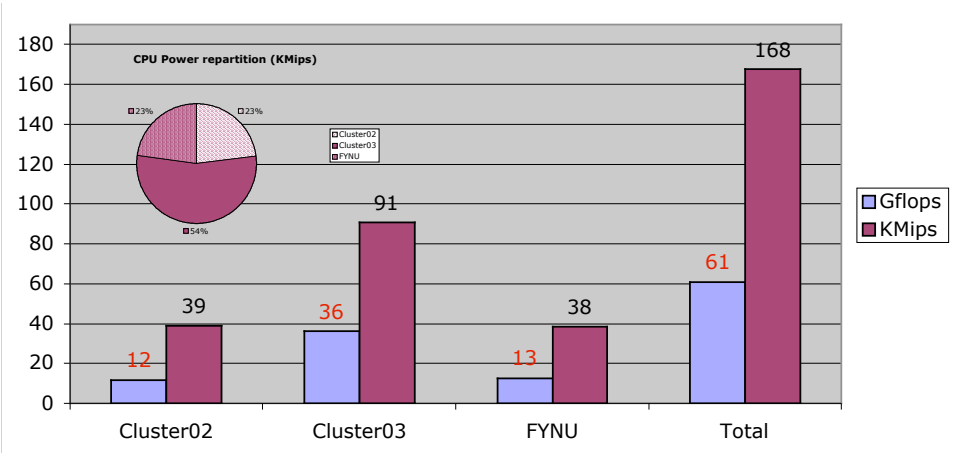


Figure B.3: Cluster CPU power in Kmips and Gflops taken from CONDOR. Cluster02 contains the twelve PIII machines. Cluster03 contains the twelve Xeon machines and FYNU stands for the desktop machines included in the CONDOR environement.

APPENDIX

C

Trigger tables

C.1 Timing trigger table at low luminosity

Following is the HLT trigger table specification used for the timing of the HLT shown described in Low luminosity is 2×10^{33} cm$^{-2}s^{-1}$

sub-trigger	parameters (threshold in GeV)
single electron	$E_T > 26$
double electron	$E_t^1 > 14.5 \wedge E_T^2 > 14.5$
single photon	$E_T > 80$
double photon	$E_T^1 > 20 \wedge E_T^2 > 35$
single muon	$E_T > 19$
double muon	$E_t^1 > 7 \wedge E_T^2 > 7$
J/Ψ trigger	
single jet	$E_T > 572$
3 jets	$E_T^i > 195,\ i = (1 \cdots 3)$
4 jets	$E_T^i > 80,\ i = (1 \cdots 4)$
jet-met	$E_T^{jet} > 180 \wedge E_T^{met} > 123$

Table C.1: Summary of the selection criteria at low luminosity

Fig. C.1 and Fig. C.2 shows the detail of the selection criteria summarised in Tab. C.1

C.2 High Level Trigger table at $L = 2 \times 10^{33}\mathrm{cm}^{-2}\mathrm{s}^{-1}$

Fig C.3 and C.4 is the trigger table used for the analysis presented in Chap 4.

