Università degli Studi di Torino

Facoltà di Scienze Matematiche, Fisiche e Naturali
Dottorato di Ricerca in Fisica - XV ciclo

Development and Performance of High Level Trigger Algorithms for the Muon Trigger of the CMS Experiment

Tutore
Prof.sa Alessandra Romero

Coordinatore
Prof. Ezio Menichetti

Candidato
Dott. Nicola Amapane

24 Gennaio 2003

Contents

Introduction

The Standard Model of particle physics (SM) is one of the most successful theories of modern physics. Its validity has been experimentally tested with striking precision over a very wide range of energies. However, the Standard Model is still unsatisfactory from several points of view. In particular, the reason of the very different masses of the force carriers of the weak and strong interactions, as well as the mechanism which gives mass to all particles, remains unexplained. The Higgs mechanism, which addresses some of these issues, predicts the existence of a new particle — the Higgs Boson — which has not yet been observed. Regardless of the correctness of this explanation, there is widespread consensus that new physics must appear at energies higher than those reached by present particle accelerators.

This has led to the design of the Large Hadron Collider (LHC), a high-energy, high-luminosity proton-proton collider, that will be installed at the European Laboratory for Nuclear Research (CERN) in Geneva (Switzerland). The first collisions are planned for the summer of 2007; at that time it will be the most powerful particle accelerator ever built. Four experiments will collect data at the LHC: two general-purpose ones (ATLAS and CMS), one dedicated to b-physics (LHC-b), and one dedicated to heavy ion studies (Alice). The LHC running conditions (high luminosity, pile-up of several events in the same bunch crossing, high bunch crossing frequency, high background rates) will be very challenging, and the trigger system will be a key element for the success of the LHC experiments.

This thesis summarises the work I did within the Muon Physics Reconstruction and Selection (PRS/mu) group of the Compact Muon Solenoid (CMS) collaboration, whose purpose is to implement the full chain of simulation, reconstruction and selection of muons from the first trigger level up to the off-line analysis. Over the last two years, the focus of this work has been the development and the validation of the complete design of the CMS High Level Trigger (HLT) system in view of the preparation of the CMS Data Acquisition and High Level Trigger Technical Design Report (TDR) [1], which was submitted in December 2002.

My work covered several aspects of the development and study of the per-

formance of the muon high level trigger system of CMS. I participated in the production of the samples of simulated events for the studies of the PRS/mu group. The efficient simulation of an inclusive background sample of events with muons in the final state, which is essential for the determination of muon trigger rates, requires a special, sophisticated procedure. I participated in the development, testing and optimisation of this procedure and during the past two years I coordinated the Monte Carlo production for the PRS/mu group. The procedure adopted for this simulation is documented in a CMS Note [2], and is described in detail in Chapter 3. I also developed the framework used to analyse these samples [3].

Moreover, I contributed to the algorithm for muon reconstruction of the Level-2 trigger, notably developing and optimising the navigation step. This algorithm is described in Section 4.5.

The muon trigger has to deal with a high background rate of muons from minimum bias events, and a threshold on the muon p_T is not sufficient to reduce it adequately. Additional selection criteria, based on the physics properties of the events, must be applied already in the trigger chain, *e.g.* exploiting the fact that muons from decays of heavy object are isolated. I studied the performance of the isolation algorithm available in the Level-1 muon trigger (Section 5.1) and I participated in the developement of several muon isolation algorithms to be applied in the HLT. These algorithms are documented in a CMS Note [4], and are described in detail in Chapter 4.

Finally, I participated in the evaluation of the HLT performance in terms of signal efficiency, background rejection and timing. These results were included in the DAQ/HLT TDR and are reported in Chapter 6.

The present thesis starts with two introductory chapters. A general introduction to the physics motivation, to the experimental challenges and to the programme of the LHC is given in Chapter 1. In Chapter 2 the CMS detector is described, with particular focus on the muon detectors.

Chapter 1

Physics with the Large Hadron Collider

The Large Hadron Collider (LHC) is a proton-proton collider under construction at CERN. With a centre-of-mass energy of 14 TeV and a design luminosity of $10^{34}\text{cm}^{-2}\text{s}^{-1}$, it is a machine of unprecedented complexity and potential.

This chapter introduces the motivation, the design and the physics programme of LHC. In Section 1.1, a short overview of the Standard Model is given, with the aim of introducing the open issues that LHC is expected to explore. The design of the LHC, the experimental environment and the requirements for LHC experiments are described in Section 1.2. Finally, an overview of the main physics studies feasible at the LHC is presented in Section 1.3.

1.1 The Standard Model

The Standard Model (SM) of fundamental interactions describes matter as composed by half-integer spin particles, the *fermions*, that can be divided in two main groups: *leptons*, including electrons, muons, taus and neutrinos, and *quarks*. The latter have fractional charge and do not freely exist in nature; they are the constituents of a wide class of particles, the *hadrons*. Examples of hadrons are protons and neutrons, which are both composed by three quarks. A classification of fermions is given in Table 1.1.

Interactions between particles are described in terms of the exchange of *bosons*, integer spin particles that mediate fundamental interactions. *Gravitation* is the interaction closest to common experience; however it is not relevant at the scales of mass and distance typical of particle physics. The bond between atoms and molecules is due to the *electromagnetic interaction*, while the *weak interaction* explains, for example, nuclear β-decays. Finally, the *strong interaction* is responsible for the confinement of quarks inside hadrons. The

Table 1.1: *Classification of the three families of fundamental fermions.*

Fermions	1^{st} fam.	2^{nd} fam.	3^{rd} fam.	Charge	Interactions
Quarks	u	c	t	$+\frac{2}{3}$	All
	d	s	b	$-\frac{1}{3}$	
Leptons	e	μ	τ	-1	Weak, E.M.
	ν_e	ν_μ	ν_τ	0	Weak

characteristics of electromagnetic, weak and strong interactions are summarised in Table 1.2.

Table 1.2: *Fundamental interactions relevant in particle physics.*

	Electromagnetic	Weak	Strong
Quantum	Photon (γ)	$W^{\pm}$ and Z	Gluons
Mass (GeV/c^2)	0	80–90	0
Coupling constant	$\alpha(Q^2 = 0) \approx \frac{1}{137}$	$G_F \approx 1.2 \times 10^{-5}\,\mathrm{GeV}^{-2}$	$\alpha_s(m_\mathrm{Z}) \approx 0.1$
Range (cm)	∞	10^{-16}	10^{-13}

The Standard Model describes these interactions with two gauge theories:

- the theory of the electroweak interaction, or Electroweak Standard Model, that unifies the electromagnetic and weak interactions;
- the theory of strong interactions or Quantum Chromodynamics (QCD).

A comprehensive description of the Electroweak Standard Model is outside the scope of this work. However, the main results are reviewed in the following section.

1.1.1 Electroweak Symmetry Breaking in the Standard Model

The theory of electromagnetic interactions is called Quantum Electrodynamics (QED), and is based on the invariance of the Lagrangian for local gauge transformations with respect to the U(1) symmetry group. This condition leads to the existence of a massless vector field, the photon.

The unification of the theory of electromagnetism and that of weak interactions is accomplished by extending the symmetry to the group SU(2)×U(1) [5], which is associated to the quantum numbers I (*weak isospin*) and Y (*hypercharge*), that satisfy the relation:

$$Q = I_3 + \frac{Y}{2}, \tag{1.1}$$

where I_3 is the third component of the weak isospin and Q is the electric charge.

The invariance for local gauge transformations with respect to the SU(2)×U(1) group introduces four massless vector fields, $W_\mu^{1,2,3}$ and B_μ, that couple to fermions with two coupling constants, g and g'. The corresponding physical fields are linear combinations of $W_\mu^{1,2,3}$ and B_μ: the charged bosons W^+ and W^- correspond to

$$\mathrm{W}_\mu^\pm = \sqrt{\frac{1}{2}}(\mathrm{W}_\mu^1 \pm i\mathrm{W}_\mu^2), \tag{1.2}$$

while the neutral bosons γ and Z correspond to

$$A_\mu = B_\mu \cos\theta_W + W_\mu^3 \sin\theta_W \tag{1.3}$$
$$Z_\mu = -B_\mu \sin\theta_W + W_\mu^3 \cos\theta_W, \tag{1.4}$$

obtained by mixing the neutral fields W_μ^3 and B_μ with a rotation defined by the *Weinberg angle* θ_W. The field A_μ is then identified with the tensor of the electromagnetic field; requiring the coupling terms to be equal, one obtains

$$g \sin\theta_W = g' \cos\theta_W = e \tag{1.5}$$

that represents the electroweak unification.

Up to this point, all particles are massless: in the SU(2)×U(1) Lagrangian a mass term for the gauge bosons would violate gauge invariance. Masses are introduced with the *Higgs mechanism*, that allows fermions and the $W^\pm$ and Z bosons to be massive, while keeping the photon massless. This is accomplished by introducing the Higgs field, a SU(2) doublet of complex scalar fields:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}. \tag{1.6}$$

The Lagrangian of this field must be invariant under SU(2)×U(1) local gauge transformations, and includes a potential term

$$V(\phi) = -\mu^2 \phi^\dagger \phi + \lambda(\phi^\dagger \phi)^2, \tag{1.7}$$

where $\mu^2 > 0$ and $\lambda > 0$, so that the potential has a minimum for

$$\phi^\dagger\phi = \frac{1}{2}(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2) = \frac{\mu^2}{2\lambda} \equiv \frac{v^2}{2}. \tag{1.8}$$

One is free to choose the values of $\phi_i = 0$ that respect this condition; the fact that the minimum is not found for $\phi_i = 0$, but for a manifold of values, is called *spontaneous symmetry breaking*.

Boson masses derive from the coupling of the boson fields with the non-zero vacuum value of the Higgs field: the potential in its fundamental state does not have the symmetry properties of the Lagrangian. However, it can be shown that the minimum for the Higgs field is invariant for U(1) transformations; the electromagnetic U(1) symmetry is unbroken and the photon remains massless.

The Higgs mechanism gives rise to three massive gauge bosons, corresponding to nine degrees of freedom. Since the initial number of independent fields is ten (three massless bosons with two polarisations states each, plus the four real ϕ_i fields), one additional scalar gauge boson should appear as a real particle. This is the Higgs boson. Its mass depends on v and λ; the value of v is related to the boson masses by the relations:

$$m_{W^\pm} = \frac{1}{2}gv; \qquad m_Z = \frac{gv}{2\cos\theta_W} \tag{1.9}$$

However the parameter λ is characteristic of the field ϕ and cannot be determined from other measurements, so the Higgs mass is unknown.

The fermion masses are generated with a similar mechanism, but appear as free parameters of the theory, six for the quarks and three for the leptons (assuming neutrinos to be massless). These bring the number of free parameters of the Standard model to 17, the others being the four independent elements of the Cabibbo-Kobayashi-Maskawa matrix that describes the mixing of quark flavours, the couplings g and g', the parameter v of the Higgs vacuum expectation value and the Higgs mass.

All these parameters, except the mass of the Higgs boson which has not yet been observed, can be determined from experimental quantities. The measurement of these quantities allows therefore a consistency check of the electroweak Standard Model, which up to now has been confirmed with very high accuracy.

1.1.2 Higgs Mass

The mass m_H of the Higgs boson is not predicted by the SM; however constraints on its value can be obtained from theoretical considerations.

First of all, the Higgs potential of Eq. (1.7) is affected by radiative corrections, which involve the mass of bosons and fermions and depend on the

renormalisation scale Λ. Radiative corrections might change the shape of the potential so that it has no more an absolute minimum; vacuum would then become unstable. The request of *vacuum stability*, *i.e.* that the λ coefficient is large enough to avoid instability up to a certain scale Λ, implies a lower bound on the Higgs mass. On the other hand, due to the running of the coupling, λ increases with the energy scale; the request that it remains finite up to a scale Λ (*triviality*) corresponds to an upper bound on m_H.

In both cases, the parameter Λ represents the scale up to which the Standard Model is taken to be valid. The theoretical bounds on m_H as a function of Λ are shown in Fig. 1.1.

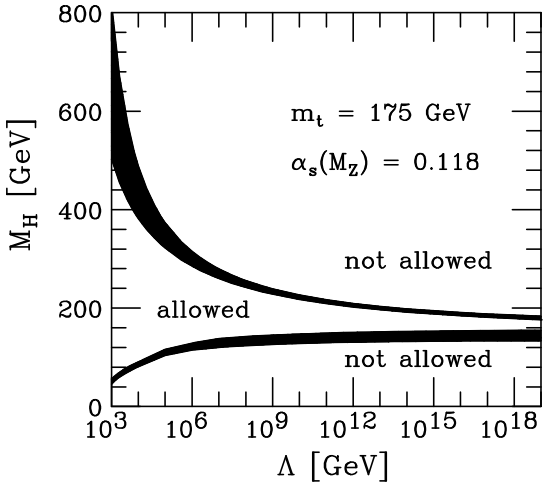


Figure 1.1: *Theoretical bounds on the Higgs mass as a function of the energy scale Λ up to which the Standard Model is valid [6].*

For the SM to remain valid up to the Planck scale ($\Lambda = 10^{19}$ GeV), the Higgs mass must be in the range 130-200 GeV/c^2. Assuming the SM to be valid only up to $\Lambda \sim 1$ TeV, the Higgs mass can be up to 700 GeV/c^2. In any case, new colliders should search for the Higgs boson up to masses of the order of ≈ 1 TeV. This is one of the guidelines for the design of the LHC. If the Higgs is not found in this range, then a more sophisticated explanation of the electroweak symmetry breaking mechanism must be found.

Experimental bounds on m_H are provided by measurements at LEP, SLC and Tevatron [7]. Direct searches excluded the region below 112.3 GeV/c^2 at 95% confidence level. Precision electroweak measurements are (logarithmically) sensitive to the Higgs mass due to radiative corrections; electroweak data can therefore be fitted taking m_H as free parameter. In Fig. 1.2, the shape of the χ^2 of the fit is shown as a function of m_H. The curve is shallow and the minimum is below the value excluded by direct searches (shaded area). One may conclude

nevertheless that the fit privileges low values of the Higgs mass. An upper limit of 170 GeV/c^2 can be set at 95% confidence level.

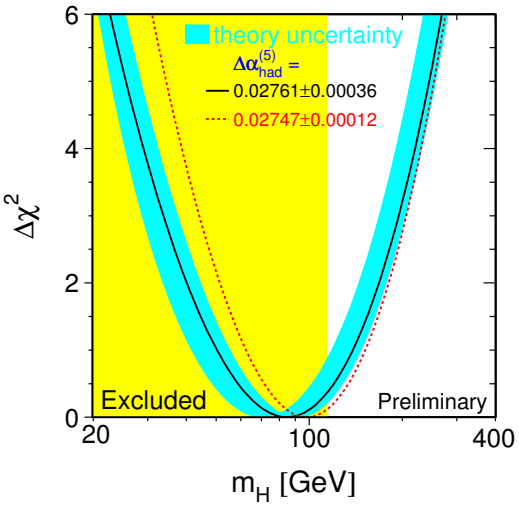


Figure 1.2: *$\Delta\chi^2$ of the fit of electroweak measurements of LEP, SLC and Tevatron as a function of the Higgs mass [7].*

1.1.3 Open Questions

Even if up to now the SM has been experimentally confirmed with very high accuracy, the Higgs boson has never been observed, and has been excluded by direct searches up to the energies accessible at LEP.

Apart from this problem, there are several reasons to think that the Standard Model is only an effective description, and that a more fundamental theory must exist. We already observed that theoretical bounds on the Higgs mass can be derived from the request that, once radiative corrections are included, the theory remains valid up to a given energy scale. It is natural to think that at higher energy scales some more general theory should be valid, possibly describing all interactions. In the Standard Model, the strong interaction is described by a SU(3) "colour" symmetry group, which however is not unified with the electroweak description. Gravity, whose strength should become comparable with that of other interactions at the Planck scale (10^{19} GeV), is not included at all. It would be appealing to find a wider symmetry that describes all interactions, and the reason why it is broken at these energy scales.

In addition, the Higgs mass suffers from divergences caused by radiative corrections which are proportional to the energy cutoff; for the Standard Model to be valid up to very high energy scales, extremely precise cancellations should be present at all perturbation levels. Such cancellations are formally possible, but there is no reason why such a fine tuning should occur (*naturalness problem*).

Other considerations of more aesthetic nature are that the Higgs is an ad-hoc addition to the Standard Model; moreover it is the only scalar particle in the theory. Also, there is no explanation for the fact that the particle masses would be significantly smaller than the energy scale up to which the theory remains valid (*hierarchy problem*). Finally, the number of free parameters of the Standard model (17, neglecting neutrino masses and mixings) appears too high not to be artificial.

Several solutions for these problems have been proposed. Among them, supersymmetry (SUSY) is an elegant theory that introduces a new symmetry between bosons and fermions. SUSY predicts that each particle has a supersymmetric partner whose spin differs by one half. The naturalness problem is solved by the fact that the loop contributions from particles and their supersymmetric partners cancel. The simplest supersymmetric model, called the Minimal Supersymmetric Standard Model (MSSM), requires at least two Higgs doublets, corresponding to five Higgs particles: two charged bosons, $H^{\pm}$, two scalar bosons, h and H, and one pseudo-scalar, A. The MSSM predicts a rich phenomenology to appear below energies of about 1 TeV; however no evidence for supersymmetry has been observed yet.

1.2 The Large Hadron Collider

1.2.1 Design of the LHC

From the discussion of the previous section we know that the main task of today's experimental particle physics is the search of the Higgs boson (or other phenomena) up to a scale of about 1 TeV. This section discusses the requirements and feasibility of a collider designed for this search, the LHC.

In a circular collider of radius R, the energy loss per turn due to synchrotron radiation is proportional to $E^4/(m^4 R)$, where E and m are respectively the energy and mass of the particles accelerated. Circular electron colliders would need enormous dimensions to reach energies of the order of 500 GeV per beam, so the natural choice for a collider with current technologies is to use beams of protons, which are almost 2000 times heavier than electrons. Protons are not elementary particles, and in hard collisions the interaction involves their constituents (quarks and gluons), which carry only a fraction of the proton energy. While this fact has the drawback that the centre-of-mass energy and

the rest frame of the hard scattering are unknown, it has the advantage that a wide range of energies can be explored with fixed-energy beams.

Since only a fraction of the proton energy is exchanged in the hard scattering, the centre-of-mass energy of the beams must be much higher than the mass of the particle that has to be produced. The cross section for different processes as a function of the centre-of-mass energies in $p-p$ collisions is shown in Fig. 1.3. It can be noted that the Higgs cross section increases steeply with the centre-

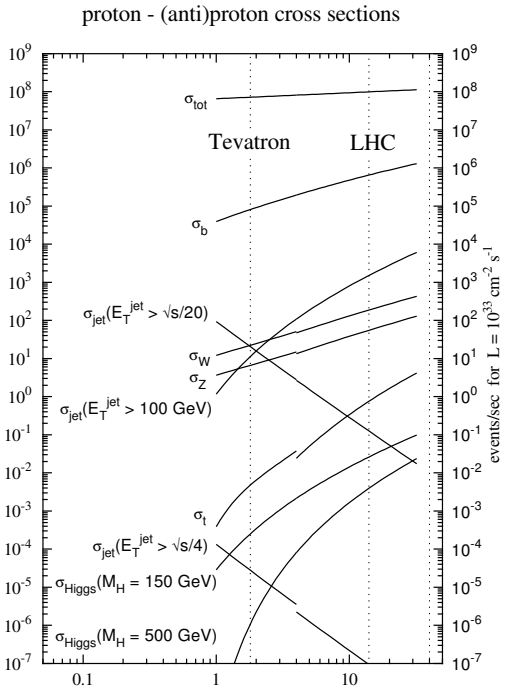


Figure 1.3: *Cross sections and event rates of several processes as a function of the centre-of-mass energy of proton-proton collisions [8].*

of-mass energy, while the total cross section (*i.e.* the background) remains almost constant. It is clear that the highest possible centre-of-mass energy should be used. The cross-section σ determines the event rate R of a given process according to the formula $R = \mathcal{L}\sigma$. The factor $\mathcal{L}$ is called *luminosity*; it represents the number of collisions per unit time and cross-sectional area of the beams. It is specific to the collider parameters and does not depend on the interaction considered:

$$\mathcal{L} = f\frac{n_1 n_2}{A}. \tag{1.10}$$

Here f is the collision frequency of bunches composed of n_1 and n_2 particles and A is the overlapping cross-sectional area of the beams.

With a centre-of-mass energy of 40 TeV, the Superconducting Super Collider (SSC), a $p-p$ collider to be constructed in the USA, would have been a perfect

Table 1.3: *LHC parameters for $p-p$ and Pb–Pb collisions.*

Parameter	$p-p$	$^{208}Pb^{82+}$
Centre-of-mass energy (TeV)	14	1148
Number of particles per bunch	1.1×10^{11}	$\sim 8\times10^{7}$
Number of bunches	2808	608
Design Luminosity ($cm^{-2}s^{-1}$)	10^{34}	2×10^{27}
Luminosity lifetime (h)	10	4.2
Bunch length (mm)	53	75
Beam radius at interaction point (μm)	15	15
Time between collisions (ns)	24.95	124.75×10^{3}
Bunch crossing rate (MHz)	40.08	0.008
Circumference (km)	26.659	
Dipole field (T)	8.3	8.3

candidate for the search of the Higgs. Unfortunately, this project was abandoned due to financial problems while its 87 km long tunnel was being excavated. The idea behind the Large Hadron Collider is to reuse the existing 27 km long LEP tunnel to install a new proton collider. Considerable financial savings are obtained from the fact that the tunnel and several infrastructures (including pre-accelerators) already exist. However, the size of the tunnel limits the centre-of-mass energy to 14 TeV, since the beams must be bent by dipole magnets whose maximum field is limited.

To compensate for the lower Higgs production cross sections, the LHC must have a very high luminosity; this has the drawback that the total event rate becomes so high that several interactions overlap in the same bunch crossing (*pile up*). The LHC will operate at a bunch crossing rate of 40 MHz and a design LHC luminosity of $10^{34}cm^{-2}s^{-1} = 10\ nb^{-1}s^{-1}$. The bunch structure is such that only about 80% of the bunches will be full [9]; given that the total non-diffractive inelastic $p-p$ cross section predicted by PYTHIA is 55 mb, on average 17.3 events will occur at every bunch crossing. With about 50 charged tracks per interaction, this pile-up poses several experimental problems, as discussed in Section 1.2.3.

In the first three years of operation, the LHC will run at a reduced luminosity of $2\times10^{33}cm^{-2}s^{-1}$. The two luminosity regimes are commonly called "High luminosity" and "Low Luminosity", respectively. The LHC will also be able to accelerate and collide beams of heavy ions such as Pb to study the de-confined state of matter, the quark-gluon plasma. The parameters of the LHC are summarised in Table 1.3.

1.2.2 Phenomenology of Proton-Proton Collisions

Protons are not elementary particles, and in high-energy inelastic proton collisions the interaction involves the partons (quarks and gluons) inside the proton. The effective centre-of-mass energy of the hard scattering, $\sqrt{\hat{s}}$, is therefore proportional to the fractional energies x_a and x_b carried by the two interacting partons: $\sqrt{\hat{s}} = \sqrt{x_a x_b s}$, where $\sqrt{s}$ is the centre-of-mass energy of the proton beams.

The momentum distribution of the partons inside the protons are called Parton Density Functions (PDFs). They are different for gluons, u and d valence quarks and low-momentum quark-antiquark pairs of all flavours produced and annihilated as virtual particles (sea quarks). They are functions of the exchanged four-momentum squared, Q^2, since for higher exchanged momenta a shorter distance scale is probed and the contribution of gluons and sea quarks becomes higher. PDFs are obtained from experimental data covering the widest possible range of (x, Q^2); their accuracy is important for several precision measurements. In Fig. 1.4, the CTEQ4M PDFs [10] at two different values of Q^2 are shown.

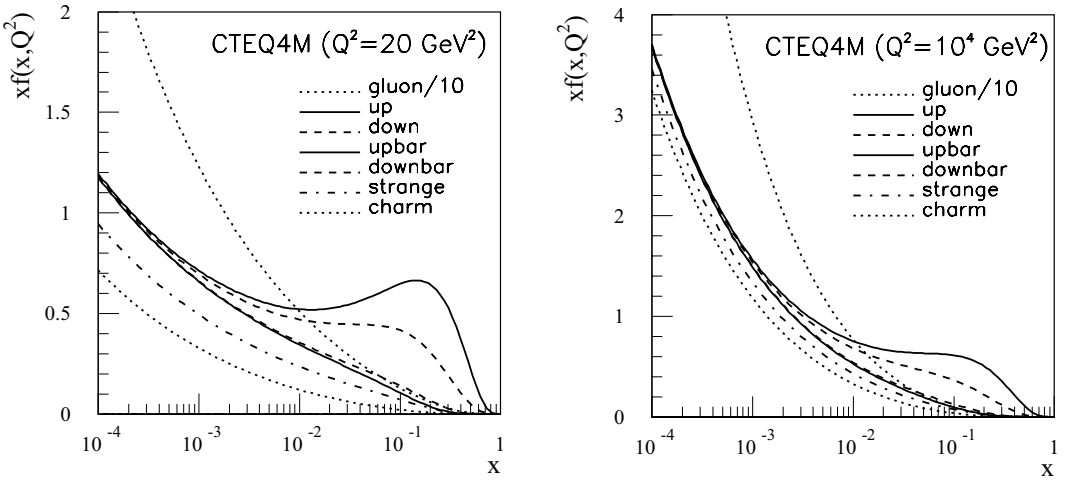


Figure 1.4: *Parton density functions for $Q^2 = 20\ GeV/c^2$ and $Q^2 = 20\ GeV/c^2$ [11].*

The fact that the two partons interact with unknown energies has two fundamental consequences. First of all the total energy of an event is unknown, because the proton remnants, that carry a sizeable fraction of the proton energy, are scattered at small angles and are mostly lost in the beam pipe, escaping detection. Experimentally, it is therefore not possible to define the total and missing energy the event, but only the total and missing transverse energies (in the plane orthogonal to the beams). Moreover, the centre of mass may be

boosted along the beam direction; it is natural to use experimental quantities that are invariant under such boosts, as the transverse momentum p_T. Angular distributions are often described in terms of the *rapidtity*, which is defined choosing the beam direction as z axis:

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z}. \tag{1.11}$$

The rapidity has the property that the shape of dN/dy distributions is invariant under boosts along the z direction. For $p \gg m$ the rapidity is approximated by the *pseudorapidity*:

$$\eta = -\ln \tan \frac{\theta}{2}, \tag{1.12}$$

where θ is the angle between the particle momentum and the z axis. The pseudorapidity has the advantage that it depends only on θ and can be also used for particles of unknown mass.

One very remarkable aspect of LHC physics is the wide cross section range of the processes under investigation: in Fig. 1.3 we see for example that the Higgs production has a cross section at least ten orders of magnitude smaller than the total inelastic cross section. In fact, the bulk of the events produced in proton collisions is either due to low-$\hat{p}_T$ scattering, where the protons collide at large distances, or to QCD high-$\hat{p}_T$ processes of the type:

$$\begin{cases} q_i\bar{q}_i \rightarrow q_k\bar{q}_k \\ q_iq_j \rightarrow q_iq_j \\ q_ig \rightarrow q_ig \\ q_i\bar{q}_i \rightarrow gg \\ gg \rightarrow q_k\bar{q}_k \\ gg \rightarrow gg \end{cases}$$

All these events are collectively called "minimum bias" and in LHC studies are in general considered uninteresting since they constitute a background for other processes, where massive particles like the Higgs are created in the hard scattering. This classification is somewhat arbitrary; for example, this definition of minimum bias events includes $b\bar{b}$ production that is of interest for b-physics studies.

1.2.3 LHC Experiments

LHC detectors will operate in a very difficult environment: the high bunch crossing frequency, the high event rate and the pile-up of several events in the same bunch crossing dictate strict requirements on the design of detectors. To cope with a bunch crossing rate of 25 ns and a pile-up of about 20 events per

crossing, the detectors should have a very fast time response and the readout electronics should also be very fast. Due to the presence of pile-up, high granularity is also required to avoid the overlap of particles in the same sensitive elements. High granularity means a large number of electronics channels, and therefore high cost. LHC detectors will also have to stand an extremely high radiation dose; special radiation-hard electronics must be used.

Additional requirements apply to the online trigger selection, that has to deal with a background rate several orders of magnitude higher than the signal rate. These requirements are discussed in Section 2.5.

Four experiments will be installed at the LHC. Two of them are devoted to specific topics: Alice to heavy ions and LHC-b to b-physics. The other two are the general-purpose experiments ATLAS and CMS. Their design differs significantly, since two very different solutions were chosen for the configuration of magnetic field: CMS uses a solenoidal field generated by a big superconducting solenoid, while ATALS uses a toroidal field produced by three sets of air-core toroids complemented by a small inner solenoid. A detector based on a toroidal magnet has the advantage that the track p_T resolution is constant as a function of pseudorapidity. A very large air-core toroid allows a good momentum resolution even without the aid of the inner tracker; however, it requires very precise detectors with excellent alignment. An iron-core solenoid, on the other hand, can generate a very intense field. The resulting system is very compact and allows calorimeters to be installed inside the magnet, improving the detection and energy measurement of electrons and photons. Precise tracking exploits both the constant field within the magnet and the field inside the return yoke. Moreover, tracks exiting the yoke point back to the interaction point, a property that can be used for track reconstruction. Multiple scattering within the yoke, however, degrades the resolution of the muon system.

Schematic pictures of CMS and ATLAS are shown in Fig. 1.5 and Fig. 1.6, respectively. The CMS experiment is described in detail in Chapter 2.

1.3 Physics with the LHC

Thanks to his high centre-of-mass energy and high luminosity, the LHC has a significant physics potential not only for the discovery of the Higgs, but also in many other fields like electroweak precision measurements, searches and b-physics. In this section, an overview of the LHC physics programme is given.

1.3.1 Higgs Search

In proton-proton collisions, several main processes contribute to the production of the Higgs boson, as illustrated in Fig. 1.7. The cross sections for the

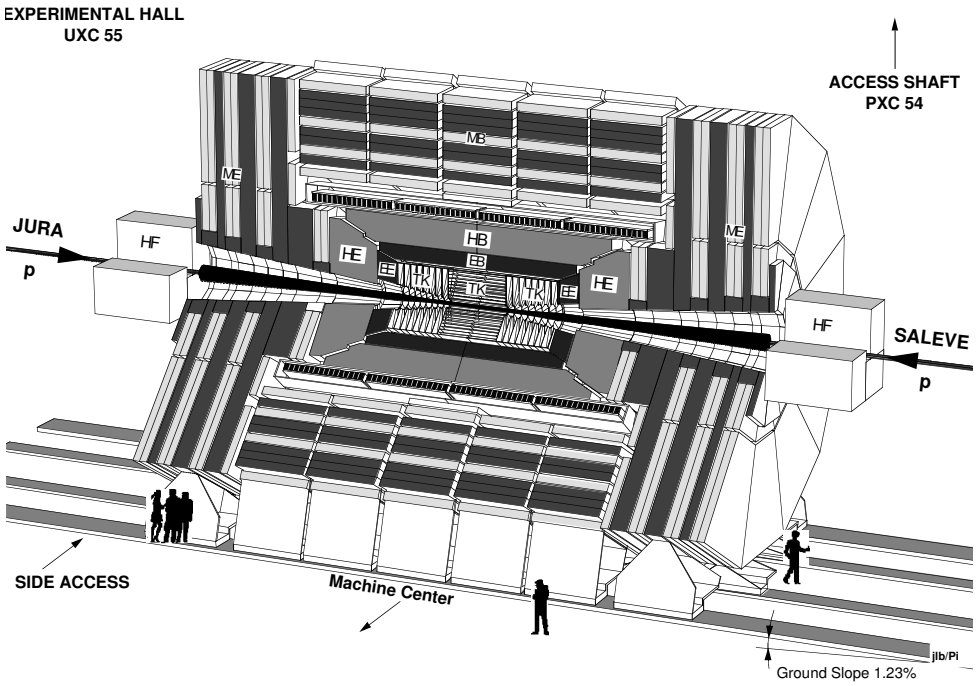


Figure 1.5: *The CMS detector.*

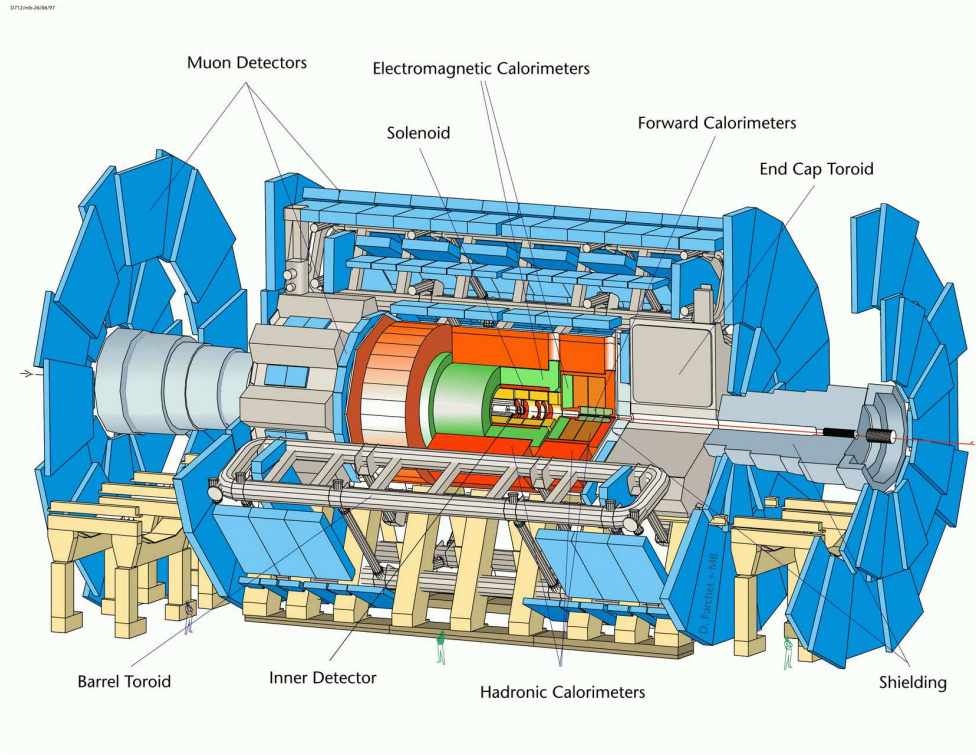


Figure 1.6: *The ATLAS detector.*

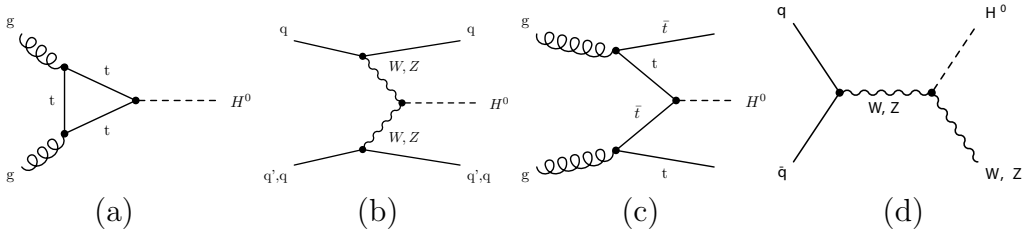


Figure 1.7: *Higgs production mechanisms at tree level in proton-proton collisions: (a) Gluon-gluon fusion; (b) W and Z fusion; (c) $t\bar{t}$ associated production; (d) W and Z associated production.*

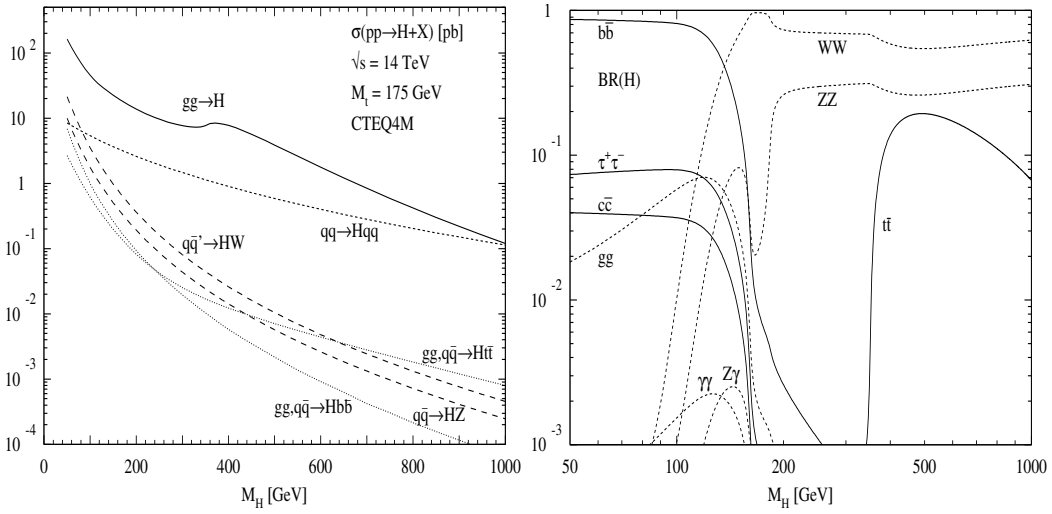


Figure 1.8: *Higgs production cross section (left) and branching ratios (right) as a function of the Higgs mass [12].*

different processes as a function of m_H are shown in Fig. 1.8. Gluon fusion is the dominant process, and only for very high Higgs masses vector does boson fusion become of comparable size. However, the presence of two forward jets is a good experimental signature that can be used to suppress backgrounds. Associated production processes have very small cross sections, except for low Higgs masses; also in this case a powerful signature is given by the additional bosons or quarks in the final state.

Cross sections are of the order of few picobarns, which at the LHC design luminosity correspond to rates of about 10^{-2} Hz. However, not all final states will be experimentally accessible, and the branching ratios of the observable channels are usually small.