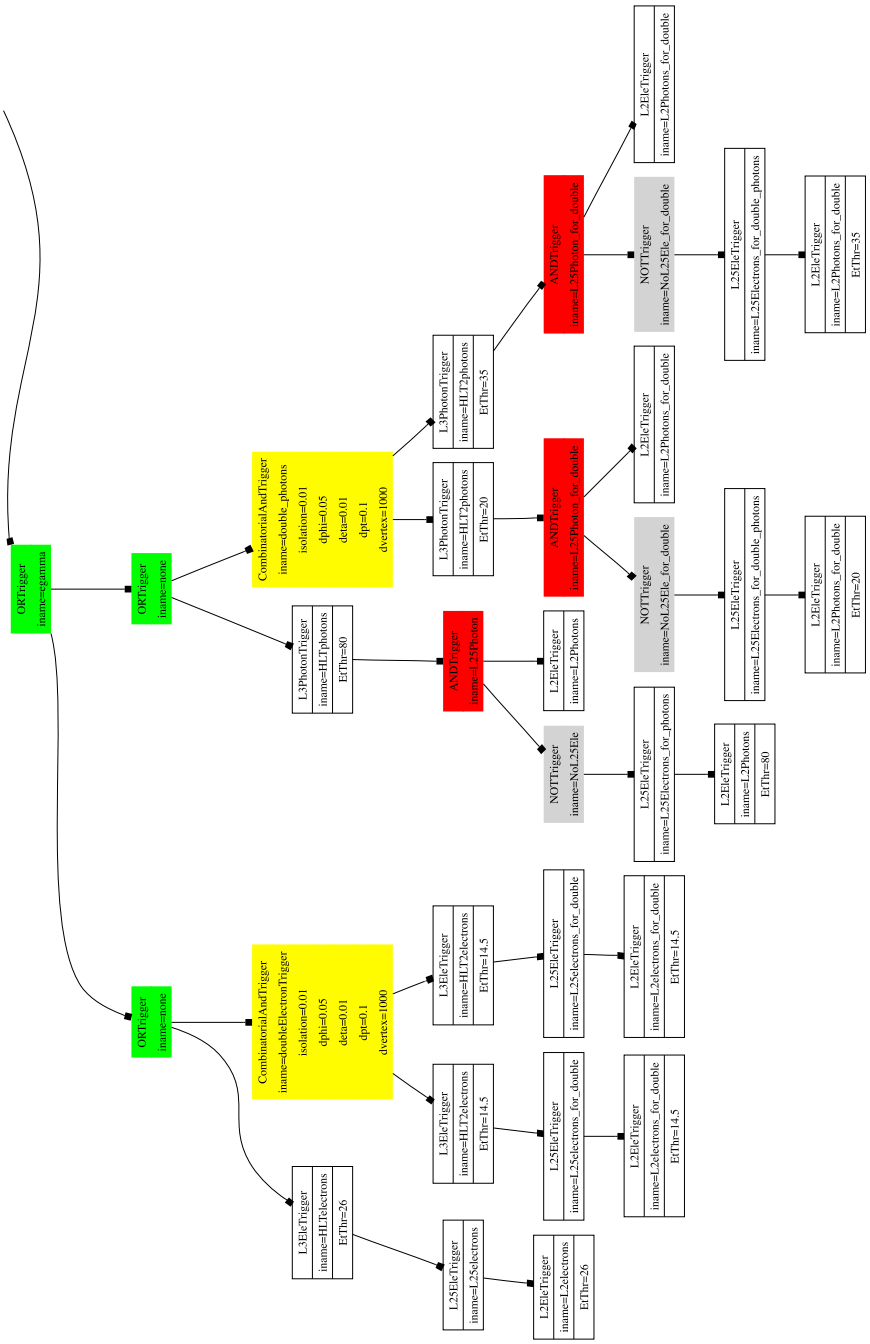


Figure C.1: Representation of the trigger tree used for the timing of the HLT. Only the left part of the tree is shown, the other part is shown in Fig. C.2