The branching ratios for several Higgs decay channels as a function of the Higgs mass are shown in Fig. 1.8. They can be interpreted on the basis of the

Higgs couplings to fermions and gauge bosons:

$$g_{Hff} = \frac{m_f}{v} \tag{1.13}$$

$$g_{HWW} = \frac{2m_W^2}{v} \tag{1.14}$$

$$g_{HZZ} = \frac{m_Z^2}{v}. \tag{1.15}$$

The coupling is proportional to the fermion masses and to the square of the boson masses; therefore the final state with the heaviest available particle dominates. For low Higgs masses ($m_H < 130$ GeV/c^2), the heaviest available fermion is the b quark, and $H \to b\bar{b}$ dominates. However, the QCD jet background is so high at LHC that it will be almost impossible to observe this decay (except maybe by exploiting associated $t\bar{t}H$ or WH production.) The most promising channel is $H \to \gamma\gamma$, which despite the very low branching ratio ($\sim 10^{-3}$) has a very clean signature. The signal should appear as a narrow peak over the continuum $(q\bar{q}, gg) \to \gamma\gamma$ background, but excellent photon energy and angular resolution are required as well as good π^0 rejection.

For larger Higgs masses, the production of WW and ZZ pairs becomes possible; the branching ratio is high, but purely hadronic final states are again not accessible. $H \to 4\ell$ is the "golden-plated" channel for Higgs search; however it has very low branching ratio and suffers from the low Higgs production cross sections above $m_H \sim 600$ GeV/c^2. The channel $H \to WW$ has the disadvantage that accessible final states ($\ell\mu\ell\mu, \ell\nu jj$) have at least one neutrino that escapes detection; however it will be a good discovery channel, especially for $m_H \approx 2m_W$ where the WW production is at threshold and the ZZ branching ratio drops to 20%. For very high Higgs masses (above 500 GeV/c^2), the cross section becomes low and semi-hadronic final states ($2\ell 2j, \ell\nu 2j$) have to be used. The Higgs width becomes also very broad, as shown in Fig. 1.9, so that the reconstruction of a mass peak becomes difficult.

1.3.2 Electroweak Physics

Several precision electroweak measurements will be possible with the two experiments ATLAS and CMS. Thanks to the high statistics available (*e.g.* 10^8 $W \to e\nu$ decays, 10^7 $t\bar{t}$ pairs produced in the first year of operation), for most measurements the statistical uncertainty will be very small. High statistics control samples will allow a good understanding of the detector response, thus reducing the systematic uncertainty.

In the following, we shortly discuss the experimental challenges of some electroweak measurements to be performed at LHC. A more comprehensive discussion can be found elsewhere [8, 13].

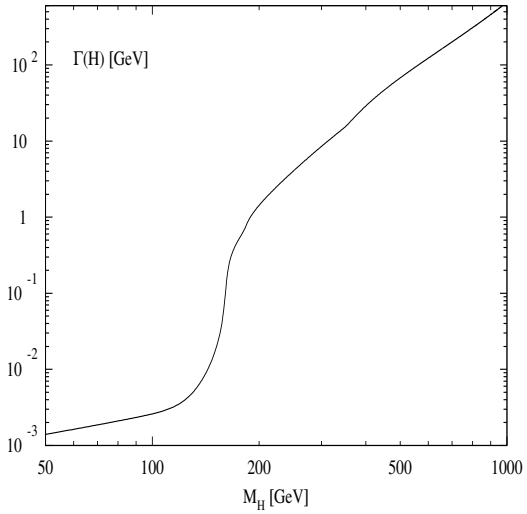


Figure 1.9: *Higgs width as a function of its mass [12].*

1.3.2.1 Measurement of the W and Top Mass

The values of the top and Higgs masses enter in the prediction of the W mass through radiative corrections. Precise measurements of m_t and m_W allow therefore to set limits on m_H and, if the Higgs is found, they will allow stringent tests of the Standard Model (SM) or its extensions like the Minimal Supersymmetric Standard Model (MSSM).

The most promising channel for the measurement of the top mass is $t\bar{t} \to W^+W^-b\bar{b}$ with one leptonic and one hadronic W decay, where the hadronic part is used to reconstruct the top mass and the leptonic part to select the event. The main source of uncertainty will be the jet energy scale, which is affected by the knowledge of fragmentation and gluon radiation and of the response of the detectors. The same sample of $t\bar{t}$ events will provide a large number of hadronic W decays to be used for the calibration of the hadron calorimeters.

The final uncertainty on m_t will be better than about 2 GeV. This will allow to constrain the Higgs mass to better than 30% but, in order not to become the dominant source of uncertainty, the W mass will have be measured with a precision of about 15 MeV.

Since the longitudinal component of the neutrino momentum cannot be measured in hadron colliders, the W mass is obtained from a fit to the distribution of the W transverse mass. The main source of uncertainty is the lepton energy and momentum scale, which should be known with a precision of $\sim 0.02\%$ to achieve the mentioned precision on m_W. This is a challenging goal that could be reached thanks to the high statistics of $Z \to \ell\ell$ decays. Other sources of systematic uncertainty include the W p_T spectrum, the W width and the proton structure functions. Since pile-up deteriorates the shape of the W transverse

mass distribution, this measurement will probably be feasible only in the low luminosity mode.

Several other measurements will be possible in the top quark sector, including rare top decays, the determination of the $t\bar{t}$ production cross section and the observation of single top production.

1.3.2.2 Drell-Yan Production of Lepton Pairs

The Drell-Yan production of lepton pairs is a process with clean signature and low experimental backgrounds. The interesting quantities are the cross section and the forward-backward asymmetry, both functions of the rapidity y and of the invariant mass $m_{\ell\ell}$ of the di-lepton system.

The measurement of the cross section can provide evidence for new physics (new resonances, contact terms etc.) and probe electroweak radiative corrections up to 1.5 TeV. The main issue in this measurement is the knowledge of the absolute luminosity which should reach a precision of 5%, an ambitious goal requiring theoretical improvements in the knowledge of the W and Z production cross section.

The measurement of the forward-backward asymmetry $\mathrm{A_{FB}}$ allows the determination of the effective electroweak mixing angle $\sin^2\theta^{lept}_{eff}$. However, it requires the knowledge of the incoming quark and anti-quark direction, which in $p-p$ colliders is difficult to determine and is affected by the uncertainty on parton density functions. Improving the LEP+SLD accuracy on $\sin^2\theta^{lept}_{eff}$ is therefore a very ambitious goal. Preliminary studies indicate that it might be possible to achieve it if the geometrical acceptance for electrons is extended up to $|\eta| < 4.9$ using the forward calorimeters.

1.3.2.3 Production of Vector Boson Pairs

At the lowest order vector boson pairs are produced in $q\bar{q}$ annihilations followed by triple gauge boson vertices. Triple Gauge Couplings (TGC) allow to test directly the non-Abelian gauge symmetry of the SM. They are described by five parameters, which can be measured with fits to the total cross sections and shapes of distributions such as that of $p_T^{\gamma,Z}$ in Wγ and WZ events. Deviations from the SM values can provide evidence for new physics, *e.g.* the presence of new heavy particles decaying to WW and ZZ pairs (including heavy Higgs). Anomalous couplings are present in many extensions of the SM, including MSSM, technicolour etc. Since the sensitivity to anomalous couplings is enhanced at high centre-of-mass energies, LHC is expected to increase significantly the precision on TGC parameters.

At high energy, the process of vector boson scattering becomes dominant. This process is particularly sensitive to the electroweak symmetry breaking

mechanism and violates unitarity at a scale of ~ 1 TeV. In the SM, this is cured with the introduction of the Higgs boson; in the absence of the Higgs new physics must couple to this channel. Whatever the scenario will be, vector boson scattering will therefore probe the mechanism of electroweak symmetry breaking. It is thus important to measure the WW scattering cross section as a function of the invariant mass of the WW system up to a scale of ~ 1 TeV.

From the experimental point of view, one can exploit the presence of the two quarks emitting the incoming bosons by requiring the signature of forward tagging jets. Other selection criteria are similar to those used for heavy SM Higgs searches. Purely hadronic final states will not be accessible due to the high QCD background; on the other hand it will be useful to reconstruct the invariant mass of the final state boson-boson system also in the absence of a resonance, which is not possible in final states with more than one neutrino.

High integrated luminosities will be necessary for these studies, due to the small cross sections and branching ratios and the high background rates.

1.3.3 Other Physics Studies

Several other topics will be covered by LHC; the most important are mentioned in the following.

Supersymmetry. As already discussed, the simplest supersymmetric model, the Minimal Supersymmetric Standard Model (MSSM), predicts the existence of five Higgs bosons: two charged bosons, $H^{\pm}$, two scalar bosons, h and H, and one pseudo-scalar, A. At tree level, all masses and couplings depend on two parameters, chosen as the mass of the A boson, m_A, and the ratio of the vacuum expectation values of the two Higgs doublets, $\tan\beta$.

In most of the parameter space, charged Higgs bosons decay predominantly to $\tau\nu$. For the neutral Higgs bosons, the decays to vector bosons are suppressed, so that the golden channels described for the case of a SM Higgs will not be observable. The dominant decays modes are those to $b\bar{b}$ and $\tau^+\tau^-$ [12], but the former is hidden by the large background of b-jets. The observation of MSSM Higgs bosons will therefore rely on the identification the leptons coming from τ decays and of τ-jets.

Additionally, supersymmetric theories predict a rich phenomenology. In general supersymmetric particles should be accessible at LHC energies and have very spectacular signatures due to cascade decays with many leptons and jets in the final state and large missing energy. If supersymmetry exists, LHC experiments will certainly be able to observe and study it.

B-Physics. In the field of b-physics, LHC will benefit from a very large $b\bar{b}$ production cross section (see Fig. 1.3). The main interest is the study of the decays of neutral B mesons, and in particular of CP violation in the $B_d^0 - \bar{B}_d^0$ and $B_s^0 - \bar{B}_s^0$ systems. B decays can be identified in leptonic final states, especially in the case of muons. However these leptons are usually soft and the identification is difficult due to the high backgrounds and pile-up. One LHC expriment, LHCb, will be dedicated to b-physics, which will be studied also by ATLAS and CMS will in the low-luminosity phase. At high luminosity, the luminosity in the LHCb collision point will be reduced by de-tuning the beams.

Heavy ions. As already mentioned, LHC will also operate as a collider of heavy ions, with an energy 30 times higher than that that of RHIC, today's most powerful ion collider. Heavy ions collisions will be studied with a dedicated detector, Alice, but also with Atlas and CMS. The goal is to investigate the de-confined state of matter, the quark-gluon plasma.

Chapter 2

The CMS Detector

CMS [14] is a general-purpose detector that will operate at the Large Hadron Collider. The first requirement for CMS was that of a compact design, which leads to the choice of a strong magnetic field (cf. Section 1.2.3). This is obtained using a 4 T superconducting solenoid.

A good and redundant muon system, a good electromagnetic calorimeter and a high quality central tracking are the design priorities of CMS. The CMS solenoid is large enough to accommodate the calorimeters, allowing precision measurements of electrons and photons. The 4 T magnetic field allows precision tracking in the all-silicon inner tracker and reduces the pile-up from soft hadrons in the muon system, which is installed inside the iron return yoke.

This chapter describes the general design of CMS and of its subdetectors. A general description of the CMS trigger system with details on the Level-1 trigger is given in Section 2.5.

2.1 Overall Design

The CMS detector is composed of a cylindrical barrel closed by two endcap disks, with a full length of 21.6 m, a diameter of 15 m and a total weight of $\sim$ 12500 tonnes.

The longitudinal view of one quarter of the CMS detector is shown in Fig. 2.1. The transversal view of the barrel region is shown in Fig. 2.2. Detectors and non-sensitive volumes are indicated with standard two-letter codes (the second being `B` for barrel and `E` for endcap) which will be used in the following.

The CMS coordinate frame is a right-handed system where the x axis points to the centre of the LHC ring, the z axis is parallel to the beam and the y axis points upwards. Reconstruction algorithms however use a spherical coordinate system based on the distance r from the z axis, the azimuthal angle ϕ with respect to the y axis and the pseudo-rapidity η defined in Eq. 1.12.

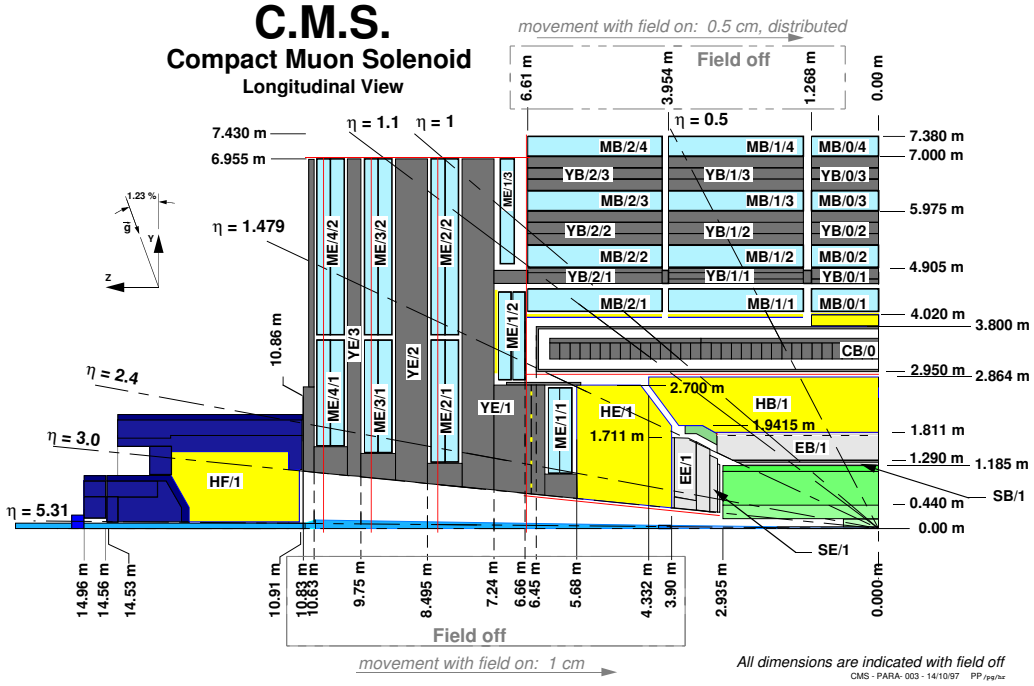


Figure 2.1: *Longitudinal view of one quarter of the CMS detector.*

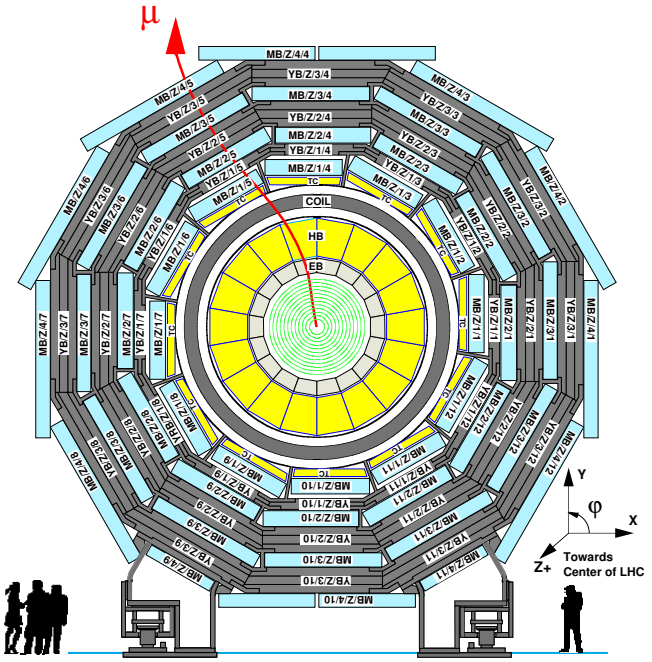


Figure 2.2: *Transverse view of the barrel region of the CMS detector. Barrel wheels are numbered Z = -2, -1, 0, 1, 2.*

The CMS design is driven by the choice of its magnet (CB), a 13 m long superconducting solenoid [15] with a diameter of 5.9 m. Cooled with liquid helium, it will generate a magnetic field of 4 T, which is kept uniform by a massive iron return yoke (YB,YE). The yoke will also host the muon system (MB,ME), composed by drift tube detectors in the barrel region and cathode strip chambers in the endcaps (up to $|\eta| < 2.4$), complemented by a system of resistive plate chambers with a coverage of $|\eta| < 2.1$. The muon system is described in detail in Section 2.4.

The calorimeters and the inner tracker are installed inside the coil. The innermost tracking detector is essential for precise vertex reconstruction and b-tagging and has to deal with a very high track density. Very fine segmentation is crucial; therefore a silicon pixel detector was chosen. In the baseline design it consists of 3 barrel layers and 2 forward disks. Outside the pixel detector, a silicon strip detector is installed, extending up to a radius of about 1.2 m. The full silicon tracker allows charged tracks reconstruction with at least 12 measurement points and a coverage of $|\eta| < 2.5$; it is described in Section 2.2.2.

Photons and electrons are measured by a homogeneous electromagnetic calorimeter (ECAL), composed by $PbWO_4$ scintillating crystals covering the region $|\eta| < 3.0$ (EB,EE). In the endcaps, it will be supplemented by a lead/silicon preshower detector, to improve the resolution in the determination of electron and photon direction and help pion rejection. The ECAL is described in Section 2.3.

Jets and energy imbalance are measured by a sampling hadronic calorimeter (HCAL) installed just before the coil. It is composed of a copper alloy and stainless steel instrumented with plastic scintillators. The barrel and endcap parts (HB,HE) have the same $|\eta|$ coverage as the ECAL, and are complemented by a very forward calorimeter (HF), which extends the coverage up to $|\eta| < 5.3$, enhancing the hermeticity of the detector and its ability to measure missing transverse energy. The HCAL is described in Section 2.3.1.

2.2 The Tracker

The goal of the inner tracker [16, 17] is to reconstruct high-p_T charged tracks in the region $|\eta| < 2.5$ with high efficiency and momentum resolution, to measure their impact parameter and to reconstruct secondary vertices. This is obtained with a detector based on several layers of silicon detectors. The first layers, which are closer to the interaction point, are crucial for the measurement of the impact parameter, and have to cope with a very high particle flux. Finely segmented pixel detectors are used in this region. The rest of the tracker is composed of single-sided and double-sided silicon strip detectors extending up to $r = 110$ cm and $|z| = 270$ cm. The pixel and silicon detectors are described

in Section 2.2.1 and Section 2.2 respectively.

2.2.1 The Pixel Detector

The pixel detector will provide high-resolution three-dimensional measurements, that will be used for charged track reconstruction. Its excellent resolution will also allow the measurement of track impact parameters, the identification of b- and τ-jets and the reconstruction of vertices in three dimensions. It consists of square n-type silicon pixels with a size of 150 μm $\times$150 μm on a n-type silicon substrate. A spatial resolution of $\sim$ 15 μm is obtained by analogic interpolation of the charge induced in nearby pixels, helped by the large Lorentz drift angle in the magnetic field. Charge sharing is enhanced in the endcaps by tilting the detectors by 20^0. The full detector consists in a total of 4.4 million pixels.

The baseline design of the pixel detector includes three barrel layers at mean radii of 4.4 cm, 7.3 cm and 10.2 cm, extending for a total length of 53 cm. Two endcap disks, extending in radius from 6 to 15 cm, will be placed on each side at z=34.5 cm and 46.5 cm. A three-dimensional representation of the pixel detector is shown in Fig. 2.3. This layout will guarantee at least two pixel hits

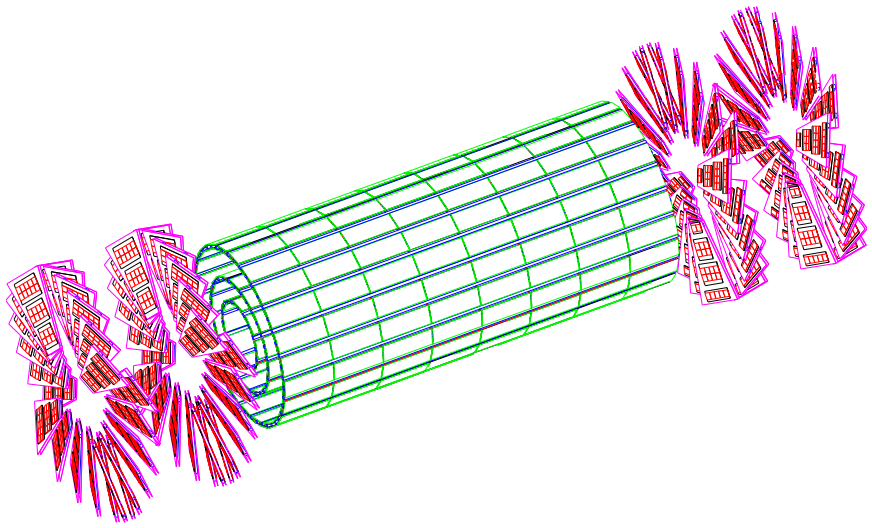


Figure 2.3: *Three-dimensional view of the full pixel detector.*

for tracks originating within $2\sigma_Z$ from the nominal collision point, up to about $|\eta| < 2.2$. Standalone track reconstruction, that requires three hits per track, will also be possible, with good efficiency in the same region. However, in the initial low-luminosity phase only two barrel layers and one end-cap disk will be installed. Standalone track reconstruction will not be possible in this case, and two-hit coverage will be limited to $|\eta| < 2.0$.

The large particle flux imposes also special requirements of radiation hardness, and it is very likely that the pixel detectors will have to be substituted during the lifetime of the experiment.

2.2.2 The Silicon Microstrip Detector

In addition to the pixel detectors, the inner tracker is composed of several layers of silicon microstrip detectors, whose layout is shown in Fig. 2.4. The inner part consists of four barrel layers and three small forward disks. The outer part consists of six barrel layers and nine forward disks.

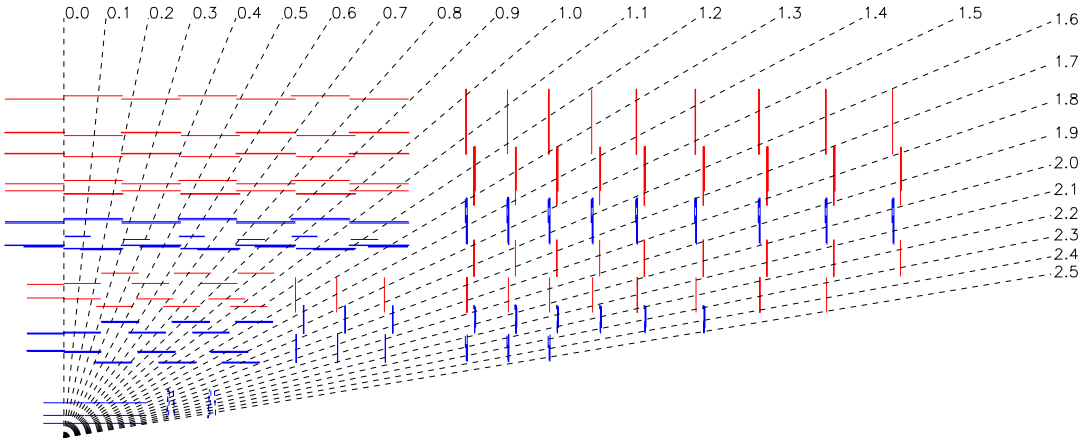


Figure 2.4: *Longitudinal view of one quarter of the silicon tracker, including the pixel detector.*

The full tracker consists of about 15000 microstrip detectors, with a pitch size ranging from about 80 to 180 μm. Some of the modules are composed by two detectors mounted back-to-back with the strips rotated by 100 mrad. These double-sided ("stereo") modules will also provide a measurement in the coordinate orthogonal to the strips.

The resolution of the full non-staged tracker for the measurement of the transverse momentum and the transverse impact parameter are shown in Fig. 2.5.

2.3 The Electromagnetic Calorimeter

The goal of the electromagnetic calorimeter (ECAL) [18] is the accurate measurement of the energy and position of electrons and photons. Its design is driven by the requirement of a 1% two-photon invariant mass resolution, in

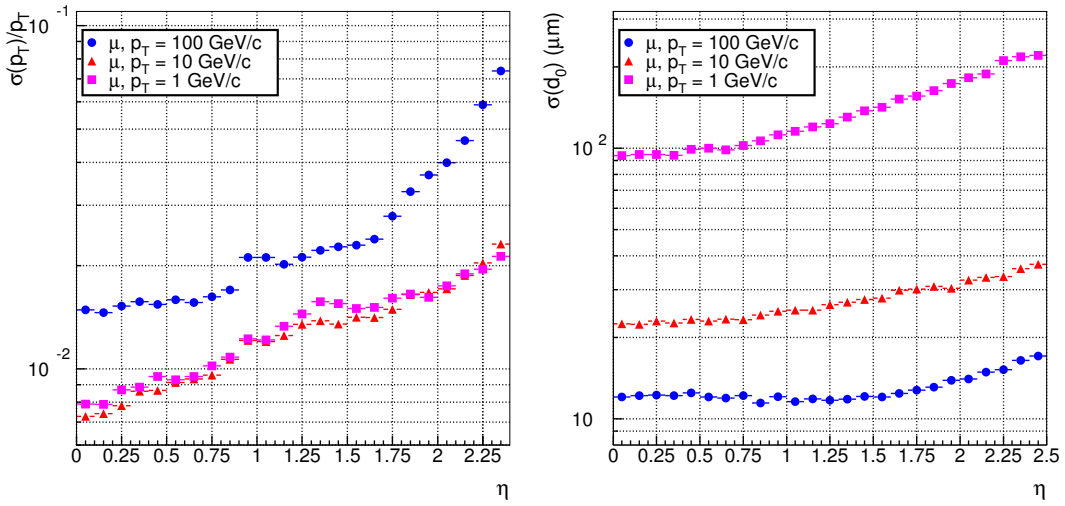


Figure 2.5: *Resolution of the tracker on the transverse momentum (left) and the transverse impact parameter (right) as a function of pseudorapidity, for single muons with transverse momentum of 1, 10 and 100 GeV/c [1].*

order to allow the observation of a low-mass Higgs in the $\gamma\gamma$ channel. A homogeneous calorimeter has therefore been chosen, composed of about 80000 finely segmented lead tungstate ($PbWO_4$) crystals. Lead tungstate is a fast, radiation-hard scintillator characterised by a small Molière radius (21.9 mm) and a short radiation length (8.9 mm), that allows good shower containment in the limited space available for the ECAL.

The length of the crystals is 230 mm in the barrel and 220 mm in the endcaps, corresponding to 25.8 and 24.7 radiation lengths respectively. Crystals are trapezoidal, with a square front face of 22×22 mm^2 in the barrel and 30×30 mm^2 in the endcaps, matching the Molière radius. Scintillator light is collected by silicon avalanche photo-diodes in the case of barrel crystals, and vacuum photo-triodes for endcaps crystals.

A preshower detector is installed in front of the endcaps, consisting of two lead radiators and two planes of silicon strip detectors, with a total radiation length of 3 X_0. It will allow rejection of photon pairs from π^0decays and improve the estimation of the direction of photons, to improve the measurement of the two-photon invariant mass.

The longitudinal view of one quarter of the ECAL is shown in Fig. 2.6. The geometric coverage extends up to $|\eta| = 3.0$.

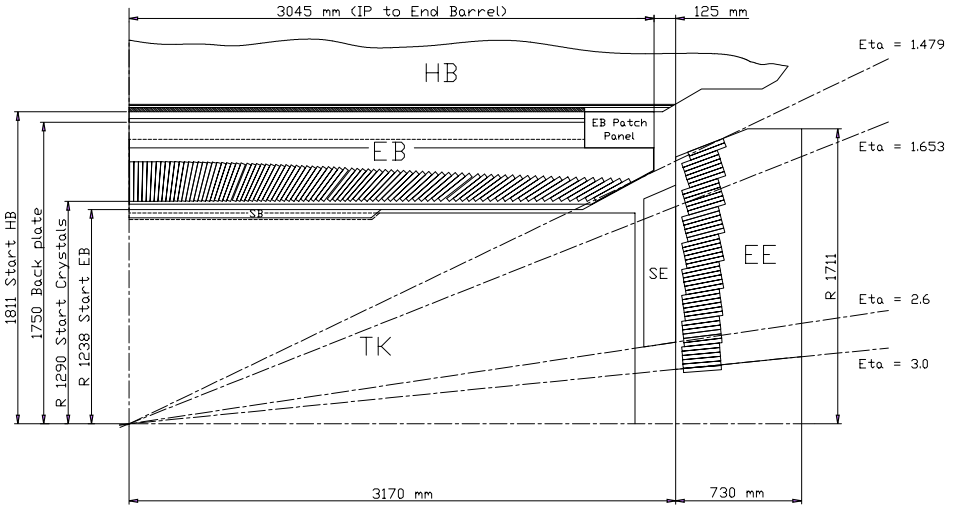


Figure 2.6: *Longitudinal view of one quarter of the ECAL.*

The energy (E) resolution of a calorimeter can be parametrised as

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{a}{\sqrt{E}}\right)^2 + \left(\frac{\sigma_n}{E}\right)^2 + c^2 \tag{2.1}$$

where a is called stochastic term and includes the effects of fluctuations in photo-statistics as well as in the shower containment, σ_n is the noise from the electronics and pile-up and c is a constant term related to the calibration of the calorimeter. The different contributions are shown in Fig. 2.7.

2.3.1 The Hadron Calorimeter

The goal of the hadron calorimeter (HCAL) [19] is is to measure the direction and energy of jets, the total transverse energy and the imbalance in the transverse energy (missing E_T). High hermeticity is required for this purpose. For this reason, the barrel and endcap parts installed inside the magnet are complemented by a very forward calorimeter which is placed outside the magnet return yokes, with a total coverage of $|\eta| < 5.3$.

The barrel and endcap HCAL cover the region $|\eta| < 3.0$. They are sampling calorimeters, whose active elements are plastic scintillators interleaved with brass absorber plates and read out by wavelength-shifting fibres. The first layer is read out separately, while all others are read out together. Both barrel and endcap are read-out in towers with a size of $\Delta\eta \times \Delta\phi = 0.087 \times 0.087$. In the barrel, full shower containment is not possible within the magnet volume, and an additional "tail catcher" is placed outside the magnet. The very forward

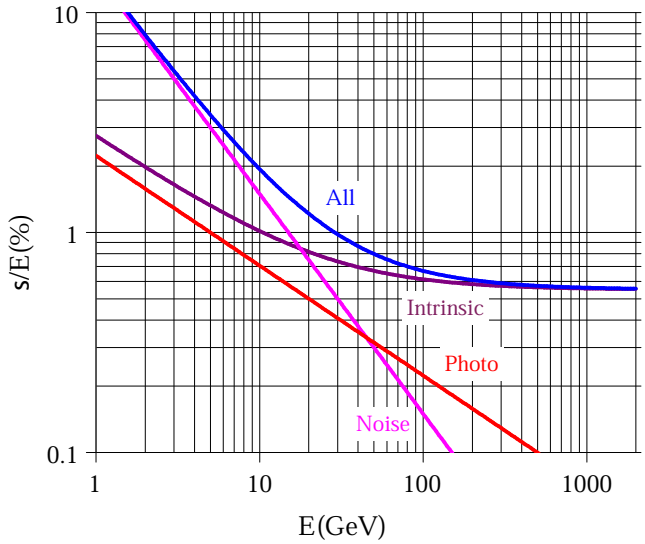


Figure 2.7: *Different contributions to the energy resolution of the ECAL [18]. The curve labelled "intrinsic" includes the shower containment and a constant term of 0.55%.*

calorimeter is placed outside the magnet yoke, 11 m from the interaction point. The active elements are quartz fibres parallel to the beam, inserted in steel absorber plates.

The energy resolution is $\sigma/E \sim 65\%\sqrt{E}\oplus 5\%$ in the barrel; $\sigma/E \sim 85\%\sqrt{E}\oplus$ 5% in the endcaps and $\sigma/E \sim 100\%\sqrt{E} \oplus 5\%$ (E in GeV) in the very forward calorimeter.

2.4 The Muon System

The goal of the muon system [20] is to identify muons and allow, in combination with the inner tracker, an accurate measurement of their transverse momenta. High-p_T muons provide a clean signature for many physics processes; therefore the muon system plays an important role in the trigger.

The muon system, shown in Fig. 2.8, is embedded in the iron return yoke of the magnet, which shields the detectors from charged particles other than muons. The magnetic field inside the plates of the yoke bends the tracks and allows the measurement of their p_T.

The muon system consists of three independent subsystems. In the barrel, where the track occupancy and the residual magnetic field are low, drift tube detectors (DT) are installed. In the endcaps, cathode strip chambers (CSC) are used, since detectors in this region have to cope with high particle rates and large residual magnetic field between the plates of the yoke. The DT and CSC

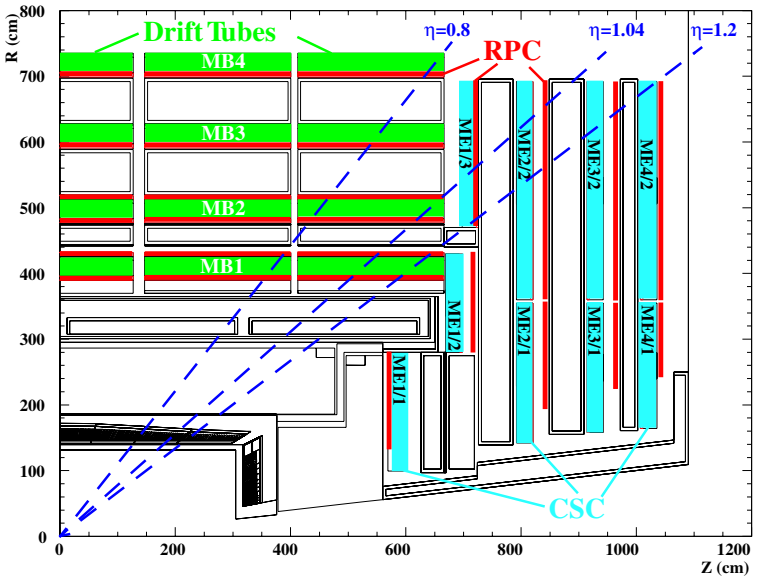


Figure 2.8: *Longitudinal view of one quarter of the muon system.*

systems cover the region $|\eta| < 2.4$. Redundancy is obtained with a system of resistive plate chambers (RPC), that are installed in both the barrel and the endcaps. RPCs have limited spatial resolution, but fast response and excellent time resolution, providing unambiguous bunch crossing identification. They are also used to complement DTs and CSCs in the measurement of the p_T. The RPC system covers the region $|\eta| < 2.1$.

2.4.1 The Drift Tube Chambers

Muon detectors in the barrel do not operate in particularly demanding conditions, since the occupancy in this region is low and the magnetic field is well contained in the iron plates of the return yoke. For this reason, drift tubes were chosen. The chamber segmentation follows that of the iron plates of the yoke, consisting in five wheels along the z-axis, each one divided in 12 sectors. Chambers are arranged in four *stations*, *i.e.* concentric cylinders, named MB1,...,MB4 as shown in Fig. 2.8. Each station consists of 12 chambers, except for MB4 where 14 chambers are present.

The basic detector element is a drift tube cell, whose section is shown in Fig. 2.9. Cells have a size of 42×13 mm. A layer of cells is obtained by two parallel aluminium planes and by "I" shaped aluminium beams which define the boundary of the cells and serve as cathodes. I-beams are insulated from the planes by a 0.5 mm thick plastic profile. The anode is a 50 μm stainless steel

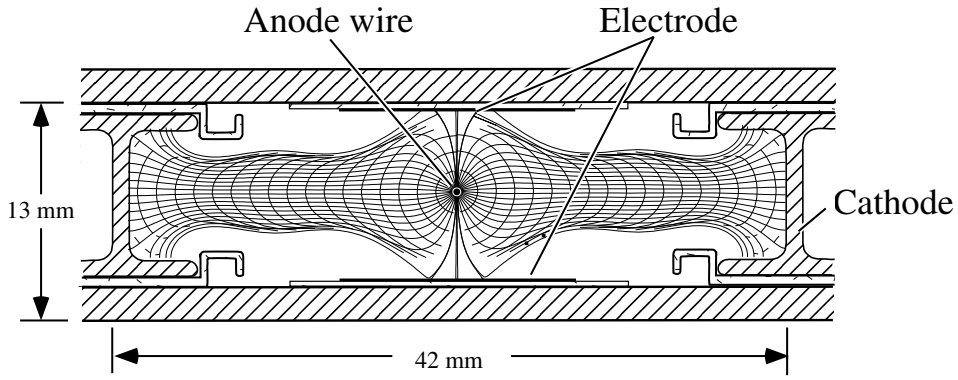


Figure 2.9: *Section of a drift tube cell.*

wire placed in the centre of the cell. The distance of the track from the wire is measured by the drift time of electrons; to improve the distance-time linearity, additional field shaping is obtained with two positively-biased insulated strips, glued on the the planes in correspondence to the wire. Typical voltages are +3600 V, +1800 V and -1200 V for the wires, the strips and the cathodes, respectively. The gas is a 85%/15% mixture of Ar/CO_2, which provides good quenching properties and a saturated drift velocity, of about 5.6 cm/μs. The maximum drift time is therefore $\sim$ 375 ns, *i.e.* 15 bunch crossings. A single cell has an efficiency of about 99.8% and a resolution of $\sim$ 180 μm.

Four staggered layers of parallel cells constitute a *super-layer*, which allows to resolve the left-right ambiguity of a single layer and provides the measurement of a two-dimensional segment. Also, it measures the bunch crossing originating a segment with no need of external input, using a generalisation of the mean-timer technique [21].

A chamber is composed by two super-layers measuring the $r-\phi$ coordinates, with the wires parallel to the beam line, and an orthogonal super-layer measuring the $r-z$ coordinates. The latter is not present in the outermost station (MB4). A cross-sectional view of a chamber is shown in Fig. 2.10.