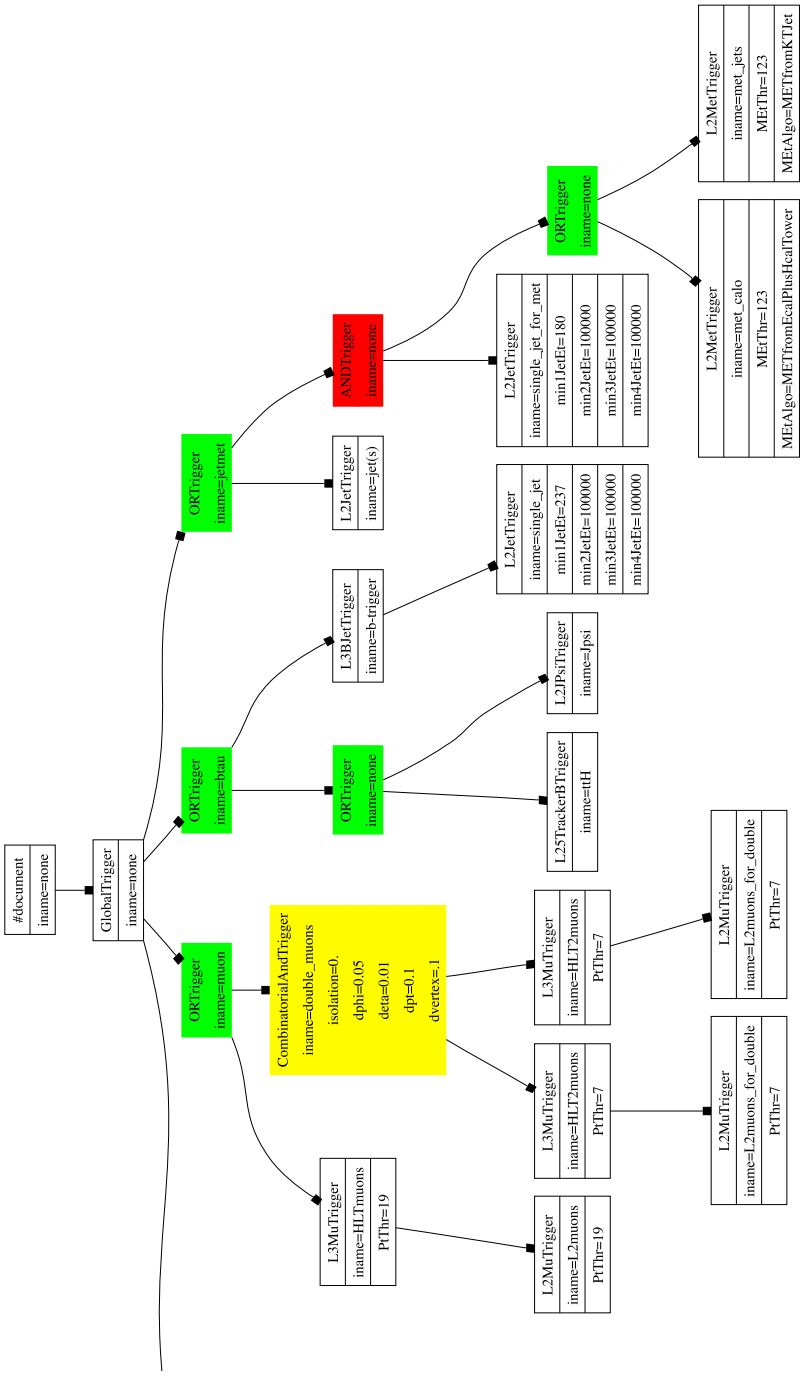


Figure C.2: Representation of the trigger tree used for the timing of the HLT. Only the right part of the tree is shown, the other part is sketched in Fig. C.1