2.4.2 The Cathode Strip Chambers

Cathode strip chambers are multi-wire proportional chambers with good spatial and time resolution, that can operate at high occupancy levels and in the presence of a large inhomogeneous magnetic field. For this reason they were adopted in the endcap region.

CSC chambers are arranged in four disks (*stations*) placed between the iron disks of the yoke and labelled ME1,...,ME4 (cf. Fig. 2.8). The innermost

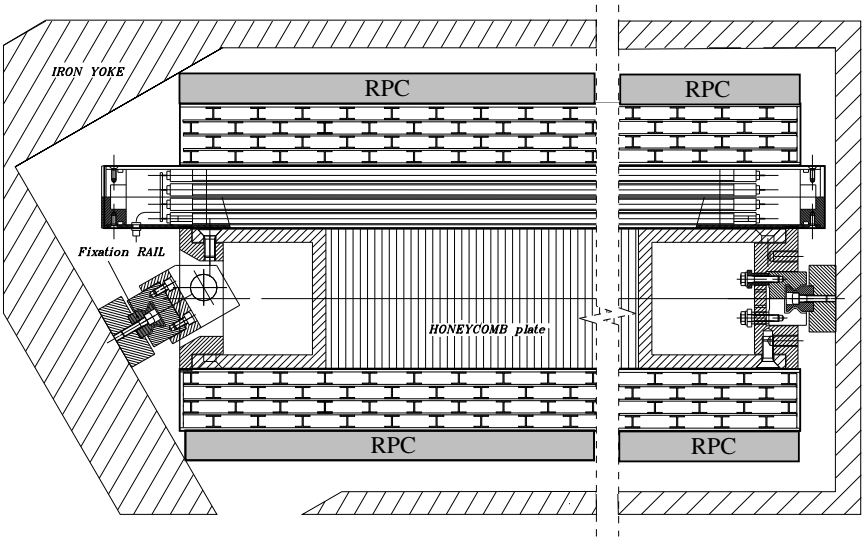


Figure 2.10: *Schematic view of a MB1/MB2 DT chamber, with two attached Resistive Plate Chambers.*

station consists of three concentric rings, the first (ME1/1) being closer to the interaction point than the other two. The other stations are composed by two disks only. The rings are formed by 18 or 36 trapezoidal chambers, which, with the exception of the outermost ring of ME1, are staggered with a small overlap in ϕ.

Chambers are composed of six layers, each consisting of an array of anode wires between two cathode planes (see Fig. 2.11). The gap is 9.5 mm thick and is filled with a 30%/50%/20% mixture of $Ar/CO_2/CF_4$. One of the two cathode planes is segmented into strips orthogonal to the wires. The avalanche produced in the gap by a crossing charged particle induces a charge in several adjacent strips, and an interpolation of the signals gives a precise spatial measurement.

Strips are radial and measure the ϕ coordinate. The orthogonal coordinate (r) is measured by the wires which, to reduce the number channels, are read out in groups of 5 to 16. The resolution is of the order of ~ 0.5 cm, to be compared with ~ 50 μm of the strip measurement.

The first disk of ME/1 has to operate in difficult conditions, as it is exposed to a high magnetic field and particle rate. A slightly different design is adopted for chambers in this disk, with wires tilted by 25° to compensate for the Lorentz drift in the magentic field. The gap is reduced to 6 mm and the number of strips is doubled above $\eta = 2.0$.

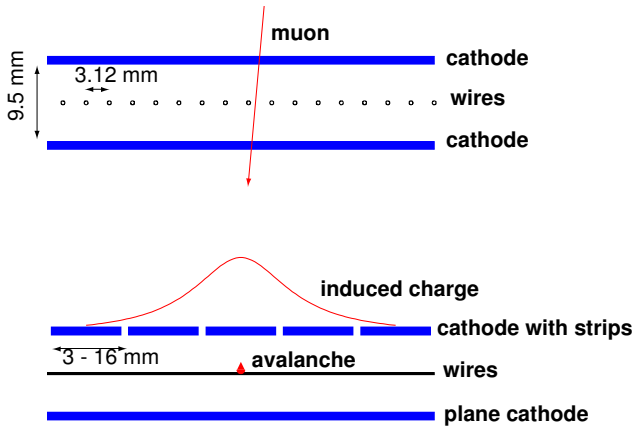


Figure 2.11: *Orthogonal sections of one CSC layer.*

2.4.3 The Resistive Plate Chambers

The RPC system is complementary to the DT and CSC systems, and adds robustness and redundancy to the muon trigger. Resistive plate chambers provide limited spatial resolution, but excellent time resolution, of the order of few nanoseconds.

In the barrel, RPC chambers follow the segmentation of DT chambers. A total of six layers of RPCs are present; the first four are attached to each side of the MB1 and MB2 DT chambers, as shown in Fig. 2.10. The other two are attached to MB3 and MB4. In the endcaps, chambers are trapezoidal; four layers are present.

The resistive plate chambers used in CMS are composed of four bakelite planes forming two coupled gaps, as shown in Fig. 2.12. The gaps are filled with a mixture of 90% $C_2H_2F_4$ (freon) and 5% i-C_4 H_{10} (isobutane). The planes are externally coated by graphite electrodes, the two innermost ones set to +9.5 kV. Insulated aluminium strips are placed in the middle, to collect the signal induced by crossing particles. This two-gap design is adopted to increase the charge induced on the strips.

In the barrel, the strips are parallel to the beam line, with a length of 80 or 120 cm. In the endcap, strips are radial with a length of 25 to 80 cm. The coordinate orthogonal to the strips is estimated as the centre of a cluster of fired adjacent strips. No measurement is available in the second coordinate, apart from the constraint coming from the strip length.

RPCs will operate in "avalanche" mode rather than in the more common "streamer" mode. This is obtained with a lower electric field, and allows to sustain higher rates. However, the gas multiplication is reduced, and improved

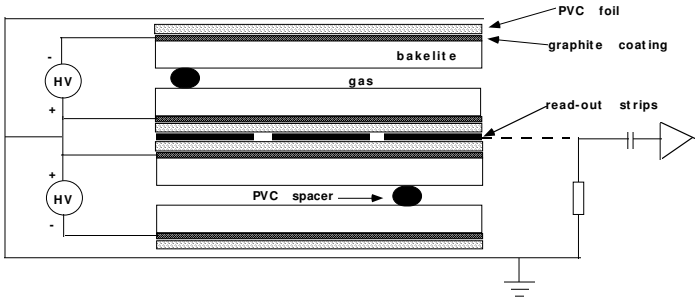


Figure 2.12: *Schema of a double-gap resistive plate chamber.*

electronic amplification is required.

RPC detectors constitute an effective and cheap alternative to scintillators. However, there are concerns about the noise rate due to discharges caused by irregularities in the bakelite surfaces. These effects can be suppressed by treating the surfaces with lineseed oil, which has been adopted for the barrel detectors. However, the high radiation dose can cause aging effects on the oil, which is therefore not used in the endcaps.

2.5 The CMS Trigger

At the LHC nominal luminosity, the total event rate is of the order of 10^9 Hz; however, the rate for interesting events is very small, as shown in Fig. 2.13. A large fraction of the corresponding selection has to be performed on-line, since the raw event size is of the order of 1 MB and storing and processing the resulting amount of data would be prohibitively difficult and expensive. The goal of the trigger system is to perform this online selection, reducing the event rate to the order of 100 Hz, that is enough to accommodate the signal channels of interest at LHC. This task is difficult not only due to the high rejection factors it requires ($\sim 10^7$), but also because the output rate is saturated already by processes like $Z \to \ell\ell$ and $W \to \ell\nu$, where high-p_T leptons are produced. The trigger must therefore be able to select events on the basis of their physics, and online selection algorithms must have a level of sophistication comparable to that of offline reconstruction.

Moreover, the time available to perform this selection is limited. Bunch crossings will occur at a rate of 40 MHz, so that a decision must be taken every 25 ns. This time is too small even to read out all raw data from the detector. The accept/reject decision will be taken in several steps (levels) of increasing refinement, where each one takes a decision using only part of the available data.

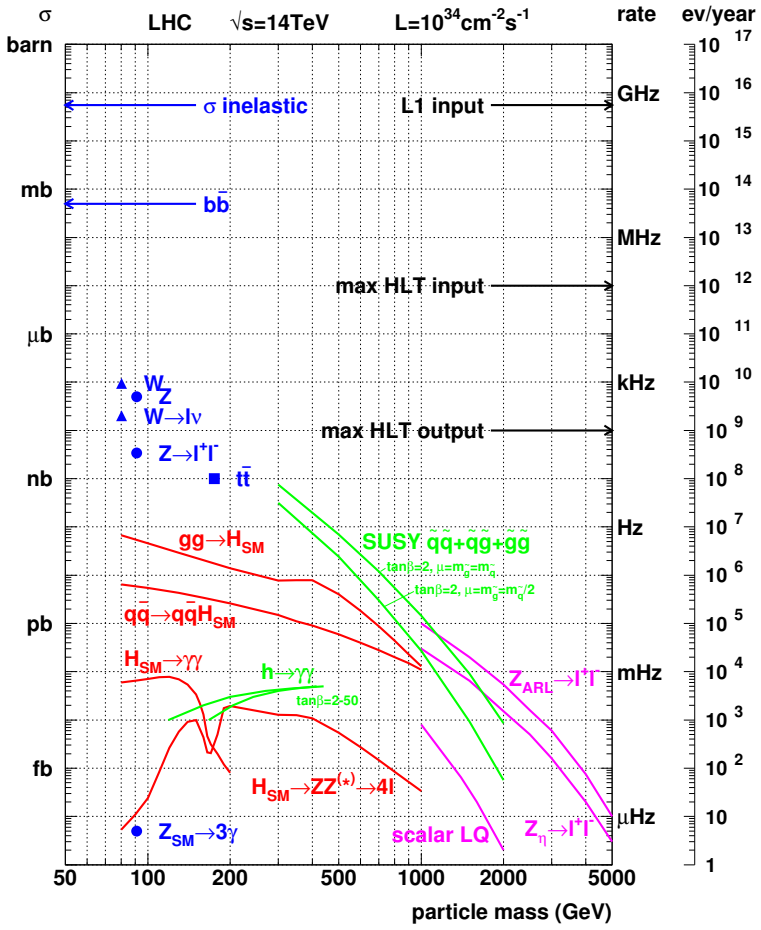


Figure 2.13: *Cross section and event rates at* $10^{34} cm^{-2} s^{-1}$ *as a function of the mass of produced objects [Courtesy G. Wrochna].*

In this way, higher trigger levels have to process fewer events and have more time available; they can go into finer detail and use more refined algorithms.

Due to the strict timing constraints, the first step of the CMS trigger, the Level-1 trigger, is implemented on dedicated programmable hardware. The Level-1 has access to the data from the calorimeters and the muon detectors with coarse granularity; on the basis of this limited information it has to reduce the input rate up to a level acceptable for the Data Acquisition system (DAQ). At startup, the DAQ system will be able to handle an event rate of up to 50 kHz, which will be increased to 100 kHz when the full LHC design luminosity is reached. Only one third of this bandwidth is allocated, the rest being used as safety margin accounting for all uncertainties in the simulation of the basic physics processes. The allocation of this bandwidth among different triggers is discussed in Section 2.5.1.4.

The selected events are passed to the High Level Trigger (HLT), which will be completely implemented on software running on a farm of commercial processors. This allows full flexibility and optimisation of the algorithms. The HLT is further subdivided in logical levels (Level-2, Level-3), although this classification is somewhat arbitrary. The implementation and performance of the muon HLT algorithms is described in Chapter 4.

The following sections briefly describe the design of the Level-1 trigger, with particular focus on the muon trigger. A more detailed discussion can be found elsewhere [22].

2.5.1 The Level-1 Trigger

The Level-1 trigger [22] is implemented on custom-built programmable hardware. It runs dead-time free and has to take an accept/reject decision for each bunch crossing, i.e. every 25 ns. This is achieved with a synchronous pipelined structure of processing elements, each taking less than 25 ns to complete. At every bunch crossing, each processing element passes its results to the next element and receives a new event to analyse. During this process, the complete detector data are stored in pipeline memories, whose depth is technically limited to 128 bunch crossings. The Level-1 decision is therefore taken after a fixed time of 3.2 μs. This time must include also the transmission time between the detector and the counting room (a cable path of up to 90 m each way) and, in the case of Drift Tube detectors, the electron drift times (up to 400 ns). The time available for calculations can therefore be as low as 1 μs.

The Level-1 trigger is divided in three subsystems: the Calorimeter Trigger, the Muon Trigger and the Global Trigger. The Muon Trigger is further subdivided in three independent systems for the DTs, CSCs and RPCs, respectively. The results of these three systems are combined by the Global Muon Trigger. A schematic view of the components of the Level-1 trigger system and of their relationships is shown in Fig. 2.14.

The Calorimeter and Muon Triggers do not perform any selection themselves. They identify "trigger objects" of different types: isolated and non-isolated electrons/photons; forward, central and τ-jets; and muons. The four best candidates of each type are selected and sent to the Global Trigger, together with the measurement of their position, transverse energy or momentum and a quality word. The Global Trigger also receives the total and missing transverse energy measurement from the Calorimeter Trigger.

The Global Trigger selects the events according to programmable trigger conditions, that can include requirements on the presence of several different objects with energies or momenta above predefined thresholds. Topological conditions and correlations between objects can be required as well. Up of 128 of these conditions can be tested in parallel, and each can be pre-scaled to

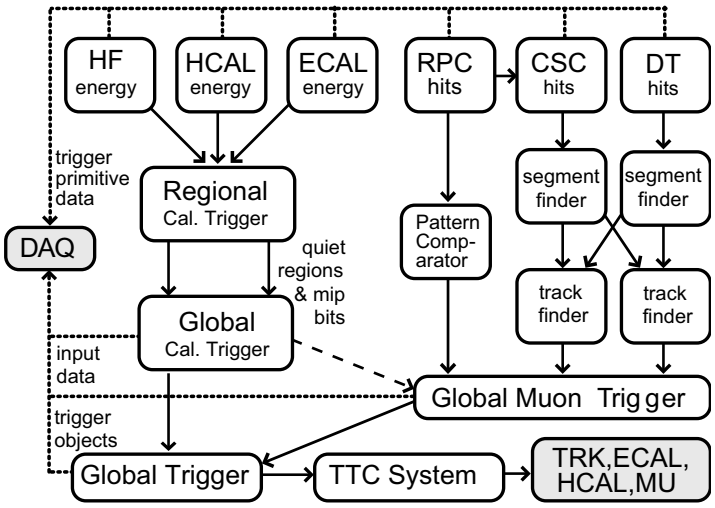


Figure 2.14: *Structure of the Level-1 trigger system.*

accept only a fraction of selected events.

The Level-1 Calorimeter and Muon Triggers are shortly described in the following. A more detailed discussion is outside the scope of this work, and the interested reader is referred to more specific publications [22].

2.5.1.1 The Level-1 Calorimeter Trigger

The calorimeter trigger identifies five types of objetcs: isolated electrons/photons, non-isolated electrons/photons, central jets, forward jets and tau jets. The measurement of the transverse energy and position of the four most energetic objects of each type is sent to the Global Trigger, together with a measurement of the total and missing transverse energy. In addition, the calorimeter trigger provides the Muon Trigger information about the activity in all calorimeter regions, to determine if the energy deposit is compatible with the passage of a muon (MIP bit) and if it is below a programmable threshold (Quiet bit).

For trigger purposes, the calorimeters are subdivided in towers with a size of $\Delta\phi \times \Delta\eta = 0.087 \times 0.087$ up to $\eta \sim 2$. At higher pseudorapidity values $\Delta\eta$ increases up to 0.35. Trigger towers match the granularity of HCAL up to $\eta > 1.74$; above that value, physical HCAL towers have twice the ϕ dimension of the trigger tower. In the barrel ECAL, each tower corresponds to 5×5 crystals, while the ECAL endcap crystals are arranged in a $x - y$ geometry, and a variable number of crystals is grouped, matching as much as possible the HCAL trigger tower boundaries. Towers are defined also in the very forward calorimeter, with a size of $\Delta\phi \times \Delta\eta = 0.348 \times 0.5$.

The trigger towers are organised in calorimeter regions, each formed by 4×4 trigger towers, with a size of about $\Delta\phi \times \Delta\eta = 0.35 \times 0.35$. Very forward calorimeter towers constitute a region by themselves, due to their size.

The data of each ECAL and HCAL trigger tower is first processed by the Trigger Primitive Generator, which is integrated in the calorimeter readout electronics. It provides bunch crossing identification based on a peak-finder algorithm, and for each tower calculates the so-called trigger primitives, *i.e.* the sum of the transverse energy and a fine grain bit. The ECAL fine grain bit provides information on the lateral extension of the electromagnetic shower, and is used in the rejection of backgrounds by the electron trigger. The HCAL fine grain bit is used to denote the compatibility of the deposit with the passage of a minimum ionising particle, and is set if the HCAL energy before conversion to the transverse scale is within a programmable range, of the order of [1.5, 2.5] GeV.

The trigger primitives are used by the different calorimeter trigger algorithms described in the following.

Photon and electron trigger. At Level-1, it is not possible to distinguish electrons and photons, which are treated together. Electron/photon candidates are found with a sliding window algorithm on 3×3 towers. Identification is based on the presence of a large energy deposition in one or two adjacent trigger towers. Requirements on the lateral and longitudinal profile of the shower are also set using the ECAL fine grain bit and the ratio of ECAL and HCAL energy deposits, respectively. A candidate is labelled isolated on the basis of the energy deposits and ECAL fine grain bits in the eight towers around the centre of the 3×3 window.

In each calorimeter region, the highest-E_T isolated and non-isolated candidates are found. The Global Calorimeter Trigger sends the four most energetic ones of each type to the Global Trigger.

Jets and τ-jets. The jet trigger uses the sum of the ECAL and HCAL transverse energies in calorimeter regions. Candidates are found with a sliding window algorithm on 3×3 regions, requiring the deposit in the central one to be higher than the deposits in the other eight.

Additionally, τ-jets are identified by their narrow profile. A jet candidate is identified as τ-jet if each of the nine regions of the window contains no more than two towers above a programmable threshold.

Jets are searched for separately in the central region ($|\eta| < 3$) and in the forward region ($3 < |\eta| < 5$), while τ jets are only identified in the central one. The Global Calorimeter Trigger sends the four most energetic candidates of each type to the Global Trigger, together with the number of jets above a

programmable threshold.

Total and missing transverse energy. The total energy is obtained by summing the transverse energy of all ECAL and HCAL calorimeter regions. The missing transverse energy is determined from the sum of the E_x and E_y components of the deposit in each region, obtained from the E_T deposits using the coordinates of the centre of the region. The total and the missing energy (absolute value and ϕ direction) are then sent to the Global Trigger.

Quiet and MIP bits. For each calorimeter region, a "Quiet" bit is defined if the transverse energy deposit in ECAL plus HCAL is below a programmable threshold. The MIP (Minimum Ionising Particle) bit is set if the quiet bit is set and if the HCAL fine grain bit is set in at least one of the 16 HCAL towers of the region. These two bits can be used by the Global Muon Trigger to select muons, as described in Section 5.1.

2.5.1.2 The Level-1 Muon Trigger

The Muon Trigger has the task to identify muons, reconstruct their position and transverse momentum and provide bunch crossing assignment with high purity and efficiency. It is described in detail in [22]. The most relevant change from what described therein is that the trigger electronics will not be installed in the forward CSC station ME1/1a, thus limiting the Level-1 trigger acceptance to $|\eta| < 2.1$ (see below).

All three detector systems described in Section 2.4 contribute to the muon trigger, which benefits from the complementary characteristics of these systems: the good spatial resolution of drift tubes and cathode strip chambers and the excellent time resolution of resistive plate chambers. The redundancy of the muon system allows a robust trigger with high efficiency and good background rejection. The wire chamber systems and the RPC system are complementary in performance and also in the backgrounds and inefficiencies they are sensitive to. The complementarity can also be used for cross-checks and to improve the understanding of the performance of each system.

The internal structure of the Muon Trigger is shown in Fig. 2.15. In the case of DTs and CSCs, the information of each chamber is first processed independently by a Local Trigger step, where track segments are reconstructed. Segments of the different stations are matched by the DT and CSC Regional Track Finders, which reconstruct muon tracks and estimate their p_T. The candidates found are then delivered to the Global Muon Trigger, with a word to indicate their quality. In the overlap region, DT and CSC segments are used by both Track Finders to allow for reconstruction of full tracks in each of the two subsystems.

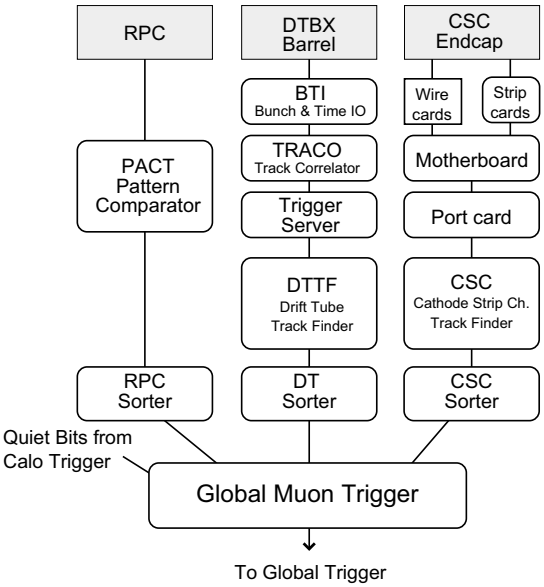


Figure 2.15: *Schematic structure of the muon trigger system.*

In the case of RPCs, the hits are collected by a Pattern Comparator Trigger (PACT), which looks for predefined patterns. The PACT provides an estimate of the p_T of the muon and its position, as well as a word to indicate its quality.

Each subsystem reconstructs up to four muon candidates. These are sent to the Global Muon Trigger that matches them and looks for the MIP and Quiet bits in the corresponding calorimeter regions. Finally, the four muons with highest p_T are sent to the Global Trigger. The track reconstruction in the different subsystems is shortly described in the following.

The DT Trigger. The DT trigger front-end is called Bunch and Track Identifier (BTI). It is directly connected to the read-out electronics, and performs a straight segment fit within a superlayer using at least three hits out of the four layers of drift cells. The fit is done with a generalisation of the mean-timer technique [21], which also returns the unknown bunch crossing originating the segment. In $r-z$ superlayers, only segments pointing to the interaction point are selected. The segments reconstructed in the two $r-\phi$ superlayers in each chamber are matched by the Track Correlator (TRACO), that improves the angular resolution thanks to the bigger lever arm. The Trigger Server (TS) selects, among all segment pairs in a chamber, the two corresponding to the highest p_T, and forwards them to the Drift Tube Track Finder (DTTF).

The DTTF matches the segments reconstructed in the four stations into a single muon track candidate, assigning the track parameters p_T, η, ϕ and a

quality word. This device is based on precomputed, memory-resident Look-Up Tables (LUT), which are used to extrapolate the segments between stations and to group matching segments. The track parameters are estimated with other LUTs on the basis of the the ϕ direction of the segments in the two innermost stations.

Finally, the candidates are sorted, and the four highest p_T muon candidates are delivered to the Global Muon Trigger.

The CSC Trigger. As for the barrel, the first trigger step in the endcaps is the local reconstruction. The CSC Local Trigger reconstructs segments independently using the strips and the wires of the six layers of each chamber. Cathode strips are used to reconstruct the ϕ coordinate and measure the transverse momentum of the track. First, the strips are digitised with half strip resolution, using a simple interpolation based on a comparator that analyses the charge in three adjacent strips. The hits in the six layers are then searched for patterns compatible with high-p_T tracks.

Anode wires are used to reconstruct the η coordinate and to provide a precise bunch crossing assignment. Wires are read out in groups of 5 to 16, to reduce the number of channels. Hit patterns compatible with a track coming from the nominal interaction point are searched for with a coincidence technique. A coincidence of two hits in different layers is needed to assign the bunch crossing, while the reconstruction of a segment requires four hits out of the six CSC planes.

Finally, the time coincidence of anode and cathode segments is required to construct three-dimensional segments. They are sent to the CSC Track Finder, which links the segments of the various stations to form a muon candidate and assigns p_T, η, ϕ and a quality word, using LUTs as in the case of the DTs.

The CSC system covers the region up to $|\eta| = 2.4$. However, the trigger electronics will not be installed in the region above $|\eta| = 2.1$ in the chamber ME1/1, effectively limiting the muon trigger geometrical acceptance to $|\eta| < 2.1$.

The RPC Trigger. No local reconstruction step is present in the RPC trigger [22], since the measurements in each chamber are simple points. The hits are collected by the Pattern Comparator Trigger (PACT), which looks for hits correlated in space and time in the four RPC stations. Hits are matched with pre-defined patterns in a large look-up table to provide identification and estimate of the p_T. The patterns allow to identify muons with at least three hits on four stations. In the case of the barrel, where six stations are present, the search is done independently for low-p_T muons in the first four layers (in MB1 and MB2) and for high-p_T ones using one layer in each station.

A ghost suppression algorithm is applied to reduce the effect of accidental

coincidences due to background hits. The four highest p_T candidates in the barrel and endcaps are then separately sent to the Global Muon Trigger.

The Global Muon Trigger. The task of the Global Muon Trigger (GMT) is to match the muon candidates from the different subsystems and combine their parameters in an optimal way. The matching is done by comparing the spatial coordinates of the segment (ϕ, η), and can be tuned to achieve the optimal balance between efficiency and background suppression. High efficiency can be obtained accepting candidates even if they are found by only one subsystem. Maximum background rejection can be obtained requiring all candidates to be reconstructed by both subsystems. The price is a lower efficiency. More refined criteria can be used; in the current implementation, candidates are accepted if they are reconstructed by two systems, otherwise they are selected on the basis of their quality word. Low-quality candidates from problematic η regions are discarded [23]. If two candidates are matched, the parameters of the track are chosen according to a programmable logic.

The GMT also assigns each candidate the MIP and Quiet bit of the corresponding calorimeter region (cf. Section 2.5.1.1). These bits can optionally be used to confirm the muon candidate and require that it is isolated. The performance of this selection is discussed in Section 5.1.

2.5.1.3 Performance of the Level-1 Muon Trigger Selection

The simplest possible Level-1 muon trigger selection is based on a threshold on the p_T of GMT tracks. The Level-1 p_T scale is defined at 90% efficiency, that is the scale where, for each given threshold, the efficiency for muons with a p_T equal to the threshold is 90% of the plateau efficiency. In other words, the p_T value reported by the trigger is such that there is only a 10% probability for the true value to be actually higher.

The resolution of the reciprocal of the transverse momentum, $1/p_T$, is shown in Fig. 2.16. The distributions are shifted with respect to zero due to the 90% scale definition. The resolution is about 17% in the barrel, 20% in the endcaps and 22% in the overlap region. Non-gaussian tails are present and, together with the limited resolution, are responsible for the "feed-through" of low p_T muons that are reconstructed at high momenta, thus passing typical trigger thresholds. This effect is important since the muon rate increases steeply for low p_T values, so that the contribution of feed-through muons to the Level-1 trigger rate is dominant even for high thresholds. This effect is visible in Fig. 2.17, where the spectrum of the generated p_T of muons passing a threshold of 14 GeV/c at low luminosity, and of 25 GeV/c at high luminosity, is shown.

This feed-through effect can be reduced only by improving the p_T resolution. This is the task of the Muon High Level triggers. The event rate corresponding

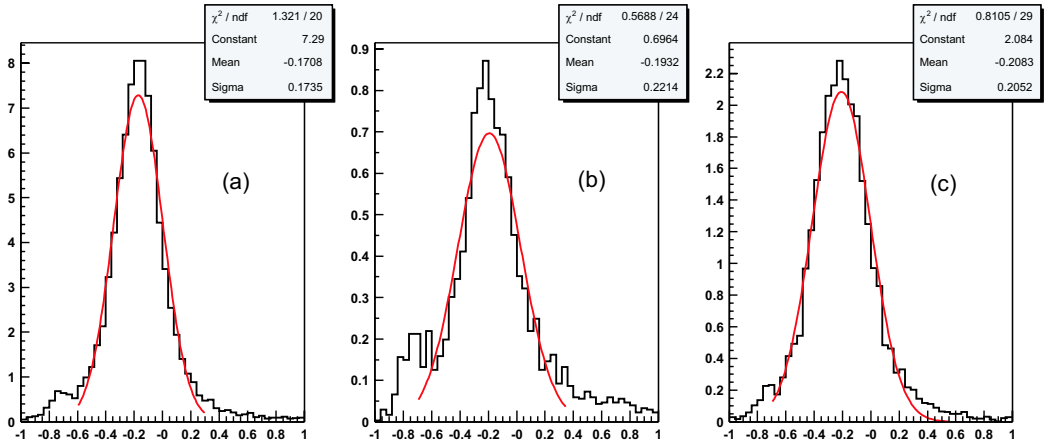


Figure 2.16: *Level-1 resolution on $1/p_T$ obtained with a sample of W decays at high luminosity for (a) the barrel region ($|\eta|$ <0.8); (b) the overlap region (0.8< $|\eta|$ <1.2); (c) the endcap region (1.2< $|\eta|$ <2.1) [1].*

to the selection of one or two muons at each trigger level (single- and di-muon selection) will be discussed in Section 4.7.

2.5.1.4 The Level-1 Trigger Table

The Level-1 trigger allows to define complex trigger algorithms based on the presence of several, different objects and on topological conditions and correlations. However, whenever possible, "inclusive" criteria should be used, to avoid biasing the sample of selected events. The simplest triggers are in general those based on the presence of one object with an E_T or p_T above a predefined threshold (single-object triggers) and those based on the presence of two objects of the same type (di-object triggers) with either symmetric or asymmetric thresholds. Other requirements are those for multiple objects of the same or different types ("mixed" and multiple-object triggers). In the case of special channels that are not efficiently selected by these simple criteria, very specific exclusive algorithms can be used. These cases are not discussed in the following.

The allocation of the Level-1 bandwidth to the different trigger streams is discussed in [1]. The choice of the Level-1 trigger thresholds is determined by the maximum event rate (bandwidth) that can be accepted by the DAQ system. The current estimate is that at startup the DAQ system will be able to handle an event rate of up to 50 kHz, which will be increased to 100 kHz when the full LHC design luminosity is reached. Only one third of this bandwidth is allocated, the rest being used as safety margin accounting for all uncertainties

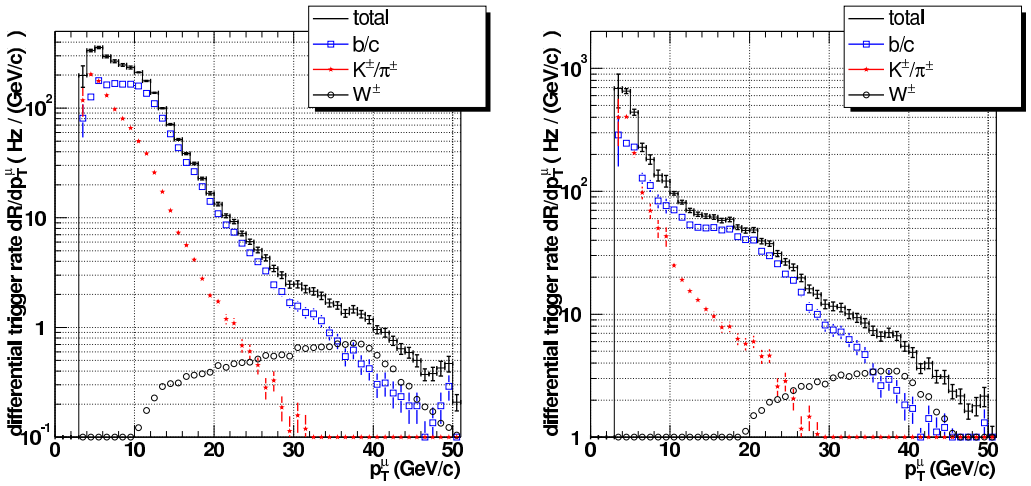


Figure 2.17: *Event rates as a function of the generated p_T for a Level-1 threshold of 14 GeV/c^2 at low luminosity (left) and of 25 GeV/c^2 at high luminosity (right). The separate contributions of b/c, K/π and W decays are shown [23].*

in the simulation of the basic physics processes, the CMS detector, and the beam conditions [1]. This bandwidth is then subdivided among the Level-1 objects described in the previous sections (muons, electrons and photons, tau jets, jets and combined channels) and for each of them between the single- and multiple-object streams.

The result is a set of thresholds called trigger table. The present Level-1 trigger tables at low and high luminosity are shown in Table 2.1 and Table 2.2, respectively.

Table 2.1: *L1 trigger table at low luminosity [1].*

Trigger	Threshold (GeV or GeV/c)	Rate (kHz)	Cumulative Rate (kHz)
Inclusive isolated electron/photon	29	3.3	3.3
Di-electrons/di-photons	17	1.3	4.3
Inclusive muon	14	2.7	7.0
Di-muons	3	0.9	7.9
Single tau-jet trigger	86	2.2	10.1
Two tau-jets	59	1.0	10.9
1-jet, 3-jets, 4-jets	177, 86, 70	3.0	12.5
Jet * ET miss	88 * 46	2.3	14.3
Electron * Jet	21 * 45	0.8	15.1
Minimum-bias (calibration)		0.9	16.0
Total			16.0

Table 2.2: *L1 trigger table at high luminosity [1].*

Trigger	Threshold (GeV or GeV/c)	Rate (kHz)	Cumulative Rate (kHz)
Inclusive isolated electron/photon	34	6.5	6.5
Di-electrons/di-photons	19	3.3	9.4
Inclusive muon	20	6.2	15.6
Di-muons	5	1.7	17.3
Single tau-jet trigger	101	5.3	22.6
Two tau-jets	67	3.6	25.0
1-jet, 3-jets, 4-jets	250, 110, 95	3.0	26.7
Jet * ET miss	113 * 70	4.5	30.4
Electron * Jet	25 * 52	1.3	31.7
Muon * Jet	15 * 40	0.8	32.5
Minimum-bias (calibration)		1.0	33.5
Total			33.5

Chapter 3

Event Simulation

As already discussed in Section 2.5, many interesting LHC processes have high-p_T muons in the final state and their selection in the trigger can be based on the request of a minimum muon p_T. However, the muon trigger has to deal with a copious background of real muons, which also may have large p_T – both *prompt* muons from b, c quarks and W or Z decays and *non-prompt* muons from pion and kaon decays.

A critical constraint on the muon selection is the maximum acceptable rate at each trigger level. The determination of the total rate as a function of the p_T threshold requires the simulation of an inclusive background sample large enough to allow good precision for rejection factors up to the order of 10^5. Since any π or K meson and b or c quark can decay into a muon, each LHC event can potentially cause a muon trigger, and the entire inelastic LHC cross section must be simulated. Moreover, the muon p_T spectrum must be simulated down to very low p_T values to take into account the triggers caused by mismeasured muons, relevant especially for the Level-1 trigger [24]. The full simulation of such a sample is very demanding, both in terms of the data volume to be stored and the CPU time needed to produce and process it. A special procedure has therefore been developed in order to obtain the maximum precision with an affordable number of simulated events.

This chapter describes the details of the event generation and detector simulation of the samples produced in the year 2002 for the study of the CMS muon trigger, which was organised in the following steps.

Event generation. The kinematic properties of the events and their stable particle content were determined using the PYTHIA [25] Monte Carlo generator, as described in Section 3.1. The output was stored in the standard HEPEVT structure in HBOOK ntuples. The procedure adopted to optimise the production of inclusive minimum bias samples is described in Section 3.1.1. Other samples of events were also produced, including signal samples and pile-

up events, as described in Section 3.1.2 and 3.1.3.

Tracking in the apparatus. The detailed simulation of the CMS detector was performed using CMSIM [26], which produces the particle *hits* in the detectors. This step is described in Section 3.2. The hits were stored in ZEBRA files.

Reformatting of the hits. This step simply consists in the conversion from the ZEBRA format used by CMSIM to the object database used by ORCA [27], the CMS object-oriented reconstruction package. From this step on, all information related to the event was stored in the same object database.

Digitisation. The response of the detectors to the particle hits was simulated with ORCA version 6, taking into account the pile-up of events in the same and contiguous bunch crossings. The final product is given by the digitised hits (*digis*), which can be used as input for the trigger simulation and reconstruction programs. Digitisation is discussed in Section 3.2.

The resulting inclusive single and di-muon rates within the kinematic acceptance of the CMS detector are given in Section 3.3.

3.1 Event Generation

The data generated for the muon Level-1 trigger and High Level Trigger (HLT) studies consist of a sample of about 10^6 fully simulated Monte Carlo events, split into several datasets corresponding to different physics processes. The bulk of the production consisted in the simulation of the main source of background, *i.e.* minimum bias events having at least one final state muon with transverse momentum high enough to reach the muon trigger system. Since the full simulation of such a sample is very demanding, a special weighting procedure was developed, as described in Section 3.1.1. In addition to the minimum bias background, other physics processes which are sources of high p_T muons were simulated: the production of both W and Z/γ^* bosons and of top quark pairs. The generation of these samples is described in Section 3.1.2, together with a number of signal channels which have been used as benchmarks for the trigger signal efficiency.

The event generation was performed using the PYTHIA 6.158 [25] Monte Carlo generator. The PYTHIA code specific to the generation of these datasets and the full set of cards used are available on the on-line CMS production database [28]. The values of the relevant PYTHIA parameters used in the simulation are given in Table 3.1. All other parameters were left to the default value.

In particular, the GRV94 [29] p.d.f., which is the default for PYTHIA 6.158, was used in the generation of all samples.