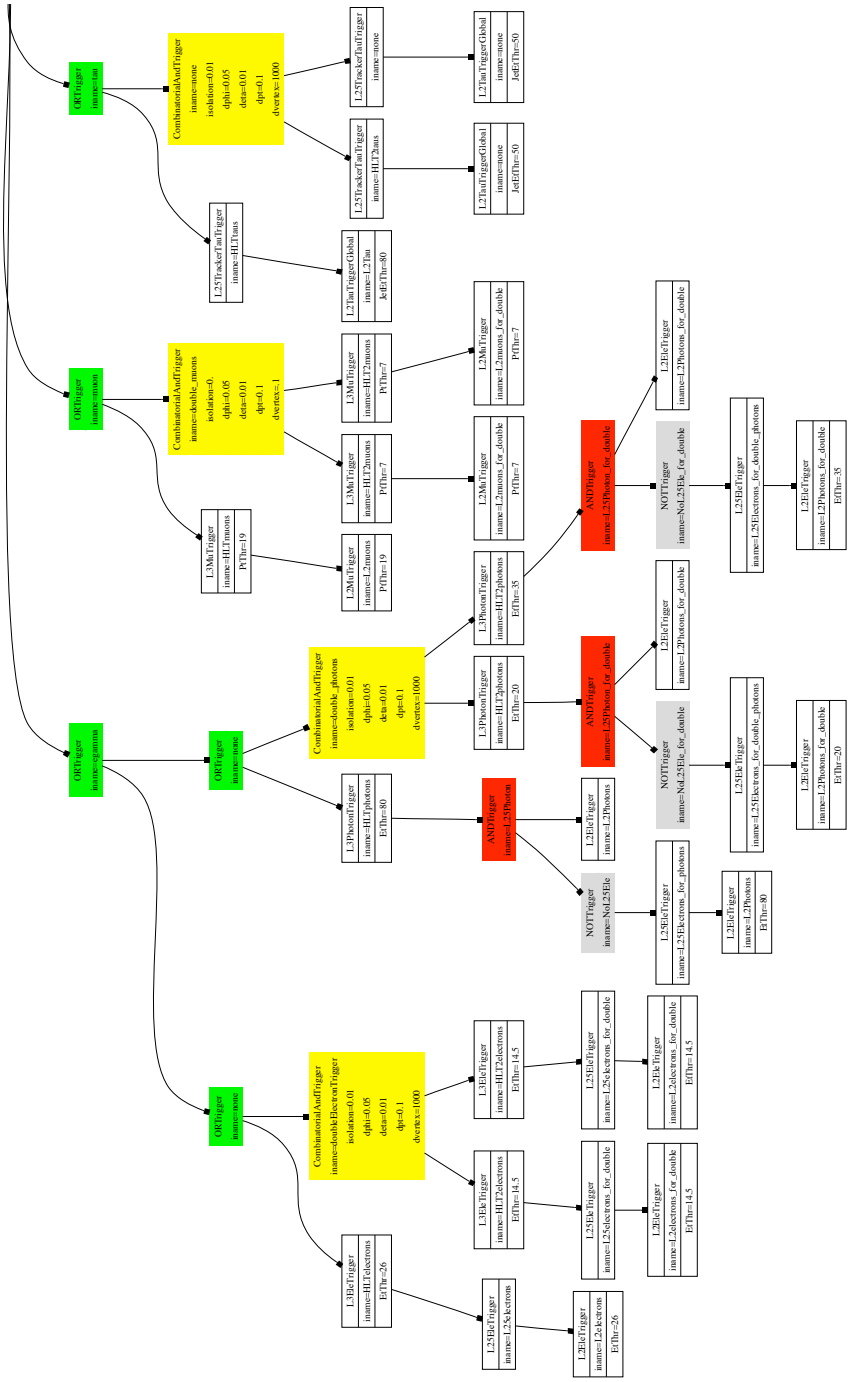


Figure C.3: Representation of the left part of the trigger tree used for the High Level Trigger analysis.In each box the iname is the instance name of the algorithm. The first name to appear in the box correspond to the class name and the parameters below those two are the specified parameters for the selection.

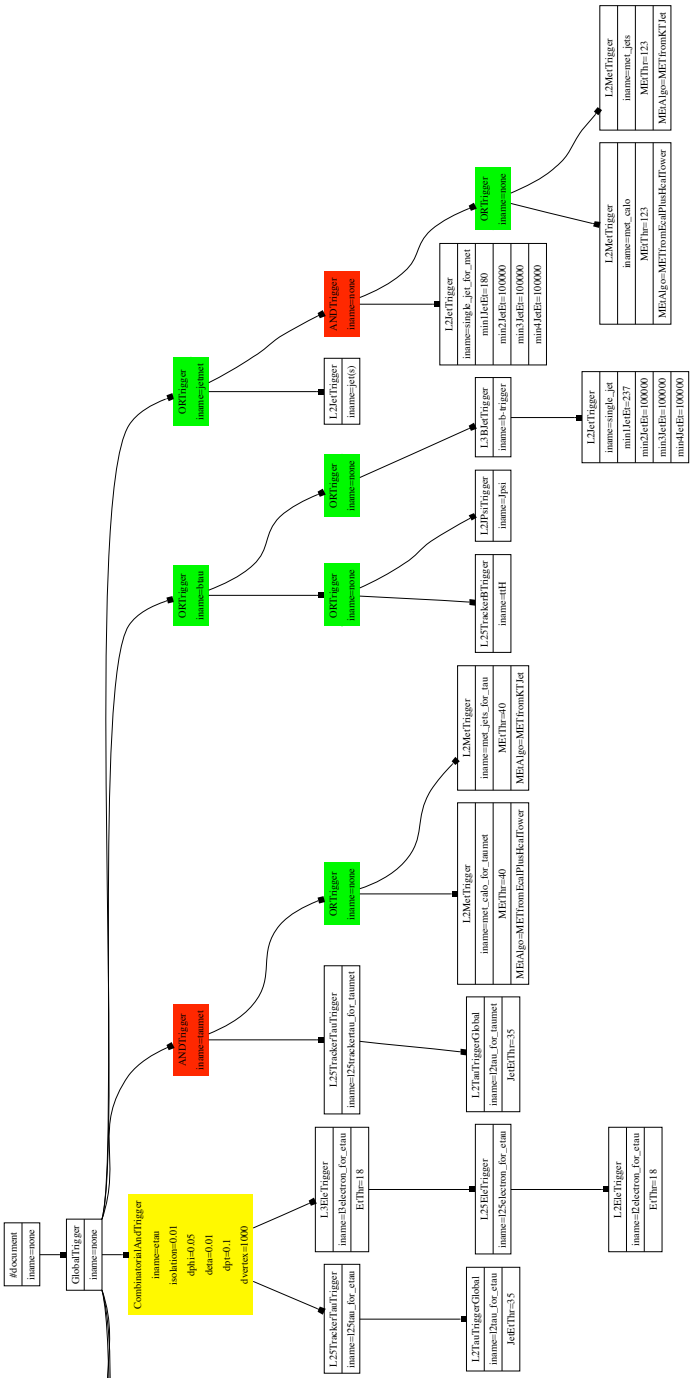


Figure C.4: Representation of the right part of the trigger tree used for the High Level Trigger analysis. In each box the iname is the instance name of the algorithm. The first name to appear in the box correspond to the class name and the parameters below those two are the specified parameters for the selection.

BIBLIOGRAPHY

[1] CMS Collaboration, CERN/LHCC 2002-26, *"The Trigger and Data Acquisition project, Volume II.Data Acquisition & High-Level Trigger Technical Design Report."*

[2] B. Clerbaux, C. Delaere, V. Lemaitre, O. van der Aa, "Search for WW decay of a Higgs boson produced in association with a fermion pair in e^+e^- collisions at LEP", ALEPH/2003-018 ALEPH-Physics/003-006 June 15, 2004

[3] R. Brout and F. Englert, Phys. Rev. Lett. 13 (1964) 321; P. Higgs, Phys. Rev. Lett. 13 (1964) 508; G. Guralnik, C. Hagen and T. Kibble, Phys. Rev. Lett. 13 (1964) 585.

[4] S. Martin, hep-ph/9709356, *"A Supersymmetry Primer"*.