Table 3.1: *Values of the PYTHIA parameters used for the generation. The values indicated in italics are equal to the PYTHIA 6.158 defaults.*

Parameter	Value	Sample	Description
`CKIN(3)`	10.0	"High-p_T" MB	$\hat{p}_T$ cut-off
	0.0	all other samples	$\hat{p}_T$ cut-off
`MSTP(81)`	*1*	All	Activate multiple interactions (M.I.)
`MSTP(82)`	*1*	b/c in MB events	Simple M.I. model
	4	All other events	Complex M.I. model
`MSTJ(11)`	3	All	Hybrid model for fragmentation

Evidence for multiple interactions within the same hadron-hadron scattering has been shown by collider data [30]. To simulate these multiple interactions, a model that assumes a double Gaussian distribution for the spatial distribution of the hadronic matter within the proton and a varying impact parameter between the two interacting partons was chosen (`MSTP(82)=4`, "complex" model in the PYTHIA terminology)[1]. However, since this complex model is very CPU-time expensive, the PYTHIA default "simple model" (`MSTP(82)=1`) was used in the dedicated generation phase where $b\bar{b}$ and $c\bar{c}$ events are produced (cf. Section 3.1.1), and where a large number of non-b, c events must be rejected in order to get a statistically enriched sample of heavy flavour events. In the simple model, the charged multiplicity of stable particles with $p_T > 1$ GeV/c is about 30% lower than in the complex model. The latter was conservatively adopted also for the minimum bias production of pile-up events (see Section 3.1.3).

Experimental data on fragmentation of jets containing c and b quarks indicate that the fragmentation function $f(z)$, where z is the fraction of the quark longitudinal momentum taken by the hadron at a given fragmentation step, is harder than in jets containing light quarks. For this reason, the hybrid model in PYTHIA (`MSTJ(11)=3`) based on the Lund fragmentation scheme for light quarks and the Peterson fragmentation function for c and b quarks was used, with the PYTHIA default values of the Peterson parameters, $\epsilon_c = 0.05$ and $\epsilon_b = 0.005$, in fair agreement with the experimental data [33].

Only events with at least one muon within the muon system acceptance were stored for the following processing step, the CMSIM detector simulation.

[1] Some indication in favour of multiple interactions models with varying impact parameter comes from the UA5 [31] and CDF [32] results on charged particle multiplicity.

3.1.1 Generation of Minimum Bias Event Samples

Minimum bias events were generated by setting `MSEL=1` in PYTHIA, which includes both QCD 2→2 parton scattering and low-p_T soft processes (where the protons are not resolved into the constituents partons, but interact as a whole). In order to efficiently generate a minimum bias sample with muons, one or more of the potential muon parents in each event was forced to decay, and the probability of the particular final state to occur was assigned as a weight. The details of this weighting procedure are described in Sections 3.1.1.1 and 3.1.1.2.

The event generation was performed in "runs", each corresponding to one PYTHIA batch job generating few hundred events. In each run, the generation was performed according to the two-step procedure described below, and the events generated in the two steps were concatenated into a single output file for each run. The first production phase was devoted to the generation of minimum bias events with only light quarks in the final state. This was done by explicitly vetoing events with b/c quarks in the hard scattering or produced from gluon splitting into $b\bar{b}$ or $c\bar{c}$. In the second phase, a dedicated generation of b/c events from all possible production mechanisms (parton fusion, flavour excitation, gluon splitting) was performed, with an integrated luminosity a factor 10 higher with respect to the first production phase[2]. This was meant to increase the statistical significance of b/c processes, which, although characterised by smaller cross sections compared to light flavour processes, can significantly contribute to trigger rates at high p_T. In addition, the total $b\bar{b}$ cross section was normalised to 500 μb, following the recommendation of dedicated studies [8] and consistently with the approach [34] followed by the PRS b/τ[3] group for the generation of $b\bar{b}$ minimum bias events.

The generation of minimum bias events was performed with inclusive kinematic requirements on the muons in the final state, in order to accurately estimate the background contribution to the accepted rate at the Level-1 and HLT trigger stages. Although single muon triggers will necessarily have rather high p_T thresholds (in the 20-40 GeV/c range), it is important to simulate low p_T muons to correctly account for feed-through effects (cf. Section 2.5.1.3) due to multiple scattering in the calorimeters and in the iron yoke. To obtain good statistical significance over a wide p_T spectrum, the generation was done in three different bins of transverse momentum. In each of these samples, events were required to have at least one muon within the acceptance of the detector, with a predefined momentum or transverse momentum cutoff:

[2] The weight of each event in the second production phase is scaled down by a factor 10, to preserve the correct normalisation of the rates.

[3] CMS working group for Physics, Reconstruction and Selection (PRS) activities related to the CMS tracker.

- "Low-p_T" sample: a pseudorapidity-dependent cut was applied, in order to take into account the different minimum p_T values needed for a muon to reach the barrel, the endcaps and the overlap region of the muon system. The minimum generated p_T was 3 GeV/c in the barrel ($|\eta| < 1.2$) and 1.8 GeV/c in the overlap region ($1.2 < |\eta| < 1.7$), while in the forward regions ($1.7 < |\eta| < 2.5$) a momentum of $p > 3.5$ GeV/c was required. To avoid overlap with the other two samples, events with muons having $p_T > 4$ GeV/c were rejected;

- "Intermediate-p_T" sample: a transverse momentum above 4 GeV/c was required within the full pseudorapidity range. In this case, the muon spectrum was simulated with no upper bound, to allow comparison with the "High-p_T" sample (see below). To avoid double counting, events with muons with $p_T > 10$ GeV/c must be excluded when analysing this sample;

- "High-p_T" sample: a transverse momentum above 10 GeV/c was required within the full pseudorapidity range. Since the fraction of minimum bias events with muons above this p_T threshold is small, the generation of this sample is more CPU-time expensive than that of the previous two samples. In order to enrich the sample with high-p_T muons and speed up the event generation, only events with a minimum transverse momentum of the parton level hard process ($\hat{p}_T$) above 10 GeV/c were generated. However, since it has been observed that the $\hat{p}_T > 10$ GeV/c cutoff introduces an unphysical p_T-dependence of the cross section when the multiple interaction complex model is used, a scaling factor was applied to the events with light flavour quarks, in order to consistently normalise the event weights with the other samples. Its value, 0.645, was determined by comparing the rate of muons from this sample with the rate of muons with $p_T > 10$ GeV/c in the intermediate-p_T sample, which was generated with no $\hat{p}_T$ cutoff. No scaling factor was necessary for the heavy flavour component, where the treatment of multiple interactions is according to the simple model.

The kinematic selection, cross section and statistics for these three samples are listed in Table 3.2, together with other parameters that are defined in the following sections.

The distribution of the generated muon momentum as a function of η in the minimum bias samples is shown in Fig. 3.1.

Table 3.2: *Parameters, cross sections and statistics of the three minimum bias samples. References to the definitions of the different columns are given in the legend below.*

$\hat{p}_T$ cut	Minimum transverse momentum of the hard process at parton level
σ_{gen}	Generated cross section
I_{int}	Integrated luminosity
N_{sel}	Number of events selected within the kinematic acceptance (after skipping K/π which interacted hadronically in the CMSIM simulation, cf. Section 3.2)
σ_{sel}	Cross section of selected events
$\bar{w}$	Average weight of selected events (cf. Section 3.1.1.2)
s	Loss of significance due to weight fluctuations (cf. eq. 3.2)
I_{int}^{eq}	Equivalent integrated luminosity (cf. Eq. 3.4)

Sample	p_T^{μ} range (GeV/c)	$\hat{p}_T$ cut (GeV/c)	σ_{gen} (mb)	I_{int} (nb^{-1})	N_{sel} $\times 10^3$	σ_{sel} (nb)	$\bar{w}$	s	I_{int}^{eq} (nb^{-1})
Low-p_T	see text	0	55.22	0.0246	146	685	0.116	1.13	0.187
Interm.-p_T	[4,10)	0	55.22	0.991	248	25.1	0.100	1.01	9.78
High-p_T	≥ 10	10	2.66	11.4	87	0.706	0.093	1.11	111

3.1.1.1 Generation of Muons in the Final State and Event Weighting Procedure

An interface to the PYTHIA generator was written, which takes over the decay of particles that can potentially produce a muon. For this purpose, all potential muon parents, such as B- or D-mesons as well as pions and kaons, are declared stable for the PYTHIA event generation step. Then a probability for a decay containing a muon within the defined acceptance region in both η and p_T is evaluated for each of these particles. The PYTHIA decay routine `PYDECY` is called twice for each of the potential muon parents, once switching off all decay modes containing muons, and once switching off all other decay modes, leaving only those with muons active. In both cases, possible muon parents might again be found among the decay products, so the procedure is recursively repeated for every particle until no muons can arise from further decays. When completed, this procedure produces an event record with several sets of decay products for the same hadron. In particular the event record contains all possible muons that could arise in the event.

For each particle that appears in the decay procedure described above, the probabilities $p_0, p_1, p_2, p_3, \ldots$ for it to yield exactly 0,1,2,3,... muons within the acceptance region are evaluated. Here are some examples:

- For a muon $p_1 = 1$ if it falls within the acceptance region, otherwise $p_0 = 1$. All other probabilities are zero;

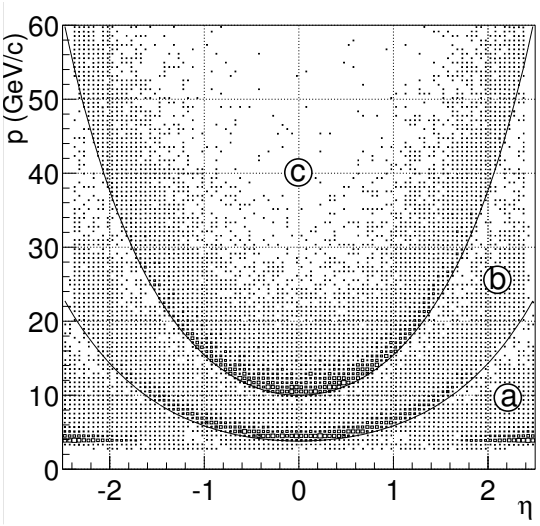


Figure 3.1: *Momentum versus pseudorapidity distribution for muons in the minimum bias samples. The contributions of the events from (a) the low-p_T, (b) the intermediate-p_T, and (c) the high-p_T weighted minimum bias samples are shown.*

- A charged pion has a 100% branching ratio into a muon and a neutrino. However, its lifetime is so long that it is most likely going to be absorbed in the calorimeters before it decays. For these long-lived particles, a decay volume is defined. For the present study, we use a cylinder with a radius of 3 m, containing the CMS calorimeters. For a pion, a probability p_{decay} is evaluated for its decay to happen inside the given detector volume. For pions with transverse momenta of a few GeV, $p_{\text{decay}} = \mathcal{O}(10^{-3})$ is typical. A decay vertex is chosen along the pion trajectory inside the volume, according to the expected exponential decay length distribution. The resulting probabilities are then $p_1 = p_{\text{decay}}$ and $p_0 = 1 - p_{\text{decay}}$, if the decay muon is within the acceptance, otherwise $p_0 = 1$ and all others are zero;

- The combined probabilities for a set of particles, *e.g.* all decay products of one parent particle, or a whole event, can be calculated from the corresponding probabilities of the individual particles. For a set of two particles (i and j) we obtain:

$$\begin{aligned} p_0 &= p_0^i p_0^j \\ p_1 &= p_1^i p_0^j + p_0^i p_1^j \\ p_2 &= p_2^i p_0^j + p_1^i p_1^j + p_0^i p_2^j \end{aligned}$$

...

- For particles that decay via more than one possible channel, all probabilities are averaged over the decay channels, weighted with the corresponding branching ratios. Here is an example for a particle with two possible decays, a and b, with corresponding branching ratios B_a and B_b:

$$p_0 = B_a p_0^a + B_b p_0^b$$
$$p_1 = B_a p_1^a + B_b p_1^b$$
$$\dots$$

Complete evaluation of these probabilities for all particles in an event leads to a set of "muon probabilities" (i.e. probability to have 0,1,2,...muons in the final state) for the event as a whole.

The sum of all probabilities for multiplicities that would satisfy the selection criteria is assigned as a weight to the event. Then one of these accepted multiplicities is chosen at random, according to their relative contributions to the event weight.

Once the number of muons for the event has been chosen, a final state is selected for the event, by distributing the muons to be generated among the particles of the event, according to the probabilities evaluated during the procedure described above. As an example, assume that an event that contains two potential muon parents (i and j) is chosen to have two muons. The total two-muon probability p_2 of the whole event as given above can be split into the contributions from 3 possible final states, as a function of the number of muons originating from each of the possible parent particles:

$$p_2 = p_2^i p_0^j + p_1^i p_1^j + p_0^i p_2^j = p_2^{20} + p_2^{11} + p_2^{02}.$$

Here p_2^{20} is the probability for the event to have two muons, which both come from the first parent particle (i); the definitions of p_2^{11} and p_2^{02} are analogous. For the selection of a final state, one of these configurations is chosen at random, according to their relative probabilities.

For the case of a particle where more than one decay channel is used, one of the channels has to be chosen. This is again done according to the probabilities for the desired muon multiplicity in each decay channel. In our example, if a particle with two decay channels (a and b) has been selected to produce 2 muons, its total 2-muon probability

$$p_2 = B_a p_2^a + B_b p_2^b$$

consists of two contributions from the two possible decays, and one of them is again chosen randomly according to their relative probabilities.

This whole procedure is repeated recursively for any daughter particle that can produce muons during decay. Once all muons have been assigned, all other particles are allowed to decay as usual, with all decay channels switched on. Each of these decays is checked for possible muons within the acceptance. In case a muon is found, the decay is repeated until no accepted muons are found.

The final result is an event satisfying the selection criteria in which all unstable particles have decayed, with no double-counting of energy. These events have all properties and distributions of the original unbiased sample, except that they come with weights, which represent the probability for the selected decay to happen.

3.1.1.2 Optimisation of Event Weights

When an event sample with a non uniform distribution of weights is used, there is a loss of statistical significance compared to a sample which contains the same number of events with uniform weights.

A figure of merit is the *equivalent number of unweighted events*, defined as the number of events in a hypothetical sample of unweighted events that would yield the same statistical uncertainty on the event rate as a given weighted sample. For a sample of N weighted events with weights $w_i, i = 1, \ldots, N$, the equivalent number of unweighted events N_{eq} can be calculated as

$$N_{\text{eq}} = \frac{(\sum_{i=1}^{N} w_i)^2}{\sum_{i=1}^{N} w_i^2}. \tag{3.1}$$

The loss of significance due to weight fluctuations is quantified as

$$s = \frac{N}{N_{eq}} = \frac{\overline{w^2}}{\overline{w}^2}. \tag{3.2}$$

In order not to use too many additional computing resources for detector simulation and further processing of the events, s should not be much greater than unity, which implies that the spread of weights in the sample should be kept small. Most minimum bias events without heavy flavours have a very small weight when forced to have one muon in the final state. Once heavy quarks (b, c or t) are produced, however, the typical event weights become considerably larger, because of the large branching fraction of B and D mesons into decay channels with muons. Events containing heavy quarks are rare and statistical fluctuations of their weight can have a big impact on the significance of the sample. In order to reduce this effect, b/c events have been generated separately, as already mentioned in Section 3.1.1, with an integrated luminosity a factor 10 higher than the rest of the events, and then combined with the light flavour sample with a weight scaled down by the same factor. The $b\bar{b}$ event weights are

further scaled by a factor which sets the total $b\bar{b}$ cross section in minimum bias processes to 500 μb as described in Section 3.1.1.

The weights of the events forced to contain muons with a minimum transverse momentum typically increase with the hard scale $\hat{p}_T$ of the parton scattering. The reason is that for such events the probability to produce b and c quarks is higher and the average momentum of the potential muon parents is larger. For the "High-p_T" sample, which is produced with $\hat{p}_T > 10$ GeV/c, the differential cross section $d\sigma/d\hat{p}_T$ used by PYTHIA to generate the events was artificially modified using the function

$$w_1 = 1 + 0.0003(\hat{p}_T/\mathrm{GeV}/c)^2 \tag{3.3}$$

so that the sample is enriched in the high-$\hat{p}_T$ component. The event weight was then divided by w_1 to get the correct normalisation and cross section. In this way, the final distribution of the weights is uniform in the $\hat{p}_T$ range of interest.

Finally, to further reduce the spread of event weights and improve the statistical significance of the sample, the abundance of events with very small weights was reduced with a procedure applied to all events with a weight w below a cut-off value $w_{min} = 0.1$. Each of these events was either selected with a probability w/w_{min} and assigned the weight w_{min}, or rejected. As a result, no event with weight below w_{min} is left and the overall composition and rate of the sample is unchanged.

The gain obtained with the weighing procedure can be quantified by defining the *equivalent luminosity* I_{int}^{eq} as the luminosity necessary to obtain the same statistical significance as a weighted sample where an integrated luminosity I_{int} was simulated:

$$I_{int}^{eq} = \frac{I_{int}}{\bar{w}s}. \tag{3.4}$$

The equivalent luminosity is much higher than the actual simulated luminosity, since events were forced to decay within the requested kinematic acceptance, resulting in an average event weight $\bar{w}$. The factor s (≥ 1) accounts for the loss of statistical significance due to weight fluctuations (cf. eq. 3.2). The values of I_{int}^{eq}, $\bar{w}$ and s for the three minimum bias samples described in Section 3.1.1 are given in Table 3.2.

The integral minimum bias rate as a function of the threshold on the muon transverse momentum obtained for a luminosity of $10^{34}\mathrm{cm}^{-2}\mathrm{s}^{-1}$ with this weighting procedure is compared in Fig. 3.2 with the one obtained from an unweighted sample of eight million minimum bias events generated with PYTHIA with `MSEL=1` and no $\hat{p}_T$ cutoff[4]. This comparison shows that the weighting pro-

[4]For the purpose of this comparison, in the weighted sample the normalisation of the b component to 500 μb total $b\bar{b}$ cross section was removed, to be consistent with the generation of the unweighted sample.

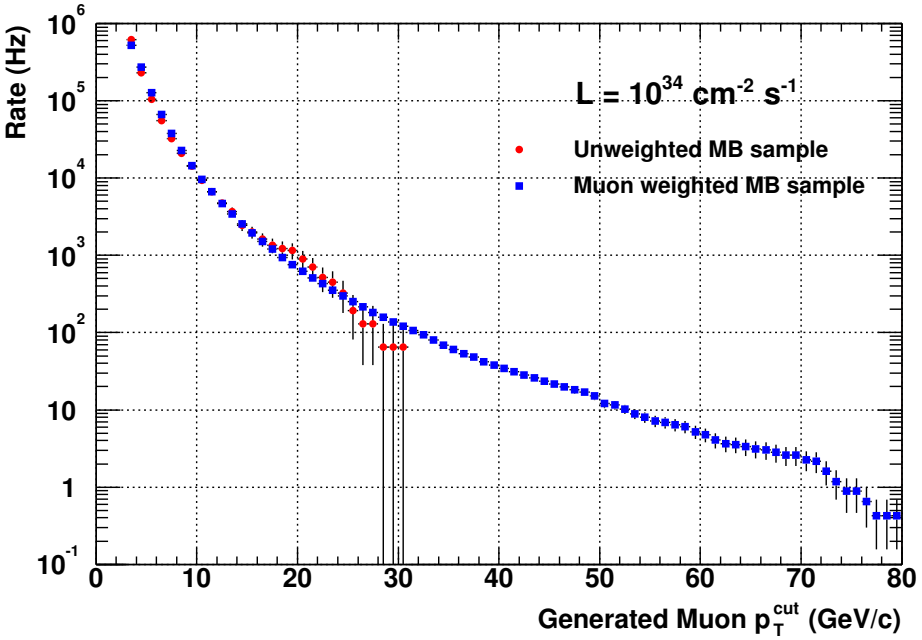


Figure 3.2: *Integral minimum bias rate for $|\eta|$ <2.1 at $\mathcal{L} = 10^{34} cm^{-2} s^{-1}$ as a function of the muon transverse momentum cutoff p_T^{cut}, from an unweighted sample of ≈ eight million minimum bias events (circles) and from the weighted muon minimum bias sample (squares). Note that, since these are integral distributions, the errors for adjacent bins are correlated.*

cedure does not significantly bias the muon rate and produces a sample that, while containing only a fraction of the events of an unweighted sample, is statistically significant up to much higher p_T thresholds.

3.1.1.3 Multi-Muon Events from Pile-Up

An important source of background events with more than one muon in the final state is the random overlap of two or more collisions with muons within the same bunch crossing. This contribution is not taken into account with the normal pile-up, which is filtered to remove muons in order to avoid multiple counting, as described in Section 3.1.3. Special procedures were therefore used to estimate di-muon rates. They are briefly mentioned in this section; however, their description is beyond the scope of this thesis, and the interested reader is referred to more specific documents [35, 23].

If detector occupancies are low, then the trigger objects caused by muons

from different collisions in the same crossing may be considered independent and the resulting trigger rates may be calculated analytically [35]. This was done for the low LHC luminosity scenario. In order to simulate multi-muon trigger rates under realistic conditions for the high-luminosity scenario, a sample of bunch crossings was constructed by properly overlaying several minimum bias collisions with muons in the final state. Pile-up events without muons were also added as discussed in Section 3.1.3. Crossing configurations with up to four collisions with muons were found to contribute significantly to the expected trigger rates and were therefore included in the sample. Computing resources were saved by constructing the crossings from the available simulated minimum bias collisions at the level of ORCA digitisation. Event weights were taken into account by defining a crossing weight proportional to the product of the weights of contributing collisions. A total of 1.10×10^5 crossings were constructed with this procedure.

An important advantage of this method over the analytic calculation of the rate of overlapping collisions with muons is that the resulting sample of crossings can be used to study any trigger condition. In particular, topological multi-object triggers and conditions based on the relative positions of the vertices of reconstructed muons or on their invariant masses can be studied.

3.1.2 Signal Samples

Two other kinds of processes were generated: the production and decay of heavy objects (W, Z/γ^*, top quark pairs), which contribute to the muon trigger rates at high p_T, and some SM and MSSM Higgs discovery channels, to be used as benchmarks of the trigger efficiency. Table 3.3 summarises the properties of these samples, which are described in the following.

3.1.2.1 W, Z/γ^*, $t\bar{t}$ Decays

The production of W and Z/γ^* bosons, as well as of top quarks pairs, represents an additional source of high-p_T muons, which has to be taken into account in the determination of trigger rates. Moreover, these events are useful as basic benchmarks for the trigger efficiency, since several LHC discovery channels will require the identification of leptons coming from W and Z decays.

The generation of single W and Z/γ^* bosons was performed via the lowest order $2 \rightarrow 1$ resonant processes (`MSEL=11,12`), with initial state QCD radiation switched on to simulate additional jets through parton showering (PYTHIA default value `MSTJ(41)=2`). In PYTHIA, this is considered the best option to simulate inclusive single-boson production with moderate transverse momentum $P_T \ll M_{W,Z}$, which is relevant for HLT muon background studies.

Table 3.3: *Generation of signal samples. N_{gen} and σ_{gen} are the number of events generated and the corresponding cross section calculated by PYTHIA. I_{int} is the generated integrated luminosity, which corresponds to the LHC running time t_{LHC}, calculated at high luminosity and assuming 100 fb^{-1} per year. N_{sel} and σ_{sel} are the number of events selected on the basis of the muon kinematics (see text) and the corresponding cross section.*

Sample	N_{gen}	σ_{gen}	I_{int}	t_{LHC}	N_{sel}	σ_{sel}
W (inclusive)	569618	185 nb	3.08 pb^{-1}	5.1 min	50000	16.2 nb
Z/γ^* (inclusive)	2268510	1003 nb	2.26 pb^{-1}	3.8 min	50000	22.1 nb
$t\bar{t}$ (inclusive)	46229	0.624 nb	74.1 pb^{-1}	2.1 h	20000	0.270 nb
$H(120)\to WW\to 2\mu 2\nu$	13581	36.3 fb	374 fb^{-1}	3.7 y	10000	26.8 fb
$H(140)\to WW\to 2\mu 2\nu$	13087	114 fb	115 fb^{-1}	1.1 y	10000	87.1 fb
$H(160)\to WW\to 2\mu 2\nu$	12722	182 fb	69.8 fb^{-1}	0.7 y	10000	143 fb
$H(180)\to WW\to 2\mu 2\nu$	12546	156 fb	80.5 fb^{-1}	0.8 y	10000	124 fb
$H(200)\to WW\to 2\mu 2\nu$	12556	104 fb	120 fb^{-1}	1.2 y	10000	83.1 fb
$H(130)\to ZZ^*\to 4\mu$	16829	0.90 fb	1.87×10^4 fb^{-1}	187 y	10000	0.53 fb
$H(150)\to ZZ^*\to 4\mu$	16030	1.69 fb	9.47×10^3 fb^{-1}	95 y	10000	1.06 fb
$H(200)\to ZZ\to 4\mu$	14938	3.45 fb	4.33×10^3 fb^{-1}	43 y	10000	2.31 fb
$H(300)\to ZZ\to 4\mu$	14147	2.19 fb	6.46×10^3 fb^{-1}	64 y	10000	1.55 fb
$H(500)\to ZZ\to 4\mu$	12709	0.89 fb	1.44×10^4 fb^{-1}	144 y	10000	0.70 fb
$H/A\to 2\tau$ ($m_A=200$)	31052	2.09 pb	14.9 fb^{-1}	133 d	10000	0.67 pb
$H/A\to 2\tau$ ($m_A=500$)	162262	39 fb	fb^{-1}		10000	2.4 fb

The generation of $t\bar{t}$ events was performed with `MSEL=6`, which simulates top quark pair production via quark or gluon fusion.

The W and Z/γ^* bosons were not forced to decay directly to muons, and all decay channels were left open. However, to optimise the generation only the events with at least one muon with $p_T > 3$ GeV/c within $|\eta| < 2.5$ were selected in the PYTHIA processing. This selection accounts for the difference between the number of generated events, N_{gen}, and the number of selected events, N_{sel}, given in Table 3.3 with the corresponding cross sections. Selected muons could originate either from direct boson decays or from the τ, b and c decay chains. This allows to use the samples for the determination of inclusive muon rates.

3.1.2.2 Higgs Samples

Samples of Standard Model Higgs boson decaying to $ZZ^{(*)} \to 4\mu$ (for m_H = 130, 150, 200, 300 and 500 GeV/c^2) and to $WW^{(*)} \to 2\mu 2\nu$ (for m_H = 120, 140, 160, 180 and 200 GeV/c^2) were also generated to demonstrate the efficiency

of the trigger selection. In this case, the decay of W and Z bosons to muons was forced in PYTHIA to obtain the desired topology. Two samples of MSSM Higgs bosons, decaying to $\tau\tau\rightarrow_{\mu\nu\bar{\nu}}$ ($b\bar{b}$ associated production mode, $\tan\beta = 20$, $m_A = 200$ and 500 GeV/c^2) were also produced using a dedicated selection code, that selected events with at least one $\tau\rightarrow\mu\nu\nu$ decay leaving the other τ decay free.

As for the samples described in the previous section, at least one muon with $p_T > 3$ GeV/c within $|\eta| < 2.5$ was requested in all events. The effect of this selection is indicated in Table 3.3.

3.1.3 Minimum Bias Sample for Pile-Up

The LHC will operate at a bunch crossing rate of 40 MHz. The bunch structure is such that only about 80% of the bunches will have collisions [9]; therefore the total inelastic pp cross section, predicted by PYTHIA to be 55 mb, results on average in 17.3 minimum bias events per bunch crossing at the LHC design luminosity of 10 $\mathrm{nb}^{-1}\mathrm{s}^{-1}$, and 3.5 events per bunch crossing at the initial luminosity of 2 $\mathrm{nb}^{-1}\mathrm{s}^{-1}$.

The realistic simulation of a triggering event in this conditions requires the simulation of the *in-time pile-up* occurring in the same bunch crossing and, for detectors with a long time window, of the *out-of-time pile-up* occurring in the contiguous bunch crossings.

Out-of-time pile-up is particularly relevant for the calorimeters, where dedicated studies [36] have shown that it is necessary to simulate the bunch crossings in a window of [-5,+3] with respect to the triggering event. At high luminosity this requires the superposition of an average of 156 minimum bias events, which, in the present study, were randomly chosen from a dedicated sample of about 2×10^5 minimum bias events.

The minimum bias pile-up sample was produced with the parameters indicated in Table 3.1, without any weighting procedure. However, since the events in the minimum bias samples used for the determination of the trigger rates are forced to contain muons (cfr. Section 3.1.1), no muon should be present in the pile-up sample to avoid an artificial increase of the di-muon rate and a bias due to the multiple occurrence of few triggering pile-up events. The pile-up sample was therefore filtered to remove all events containing potentially triggering muons (*i.e.* with $p_T >$ 1.0 GeV/c in $|\eta| < 2.0$ and $p > 3.0$ GeV for $2.0 < |\eta| < 2.5$).

Muon detectors are also sensitive to thermalised neutrons (produced in collisions up to millions of bunch crossings earlier), which, when captured by a nucleus, may yield a gamma-ray that converts into an electron-positron pair close enough to an active gas layer to produce a signal. Although parameterisations of this long-time pile-up exist, they were not enabled in this simulation,

mainly due to the excessive CPU time spent on digitising a large number of random hits. A dedicated Level-1 trigger study showed negligible impact of this background on the performance of the Level-1 Track-Finder for the CSC system [37]. The neutron background effect is also expected to be negligible on the performance of the Level-1 Track-Finder for the Drift Tube system, where the neutron flux is expected to be 10–100 times smaller [38]. On the other hand, the output of the Level-1 Pattern Comparator Trigger of the RPC system is sensitive to the neutron background in a way that is dependent on the intrinsic detector noise assumed [39], and further dedicated Level-1 trigger studies are underway.

The details of the digitisation procedure, where the pile-up is added to triggering events, are described in Section 3.2.

3.2 Detector Simulation

Particles were tracked through the CMS detector using CMSIM 125 [26], a simulation program based on GEANT3 [40] which includes the detailed geometry of the CMS detector. During this step, the collision point was taken to have a Gaussian distribution centered on the nominal interaction point, with $\sigma = 5.3$ cm along the beam line and $\sigma = 1.5$ mm in each of the transverse coordinates.

Multiple scattering, bremsstrahlung, Compton scattering and pair-production processes were activated in the GEANT simulation with low energy cutoffs in the last 4 cm of iron layers before the muon chambers, in order to have a realistic simulation of the delta-rays and shower processes in the chambers. Hadronic interactions at energies as low as 1 MeV were simulated with the GCALOR hadronic package [41]. The values of the energy cuts set in GEANT for tracking particles in the apparatus are summarised in Table 3.4.

A special treatment was necessary for the weighted minimum bias samples, due to the presence of the non-prompt muons generated by PYTHIA with the procedure described in Section 3.1.1. A routine was therefore implemented in CMSIM[5], where the pions and kaons that were forced to decay at the generation step were tracked through the detector and their decay point was chosen according to the lifetime assigned in the generation stage. The direction of the decay products was assigned taking into account the bending of the parent particle in the magnetic field. However, during the tracking, pions and kaons were allowed to undergo hadronic interactions before the predicted decay point; in this case no muon was generated.

[5]This routine has been included in the special CMSIM release 125.1 and can be activated with the card `STRD`.

Table 3.4: *Energy cut-off values set in GEANT for tracking particles in the apparatus. "Special" cut values were applied in the Tracker region and in the 4 cm thick material regions before muon detector active volumes. "Normal" values were adopted elsewhere in the apparatus.*

GEANT parameter	cut value *normal*	cut value *special*	GEANT parameter	cut value *normal*	cut value *special*
CUTGAM	1 MeV	10 keV	CUTELE	1 MeV	10 keV
CUTNEU	10 MeV	10 MeV	CUTHAD	10 MeV	100 keV
CUTMUO	10 MeV	100 keV			
BCUTE	1 MeV	10 keV	BCUTM	1 MeV	10 keV
DCUTE	10 TeV	10 keV	DCUTM	10 TeV	10 keV

After the processing with CMSIM, the simulated hits were read by ORCA and stored in an object database. Weighted minimum bias events containing no muons because of the hadronic interaction of the parent pions and kaons before the predicted decay point were skipped to save CPU time and disk space in the following steps. The number of selected events for the three weighted minimum bias samples is indicated in Table 3.2 in the column N_{sel}.

The final step in the production chain was the *digitisation*, where the response of the detector to the hits of the triggering event and of the pile-up was simulated with ORCA. The results are the *digis*, which are equivalent to the raw data collected by the detector for real events. Events were independently digitised for two LHC luminosity scenarios: $\mathcal{L} = 2 \times 10^{33}\text{cm}^{-2}\text{s}^{-1}$ ("Low luminosity") and $\mathcal{L} = 10^{34}\text{cm}^{-2}\text{s}^{-1}$ ("High luminosity"). Pile-up events were added to the triggering event via a random selection from the sample described in Section 3.1.3. The number of events added per bunch crossing was determined according to the Poisson distribution, with averages of 17.3 and 3.5 at high and low luminosity, respectively. Out-of-time pile-up in a window of [-5,+3] bunch crossings was included in the digitisation of calorimeters and of the CSC detectors. Out-of-time pile-up does not significantly affect the CMS tracker, the Drift Tubes and the RPC detectors and was not taken into account for these systems.

The effect of the superposition of 2 triggering interactions in the same bunch crossing was studied using the "mixed" sample described in Section 3.1.1.3. On the other hand, the effect of events with real muons occurring in off-time bunch crossings with respect to the triggering event was neglected in this study, since the probability for a muon to give a trigger at the wrong bunch crossing has been shown to be only 2% in the case of the Level-1 drift tube track-finder [22], and is expected to be smaller in the cathode strip chamber system.

A sizeable fraction of minimum bias events containing very low p_T muons does not lead to a trigger. To save resources, the digitisation of the low-p_T and intermediate-p_T minimum bias samples was done in two steps. Initially only the muon detectors and the calorimeters were digitised. The digitisation of the tracker was performed separately for events with at least one muon reconstructed by the Level-1 trigger.

The multiple-step event selection procedure described above affects the final event weight to be applied for normalisation at the analysis stage. Let N_{GEN} be the number of generated events, N_{HIT} those written to the hit database, N_{DIGI} the events after muon and calorimeter digitisation, N_{FILTER} those passing Level-1 filtering (if any) and N_{ANA} those after tracker digitisation and event reconstruction. The case of a non negligible event loss due to random failures (e.g. crashes due to data access problems) in the processing is reflected by $N_{HIT} \neq N_{DIGI}$ and $N_{FILTER} \neq N_{ANA}$. The rate contribution r_i corresponding to an event with the weight w_i is then given by:

$$r_i = w_i \frac{\sigma \mathcal{L}}{f N_{GEN}} \ , \tag{3.5}$$

where $f = (N_{DIGI}/N_{HIT}) \cdot (N_{ANA}/N_{FILTER})$ is the fraction of processed events, σ is the cross section of the data sample and $\mathcal{L}$ is the LHC luminosity. The formula is also valid for signal events where weights are equal to 1 and no Level-1 filtering step was applied ($N_{FILTER} = N_{DIGI}$).

3.3 Rates

In Fig. 3.3 the single-muon differential cross sections of the three minimum bias datasets are shown as a function of the muon p_T along with those of the W, Z and $t\bar{t}$ samples. The three weighted minimum bias samples are combined together as described in Section 3.2. The Higgs samples described in Section 3.1.2.2 have much smaller cross sections and are not shown in this figure.

The resulting integral rate of single muons obtained from PYTHIA as a function of the p_T threshold is shown in Fig. 3.4. Rates are computed for a luminosity of $\mathcal{L} = 10^{34}\text{cm}^{-2}\text{s}^{-1}$ (1 nb $\equiv$ 10 Hz). The muons are restricted to the muon trigger coverage of $|\eta|$ <2.1. The breakdown in terms of the muon parent particles is also shown.

From Fig. 3.4 it can be seen that the inclusive single muon rates for p_T thresholds up to 5 GeV/c are completely dominated by non-prompt muons from charged K and π decays; for p_T thresholds between 5 and 25 GeV/c the dominant contribution is bottom and charm quark decays, and only above $p_T > 25$ GeV/c does the contribution from W and Z decays become important.