[5] Y. Okada, KEK-TH-508, hep-ph/9701259, *"Phenomenolgy of the Higgs sector in superymmetric standard model"*

[6] M. Carena, S. Heinemeyer, C.E.M Wagner, G.Weiglein, hep-ph/9912223, *"Suggestions for Improved Benchmark scenarios for Higgs-Boson Searches at LEP2"*

[7] ALEPH, DELPH, L3 and OPAL Collaboartion, and the LEP Higgs Working Group, hep-ex/0107030, *"Searches for the Neutral Higgs Bosons of theMSSM: Preliminary Combined Results Using LEP Data Collected at Energies up to 209 GeV"*

[8] Pedro A. Movilla Fernandez, hep-ex/0409001, *"Top quark Mass Measurements at CDF"*.

[9] S. Heinemeyer, S. Kraml, W. Porod, G. Weiglein, hep-ph/0409063 *"Impact of a precise top mass measurement."*

[10] M. L. Perl et al.,Phys. Rev. Lett. 35 (1975) 1489, *"Evidence for Anomalous Lepton Production in e^+e^- Annihilation."*

[11] S. Eidelman, et al., Physics Letters B 592 (2004) 1, *"Particle Physics Booklet."*

[12] Danek Kotlinski, Nuclear instruments and methods in phyiscs research, section A 477 (2002) 446-450, *"The design of the CMS pixel detector system."*

[13] S. Cucciarelli, M. Konecki, D. Kotlinski, T.Todorov, CMS Note 2003-026 *"Track Parameter Evaluation and Primary Vertex Finding with the Pixel Detector."*

[14] CMS Collaboration, CERN/LHCC 98-06, *"The Tracker Project, Technical Design Report"*

[15] CMS Collaboration,CERN/LHCC 2000-016, *"Addendum to CMS Tracker TDR"*

[16] Nucl. Instrum. and Methods. A262 (1987) 444 R. Fruhwirth, *"Application of Kalman filtering to track and vertex fitting"*

[17] CMS Collaboration, CERN/LHCC 1997-13, *"The Electromagnetic Calorimeter Project, Technical Design Report"*

[18] E. Tournefier, CMS CR 2000-010, *"The Preshower Detector of CMS at LHC"*

[19] David Kirkby, LEP L3 Technical Report 1816, 24 Juillet 1995 *"π_0 γ discrimination in L3, using shower-shape analysis"*

[20] J. Schwindling, B.M. Mansoulii, O. Couet, *"MLPfit 1.40 Neural network"*

[21] CMS Collaboration, CERN/LHCC 1997-31, *"The Hadron Caolrimeter Project, Technical Design Report."*

[22] Vasken Hagopian, CMS CR 1999-001, *"The Compact Muon Solenoid (CMS) Hadron Calorimeter."*

[23] Vasken Hagopian, Nuclear Physics B Proceedings supplements, *"The Hadron Calorimeter of the compact muon solenoid (CMS)."*

[24] CMS Collaboration, CERN/LHCC 1997-32, *"The Muon Project, Technical Design Report."*

[25] M. Benedotti et al., Nucl. Instrum. and Methods. A410 (1998) 133-147, *"Performance of the drift tubes for the barrel muon chambers of the CMS detector at LHC."*

[26] M. M. Baarmand, Yu. Bonushkin, D. Chrisman, S. Durkin, T. Ferguson, P. Giacomelli, L. Gorn, W. Gorn, J. Hauser, J. Hirschfelder et al., Nucl. Instrum. and Methods. A402 (1998) 36-52, *"Tests of cathode strip chamber prototypes."*

[27] W. Badgett, S. Dasu, M. Jaworski, J. Lackey and W. H. Smith, CMS NOTE 1999-026, *"CMS Calorimeter Level 1 Regional Trigger Electron Identification."*

[28] N. Neumeister, CHEP Interlaken 2004 Proceedings, *"Muon Reconstruction Software in CMS."*

[29] G. Bruno et al, CMS Note 2002-043, *"Local reconstruction in the muon detectors."*

[30] E. Meschi, T. Monteiro, C. Seez, P. Vikas, CMS Note 2001-034, *"Electron Reconstruction in the CMS Electromanetic Calorimeter."*

[31] G. Daskalakis, K. Lassila-Perini, CMS Note 2002-039 *"Jet rejection using the pixel matching for the low and the high luminosity"*

[32] J.E. Huth et al., in Research Directions for the Decade, Proceedings of the Summer Study on High Energy Physics, Snowmass, Colorado, 1990, p. 134.