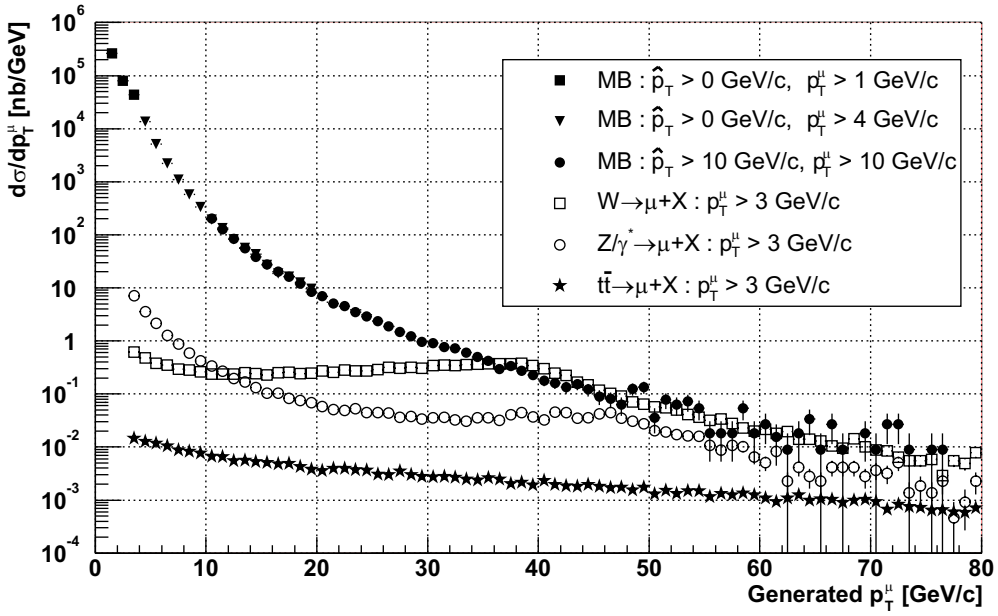


Figure 3.3: *Differential cross section for events with at least one muon within the trigger coverage ($|\eta|$ <2.1) for the various datasets.*

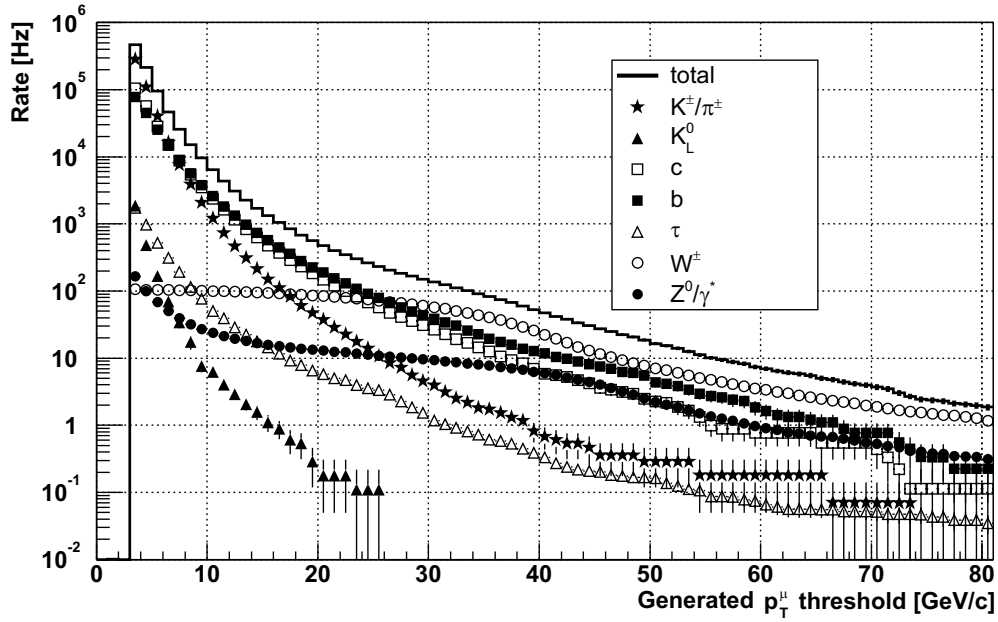


Figure 3.4: *Inclusive integral rate of single muons from PYTHIA as a function of the muon p_T threshold for a luminosity of $10^{34} cm^{-2} s^{-1}$. The breakdown of the rate on each source of muons is shown.*

Fig. 3.5 shows the integrated rate of di-muons at the generator level, for a luminosity of $\mathcal{L} = 10^{34}\text{cm}^{-2}\text{s}^{-1}$. For low p_T thresholds the inclusive di-muon rate is dominated by minimum bias events, while for p_T thresholds above 13 GeV/c decays of Z bosons dominate.

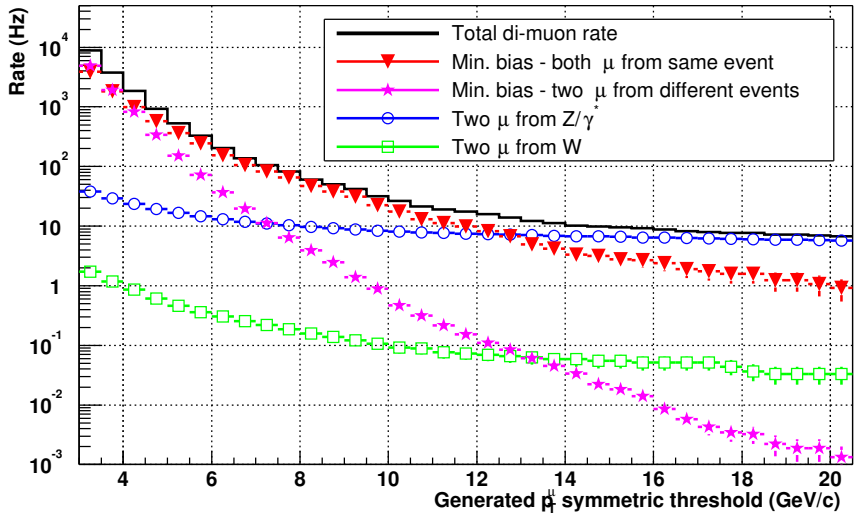


Figure 3.5: *Inclusive integral rate of di-muons from PYTHIA as a function of the (symmetric) muon p_T threshold for a luminosity of $10^{34}cm^{-2}s^{-1}$. The breakdown of the rate into the various muon sources is shown.*

3.4 Conclusions

This chapter reports the details of the event generation and full detector simulation of the signal and background samples produced in the year 2002 for the study of the CMS muon High Level Trigger performance. Particular emphasis was given to the description of the procedure adopted for the production of inclusive muon samples in order to get the maximum significance with an affordable number of simulated events. This procedure has been used to generate and simulate about 8×10^5 weighted minimum bias events that cover the p_T range 3–70 GeV/c. Obtaining the same statistical significance from an unweighted production using similar p_T bins and selection requirements would have required the generation of about 850 millions of events, and would not have been affordable with the existing computing resources.

The validation, optimisation and running of this production, which was documented in a CMS Note [2], were an integral part of my work in the PRS/mu

group. The samples produced were used for the results included in the CMS DAQ TDR [1] and, except where differently stated, are used for the studies presented in the following chapters.

The experience acquired suggests some possible improvements for similar productions in the future:

- The parton density function used was the default one for PYTHIA 6.128, *i.e.* GRV94. The use of other parton density functions which include more recent collider data should be considered. A detailed discussion on p.d.f.s can be found in [8] and references therein;

- The use of the simple multiple interaction model for the b, c component of the minimum bias samples (cf. Section 3.1) may affect some specific studies, in particular on muon isolation. This issue is discussed in Section 5.3;

- The weight rejecting procedure described in Section 3.1.1.2 is implemented to give uniform event weights in the sample. The same minimum accepted weight $w_{min} = 0.1$ was used for all three weighted minimum bias samples. This choice is probably not optimal, since the intermediate- and high-p_T samples have on average smaller weights, because of the small probability of having high-p_T muons in the final state. Smaller values of w_{min} should therefore be used for these samples;

- Several filtering steps were used in the production chain, in order to discard non-interesting events as soon as possible and save CPU time and disk space in the subsequent steps. However, it should be noted that all filters that discard events on the basis of reconstructed information (*e.g.* requiring the presence of at least one Level-1 muon candidate) bias the filtered sample so that it is not possible to get correct generator-level distributions. A mechanism to access the generation information of rejected events in the filtered samples would therefore be very useful.

Chapter 4

HLT Muon Reconstruction

The High Level Trigger (HLT) is the second step of the online CMS selection chain. Its goal is to reduce the event rate from the maximum Level-1 output rate of 100 kHz to a rate of O(100 Hz). This is the maximum acceptable rate for storage and further processing, given that the average event size after zero suppression is of the order of 1 MB.

The HLT is fully implemented on software running on a single farm of commercial processors. The advantages of this choice are multiple. A fully programmable system allows complete flexibility of the algorithms, which are only limited by the maximum available CPU time and data bandwidth. This flexibility will also allow to adapt the system to unforeseen conditions or developments. The use of standard commercial components leads to significant economies in the costs of hardware, development and maintenance. It also allows to benefit from future advances of technology, especially for what regards CPU speeds and network bandwidth. The use of standard software techniques and languages will also make it possible to benefit from the continuous improvements in the reconstruction software. In particular, the code used in the HLT, which will have access to "offline-quality" data from any part of the detector, will be identical to the code used in the offline analysis.

While the HLT runs on a single processor farm and does not have an internal architecture of separate trigger levels, background events should be discarded as soon as possible. It is therefore useful to organise the selection in a chain of logical steps that consist of progressively more sophisticated and CPU-time consuming algorithms. Selection criteria are applied at the end of each step, in order to reduce the event rate to a level acceptable for the following one. It is customary to identify two steps, the Level-2, where the data of the calorimeters and muon detectors are used, and the Level-3, which also utilises the information from the inner tracker, for which reconstruction is more time consuming. Obviously, this conventional separation is somewhat arbitrary; for example intermediate selection steps, where only part of the tracker data is used, are

possible if they are executed within an acceptable CPU time.

A full description of the HLT has recently been published [1]. This chapter describes the current implementation of the prototype of the Level-2 and Level-3 muon trigger reconstruction. The Level-2 step, described in Section 4.4, uses the full granularity of the muon detectors to improve the Level-1 muon trigger response. The selection is based on a threshold on the muon transverse momentum, and has to achieve a reduction factor of about 10. The Level-3, described in Section 4.6, further improves track reconstruction to offline quality including data from the tracker. Again, the selection is based on a threshold on the muon p_T, which however benefits from the improved resolution. Additional rate rejection is obtained at both levels with isolation algorithms, which are the subject of Chapter 5.

The following section describes the general design principles of the HLT reconstruction software and introduces the common concepts and terminology used throughout the chapter.

4.1 Software Design and Framework

The HLT algorithms will be fully implemented on software running on commercial processors. This allows to benefit from the wide availability, diffuse knowledge and constant technological improvement of standard components, not only for the hardware but also in the field of software. In particular, the CMS trigger and reconstruction software, ORCA (Object-oriented Reconstruction for CMS Analysis) [27], uses the modern technology of object-oriented programming, implemented in the C++ language. Proper object-oriented design allows to develop software components that are modular and reusable, and that can be easily prototyped, improved and replaced.

ORCA is implemented in the COBRA [42] framework, which provides basic computing services (data access, user interface, flow control, etc.) and utilities (mathematical algorithms, histogrammers, etc.). The presence of a framework allows for homogeneous implementation of the basic design mechanisms of ORCA. In particular, the mechanism of *on-demand* reconstruction consists in the fact that a given algorithm is performed only when (and if) the corresponding piece of information is requested.

The HLT reconstruction algorithms are implemented using the concept of *regional reconstruction*, that is the ability to reconstruct an object using the information coming only from a limited region of the detector. As an example, only the data of muon detectors in a region identified by a Level-1 muon candidate have to be processed in the Level-2 reconstruction, which in turn defines a small region of the tracker that is used for the Level-3 reconstruction. This allows to avoid global reconstruction steps, where the full information of each

detector has to be accessed and analysed, and leads to significant CPU savings. However, it also implies that objects that are not reconstructed in a given step are not recovered in the following ones in the case the event is anyhow selected.

A general principle of object-oriented programming is the distinction between component interfaces (that determine the relationships with other components) and their actual implementation. Proper design allows to implement generic algorithms that are abstract enough to cover a variety of similar situations. Track reconstruction is a typical example. For the results presented in this work, the reconstruction of tracks is performed with a generic method based on the Kalman filter technique, which is described in Section 4.2. The Kalman filter is an iterative method where at each step a new measurement is compared with a prediction based on the result of the previous step. The algorithmic part of this method is implemented in a generic way, even if the measurements produced by the different detectors are very different: for example, they are two- or three-dimensional segments in the DTs or CSCs and three-dimensional points in the RPCs and in the tracker[1]. Regardless of the differences, they can be used together in a track fit because they are implemented with a common interface, called the *RecHit*, and the fitting algorithm does not have to know the differences of the corresponding detectors. All the detector-specific aspects of the RecHits are instead handled in the components in charge of their creation. This operation is called *local reconstruction* and, for the case of muon detectors, it is described in Section 4.3.

The same design principle is used for other components; in particular for the search of the RecHits to be included in a track fit within the three muon subsystems, as described in Section 4.5.

4.2 The Kalman Filter Method

The Kalman Filter [43] is a recursive method for the fit of a discrete set of data. This method is particularly suited for track fitting, since it allows to implement reconstruction code which is independent of the number of measurements available. The problem consists in the determination of an estimate of a generic state vector $\boldsymbol{x}$ given a set of measurements $\boldsymbol{z}_k$, that are assumed to have the form

$$\boldsymbol{z}_k = H_k \boldsymbol{x}_k + \boldsymbol{v}_k. \tag{4.1}$$

Here $\boldsymbol{v}_k$ represent the measurement noise, and H_k is a matrix that relates the state to the measurement at a given step k.

[1] For detectors that measure only one coordinate, like RPCs and single-side silicon strip detectors, the second coordinate is assigned as the centre of the chamber or module, with an uncertainty equal to its length divided by $\sqrt{12}$.

Let us define $\boldsymbol{x}_k^-$ to be an *a priori* state estimate at the step k and $\boldsymbol{z}_k^+$ to be the *a posteriori* state estimate, obtained using the measurement $\boldsymbol{z}_k$. *A priori* and *a posteriori* errors are defined as:

$$\boldsymbol{e}_k^- = \boldsymbol{x}_k - \boldsymbol{x}_k^-; \qquad \boldsymbol{e}_k^+ = \boldsymbol{x}_k - \boldsymbol{x}_k^+ \tag{4.2}$$

and the corresponding covariance matrices are:

$$P_k^- = E[\boldsymbol{e}_k^- \boldsymbol{e}_k^{-T}]; \qquad P_k^+ = E[\boldsymbol{e}_k^+ \boldsymbol{e}_k^{+T}]. \tag{4.3}$$

The *a posteriori* state estimate $\boldsymbol{x}_k^+$ is given by a linear combination of the *a priori* estimate $\boldsymbol{x}_k^-$ and the weighted difference between the measurement $\boldsymbol{z}_k$ and a measurement prediction $H_k \boldsymbol{x}_k^-$:

$$\boldsymbol{x}_k^+ = \boldsymbol{x}_k^- + K_k(\boldsymbol{z}_k - H_k \boldsymbol{x}_k^-). \tag{4.4}$$

The term in parenthesis is called "residual" and reflects the discrepancy between the prediction and the actual measurement. The matrix K_k is called "gain" and is chosen minimising the *a posteriori* error covariance P_k^+ [44]:

$$K_k = \frac{P_k^- H_k^T}{H_k P_k^- H_k^T + R_k}, \tag{4.5}$$

where R_k is the measurement error covariance matrix for $\boldsymbol{z}_k$. We can notice that for small errors the *a posteriori* estimate is dominated by the measurement:

$$\lim_{R_k \to 0} K_k = H_k^{-1}. \tag{4.6}$$

On the other hand, if the *a priori* error covariance is small, the measurement becomes less important in the estimate:

$$\lim_{P_k^- \to 0} K_k = 0. \tag{4.7}$$

The *a posteriori* error covariance estimate is obtained as

$$P_k^+ = (1 - K_k H_k) P_k^-. \tag{4.8}$$

The iterative method starts from an externally provided estimate of the initial state $\boldsymbol{x}_0$ and its covariance matrix P_0, called *seed*. Each step k consists of two phases: first a "prediction" of $\boldsymbol{x}_k^-$ and P_k^- is obtained by projecting the previous step and covariance matrix; then the state is "updated" using Eq. 4.5, Eq. 4.4 and Eq. 4.8 to obtain $\boldsymbol{x}_k^+$ and P_k^+. A new iteration is then started using the updated state and covariance matrix.

In the case of track reconstruction, the state vector is defined as the position and momentum relative to a given surface (local coordinate frame):

$$\boldsymbol{x} = \begin{pmatrix} q/p \\ \tan\phi \\ \tan\theta \\ x \\ y \end{pmatrix} \tag{4.9}$$

where q is the charge, p is the momentum and ϕ, θ, x and y identify the track direction and position on the surface.

The first step consists in the seed generation. For muon tracks, the seed is generated from the parameters of Level-1 candidates. The seed consists of the initial state vector and of an estimate of its covariance matrix.

Each step is then decomposed into two parts: the prediction of the state vector and of the error covariance matrix on the surface of the next measurement to be included, and the update including the information from that measurement. They are encapsulated in two basic components, called *propagator* and *updator*. The updator includes all the algebra of the fit, so that the complexity of fitting is reduced to providing an analytic or numeric algorithm to propagate tracks and their errors. The propagator used in the muon system is implemented using the FORTRAN package GEANE [45], which is able to extrapolate a state vector and its covariance matrix in a non-constant magnetic field, taking into account the effect of energy loss and multiple scattering in the material traversed by the track. An external steering code is also needed to look for the measurements to be included in the fit. The algorithm for optimised hit search in the muon detectors is described in Section 4.5.

The result of a Kalman filter fit is a state on the surface of the last measurement, which includes all available information. However, the trajectory parameters calculated at other points of the trajectory do not include the information of all measurements. A special procedure is used to update the parameters so that they include all the measurements at every measured surface. In the Kalman filter terminology, this procedure is called *"smoothing"*.

4.3 Local Reconstruction

Local reconstruction is the first step of reconstruction, and is performed at the level of individual detectors. The results are track segments in the DT or CSC systems, or three-dimensional points for the RPCs. Despite being different, all these results represent one measurement as described in the previous section; therefore they are generally referred as "reconstructed hits" (RecHits) and implemented with the same interface. The local reconstruction in the three muon

subsystem has been extensively documented in [46], and is briefly described in this section.

DT. The position of hits in single drift cells is estimated from TDC measurements. This is done in two steps: initially, an average value for the drift velocity is used to fit a two-dimensional segment in the superlayer. The unknown bunch crossing that originated the hits is a parameter of the fit, which uses the mean-timer technique [21]. The parameters obtained are then used to determine the correct effective drift velocity, refine the hit position and the fit. In each chamber, the segments reconstructed in the two $r-\phi$ superlayers are then refitted together, and the result is combined with the segment in the $r-z$ superlayer to produce a three-dimensional segment. The direction resolution in the $r-\phi$ plane is about 0.9 mrad. In the $r-z$ (non-bending) plane, it is about 9 to 13 mrad for tracks in $|\eta| < 0.9$.

CSC. Each CSC plane measures a point in two dimensions. One coordinate is measured by the wires, which are read out in bunches resulting in a limited precision. The other coordinate is measured by the strips, where the charge distribution of a cluster of three neighbouring strips is fitted to the so-called "Gatti function" to obtain a precise position measurement. The hits in a chamber are used to fit a three-dimensional straight line segment. The direction resolution of the segment varies from 7 to 11 mrad in ϕ and from 50 to 120 mrad in θ for 50 GeV muons.

RPC. The hits produced by the RPCs are three-dimensional points. They are obtained by clustering the strips and calculating the centre of gravity of the area covered by the strips in the cluster (*i.e.* the width of the strips times their full length). Uncertainties are computed assuming that the hit can have happened anywhere in this area with flat probability, *e.g.* in the simplest case of a rectangular area they are equal to the length of each side divided by $\sqrt{12}$.

According to the principle of regional reconstruction, the local reconstruction is not performed in a single step for the full detector. At each step of the Kalman filter fit, the current state vector is extrapolated to the next layer of detectors, and only compatible detectors are requested to perform local reconstruction. This operation is called "navigation" in the detectors, and for the case of the muon systems is described in Section 4.5.

4.4 The Level-2 Reconstruction

The Level-2 selection consists of a cut on the transverse momentum of muons reconstructed using the muon chambers alone. The Level-2 reconstruction is based on the Kalman filter method described in Section 4.2. It consists of several steps, described in the following.

Seed generation. The Kalman filter method starts from the determination of a trajectory seed, that in general can be built from any measurement in the muon detector (*internal seeding*). However, for trigger purposes one can take advantage from the fact that the Level-1 trigger provides muon candidates that can be used to generate seeds. This *external seeding* allows significant time savings, in particular because it does not require local reconstruction in the full detector. On the other hand, this limits the efficiency of the HLT, which cannot reconstruct muons not already reconstructed by the Level-1.

To create the seed, the parameters and uncertainties of Level-1 candidates, which are given at the second barrel or endcap station, are extrapolated backwards, before the first muon station. The GEANE-based propagator mentioned in the previous section is used for this purpose.

RecHit collection and seed refinement. Hits to be included in the fit are looked for iteratively. At each step, the trajectory is extrapolated to a layer of detectors, where local reconstruction is performed in the chambers compatible with the extrapolated state. The search for the compatible detectors (*navigation*) is done in an optimised way, in order to minimise the number of extrapolations and of detectors queried. If measurements are found, they are tested for compatibility with the current state with a χ^2 test; if they are compatible, the state is updated using the Kalman filter method. In the DT, the RecHits used in the fit are full segments, as described in the previous section. Inside the CSC chambers, the magnetic field is not negligible, so that segments describe the trajectory only in an approximate way. Thus, the points that constitute CSC segments are used as individual RecHits.

This procedure is repeated until the fit reaches the last station. The result is a state at the outermost muon station, updated with all available measurements.

Fit. The improved state from the previous step is used as seed to the actual fit, which is performed in a similar way going inwards to the detector centre. A tighter χ^2 cut for the inclusion of hits is used in this case. The final estimate of the state is obtained at the innermost muon station.

Trajectory cleaning and smoothing. When all trajectories have been built, ghost tracks, *i.e.* multiple tracks that correspond to the same muon, are looked for and eliminated on the basis of the χ^2 of their fit (cleaning).

Since the Kalman filter is performed incrementally, the full information of all measurement is included only at the last step. The trajectory smoothing (cf. Section 4.2) consists of recalculating the track parameters at each measurement point, so that they include the information of all measurements.

Vertex extrapolation and constraint Finally, the trajectory is extrapolated back to the point of closest approach to the beam line, and the nominal interaction point with its spread is included in the track fit. This fit can fail in case the muon is non-prompt and its trajectory does not point to the collision vertex. If so, the trajectory is discarded.

The performance of the Level-2 reconstruction in terms of algorithmic efficiency and transverse momentum resolution is discussed in Section 4.7.

4.5 Navigation in the Muon Detectors

The track fitting method described in the previous sections does not require the full set of reconstructed hits to be available before the fit. Instead, at each step the track parameters are used to identify the detectors that most probably contain the next RecHit. Local reconstruction is performed in these detectors only, realising the principle of regional reconstruction.

The algorithmic problem of finding the next detector crossed by one trajectory, given its parameters at a given point, is called *navigation*. While completely general solutions exist, the layout of the CMS tracking detectors is such that specialised, highly optimised solutions can be implemented, both in the muon system and in the inner tracker. The efficiency and speed of this operation are fundamental: the navigation is the most time consuming part of the track fit, and a very fast implementation is necessary to allow track reconstruction within the timing constraints of the trigger. Also, optimised navigation minimises the number of detectors queried for local reconstruction, reducing the amount of data that has to be actually accessed and analysed.

The possibility of an optimised implementation derives from the fact that the tracking detectors in CMS are arranged in layers with a defined distance from the detector centre, so that a track coming from the interaction point always crosses them in a defined sequence. Moreover, individual detectors within the layers are arranged in a (quasi-) periodic way, and can be organised in substructures with a simple geometrical shape, like disks or rods.

The problem of navigation is therefore solved by organising the detectors in a hierarchical structure. Each element in the hierarchy is described geometrically in terms of an elementary surface that approximates the surface of its constituents (detectors or groups of detectors). The element can be queried to find which detectors are compatible with a given trajectory; in this case, the trajectory is extrapolated to the surface of the element and compatible constituents are selected taking into account the uncertainty on the extrapolated track position. Selected constituents are queried in the same way, so that the hierarchy is traversed vertically, up to the level of individual detectors. At this point, local reconstruction is performed on the selected detectors. As a result, the reconstructed hits are returned together with the trajectory state extrapolated to the corresponding detector surface. Both the RecHits and the state are directly used for the Kalman filter update step (cf. Section 4.2). This procedure allows to limit the number of track extrapolations to the absolute minimum, especially in the case of extrapolations between stations in the muon system, where the trajectory crosses the iron return joke. These constitute the time-consuming part of the navigation. The propagator used within the muon system is the same used for the track fit, GEANE (cf. Section 4.2).

ORCA provides a framework and a set of base classes for the implementation of these hierarchies. Such classes are shared by all tracking detectors, to allow a consistent behaviour. Navigation in each detector system is implemented by specialising these classes, *i.e.* grouping the detectors in the optimal way and defining the rules to determine how the search is performed within each hierarchy level. The case of the muon system is described in the following.

The first level of the hierarchy is a full layer of detectors; in the barrel muon system it consists of a cylinder, corresponding to either one DT or one RPC station. In the endcap, layers are flat vertical disks corresponding to either one CSC or one RPC station, except for the first station (ME1), which is split in two layers at different positions in z (cf. Fig. 2.1). Layers are sorted by their distance from the detector centre (*i.e.* by radius in the barrel and by z in the endcaps). This is the order in which they are traversed by a trajectory coming from the interaction point, so that, at each fitting step, RecHits are looked for in the next layer in the list (provided that the track direction with its error is within the layer's η boundaries.)

Inside layers, detectors are organised in groups with a common surface. Since an extrapolation to each detector surface is needed to test the compatibility with a track, the most efficient organisation is obtained when individual detectors with a common surface are grouped together. The organisation is different for the barrel and the endcaps.

Muon barrel layers. From the mechanical point of view, the muon barrel is constituted of five wheels of 12 or 14 chambers each. However, as discussed above, it is preferable to have chambers with the same flat surface grouped together. For this reason, barrel muon layers are constituted of rods, each consisting of five chambers. There are 12 rods in each layer, except for MB4, where the rods are 14. Chambers in a rod lie on the same plane, are contiguous in z and are all at the same r and ϕ coordinates. Rods in a layer are quasi-periodic in ϕ and, in some cases, slightly overlap in this coordinate.

When a layer is queried for the RecHits compatible with a given track, a first extrapolation is done to the cylinder representing the layer's surface. The extrapolated position in ϕ is used to determine the closest rod, which is selected. If the extrapolated error, scaled by an adjustable factor, exceeds the rod's border in ϕ, the neighbouring rods are selected as well. Then, selected rods are individually queried for compatible chambers. A short extrapolation is made from the layer's cylinder to the rod plane, and one or more chambers are selected according to the z position of the extrapolated track and its uncertainty.

Muon endcap layers. Endcap layers are organised in rings of chambers. All chambers in one ring are located at the same r and are periodic in ϕ. In all endcap layers, with the exception of ME1/3 CSCs and ME1/2 RPCs, the chambers in a ring are staggered in z. In this case, the ring's nominal surface is a flat disk placed between the two planes of chambers.

When the layers is queried for the measurements compatible with a track, a first extrapolation is done to the plane representing the layer's surface. The extrapolated position in r is used to determine the closest disk, which is selected. If the extrapolated error, scaled by an adjustable factor, exceeds the disk's border in r, the neighbouring disks are selected as well. Then, selected disks are individually queried for compatible chambers. This is done using the ϕ coordinate and the uncertainty of the previously extrapolated state. In the case of staggered chambers in a disk, one more short extrapolation from the ring's central surface to the surface of each chamber is needed to obtain the state on the detector surface as required by the track fitting procedure.

It should be noted that in the case of staggered chambers the final step is not fully optimised, since if more than two chambers in one disk are selected, the short extrapolation is repeated unnecessarily. The optimisation of this case was left for a future improvement of the algorithm. It can be obtained describing a disk of staggered detectors with two parallel surfaces instead than with just one, so that only one short extrapolation is done in every case. The expected speed-up is discussed in the following.

The implementation of the navigation described here guarantees by con-

struction the minimum number of extrapolation between layers. These are the "long" extrapolations that cross the iron yoke, that are the most CPU-time consuming ones. The result of this optimisation alone was a reduction of the total Level-2 reconstruction time by a factor ~ 6.5 [47] with respect to a previously existing, non-optimised prototype. This speed-up was necessary to meet the timing requirements of the trigger.

While long extrapolations are fully optimised, short ones within layers are optimised only in the barrel. The resulting number of calls to the GEANE routine per event as a function of pseudorapidity is shown in Fig. 4.1. In the

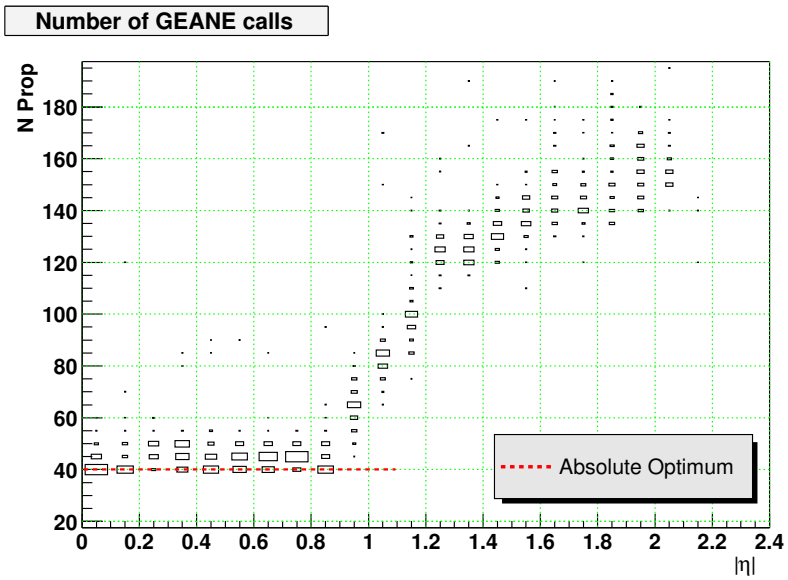


Figure 4.1: *Number of calls to the GEANE routine per event as a function of the muons pseudorapidity. The dotted line represents the absolute minimum for the reconstruction of a muon. The effect of the full optimisation in the barrel region is evident.*

barrel region, it is almost equal to 40. This is the absolute minimum for the full reconstruction of one muon, considering that in the barrel there are ten layers (six of RPCs and four of DTs), that for each layer the extrapolation is divided in two steps (first to the layer's cylinder, then to the plane of the selected rod) and that the detector is traversed twice, the first time inside-out for the seed refinement, and the second outside-in for the final fit. These 40 extrapolations correspond to traversing the detector in the two direction stopping each time in 20 places; no path is repeated unnecessarily.

Also in the case of the endcaps the minimum number of calls is 40, but much more are actually done. The additional extrapolations are the short ones between the staggered chambers within disks, and the very short ones between the

six CSC planes inside each chamber during the final fit[2]. Short extrapolations account for about 40% of the total CPU time spent in extrapolations during the fit. A detailed study of the CPU-time spent in the Level-2 reconstruction is presented in Section 6.3.2.

Possible improvements include the already discussed optimisation of short extrapolations between staggered endcap chambers, as well as the use of tuned propagators. In particular, it was proposed to use a simplified treatment of material effects for short extrapolations that do not cross iron volumes, as in the case of extrapolations inside each CSC chamber, within layers and between contiguous RPC and DT/CSC layers. The speed-up achievable is discussed in Section 4.4.

4.6 The Level-3 Muon Reconstruction

At Level-3, muon tracks are reconstructed using data from the full tracker system. The selection is again based on a cut on the reconstructed transverse momentum, which is now measured with much higher precision.

The Level-3 reconstruction starts with the definition of a "region of interest" in the silicon detectors, based on the parameters and uncertainties of the Level-2 track after the vertex constraint. Regional track reconstruction is performed using tracker hits within this region. Seeds are generated from pairs of reconstructed hits from the pixel detectors or the double-sided silicon strip layers. A beam spot constraint allows an estimate of the trajectory parameters of these seeds, whose number is limited by a cut on their minimum transverse momentum. The seeds are then used to perform reconstruction within the region of interest using the Kalman filter method. The procedure adopted is similar to that described for the Level-2 reconstruction. Starting from the seeds, compatible RecHits are searched for in the following layers. If more that one is found, a new trajectory is created. These trajectories are grown in parallel, but to avoid a combinatorial explosion their number is limited by selecting the best 30 on the basis of their χ^2. In the case no hit is found on a layer, the search is repeated in the following one. The track is discarded if no hit is found in more than four consecutive layers.

After the trajectories for all seeds have been generated, the trajectories with the lowest χ^2 are selected among those that share more than half of their hits. Finally, the trajectories are smoothed (cf. Section 4.2), including also the hits of the Level-2 track reconstructed in the muon chambers.

The geometry of the tracker is more complex than that of muon detectors; however the navigation in the tracker is implemented as a specialisation of

[2] As described in Section 4.4, each of the six points measured in a chamber is included in the fit independently.

the same base classes used for the muon navigation. The propagator used is different though, since inside the tracker volume the magnetic field is almost constant and the amount of material is relatively small. The propagation is thus performed analytically and material effects (energy loss and multiple scattering) are treated in a thin-layer approximation.

4.7 Performance and Rates

The selection of muons in the HLT is based on the Level-2 and Level-3 reconstruction, with additional requirements on the quality of track candidates, which are necessary to reduce trigger rates from poorly measured muons. In particular, low quality Level-1 CSC candidates are not considered in the seed generation if they are not matched to RPC candidates by the Global Muon Trigger (cf. Section 2.5.1.2). At Level-2, reconstructed tracks in the barrel region must include at least one DT segment and more than three RecHits in the DT and RPC detectors. At Level-3, a track must include at least 5 silicon hits.

The efficiency for the reconstruction of single muons generated with flat p_T in the range 5–100 GeV/c is shown in Fig. 4.2 as a function of the generated pseudorapidity. Since each level uses the results of the previous one as seed,

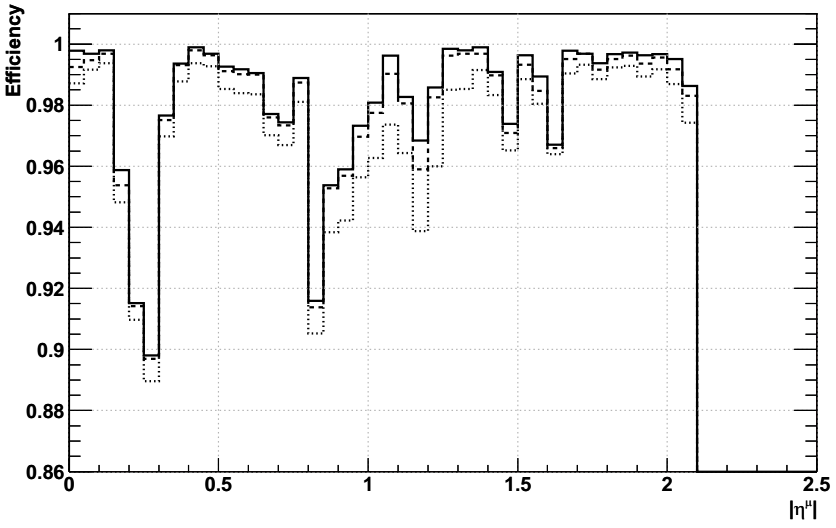


Figure 4.2: *Efficiency for reconstructing muons in a single muon sample digitised without pile-up for Level-1 (solid), Level-2 (dashed), and Level-3 (dotted) triggers as a function of the generated η [1].*

the efficiencies are slightly smaller at each level. The overall efficiency is above 97%, but drops to about 90% in correspondence of the gaps between the wheels of muon detectors.

The selection of muons is based on a cut on their transverse momentum. The resolution on the Level-2 measurement of the reciprocal of the transverse momentum, $1/p_T$, is shown in Fig. 4.3 for muons coming from W decays digitised at high luminosity. The resolution is about 10% in the barrel region, 16% in

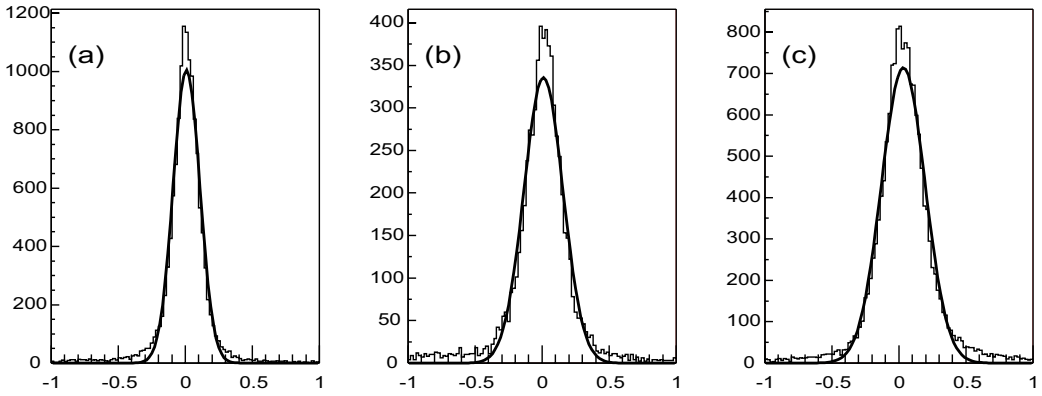


Figure 4.3: *Level-2 resolution on $1/p_T$ obtained with a sample of W decays at high luminosity for (a) the barrel region ($|\eta|$ <0.8); (b) the overlap region (0.8< $|\eta|$ <1.2); (c) the endcap region (1.2< $|\eta|$ <2.1) [1].*

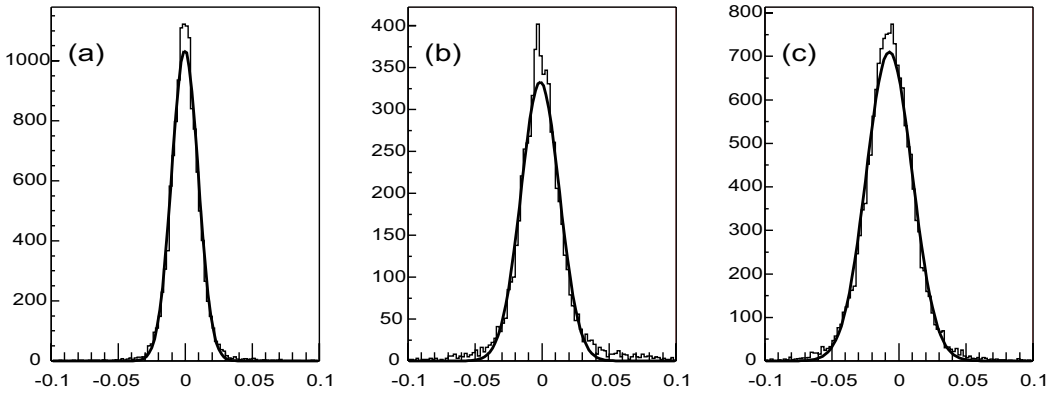


Figure 4.4: *Level-3 resolution on $1/p_T$ obtained with a sample of W decays at high luminosity for (a) the barrel region ($|\eta|$ <0.8); (b) the overlap region (0.8< $|\eta|$ <1.2); (c) the endcap region (1.2< $|\eta|$ <2.1) [1].*

the endcap region and 15% in the overlap. The resolution obtained at Level-3 is about ten times better, as shown in Fig. 4.4, and amounts to about 1.0% in the barrel region, 1.7% in the endcap region and 1.4% in the overlap.

The improvement in the resolution is reflected in the efficiency turn-on curves shown in Fig. 4.5. They represent the efficiency to select muons by setting a

given p_T threshold, as a function of the generated muon p_T. The same threshold is used at all trigger levels[3]. It can be seen that the improved resolution sharpens the curves, thus improving the trigger selection.

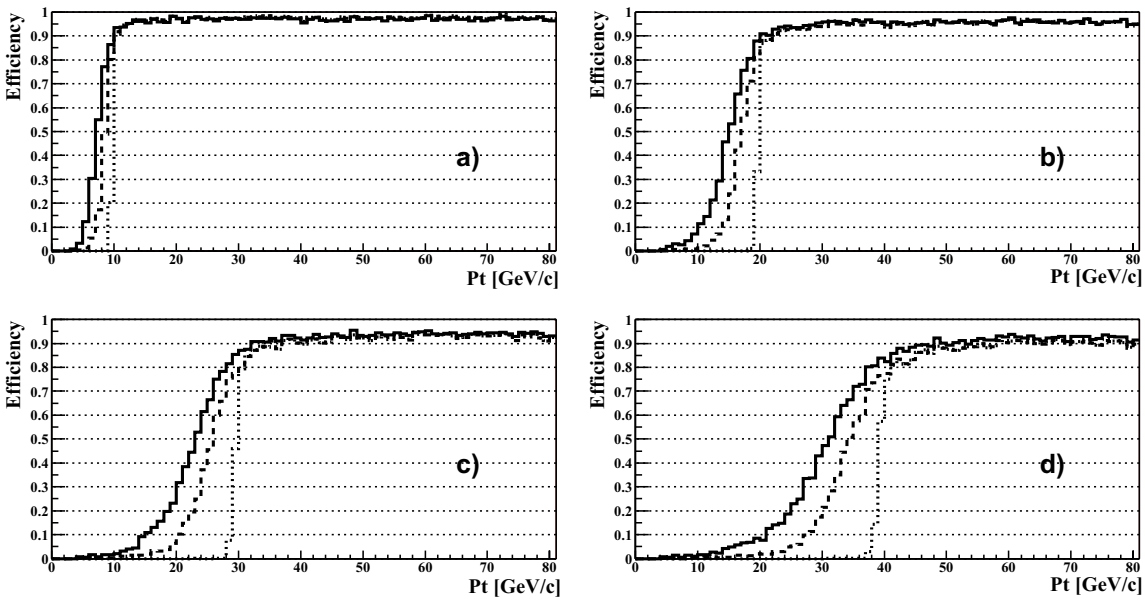


Figure 4.5: *Efficiency for the selection of single muons without pile-up with the Level-1 (solid), Level-2 (dashed) and Level-3 (dotted) triggers as a function of the generated p_T, for trigger thresholds of (a) 10 GeV/c, (b) 20 GeV/c, 30 GeV/c and (d) 40 GeV/c [1].*

The rate obtained with a cut on the muon p_T is shown in Fig. 4.6 as a function of the threshold at Level-1, Level-2 and Level-3, for the low and high LHC luminosity scenarios. Also shown is the inclusive generated rate. The input sample is the inclusive sample of minimum bias, W, Z and $t\bar{t}$ events described in Section 3. It can be seen that the Level-1 rate is much higher than the generated one. The excess is not due to ghosts or fake muons. These are real muons with low p_T, which are reconstructed at higher p_T due to the limited resolution of the Level-1 reconstruction and due to the presence of non-Gaussian tails in the p_T resolution (cf. Section 2.5.1.3). This "feed-through" effect is so large because the generated rate grows very rapidly with decreasing transverse momenta. The improved Level-2 resolution allows a significant reduction of the rate, which is however still well above the generator rate, for the same reason.

[3]For consistency with the Level-1 definition, the Level-2 and Level-3 thresholds are converted to a 90% efficiency scale relative to the plateau efficiency. Cf. the definition in Section 2.5.1.3.

The full resolution of the Level-3 reconstruction allows to reduce the trigger rate to the generated level.

The rates for the symmetric di-muon trigger, where at least two muons in each event are required to be above the same p_T threshold, are shown in Fig. 4.7. Also in this case, the Level-1 rate is above the generated rate, mostly due to the feed-through effect, but, for a small part, also to the contribution of events with one real muon and one ghost. Useful thresholds are in the range 4–15 GeV/c^2; for higher values the additional rate reduction is small.

4.8 Conclusions

This chapter describes the current prototype of the Level-2 and Level-3 muon trigger reconstruction and selection. The inclusive rates obtained with single- and di-muon trigger selections are discussed. Level-1 and Level-2 rates suffer from the feed-through of low-p_T muons due to their limited p_T resolution. This problem is not present at Level-3, which has a much better p_T resolution. At this level, the trigger rates are close to the generated ones. However, in the interesting range of thresholds (15-30 GeV/c), these rates are still high, if we consider that the maximum acceptable HLT output rate will be of the order of 100 Hz. The muon rate must therefore be reduced with additional criteria, that must select events on the basis of their physics origin. A possibility is to exploit muon isolation, which is the subject of the next chapter.

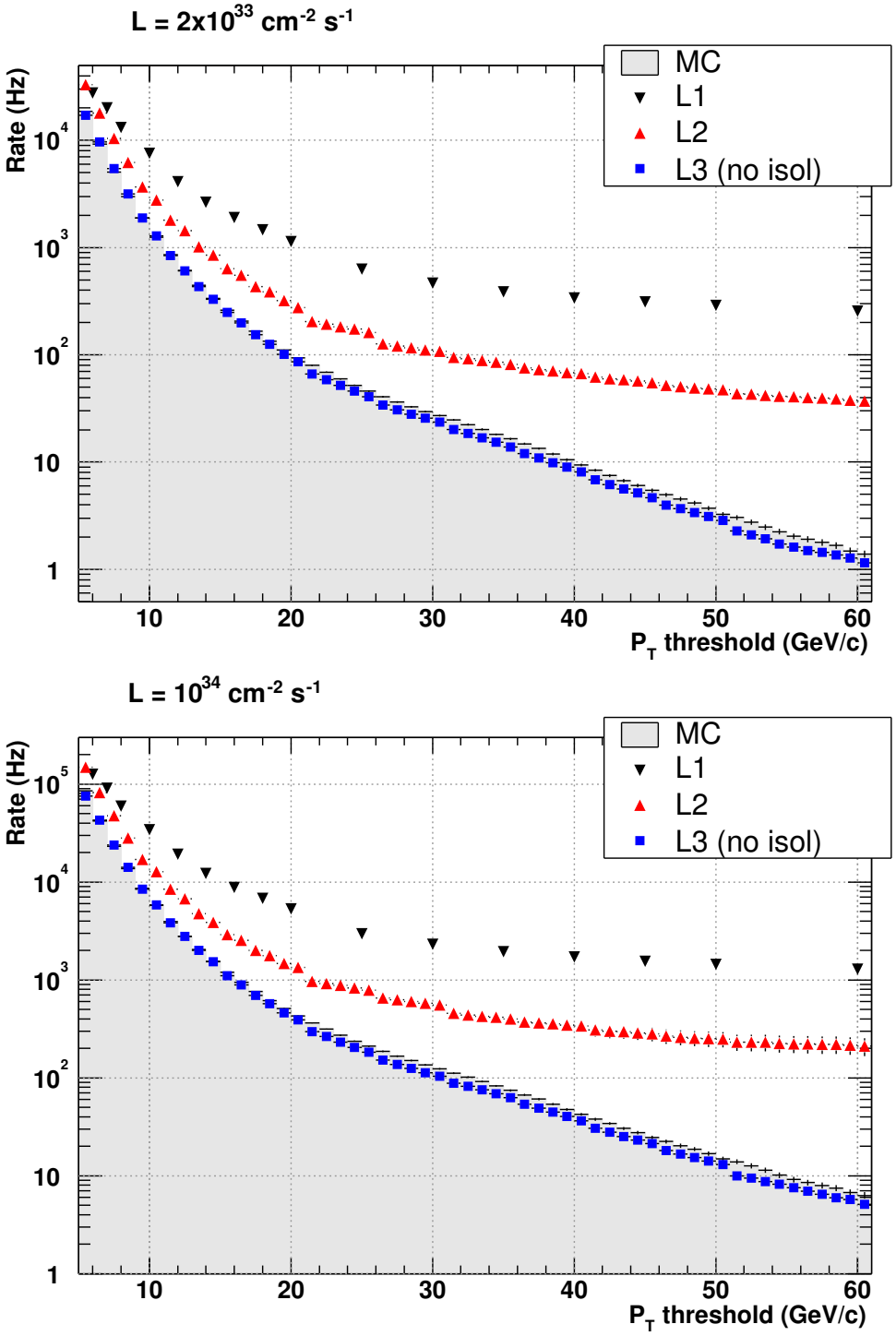


Figure 4.6: *Single muon trigger rates as a function of the p_T threshold at low luminosity (top) and high luminosity (bottom). The curve labelled "MC" represents the generated rate as a function of the generated p_T of the muons.*

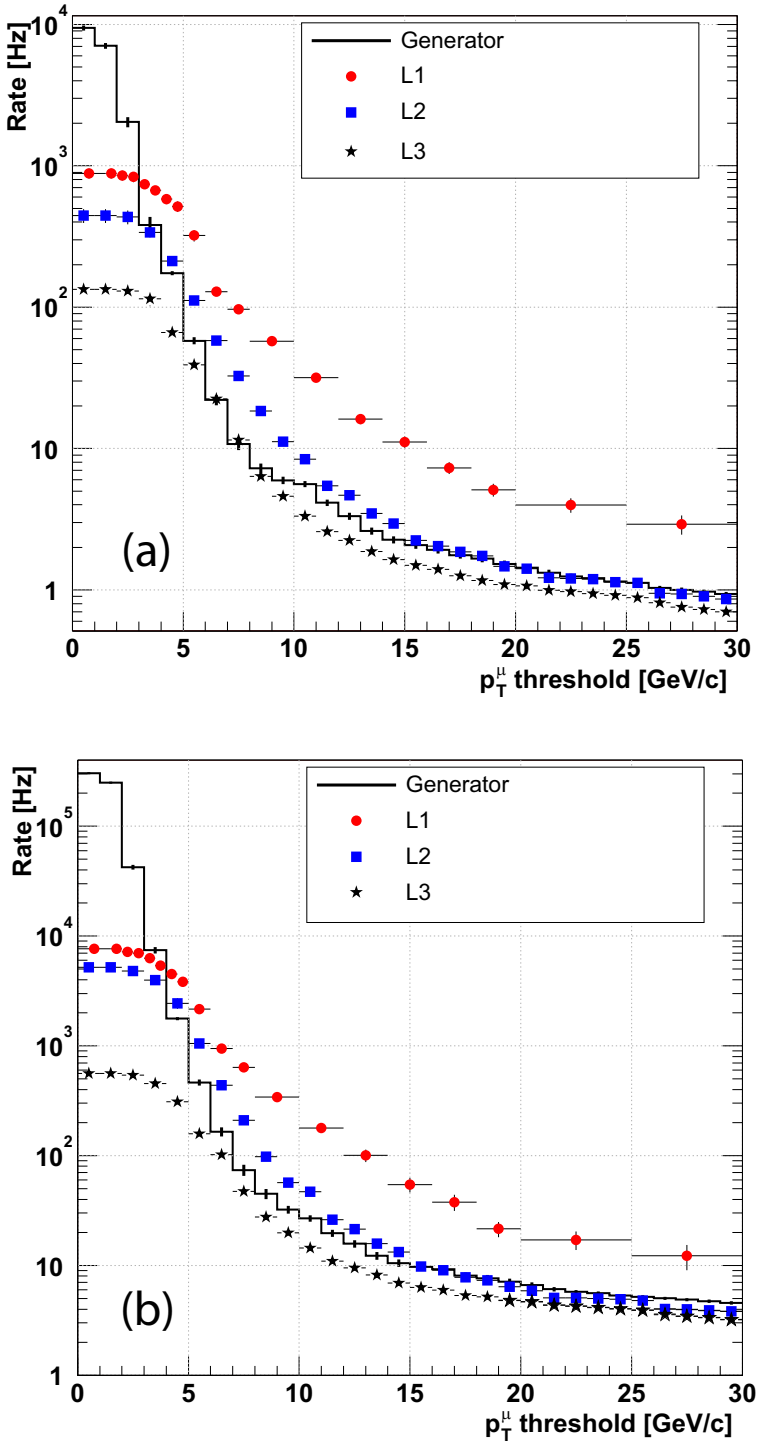


Figure 4.7: *Di-muon trigger rates as a function of the symmetric di-muon threshold at low luminosity (top) and high luminosity (bottom) [1].*

Chapter 5

Muon Isolation

As discussed in the previous chapter, a general criterion for LHC triggers is the requirement of the presence of high-p_T objects like leptons and jets. Typical thresholds for the muon trigger are of the order of 15–25 GeV/c [48]; higher values affect the trigger efficiency for the selection of interesting events. Unfortunately, in this range the muon background from minimum-bias events is still quite high, as shown in Fig. 4.6. The largest source of background below p_T thresholds of about 30 GeV/c is the decay chain of b, c quarks; the rejection of these muons is difficult since they are real and prompt, *i.e.* produced close to the interaction point.

A way to separate muons from b, c decays from those from signal events relies on the fact that b and c quarks are produced in jets while muons from heavy object decays (like $W \rightarrow \mu\nu$) are *isolated* – *i.e.* not surrounded by other particles, except for those from pile-up collisions. Isolation algorithms can be implemented to discriminate events on the basis of any measured quantity related to the energy flow in a region around the direction of the muon. In particular, the energy deposits in the calorimeters and the transverse momentum of particles reconstructed by the tracking system can be used.

As described in Section 2.5.1.1, the Level-1 muon trigger has the possibility to use calorimeter information with coarse granularity to identify isolated muons. A preliminary study on the tuning and performance of this selection is described in Section 5.1.

More sophisticated isolation algorithms can be implemented for the High Level Triggers. Several algorithms using the full granularity of the calorimeters and the reconstruction of tracks at Level-2 and Level-3 have been implemented and are described in Section 5.2.

5.1 Muon Isolation at Level-1

As described in Section 2.5.1.1, the Level-1 calorimeter trigger provides two bits of information to the Level-1 Muon Trigger for each calorimeter region (4×4 towers, *i.e.* $\Delta\eta \times \Delta\phi = 0.35 \times 0.35$). The "MIP" bit denotes an energy deposit compatible with the passage of a minimum ionising particle in the region. The "Quiet" bit is assigned if the transverse energy deposit in the region is below a programmable threshold, and can be used to determine if a given muon candidate is isolated.

As a first step, the Level-1 Muon Trigger extrapolates the coordinates of muon candidates either to the calorimeter surface or to the vertex, in order to find the corresponding calorimeter region. The implementation of the extrapolation logic in the trigger hardware is described elsewhere [23]. In the case of the Quiet bit, the extrapolation is done to the vertex, since the muon direction at the vertex is the best estimate of the axis of a hypothetical cone accompanying the muon. Several contiguous calorimeter regions in a rectangular window (1×1, 1×2, 2×2,... regions) can be queried for the state of the Quiet bit. The muon candidate is labelled as isolated if all regions in the window have the Quiet bit set. When DT, CSC and RPC candidates are merged to create one Global Muon Trigger candidate, the resulting Quiet bit is calculated as the logical AND of the bits of the candidates used.

A very preliminary study on Level-1 isolation using the Quiet bit is described in [49]. The results reported in this section were obtained with a more realistic simulation, which includes the detailed description of the calorimeter and muon triggers, and in particular of the muon extrapolation logic. A sample of $W \to \mu\nu$ decays and one of minimum bias events were used as benchmark signal and background, respectively. They were generated in the year 2000 with a procedure [50] similar to that described in Chapter 3. The trigger simulation was performed with ORCA version 5.

The Level-1 isolation algorithm has only two parameters subject to optimisation: the value of the threshold on the transverse energy deposit in the calorimeter regions and the size of the isolation window. Wider windows have more probability to include the jets accompanying non-isolated muons. However, the probability to discard isolated muons increases as well, due to the presence of pile-up, especially at high luminosity. For the high luminosity case, the best results were obtained with a 2×2 window, for a Quiet threshold of 5 GeV. The resulting efficiency as a function of the generated muon p_T is shown in Fig. 5.1a, separately for direct $W \to \mu\nu$ decays and for the minimum bias background. It can be noted that this isolation algorithm is very effective in rejecting high-p_T minimum bias muons, and that the efficiency of selecting signal muons is high. However, the algorithm is not effective against low-p_T minimum bias muons. These muons in fact are usually accompanied by soft, wide jets

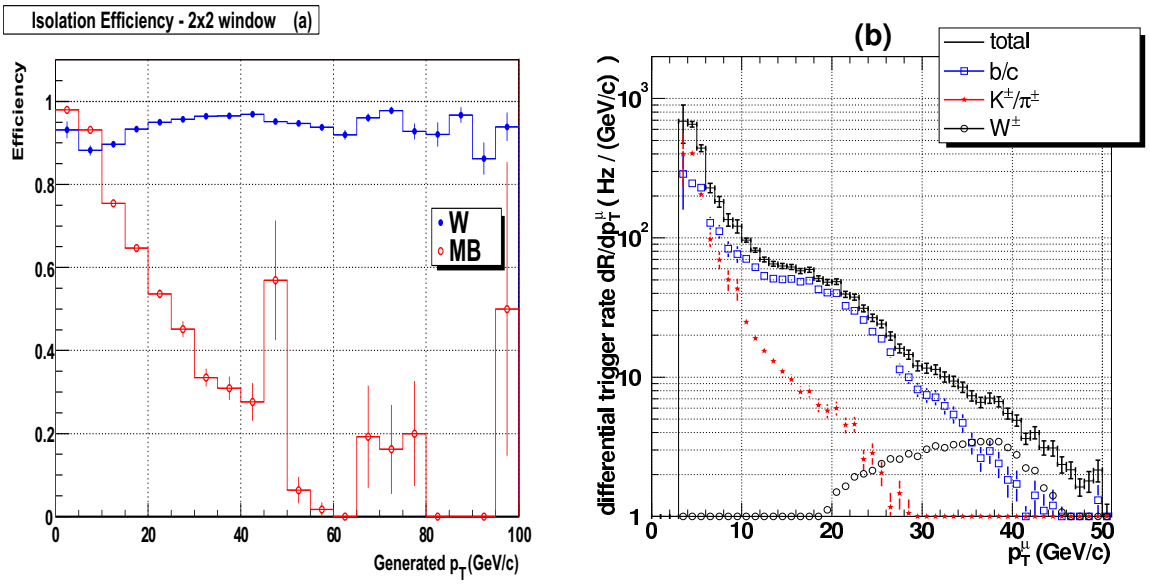


Figure 5.1: *(a) Efficiency of the Level-1 isolation algorithm for muons coming from direct W decays and for minimum bias events, as a function of the generated muon p_T. Only muons passing a Level-1 p_T threshold of 25 GeV/c are included. (b) Spectrum of the muons passing a Level-1 p_T threshold of 25 GeV/c [23], cf. Section 2.5.1.3. It can be noted that the spectrum is dominated by low-p_T muons, against which isolation is not effective.*

that leave small deposits in the calorimeters. The spectrum of the muons passing a Level-1 p_T threshold of 25 GeV/c is shown in Fig. 5.1b. As discussed in Section 2.5.1.3, it is dominated by very low-p_T muons due to the combined effect of non-Gaussian tails in the p_T resolution of Level-1 muons and the steeply falling behaviour of the p_T spectrum of minimum bias events. Therefore, the rate reduction obtained with the Level-1 isolation algorithm on the inclusive single-muon stream is negligible.

This problem could be mitigated only by identifying muons with highly overestimated p_T. In principle the MIP bit, that indicates the passage of a minimum ionising particle in the region, can be used for this purpose: a wrong p_T assignment will result in the extrapolation to the wrong region, where the absence of a MIP bit will indicate that the muon p_T is not correctly estimated.

A calorimeter trigger region is composed by $4 \times 4 = 16$ trigger towers. The MIP bit is set if the energy deposit in at least one of the 16 HCAL trigger towers in the region lies within a programmable range (cf. Section 2.5.1.1). Typical ranges are of the order of 1 to 2.5 GeV. The muon trigger extrapolates the coordinates of muon candidates to the calorimeter surface and associates them to the MIP bit set in the corresponding region. When DT, CSC and RPC

candidates are merged to create Global Muon Trigger candidates, the resulting MIP bit is calculated as the logical OR of the bits of the candidates used. While the muon propagation and assignment logic is already implemented [23], no simulation of the MIP assignment in the HCAL trigger is yet existing. However, some considerations on the feasibility of this selection can be made.

The probability that the propagation of a muon candidate reaches a different ϕ calorimeter region from that actually traversed by the muon is shown in Fig. 5.2 as a function of the muon p_T, for muons passing a 25 GeV/c Level-1 threshold. Muons below 25 GeV/c in this figure represent the mismeasured muons that have to be rejected. This plot gives the maximum rejection factor

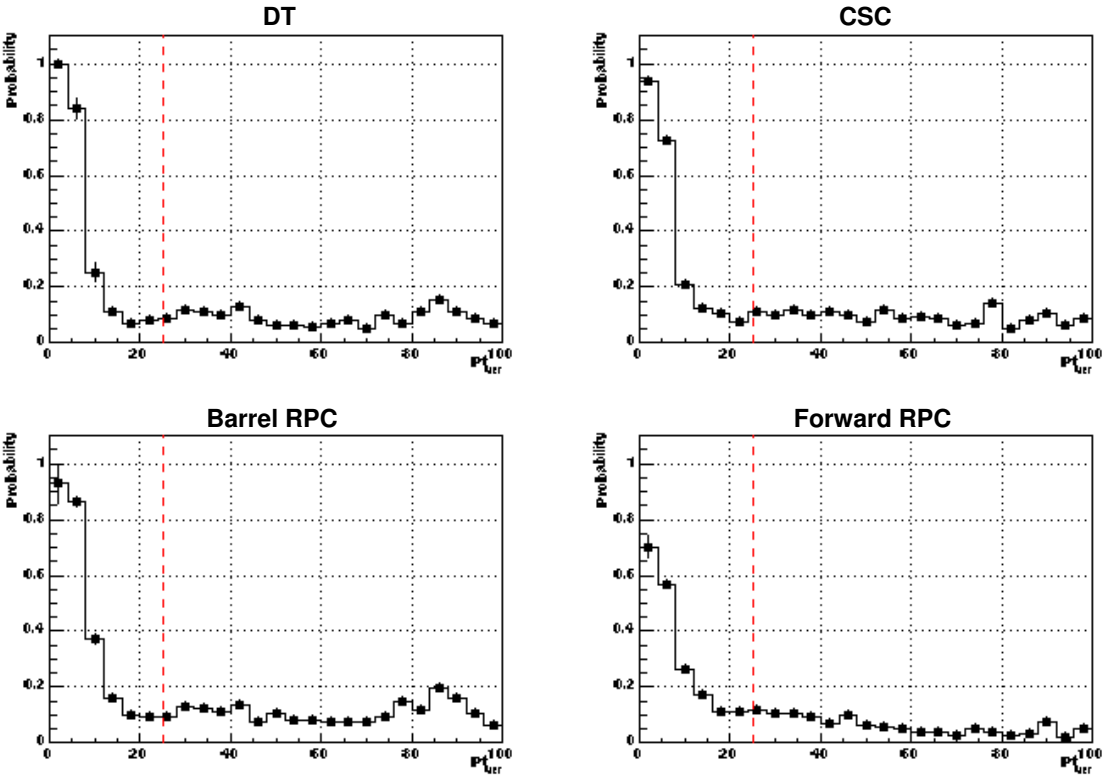


Figure 5.2: *Probability that the propagation of the muon candidates of the DT, CSC, barrel and forward RPC systems reaches a different calorimeter region in ϕ than that actually traversed by the muon, as a function of the generated muon p_T. Only muons passing a 25 GeV/c Level-1 threshold are included.*

that can be obtained using the MIP bit. For all muon subsystems, it is sizeable ($1.2 \sim 2$) for muons below 10 GeV/c.

However, the assignment of the MIP bit is affected by the presence of pile-up and detector and noise in the read-out and trigger electronics. This effect is

amplified by the fact that the region's MIP bit is defined as the logical OR of the MIP bits of the 16 independent towers that compose the region. At the time of the writing of this thesis, it is not clear what the actual noise occupancy of the HCAL trigger primitive used to define the MIP will be. An estimate of the total MIP occupancy due to pile-up and noise [51] at high luminosity indicates that for typical thresholds, *e.g.* [1.0,2.5] GeV, as much as $\sim 15\%$ of the regions can have a spurious MIP, spoiling the rejection power of the algorithm.

The results obtained with a simple simulation of the MIP assignment, which does not include a detailed simulation of the HCAL readout and of trigger electronics and noise, is shown in Fig. 5.3 as a function of the generated muon p_T, for muons passing a 25 GeV/c Level-1 threshold. Muons below 25 GeV/c in this plot are the mismeasured muons that have to be rejected, that represent the dominant contribution to the Level-1 rate. It can be seen that the algorithm

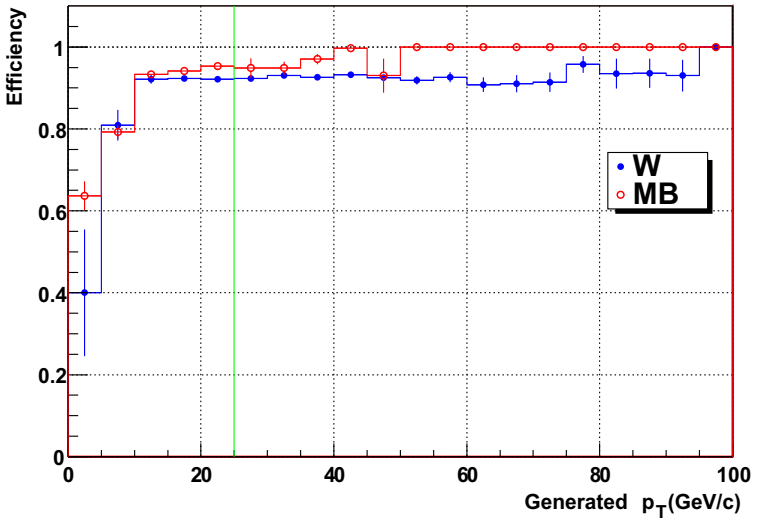


Figure 5.3: *Efficiency obtained with a simple simulation of the MIP selection on muons coming from direct W decays and from minimum bias events as a function of the generated muon p_T, at high luminosity. Only muons passing a 25 GeV/c Level-1 threshold are included.*

gives a reasonable rejection of these muons, which is similar for both signal and background. However, a more realistic simulation of the HCAL electronics and MIP assignment is necessary to assess the actual performance of this algorithm.

In conclusion, as of today it is not clear if the MIP will allow a significant rejection of low-p_T muons, and anyhow its performance will be strongly influenced by pile-up and detetcor and electronics noise. The Level-1 isolation algorithm, on the other hand, does not significantly reduce the inclusive muon trigger rates, due to the very soft p_T spectrum of muons selected by the Level-1 trigger. There is therefore no advantage in requiring isolation at such an early

trigger step, and it is preferable to apply isolation algorithms in the HLT, where the full granularity of the calorimeters can be exploited. The Level-1 isolation algorithm could be effective for selecting well isolated muons in specific trigger streams, where the backgrounds are non-isolated muons with high p_T. However, there is currently no plan for such exclusive triggers at Level-1.

5.2 Muon Isolation in the HLT

The flexibility of the software implementation of the HLT allows the development of sophisticated isolation algorithms. The following sections describe three different algorithms implemented using the transverse energy deposit in calorimeters and the transverse momenta of the tracks reconstructed in the tracking system. They are referred to as calorimeter, pixel and tracker isolation in the following.

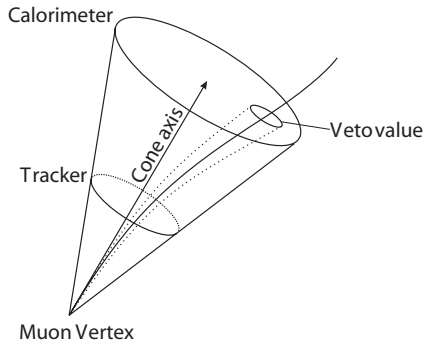
Calorimeter isolation. The transverse energy measured in the towers of the hadronic calorimeter (HCAL) is combined with the reconstructed transverse energy deposit in the electromagnetic calorimeter (ECAL). Since it is based on the calorimeters, this algorithm becomes less effective at high luminosity because of pile-up. This algorithm is described in detail in Section 5.2.3.

Pixel isolation. The pixel detector (cf. Section 2.2.1) in its baseline design consists of 3 cylinders of pixel detectors surrounding the beam pipe, supplemented by two disks in each endcap. The disks and cylinders are arranged so that each track coming from the nominal beam crossing point with pseudorapidity $|\eta| \leq 2.5$ crosses at least 3 layers of pixel detectors, thus allowing track reconstruction even without the aid of the silicon strip detectors. The pixel isolation algorithm, described in Section 5.2.2, is based on the measurement of the sum of the transverse momenta of tracks reconstructed around the muon. Only tracks originating from the same collision vertex as the muon are used in the sum, so that the algorithm is less sensitive to pile-up than calorimeter isolation.

Tracker isolation. While pixel-only reconstruction algorithms are very fast, the reconstruction with the full tracking system (where the average number of hits per track is 12) is more robust, efficient and accurate. The typical full tracker transverse momentum resolution for low-p_T tracks ($\mathcal{O}(1$ GeV)) is $\sigma(p_T)/p_T = 1\%$, to be compared with the 8% resolution obtained with the pixel-only reconstruction. Apart from this, the tracker isolation algorithm, which is described in detail in Section 5.2.4, is similar to the pixel isolation one.

The isolation algorithms are implemented in a consistent framework since they all share the same logic. The two main building blocks are the extraction of the signal deposited around the muon and the comparison of the result with a predefined threshold, which can be η-dependent in order to guarantee a flat efficiency in η. The signal extraction is specific to the algorithm and detector used; in the philosophy of object-oriented software, it is implemented in different interchangable modules for the different algorithms. The rest of the code, including the part used for the optimisation procedure described in Section 5.2.1, is common.

Energy deposits or tracks around the muon are collected in a cone defined by the condition $\Delta R \leq \Delta R_{MAX}$, where $\Delta R = \sqrt{\Delta\eta^2 + \Delta\varphi^2}$, $\Delta\eta$, $\Delta\varphi$ being the distances in pseudorapidity and azimuthal angle between the deposit and the cone axis. The cone axis is chosen according to the muon direction with a procedure that is tailored to the specific properties of each algorithm. The muon itself contributes to the detector measurement inside the cone; this contribution (called *veto value* in the following) can be subtracted to improve the discriminating power of the isolation algorithm. A schematic image of the isolation cone is shown in Fig. 5.4. Both the ΔR of the cone and the threshold



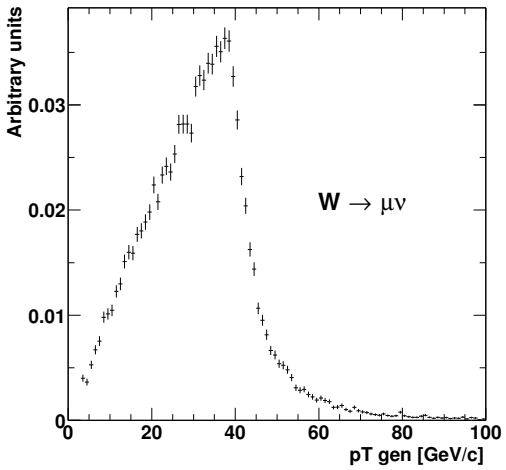


Figure 5.4: *(Left) Schematic illustration of the isolation cone. In the case of ECAL and for pixel and tracker algorithms, the "cone axis" coincides with the muon direction at the vertex. In the case of HCAL, the cone axis is the direction of the centre of the tower pointed to by the muon direction at the vertex. The "veto value" is used to subtract the contribution of the muon from the detector measurement in the cone.*

Figure 5.5: *(Right) The p_T spectrum of muons from direct W decays.*

can be optimised in order to maximise the rejection of the background while keeping the efficiency on the reference signal above a predefined value. This optimisation is described in the following section.

The algorithms described here are designed to run both in the trigger chain and in the offline selection. For this reason, they have been implemented to work over the full acceptance of the muon system ($|\eta| < 2.4$), even if at LHC startup the trigger electronics will not be installed in the forward Cathode Strip Chamber (CSC) station ME1/1a, thus limiting the muon trigger acceptance to $|\eta| < 2.1$ (cf. Section 2.5.1.2).

5.2.1 Optimisation of the Algorithms

The parameters of the isolation algorithms were optimised by maximising the rejection of the events in a *reference background* sample while keeping the efficiency on a *reference signal* channel with isolated muons above a predefined nominal value. Such predefined values of the reference signal efficiency (which we will refer to as *nominal efficiency* in the following) can be used to tune the algorithm performance for a particular study, *i.e.* to choose the balance between the background rejection and the efficiency on the signal of interest.

In this study, the direct $W \to \mu\nu$ decay was chosen as reference signal since it contains well isolated muons with a p_T spectrum in the range relevant to the expected HLT p_T thresholds (see Fig. 5.5). The reference background sample is composed of events with muons coming from minimum bias collisions. Both the reference signal and the reference background samples should include only events with muons above typical trigger thresholds, since the performance of isolation algorithms depends on the p_T of the muon and no optimisation is needed for low-p_T muons which should be discarded anyhow by the trigger. Therefore only events with muons with transverse momentum above 16 (22) GeV/c are considered for the optimisation at low (high) luminosity. It should be noted that this selection is applied to the p_T of the Monte Carlo muons, and not to the p_T measured by the trigger. This is necessary since, for whatever p_T threshold, the Level-1 and Level-2 trigger selection is highly contaminated by the feed-through of low-p_T muons (cf. Section 2.5.1.3), which should not be included in the reference signals.

The parameters subject to optimisation are the size of the cone and the value of the threshold. Since some detector effects (resolution, noise) are η-dependent, to obtain a flat efficiency versus η on the reference signal the optimisation is done independently in each of the pseudorapidity bins listed in Table 5.1, which correspond to the η segmentation of the calorimeter towers.

In order to find the optimal cone sizes, we define a set of cones of different size (Table 5.2).

First, the $\sum E_T$ or $\sum p_T$ inside all these cones is determined for the muons

Table 5.1: *The maximum η values for the pseudorapidity bins used for the optimisation of the size and threshold of the cones. The first bin starts at η=0. Bins for negative values of η are obtained by mirror reflection.*

η-bin	1	2	3	4	5	6	7	8	9
η_{MAX}	0.087	0.174	0.261	0.348	0.435	0.552	0.609	0.696	0.783
η-bin	10	11	12	13	14	15	16	17	18
η_{MAX}	0.87	0.957	1.044	1.131	1.218	1.305	1.392	1.479	1.566
η-bin	19	20	21	22	23	24	25	26	
η_{MAX}	1.653	1.74	1.83	1.93	2.043	2.172	2.322	2.5	

Table 5.2: *The set of cones used to optimise the cone size.*

Cone	1	2	3	4	5	6	7	8	9	10	11	12	13	14
ΔR_{MAX}	0.02	0.045	0.09	0.13	0.17	0.2	0.24	0.28	0.32	0.38	0.45	0.5	0.6	0.7

in the reference signal sample. For each cone and each η-bin, the threshold corresponding to a set of nominal efficiency values on the reference signal (typically 75%,...,90%,...95%,..97%...) is determined. The optimal cone for a given nominal efficiency is defined as the one which gives the maximal rejection on the reference background sample in the full η range. The result of this optimisation procedure is that for any predefined nominal efficiency a cone size is chosen, with thresholds defined in bins of pseudorapidity.

It is important to mention that the optimisation is affected by two factors: the LHC luminosity and the p_T range of the muons used. The luminosity affects the average energy deposited per unit of ΔR in each event, so that the thresholds for a given nominal efficiency must be scaled with luminosity. This dependence is not critical: if the actual luminosity is lower than the value used for the optimisation, *e.g.* due to the decrease of luminosity during a run, the signal efficiency will simply be higher than the nominal value. Results obtained after optimising the algorithms for the high and low luminosity LHC operating modes are presented for all algorithms described in the following. As already mentioned, the rejection power is highly dependent on the muon p_T, which is correlated to the energy of the accompanying jet [49]. It is therefore important to optimise the threshold for the p_T range of interest, in order to obtain the desired efficiency for useful muons (*i.e.* those actually selected by the trigger p_T cut). However, it is still possible to use the algorithm for muons outside the

optimised p_T range; the corresponding signal efficiency will be slightly different than the nominal value.

5.2.2 Calorimeter Isolation

The calorimeter isolation algorithm uses as input the direction and momentum of the Level-2 reconstructed muon [24] at the vertex (*i.e.* the point of closest approach of the muon track to the beam line in the plane transverse to the beam). The muon direction at the vertex is the best approximation of the direction of a possible accompanying jet and is used in the definition of the cone axis as described in the following.

The extraction of the energy deposits is done independently in the ECAL and the HCAL. In the case of ECAL the measured quantity is the $\sum E_T$ in the crystals around the muon direction at the vertex. In the case of HCAL, which has a much coarser segmentation than the ECAL, the cone axis is defined instead as the centre of the tower to which the muon direction at the vertex points; the measured quantity is the $\sum E_T$ of the towers with centres inside the cone. This guarantees that the same number of towers contributes to all cones of a given size at a given pseudorapidity.