[33] G. Bagliesi, S. Gennai and G. Sguazzoni, CMS Note 2003-018, *"A L2 Trigger for Tau Hadronic Decays with Tracker Isolation in the Low Luminosity Scenario."*

[34] C. Seez, D. Barney, CMS IN 2001-023, *"Summary of ECAL-egamma Autumn 2000 Monte-Carlo Data Production Parameters and Specifications"*

[35] V. Innocente, L. Silvestris, D. Stickland, Elsevier, Comput. Phys. Commun. 140 (2001) 31-44, *"CMS software architecture Software framework, services and persistency in high level trigger, reconstruction and analysis."*

[36] CMS Collaboration, V. Innocente, Elsevier, Comput. Phys. Commun. 110 (1998) 192-197, *"CMS reconstruction and analysis an object oriented approach."*

[37] C. Charlot.,CMS CR 1999-018, *"On Demand Reconstruction within ORCA"*

[38] D. R. Musser, Software Practice and Experience 27(8) (1997) 983, *Introspective Sorting and Selection Algorithms*,

[39] Erich Gamma, Richard Helm, Ralph Johnson, John Vlissides, Addison-Wesley Professional, 1st edition January 15, 1995, *"Design Patterns."*,

[40] M. Spira, `http://people.web.psi.ch/spira/proglist.html`, *"HQQ v1.1"*

[41] A. Djouadi, J. Kalinowski, M. Spira, hep-ph/9704448, *"HDECAY: a Program for Higgs Boson Decays in the Standard Model and its Supersymmetric Extension".*

[42] S. Dittmaier, M. Kramer, M. Spira, hep-ph/0309204, *"Higgs Radiation off Bottom Quarks at the Tevatron and LHC."*

[43] M.L. Mangano, M. Moretti, F. Piccinini, R. Pittau, A.D. Polosa, hep-ph/0206293, *"ALPGEN, a generator for hard multiparton processes in hadronic collisions."*

[44] S.R. Slabospitsky, L. Sonnenschein, Comput. Phys. Commun. 148 (2002) 87, hep-ph/0201292 *"TopReX generator. Short manual."*

[45] T. Sjstrand, P. Edn, C. Friberg, L. Lnnblad, G. Miu, S. Mrenna and E. Norrbin, Computer Phys. Commun. 135 (2001) 238, (LU TP 00-30, hep-ph/0010017)

[46] E. Boos, V. Bunichev, M. Dubinin, L. Dudko, V. Ilyin, A. Kryukov, V. Edneral, V. Savrin, A. Semenov, A. Sherstnev, hep-ph/0403113, *"CompHEP 4.4 - Automatic Computations from Lagrangians to Events".*

[47] Zech G, Nucl. Instrum. Methods A 277 (1988) 608.

[48] T. Plehn, D. Rainwater and D. Zeppenfeld, FERMILAB-Pub-99/290-T, *"A method for identify $H \to \tau\tau \to e^{\pm}\mu^{\mp}P_T$ at the CERN LHC".*

[49] G. Senegri, F. Palla, CMS Note 2002-046, *"Lifetime based-tagging with CMS".*

[50] A. Djouadi, J. Kalinowski and M. Spira, Comput. Phys. Commun. 108 (1998) 56 and The LEP Higgs working group, LHWG note 2001-08.

[51] J.C. Romao, S. Andringa, Eur. Phys. J. C7 (1999) 631-642, *"Vector boson decays of the Higgs boson".*

[52] G. Ganis and P. Janot, Eds. G. Altarelli, T. Sjöstrand and F. Zwirner, CERN 96-01 (1996), Vol 2, 309. *"The HZHA Generator" in "Physics at LEP2".*

[53] Tomasz Pierzchala, private communication. November 2003.