In order to reject pile-up deposits, the HCAL towers with reconstructed transverse energy (E_T) below 0.5 GeV and the ECAL crystals with reconstructed E_T below 0.2 GeV are neglected. To avoid electronic and detector noise, an additional energy (E) threshold of 0.12 GeV in barrel ECAL, 0.45 GeV in endcap ECAL and 0.6 GeV in HCAL is applied. These values correspond to 3 standard deviations of the nominal noise level.

To subtract the energy deposited in the cone by the muon itself, the muon trajectory is extrapolated to the boundary between ECAL and HCAL using the package GEANE [45]. The distance between the extrapolated muon position and the centre of the calorimeter cell traversed by the muon in the ECAL or HCAL is shown in Fig. 5.6. In the ECAL the transverse energy of crystals in a small area of $\Delta R \leq 0.07$ around the muon extrapolation is subtracted from the cone measurement. In the HCAL the transverse energy of a single tower is subtracted, chosen as the tower with highest deposit among those with centre at $\Delta R \leq 0.1$ from the muon extrapolated point.

The actual isolation variable is constructed from both HCAL and ECAL deposits in the cones (after subtraction of the veto value) using a weighting parameter α:

$$E_T^{WEIGHT} = \alpha \cdot \sum {E_T}^{ECAL} + \sum {E_T}^{HCAL}. \tag{5.1}$$

The optimal weighting parameter has been found to be $\alpha = 1.5$ (see below). Typical spectra of E_T^{WEIGHT} for the reference signal sample are shown in Fig. 5.7 for cone number 6 in different η-bins. For each cone and η-bin, the threshold is

given by the point where the integral of the normalised spectrum is equal to the nominal efficiency value[1]. The resulting thresholds for one of the cones and for some values of the nominal efficiency are shown in Fig. 5.8 and 5.9 for the low and high luminosity cases, respectively. As expected, the thresholds are higher for high luminosity and vary with pseudorapidity.

Figure 5.10 shows the η-averaged efficiency for each cone and nominal efficiency. The efficiency is by construction above the nominal value. For smaller cone sizes the curves for different nominal efficiencies tend to converge. This is due to the fact that in small cones around an isolated muon the deposit can be small or zero, so that the optimised threshold is the same regardless of the nominal efficiency.

To determine the optimal cone sizes, the same thresholds are then applied to the reference background sample. Figure 5.11 shows the background efficiencies obtained using different cones. The optimal cone for a given nominal efficiency is given by the minimum of each curve. The results of this optimisation are summarised in Table 5.3.

Table 5.3: *Optimal cones for calorimetric isolation as a function of the nominal efficiency. Numbers correspond to the cones defined in Table 5.2.*

Nominal efficiency	0.75	0.80	0.85	0.90	0.95	0.97	0.98	0.99
At $\mathcal{L}$ =2×10^{33}cm^{-2}s^{-1}	10	9	8	8	8	7	7	7
At $\mathcal{L}$ =10^{34}cm^{-2}s^{-1}	8	8	8	8	7	6	6	6

The same cone size is chosen for a given nominal efficiency for all η-bins. It was checked that the independent optimisation of the cone size in a few η ranges leads to a slight improvement in the rejection factor; however we are not confident that such optimisation is significant given the available statistics, and we do not use it in the following.

With the optimised cones of Table 5.3 the performance of the algorithm can be studied. In Fig. 5.12 the isolation efficiency for the reference signal is shown as a function of the p_T of the generated muon. The efficiency is slightly decreasing for low $p_T$values. However the efficiency is rather flat for muons passing typical trigger thresholds[2]. The efficiency is by construction flat in pseudorapidity, as shown in Fig. 5.13.

Similar plots can be made for the minimum bias background. Figure 5.14 shows that the efficiency for the reference background is flat as a function of η.

[1]Due to the finite size of the bins used in this procedure, the resulting signal efficiency may be *above* the nominal value.

[2]Note that the nominal efficiencies are obtained convoluting these curves with the actual p_T spectra of the reference signal (Fig. 5.5).

However the background efficiency strongly depends on the muon p_T (Fig. 5.15). As already mentioned, this is a general feature of isolation algorithms, due to the correlation between the p_T of the muon and the energy of the accompanying jet.

The performance of the algorithm can be summarised by plotting the efficiency for minimum bias events as a function of the efficiency for signal events. The background rejection efficiency is highly dependent on the muon p_T and such plots should have a cutoff for the minimal p_T taken into account (lower p_T values are supposed to be discarded by the muon trigger cut). The curves for a few different p_T cutoff values are shown in Fig. 5.16.

Such plots allow the comparison of the final performance of different algorithms. In particular, the effect of different values for the weighting parameter α is presented in Fig. 5.17, which shows that $\alpha = 1.5$ is close to optimal in most of the cases.

5.2.3 Pixel Isolation

The pixel isolation algorithm uses as detector measurement the $\sum p_T$ of tracks reconstructed in a cone around the muon by the pixel detector alone, neglecting the p_T of the muon itself.

The pixel reconstruction algorithm [52] looks for pixel hits compatible with tracks with transverse momenta as low as 1 GeV/c. The track candidates are used to fit primary vertices; track candidates with no association to reconstructed vertices are rejected. The algorithm returns a list of vertices with the corresponding tracks and their momenta. Vertices are sorted by the $\sum p_T$ of tracks assigned to them, allowing the identification of the primary vertex of the hardest interaction (usually the most interesting one in the bunch crossing).

The main drawback of this isolation algorithm is that the pixel reconstruction relies on the reconstruction of three hits out of the three layers available, and therefore it is very sensitive to detector, geometrical and electronic read-out inefficiencies. Moreover, it will not be available in staging scenarios where only two pixel layers will be installed.

5.2.3.1 Pixel Isolation with Level-3 Muons

An isolation algorithm using pixel tracks is powerful if applied on Level-3 muons, which are reconstructed with enough precision to provide a useful estimate of the muon vertex (*i.e.* the point of closest approach of the muon trajectory to the beam line). It is therefore possible to require that all pixel tracks contributing to the energy measurement in the cone come from the same primary vertex as the muon.

Figure 5.18a shows the distance ΔR between the direction of the muon and the direction of the closest pixel track. The peak for $\Delta R < 0.015$ indicates a very good matching between the Level-3 reconstructed muon and the corresponding pixel track (hereafter called *pixel muon*). Figure 5.18b shows the distance along the beam line (the Z coordinate in the CMS reference frame) between the vertex associated to the pixel muon and the Level-3 muon vertex. A cut of $|\Delta Z|$ <0.2 cm for pixel tracks contributing to $\sum p_T$ has been applied.

A set of thresholds was created with the same procedure described in the previous section for the calorimeter isolation algorithm. Typical spectra of $\sum p_T$ are shown in Fig. 5.19 for cone number 6 in few selected η-bins. The thresholds assigned to the set of predefined cones are shown in Fig. 5.20 and 5.21 for a few different nominal efficiency values. The relative difference between high and low luminosity is smaller than in the case of calorimeter isolation, due to the requirement that the contributing tracks come from the same primary vertex.

The efficiencies for reference signal events are shown in Fig. 5.22. For small cones, the lack of tracks in the vicinity of the muon makes the determination of a threshold impossible for any nominal efficiency, and the curves for different nominal efficiencies overlap. This effect is more pronounced here than in the case of calorimeter isolation.

Following the procedure already described for calorimeter isolation, the computed thresholds are used to determine the best cone for each nominal efficiency value (Fig. 5.23). The results are summarised in Table 5.4.

Table 5.4: *Optimal cones for pixel isolation as a function of the nominal efficiency. Numbers correspond to the cones defined in Table 5.2.*

Nominal efficiency	0.75	0.80	0.85	0.90	0.95	0.97	0.98	0.99
At $\mathcal{L} = 2 \times 10^{33}\text{cm}^{-2}\text{s}^{-1}$	11	10	9	9	7	6	6	6
At $\mathcal{L} = 10^{34}\text{cm}^{-2}\text{s}^{-1}$	10	10	9	8	7	6	6	6

In Figs. 5.24 and 5.25 the resulting efficiencies for the reference signal sample are shown as a function of the pseudorapidity and p_T of the generated muon. As discussed earlier, the algorithm ensures by construction that the resulting efficiency is above the nominal efficiency and that it is flat in η. The efficiency is also flat as a function of p_T except for a small drop at high luminosity for low nominal efficiency values and low-p_T muons.

The efficiency for the reference background (Fig. 5.26) is fairly flat as a function of pseudorapidity, except for an increase at high η values. The reason is that the pixel reconstruction requires the presence of three pixel hits for each track, while very forward tracks do not cross all three pixel layers when their vertex is displaced from the detector centre. As for calorimeter isolation, and

for the same reason, the background efficiency depends strongly on the p_T of the muon, as shown in Fig. 5.27.

The background efficiency obtained as a function of the signal efficiency is shown in Fig. 5.28.

5.2.3.2 Pixel Isolation with Level-2 Muons

The pixel reconstruction algorithm can also be used at Level-2, using the Level-2 muon to define the cone axis. However the Level-2 muon reconstruction does not provide a precise enough measurement of the muon vertex (see Fig. 5.29), to be used to select the pixel tracks that contribute to the cone.

An estimate for the primary vertex of the "most interesting event" in the bunch crossing is given by the pixel reconstruction algorithm itself. In Fig. 5.30 the distance along the beam axis between the Monte Carlo primary vertex of the event containing the triggering muon and the vertex reconstructed by the pixel reconstruction algorithm in the minimum bias sample is shown. The peak at zero corresponds to correct associations. However, the association efficiency is not satisfactory (even at low luminosity), showing that this estimate cannot be used effectively. Another possibility could be to estimate the position of the muon vertex using the pixel track closest to the muon. In Fig. 5.31 the distance ΔR between the direction of the Level-2 muon and the direction of the closest pixel track, obtained from the minimum bias sample, is presented. This distance is rather large ($\Delta R \leq 0.2$) due to the limited resolution of the Level-2 muon. Even though the pixel track found using this procedure usually points to the correct primary vertex, the efficiency of the association is again too low. We therefore conclude that it not possible to use efficiently the primary vertex constraint on the tracks contributing to the cone for Level-2 pixel isolation.

The optimisation procedure for pixel isolation with Level-2 muon differs therefore from what is described in the previous section because of the lack of the primary vertex constraint and because the pixel muon is searched in a much wider ΔR around the direction of the Level-2 muon. The results are shown in Fig. 5.32. The background rejection at low luminosity is comparable with what obtained with pixel isolation at Level-3. However, at high luminosity the rejection is smaller due to the missing vertex constraint. This algorithm will not be discussed further in the following.

5.2.4 Tracker Isolation

The tracker isolation algorithm uses the $\sum p_T$ of fully reconstructed tracks in a cone around the direction of the Level-3 muon, neglecting the contribution from the muon itself.

Tracks are reconstructed using regional tracking, *i.e.* track seeds are created using pairs of pixel hits in a region of interest defined by a vertex constraint, by constraints on the track direction at the vertex and by the minimum transverse momentum for the tracks to be reconstructed. The direction of the tracks is constrained to be in a region of width $\Delta\eta \times \Delta\varphi$ (determined by the size of isolation cone) around the cone axis, defined as the direction of the Level-3 muon. The vertex constraint is specified by the radius r and the half-width Δz of the cylinder around the beam line which contains the impact point of the Level-3 muon. The transverse and longitudinal displacement of the Level-3 muon vertex from the primary vertex is shown in Fig. 5.33, which allows to choose the values $r = 0.1$ cm and $\Delta z = 0.2$ cm.

Another important parameter is the minimal transverse momentum required for the tracks contributing to the isolation cone. The optimal value for this parameter has been found to be 0.8 GeV/c (see below).

Since very high quality reconstruction of the tracks contributing to the isolation cone is not essential, the track fitting was stopped as soon as five hits were included in the fit. This allows a significant speed-up of the algorithm. However, in the case of low-p_T particles there is some probability of reconstructing ghost tracks. In order to reject them, the value of the χ^2 of the track fit was required to be less than 8 for the tracks with only two pixel hits supplemented by three silicon strip hits. This requirement is not necessary if one more hit is used in the track fit.

Once tracks are reconstructed, the same algorithm used for the pixel isolation with Level-3 muon is used. Typical spectra of $\sum p_T$ are shown in Fig. 5.34 for cone number 6 in a few selected η-bins. The thresholds assigned to the set of predefined cones are shown in Fig. 5.35 and 5.36 for a few different nominal efficiency values. The thresholds are similar to those used by the Level-3 pixel isolation algorithm.

The efficiencies for reference signal events are shown in Fig. 5.37. As for the case of pixel isolation, for small cones the lack of tracks in the muon neighbourhood makes the determination of a threshold impossible for any nominal efficiency, and the curves overlap.

Following the procedure already described, the computed thresholds are used to determine the best cone for each nominal efficiency value (Fig. 5.38). The results are summarised in Table 5.5.

In Fig. 5.39 and 5.40 the resulting efficiencies for the reference signal sample are shown as a function of the pseudorapidity and p_T of the generated muon. The algorithm ensures by construction that the resulting efficiency is above the nominal efficiency and that it is flat in η. As for the case of the Level-3 pixel isolation algorithm, the efficiency is also fairly flat as a function of p_T.

The efficiency for the reference background (Fig. 5.41) is flatter as a function

Table 5.5: *Optimal cones for tracker isolation as a function of the nominal efficiency. Numbers correspond to the cones defined in Table 5.2.*

Nominal efficiency	0.75	0.80	0.85	0.90	0.95	0.97	0.98	0.99
At $\mathcal{L} = 2 \times 10^{33}\text{cm}^{-2}\text{s}^{-1}$	11	10	9	9	7	7	7	5
At $\mathcal{L} = 10^{34}\text{cm}^{-2}\text{s}^{-1}$	10	10	9	8	7	6	6	6

of pseudorapidity than in the case of the pixel isolation algorithm. The reason is that with full tracker reconstruction tracks can be reconstructed with only two pixel hits, while in the case of pixel reconstruction three hits out of the three pixel layers are necessary.

The background efficiency as the function of the p_T of the muon is shown in Fig. 5.42. The summary of the algorithm performance, *i.e.* the background efficiency obtained as a function of the signal efficiency, is shown in Fig. 5.43.

As already mentioned, the performance of this algorithm is influenced by the cut on the p_T of tracks contributing to the isolation cone. The effect of varying this cut is shown in Fig. 5.44, where the efficiency for the reference signal is shown as a function of the minimal p_T of contributing tracks for a few different values of nominal efficiency. This plot was obtained by repeating the threshold optimisation for each of the p_T cut values used. The dependence is not very strong, and the best performance is obtained for p_T cuts in the range 0.7–0.9 GeV/c. Hence, the value 0.8 GeV/c was chosen.

5.2.5 Performance

The multi-tiered structure of the CMS trigger allows for progressively more powerful muon isolation algorithms at each trigger level as more information becomes available. At Level-2, data from calorimeters are already available and the calorimeter isolation can be applied. It is also possible to process data from the pixel detector and to apply pixel isolation to the Level-2 candidate. At Level-3, candidates reconstructed using the full tracker information can be applied both pixel and tracker isolation. More than one isolation algorithm can be applied to the same muon in the different HLT steps; the best strategy (*i.e.* which algorithms are included in the HLT chain and in which order) should be determined by a trade-off between algorithm rejection power, efficiency and speed.

This section discusses the performance of the algorithms applied independently or in cascade in the trigger chain, in terms of signal efficiency and reduction in inclusive muon trigger rates. The performance in the selection of several important processes is discussed in the next chapter.

As a general remark, the performance of isolation algorithms is affected by the luminosity. The presence of pile-up forces the thresholds to be higher for the same nominal efficiency; this mostly affects the calorimeter isolation, while for tracking algorithms the effect of pile-up is reduced by the requirement that contributing tracks come from the same primary vertex as the muon.

5.2.5.1 Rate Reduction

The effect of isolation algorithms on Level-2 and Level-3 single-muon inclusive trigger rates is shown in Fig. 5.45. The Level-2 rate is above the generator rate for any trigger threshold, due to the feed-through of low-p_T muons which contaminate the Level-2 p_T spectrum (cf. Section 2.5.1.3). As already mentioned, isolation algorithms are not effective against these muons; therefore the Level-2 calorimeter isolation algorithm gives only a small reduction of the inclusive Level-2 rate. The Level-3 rate is not affected by low-p_T contamination, and Level-3 isolation algorithms show a significant rejection.

While the application of several isolation algorithms in sequence provides some additional rate reduction, the overall rejection is lower than the product of the rejection factors obtained with individual algorithms, since the quantities measured by the different isolation algorithms are not independent. In particular, no significant gain is given by the combined use of both the pixel and tracker isolation algorithms at Level-3. Hence, if the CPU timing is critical, a good compromise is to apply the calorimeter isolation algorithm at Level-2 and either the pixel or tracker isolation algorithms at Level-3.

The overall rate reduction is significant for thresholds below 30 GeV/c. Above that value the rate is dominated by isolated muons coming from W decays, as shown in Fig. 5.46, where the contributions of the different sources of muons to the Level-3 rate at high luminosity before and after the application of the isolation algorithms is presented. It can be noted that isolation algorithms strongly suppress the contributions from b, c, K, and π decays.

For the di-muon selection, a good balance between efficiency and rejection is obtained requiring that at least one of the two muons is isolated. The resulting rates are shown in Fig. 5.47.

5.2.5.2 Signal Efficiency

The efficiency of each isolation algorithm on the reference signal is by construction equal to the nominal efficiency. To study the correlation between the signal efficiencies, the algorithms were applied in sequence, setting the nominal efficiency of each to 97%. The total efficiency on the reference signal is given in Table 5.6. Since tracker and pixel algorithms exploit the presence of the same tracks around the muon, their signal inefficiencies are rather correlated. How-

ever the signal inefficiency of calorimeter isolation is largely uncorrelated from those of the pixel and tracker algorithms. This can be explained by the fact that the sources of inefficiency in the calorimeter and tracker algorithms are different. For example, pile-up in the calorimeters can come from an interaction vertex far from the one producing the muon, so that it does not affect the region around the muon in the tracking system. Noise and ghosts, which affect the calorimeter and tracker algorithms respectively, are also independent.

Table 5.6: *Total efficiency of combinations (logical AND) of isolation algorithms on the reference signal. Algorithms included are Level-2 calorimeter isolation ("Calo"), Level-3 pixel isolation ("Pixel") and Level-3 tracker isolation ("Tracker"), all set to* 97% *nominal efficiency.*

Algorithms	Calo+Pixel	Calo+Tracker	Pixel+Tracker	Calo+Pixel+Tracker
ε (Low luminosity)	0.947	0.947	0.960	0.937
ε (High luminosity)	0.946	0.946	0.960	0.935

The algorithms presented in this chapter are optimised using a sophisticate procedure, whose goal is both to get the maximal performance and to guarantee a flat signal efficiency as a function of pseudorapidity. It is however interesting to compare the results obtained with this optimisation with the performance of a simpler algorithm. As an example, the Level-3 pixel isolation algorithm has been compared with a simple algorithm based on the analysis of the number of tracks in the cone around the muon. Algorithms requiring no track, no more than one track, or no more than two tracks in a cone with a size in the range $\Delta R_{MAX} = 0.09 - 0.6$ have been implemented. Only pixel tracks coming from a primary vertex reconstructed by the pixel reconstruction algorithm closer than 0.2 cm to the Level-3 muon vertex are taken into account. The tracks in a small cone of $\Delta R < 0.015$ around the muon are excluded, to remove the contribution from the muon itself. The performance of such algorithms, in terms of background versus signal efficiency, is shown in Fig. 5.48, which demonstrates that the Level-3 pixel isolation algorithm always gives a better rejection for the same signal efficiency.

5.3 Conclusions

Isolation is a powerful criterion to reduce the background of muons coming from kaons, pions and b, c quarks. This chapter shows the implementation and performance of isolation algorithms in the Level-1 trigger and in the HLT.

A general conclusion is that the spectrum of input muons affects the performance of isolation algorithms. Low-p_T background muons, in fact, are usually accompanied by wide, low energetic jets that escape measurement. For this reason, the Level-1 trigger isolation algorithm does not allow a significant rate reduction, since the Level-1 rate is dominated by low-p_T muons for any trigger threshold.

Three isolation algorithms were implemented to be used in the HLT selection or in the offline analysis, and a complete description of these algorithms was published in a CMS Note [4]. They are based on reconstructed data from the calorimeters, the pixel detector and the full tracker. Their performance is tunable using a parameter called "nominal efficiency", which represents the expected efficiency on a reference signal sample, and are designed to provide flat signal efficiency as a function of pseudorapidity. It is shown that these algorithms give a considerable rate reduction in the High Level trigger, while keeping very high efficiency for isolated muons from signal events.

The performance of the HLT isolation algorithms was tested with the samples described in Chapter 3. It is worth to recall that, as described in Section 3.1, these samples were generated using the PYTHIA complex multiple interaction model, except for the b, c quark component of the minimum bias sample. For these events, the simple multiple interaction model was used, which predicts lower charged track multiplicity. This is expected to reduce the rejection power of isolation algorithms in the simulation, so that the results shown in this chapter should be considered conservative. Further studies are needed to determine the effect of the model for multiple interactions on the background rejection.

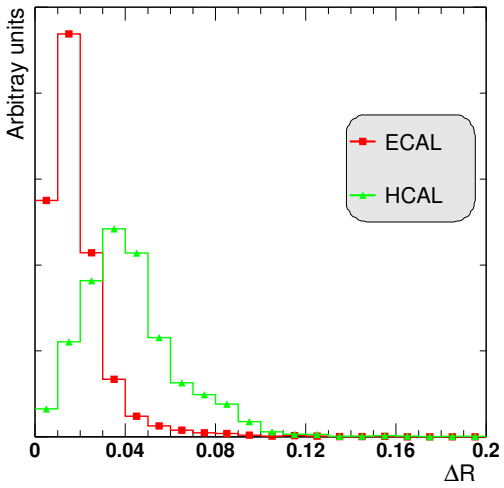


Figure 5.6: *Distance ΔR between the extrapolated position of the muon to the ECAL/HCAL boundary and the centre of the calorimeter cell travesed by the muon in the ECAL or HCAL.*

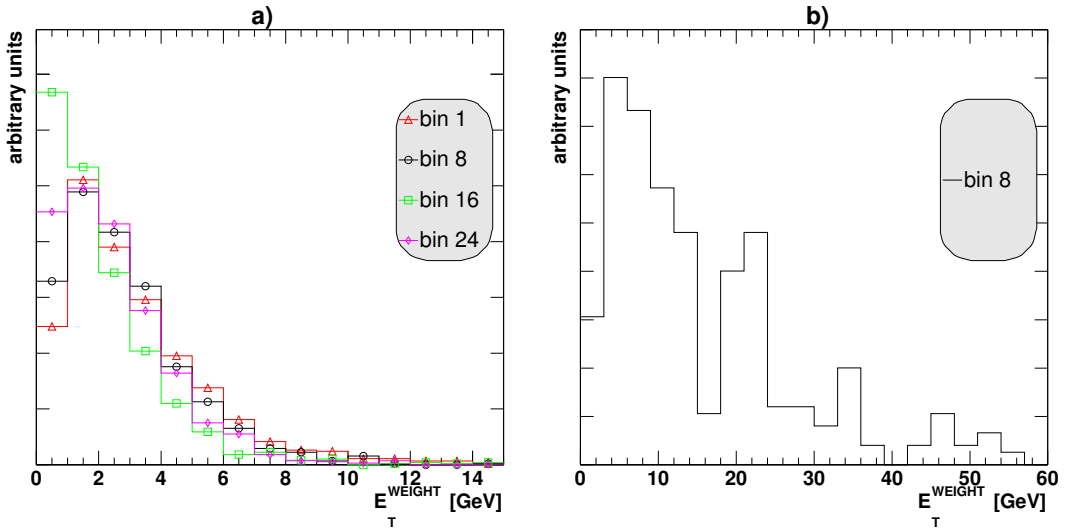


Figure 5.7: *Example of the spectra of E_T^{WEIGHT} in the calorimeters after subtraction of the veto value for cone 6 for the reference signal sample at high luminosity in η-bins 1, 8, 16 and 24 (a). For comparison the distribution of E_T^{WEIGHT} is shown for the reference background sample (cone 6, η-bin 8) (b).*

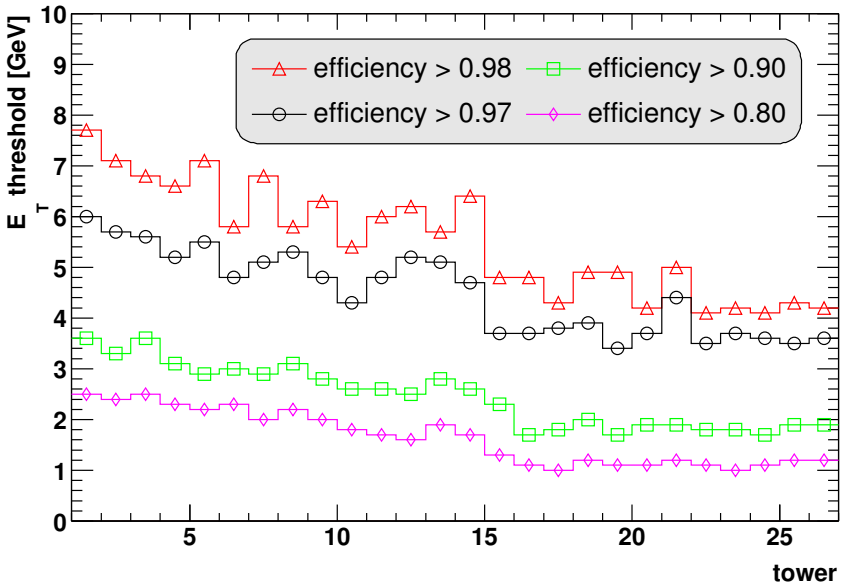


Figure 5.8: *Threshold values of the calorimeter isolation algorithm at low luminosity for cone 6 and nominal efficiencies 0.8, 0.9, 0.97 and 0.98.*

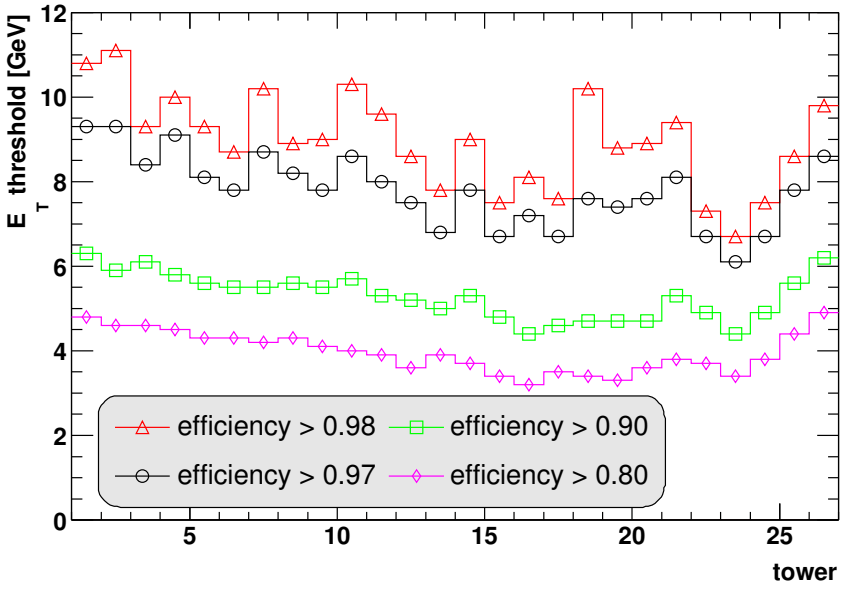


Figure 5.9: *Threshold values of the calorimeter isolation algorithm at high luminosity for cone 6 and nominal efficiencies 0.8, 0.9, 0.97 and 0.98.*

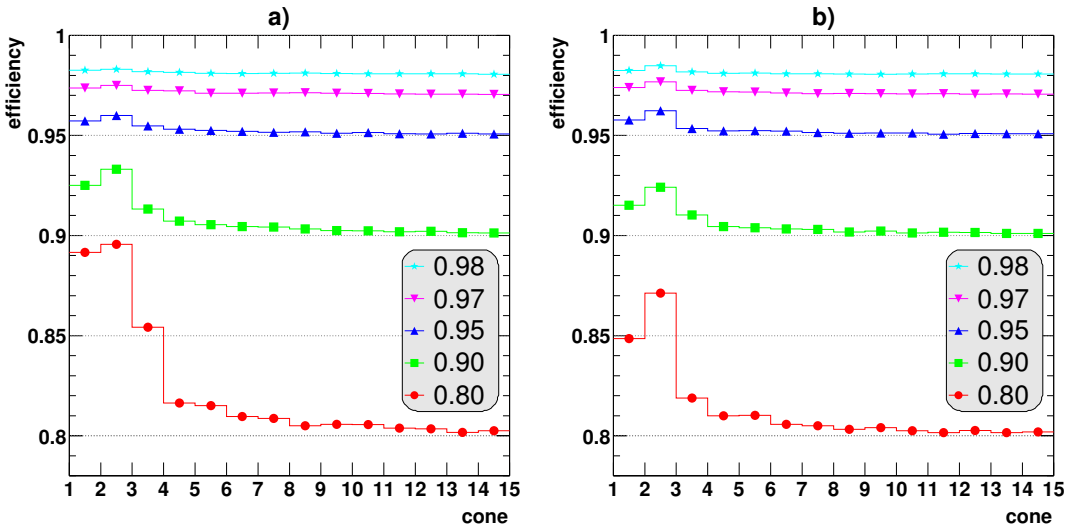


Figure 5.10: *Reference signal efficiency of the calorimeter isolation algorithm for different cone sizes and nominal efficiency values, at low (a) and high (b) luminosity.*

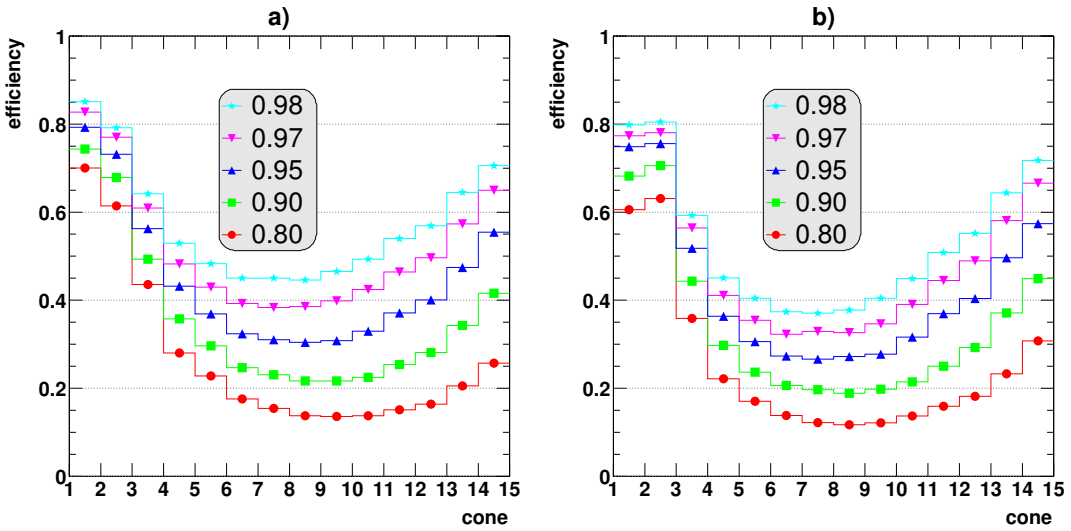


Figure 5.11: *Efficiency of the calorimeter isolation algorithm for the reference background sample for different cone sizes and nominal efficiency values, at low (a) and high (b) luminosity. The minimum of each curve indicates the optimal cone for the corresponding nominal efficiency.*

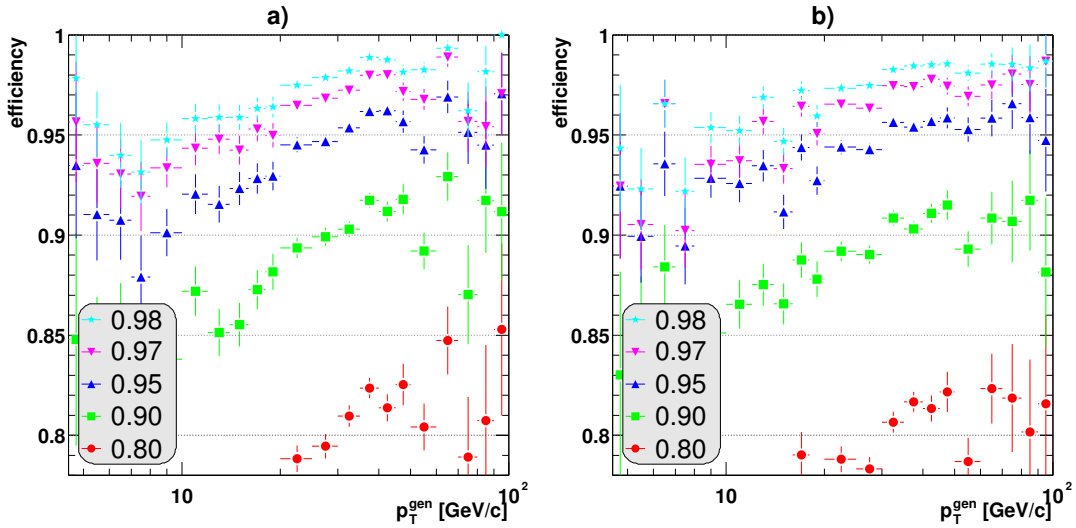


Figure 5.12: *Signal efficiency of the calorimeter isolation algorithm as a function of the p_T of the muon for different nominal efficiency values, at low (a) and high (b) luminosity. The nominal efficiencies are obtained by convoluting these curves with the actual p_T spectra of the reference signal.*

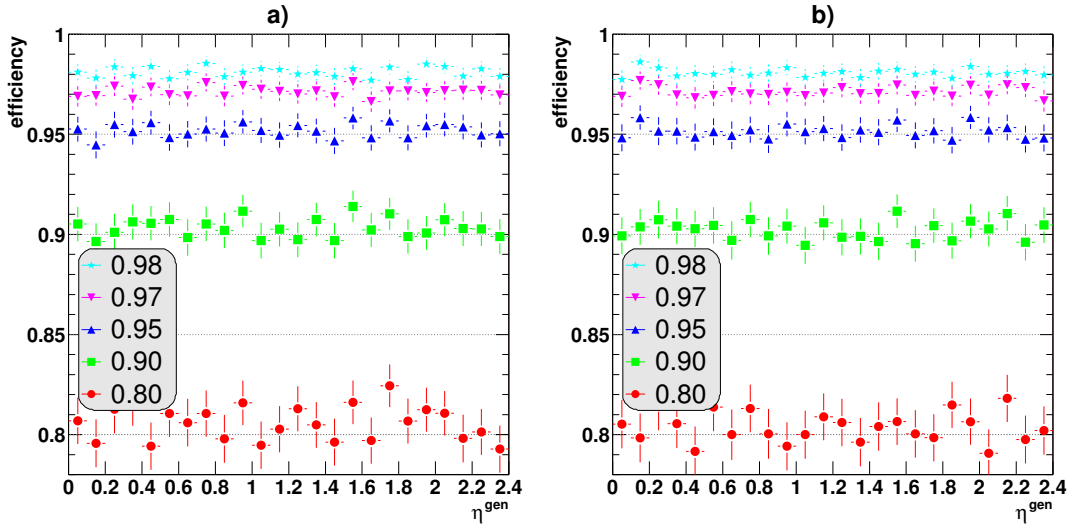


Figure 5.13: *Reference signal efficiency of the calorimeter isolation algorithm as a function of the pseudorapidity of the muon for different nominal efficiency values, at low (a) and high (b) luminosity.*

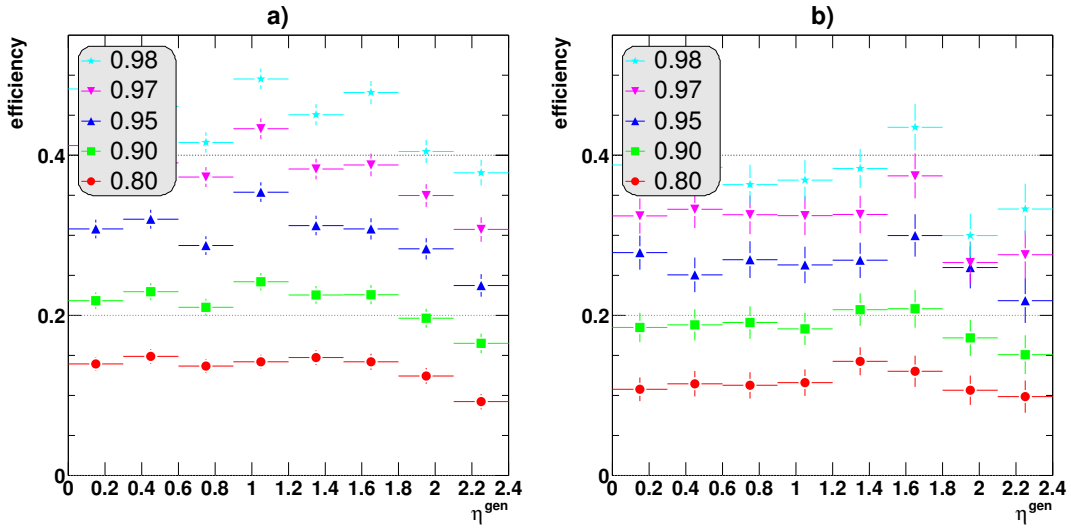


Figure 5.14: *Efficiency of the calorimeter isolation algorithm for the reference background sample as a function of the pseudorapidity of the muon for different nominal efficiency values, at low (a) and high (b) luminosity.*