[54] S. Jadach, W. Placzek and B.F.L. Ward, Phys. Lett. B390 (1997) 298.

[55] S. Jadach and Z. Wąs, Comput. Phys. Commun. 36 (1985) 191.

[56] J.A.M. Vermaseren, in "Proceedings of the IVth International Workshop on Gamma Gamma interactions", Eds. G. Cochard and P. Kessler, Springer Verlag, 1980.

[57] M. Skrzypek, S. Jadach, W. Placzek and Z. Wąs, Comput. Phys. Commun. 94 (1996) 216.

[58] S. Jadach, B.F.L. Ward, and Z. Wąs, Comput. Phys. Commun. 130 (2000) 260.

[59] "GEANT Detector description and simulation too", CERN Program Library, CERN-W5013 (1993).

[60] R. Cavanaugh, C. Delaere, G. Dissertori, K. Huettmann, V. Lemaitre, O. van der Aa, "ALPHA++ user's manual. ALEPH OO Physics Analysis Package version 3.4", CERN-ALEPH-2002-033; CERN-ALEPH-SOFTWRE-2002-002. (2002) 49p.

[61] C. Delaere, V. Lemaitre "An Alternative invisible Higgs search performed with ALPHA++", CERN-ALEPH-2002-035; CERN-ALEPH-SOFTWRE-2002-003. (2002) 32p. and V. Lemaitre, C. Delaere, "Comparison between an event selection program performed with ALPHA and ALPHA++", CERN-ALEPH-2000-093; CERN-ALEPH-SOFTWRE-2000-034. (2000) 49p.

[62] C. Delaere, T. Keutgen, G. Leibenguth, V. Lemaitre, A. Ninane, O. van der Aa, Dell Power Solutions, (August 2003), *"Solving High-Energy Physics Using a Dell HPC Cluster at the Université catholique de Louvain".*

[63] Jim Basney, Miron Livny, and Todd Tannenbaum, HPCU news, Volume 1(2), June 1997, *"High Throughput Computing with Condor".*,

[64] G. Stenuit et al, Mod. Phys. Lett. B, Vol. 17, Nos 10,11 & 12 (2003) 537-547, *"Vortex Configurations in Mesoscopic Superconducting Nanowires".*

Acknowledgements

Comme on a pu le constater au long de cette dissertation, la physique des hautes énergies nécessite la mise en oeuvre de détecteurs d'une telle complexité que seules d'importantes collaborations sont capables de les élaborer. En tant que doctorant, s'insérer dans ces collaborations est un défi en soi, connaissant les contraintes et les dead-lines omniprésentes auxquelles sont confrontés les membres du groupe. Dans ce contexte, je tiens à remercier particulièrement mon promoteur Vincent Lemaître qui, outre sa disponibilité, son support et son enthousiasme, m'a accepté comme premier doctorant et m'a permis de faire de fréquents voyages au CERN. Merci également à Ghislain Grégoire d'avoir supporté quelques longs séjours au CERN durant les mois de juillet et août. Je tiens aussi à remercier Sasha Nikitenko pour ses conseils éclairés, en particulier lors de la rédaction du DAQ-TDR, ainsi que Emilio Meschi qui, au début de ma thèse m'a guidé dans les arcannes du software du groupe e-gamma. Merci à Chris Seez qui a été scientifiquement présent aussi bien au début qu'a à la fin de ma thèse. Merci à Paris Sphicas pour ses encouragements initiaux et ses critiques finales. Je tiens aussi à remercier Christophe Delaere pour les projets software que nous avons initiés et réalisés ensemble lors d'une collaboration complémentaire et efficace. Merci à Alain Ninane de nous avoir laissé "chipoter" dans l'environnement informatique de FYNU. Enfin, merci à tous ceux qui, de près ou de loin, ont contribué à ce que cette thèse devienne une réalité; ils se reconnaîtrons certainenement.