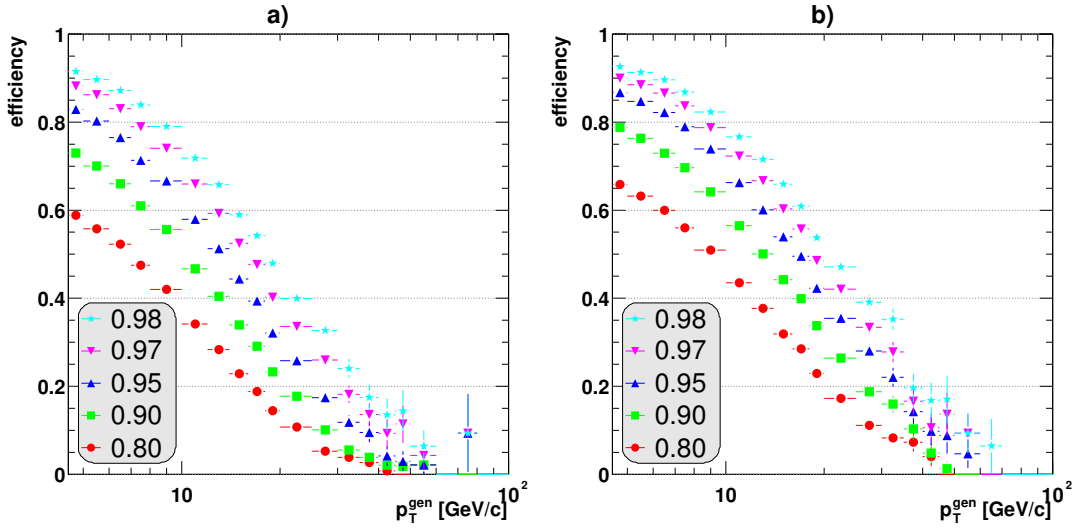


Figure 5.15: *Background efficiency of the calorimeter isolation algorithm for the full sample of minimum bias events as a function of the p_T of the muon and for different nominal efficiency values, at low (a) and high (b) luminosity.*

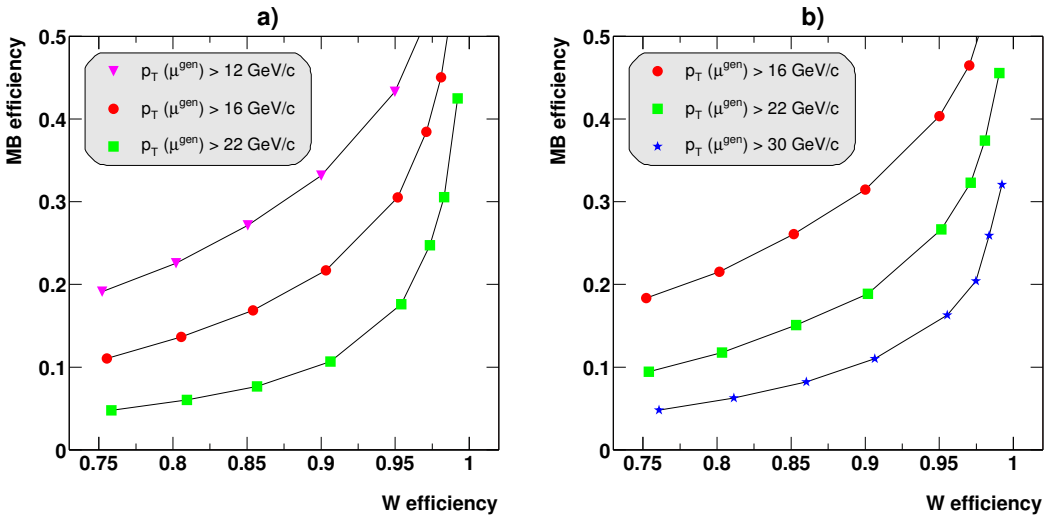


Figure 5.16: *Efficiency of the calorimeter isolation algorithm for minimum bias events as a function of the efficiency for direct $W \rightarrow \mu\nu$ decays at low (a) and high (b) luminosity. The different curves correspond to different cuts on the p_T of the generated muon; the intermediate one corresponds to the definition of the reference signal and background samples.*

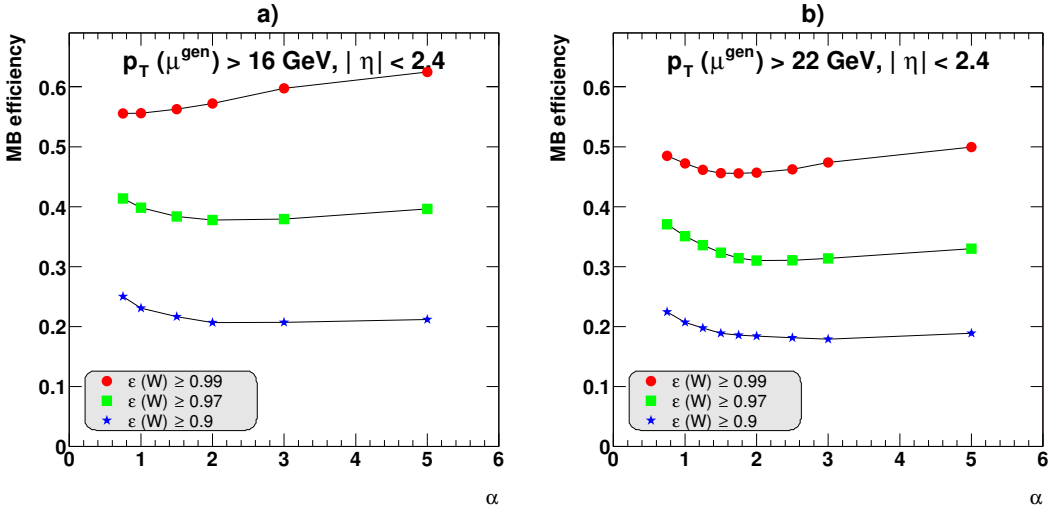


Figure 5.17: *Efficiency of the calorimeter isolation algorithm on the reference background sample as a function of the ECAL-HCAL weighting parameter α at low (a) and high (b) luminosity.*

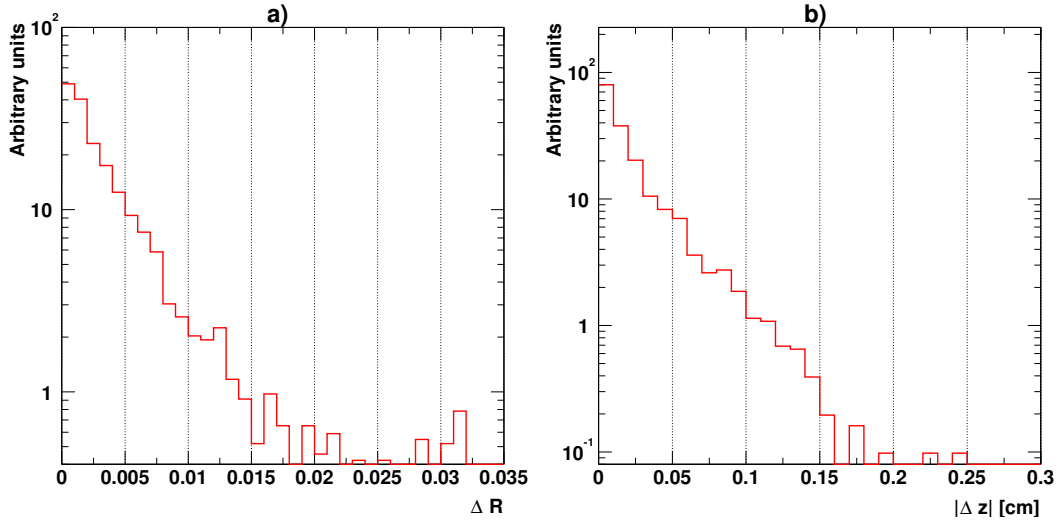


Figure 5.18: *Identification of the "pixel muon" at Level-3, for the reference background at high luminosity. (a) ΔR distance between the direction of the Level-3 muon and the direction of the closest pixel track. (b) Distance along the beam line between the vertex associated to the pixel muon and the Level-3 muon vertex.*

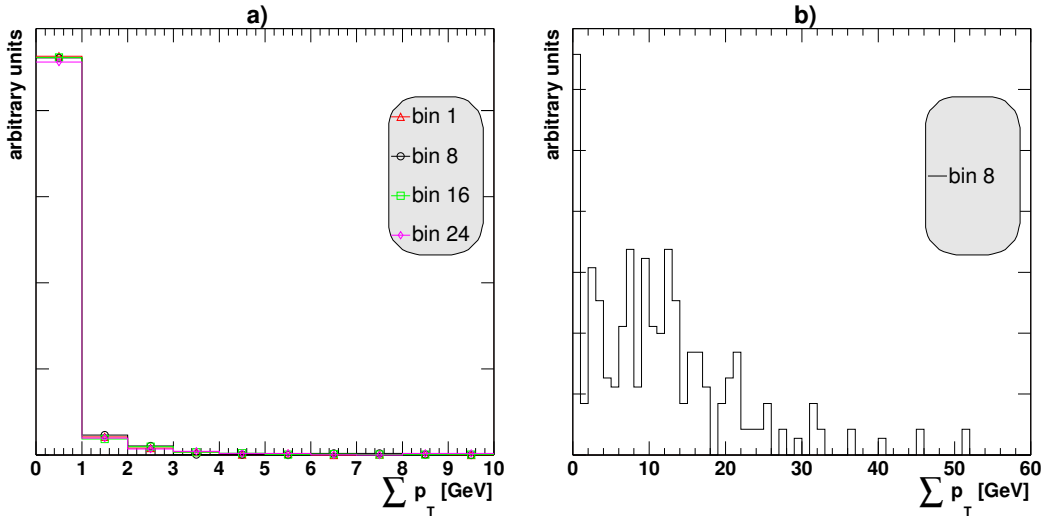


Figure 5.19: *Example of the spectra of $\sum p_T$ of pixel tracks after subtraction of the veto value in cone 6 for the reference signal sample at high luminosity in η-bins 1, 8, 16 and 24 (a). For comparison the distribution of $\sum p_T$ is shown for the reference background sample (cone 6, η-bin 8) (b).*

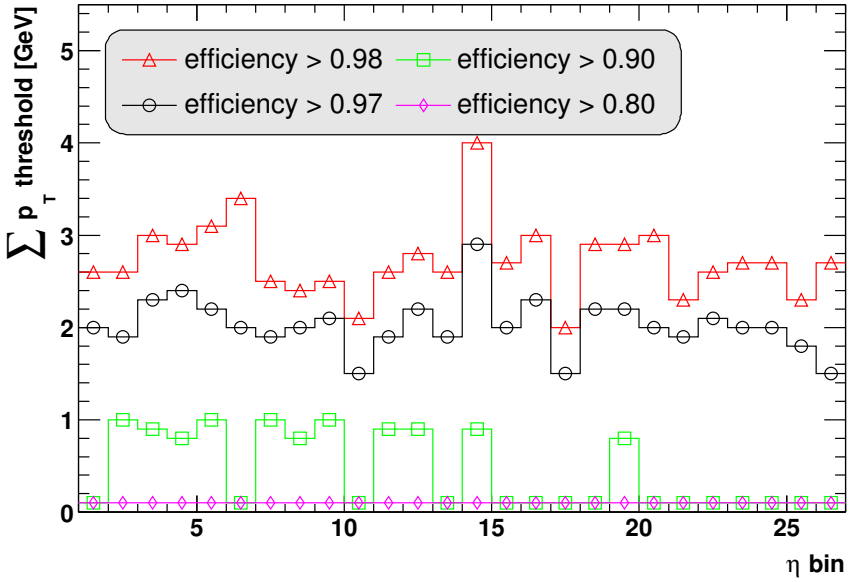


Figure 5.20: *Threshold values of the Level-3 pixel isolation algorithm at low luminosity for cone 6 and nominal efficiencies 0.8, 0.9, 0.97 and 0.98.*

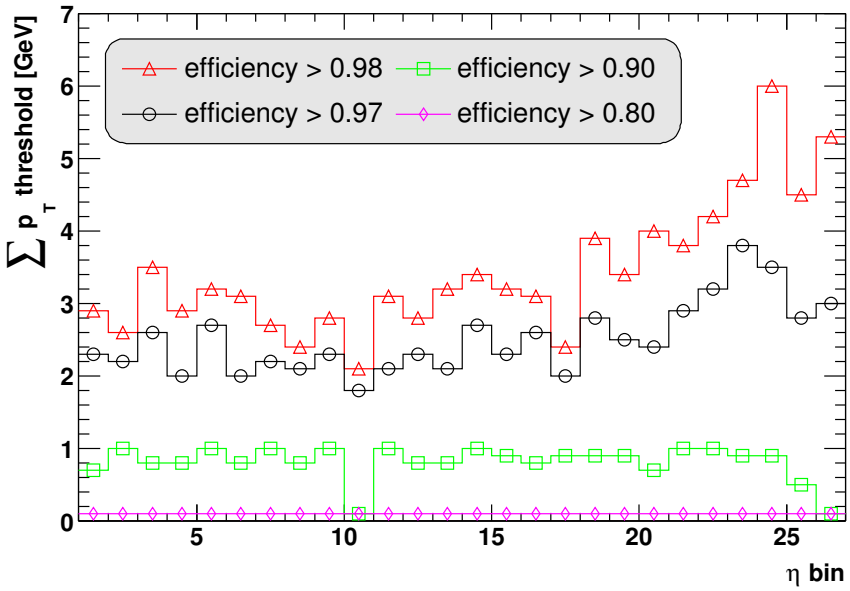


Figure 5.21: *Threshold values of the Level-3 pixel isolation algorithm at high luminosity for cone 6 and nominal efficiencies 0.8, 0.9, 0.97 and 0.98.*

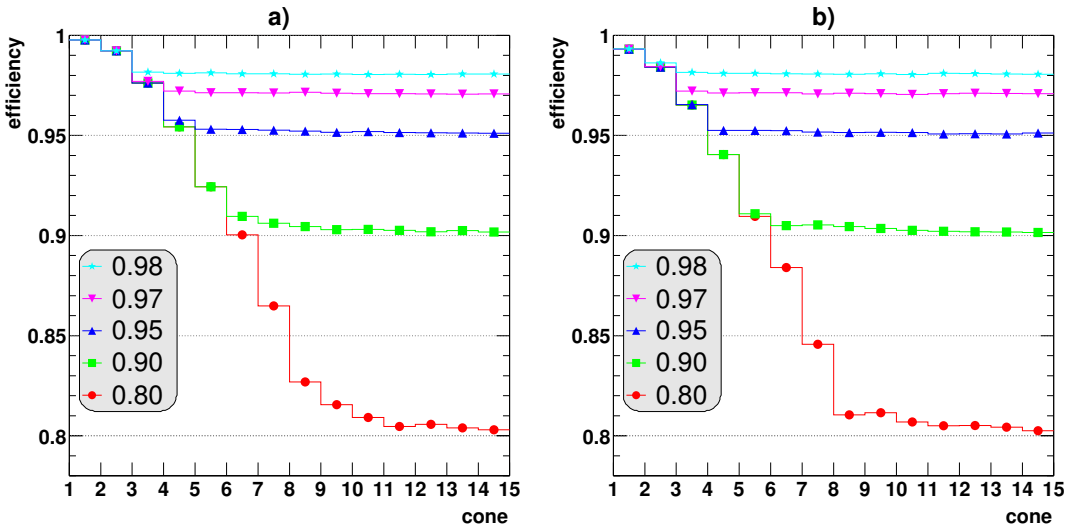


Figure 5.22: *Reference signal efficiency of the Level-3 pixel isolation algorithm for different cone sizes and nominal efficiency values, at low (a) and high (b) luminosity.*

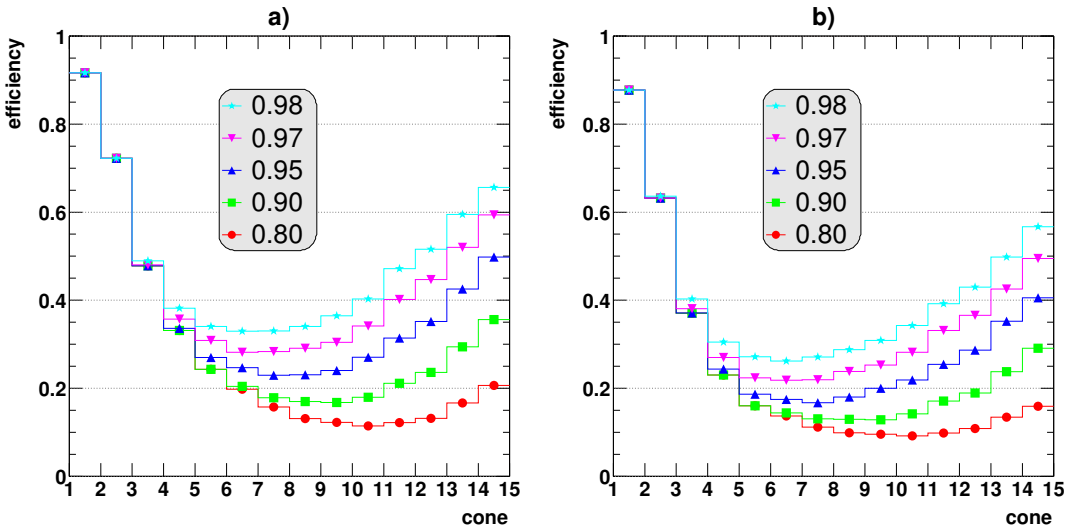


Figure 5.23: *Efficiency of the Level-3 pixel isolation algorithm for the reference background sample for different cone sizes and nominal efficiency values, at low (a) and high (b) luminosity. The minimum of each curve indicates the optimal cone for the corresponding nominal efficiency.*

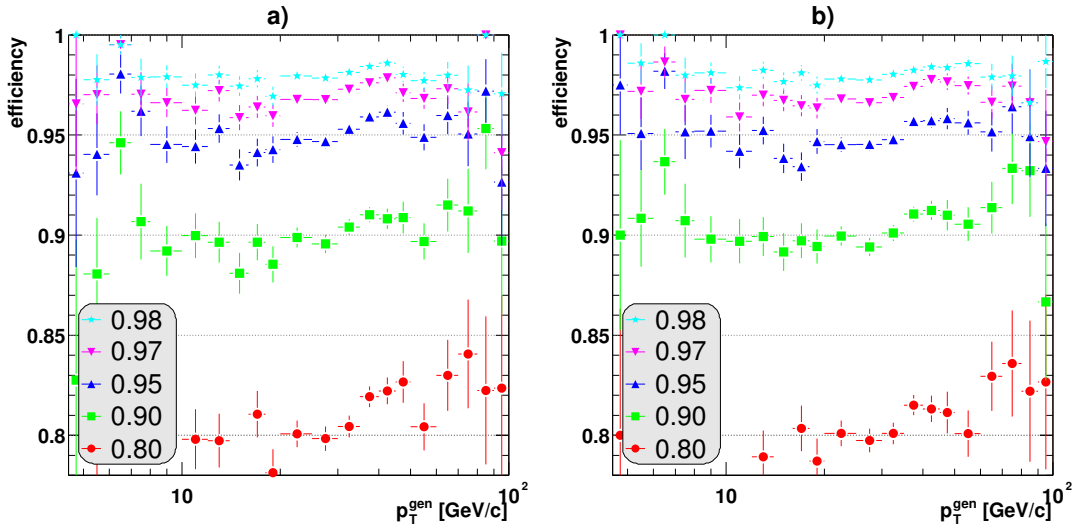


Figure 5.24: *Signal efficiency of the Level-3 pixel isolation algorithm as a function of the p_T of the muon for different nominal efficiency values, at low (a) and high (b) luminosity. The nominal efficiencies are obtained by convoluting these curves with the actual p_T spectra of the reference signal.*

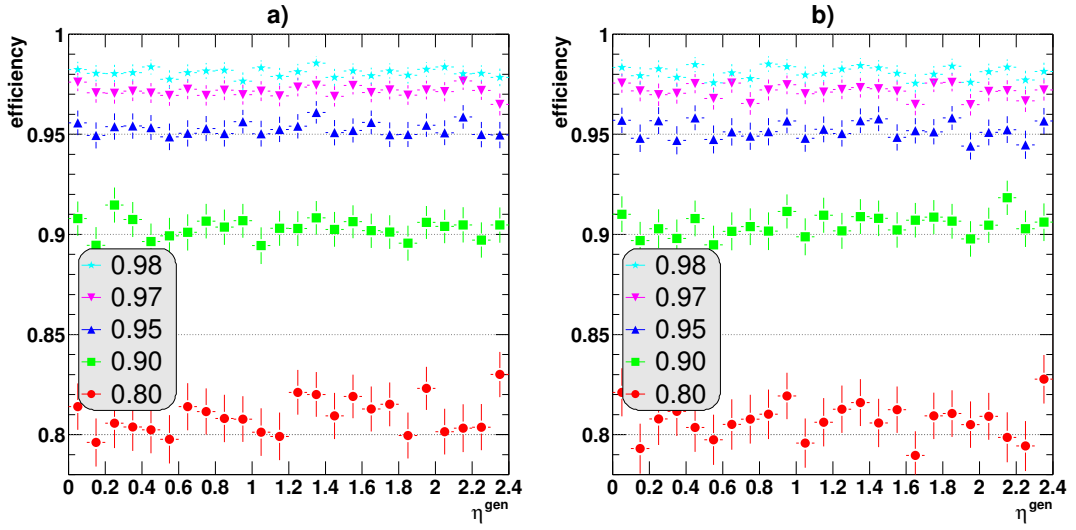


Figure 5.25: *Reference signal efficiency of the Level-3 pixel isolation algorithm as a function of the pseudorapidity of the muon for different nominal efficiency values, at low (a) and high (b) luminosity.*

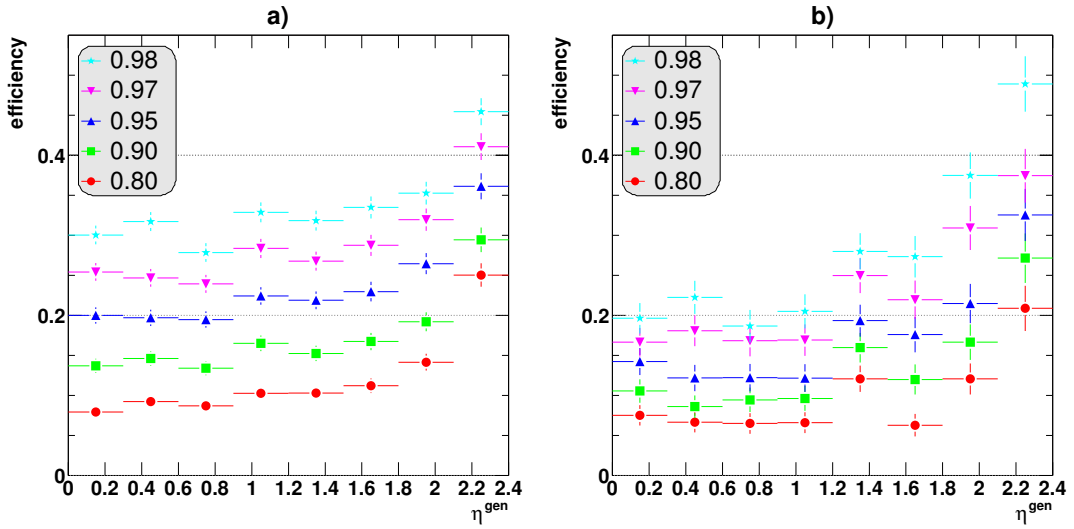


Figure 5.26: *Efficiency of the Level-3 pixel isolation algorithm for the reference background sample as a function of the pseudorapidity of the muon for different nominal efficiency values, at low (a) and high (b) luminosity.*

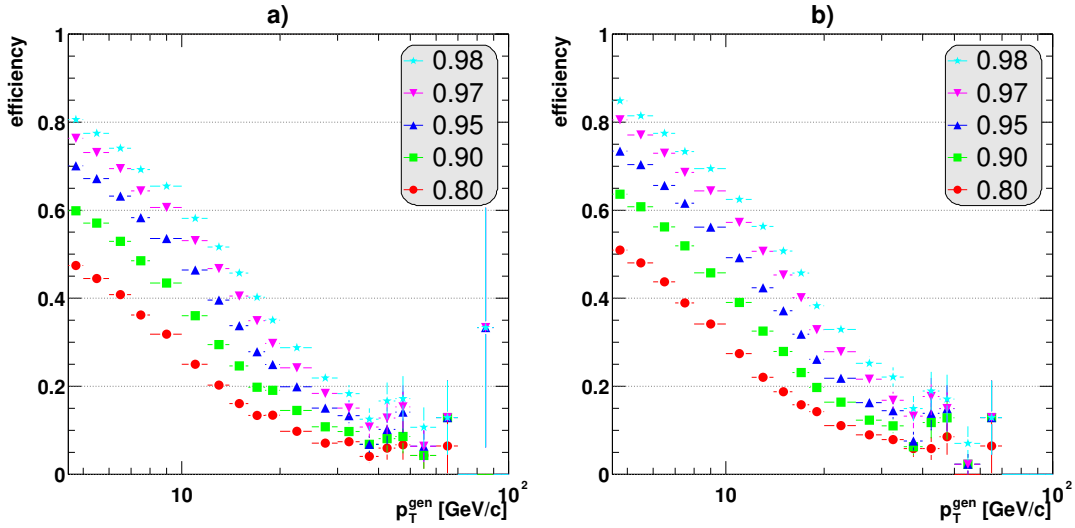


Figure 5.27: *Background efficiency of the Level-3 pixel isolation algorithm for the full sample of minimum bias events as a function of the p_T of the muon and for different nominal efficiency values, at low (a) and high (b) luminosity.*

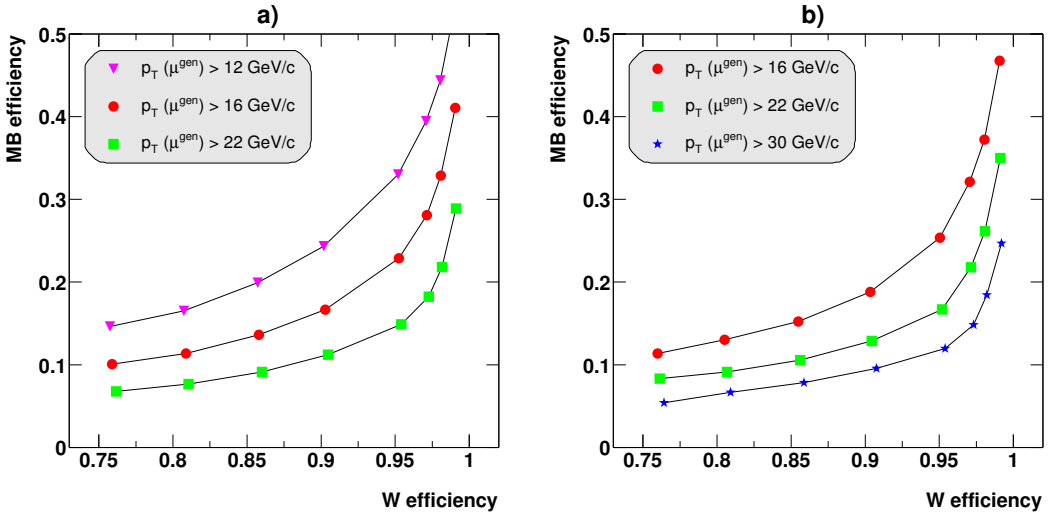


Figure 5.28: *Efficiency of the Level-3 pixel isolation algorithm for minimum bias events as a function of the efficiency for direct $W \rightarrow \mu\nu$ decays at low (a) and high (b) luminosity. The different curves correspond to different cuts on the p_T of the generated muon; the intermediate one corresponds to the definition of the reference signal and background samples.*

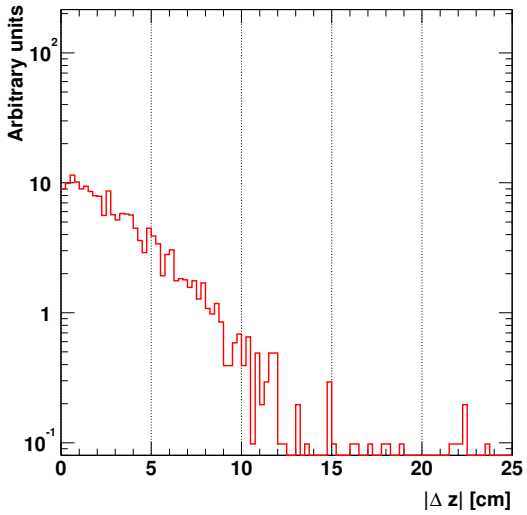


Figure 5.29: *Distance along the beam line between the Monte Carlo primary vertex of the interaction containing the muon and the Level-2 muon vertex (reference background sample at high luminosity).*

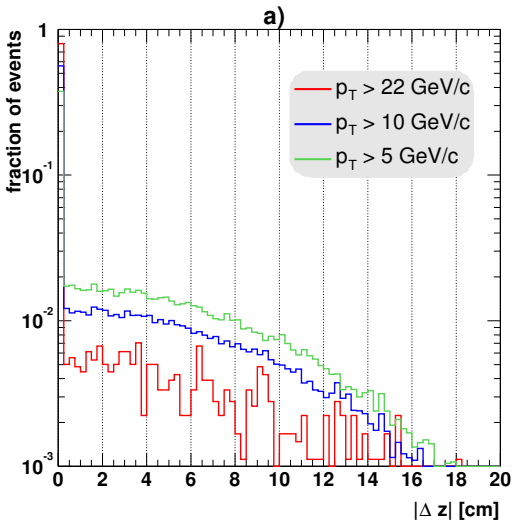


Figure 5.30: *Reconstruction of the primary vertex by the pixel reconstruction algorithm for minimum bias events at high luminosity. (a) Distance along the beam line between the Monte Carlo primary vertex of the interaction containing the muon and the primary vertex reconstructed using the pixel reconstruction algorithm.*

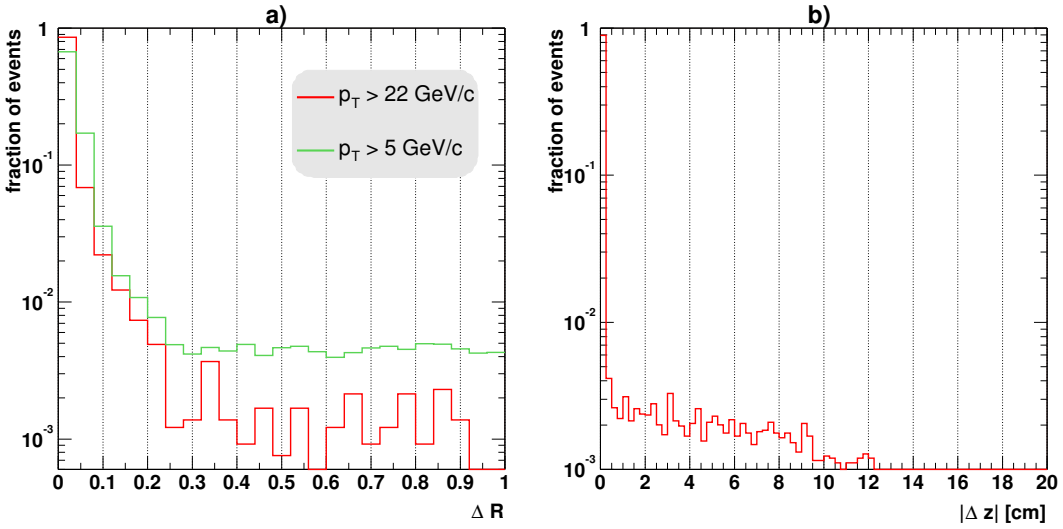


Figure 5.31: *(a) ΔR distance between the direction of the Level-2 muon and the direction of the closest pixel track at high luminosity (minimum bias sample at high luminosity). (b) Distance along the beam line between the pixel track closest in ΔR to the Level-2 muon and the Monte Carlo primary vertex (reference signal at high luminosity).*

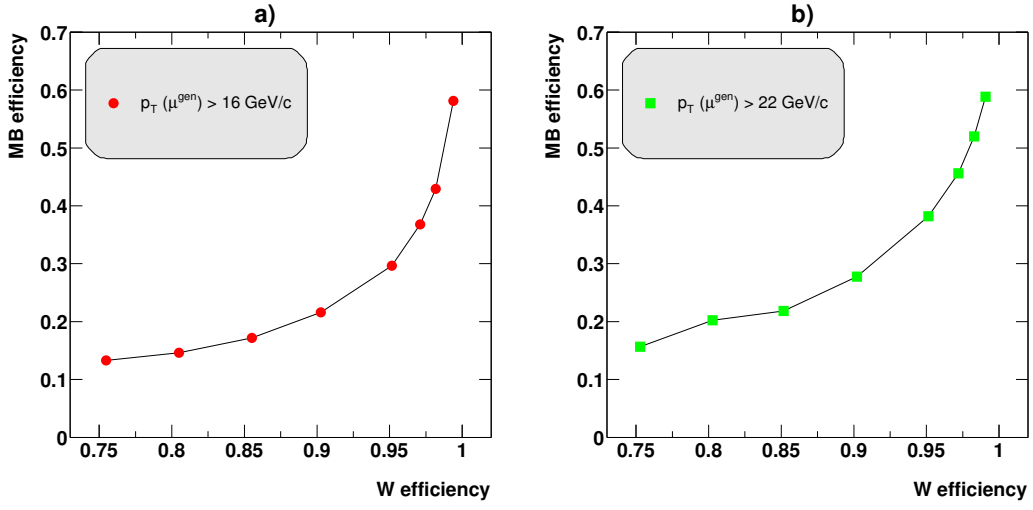


Figure 5.32: *Efficiency of the Level-2 pixel isolation algorithm for minimum bias events as a function of the efficiency on the reference signal at low (a) and high (b) luminosity.*

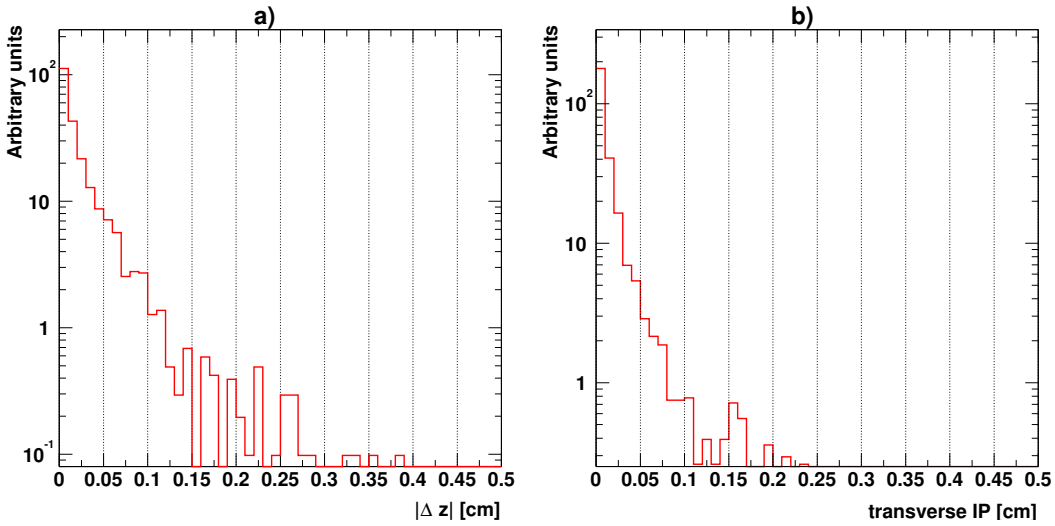


Figure 5.33: *Distance of the reconstructed Level-3 muon from the Monte Carlo primary vertex (reference background at high luminosity): Distance along the beam line (a) and transverse impact parameter (b).*

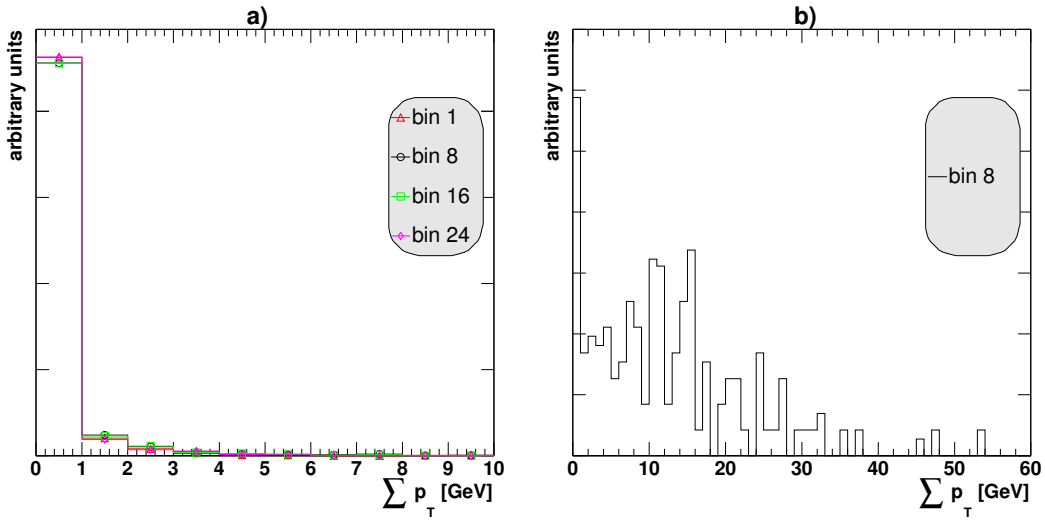


Figure 5.34: *Example of the spectra of $\sum p_T$ of tracker tracks after subtraction of the veto value, in cone 6 for the reference signal sample at high luminosity in η-bins 1, 8, 16 and 24 (a). For comparison the distribution of $\sum p_T$ is shown for the reference background sample (cone 6, η-bin 8) (b).*

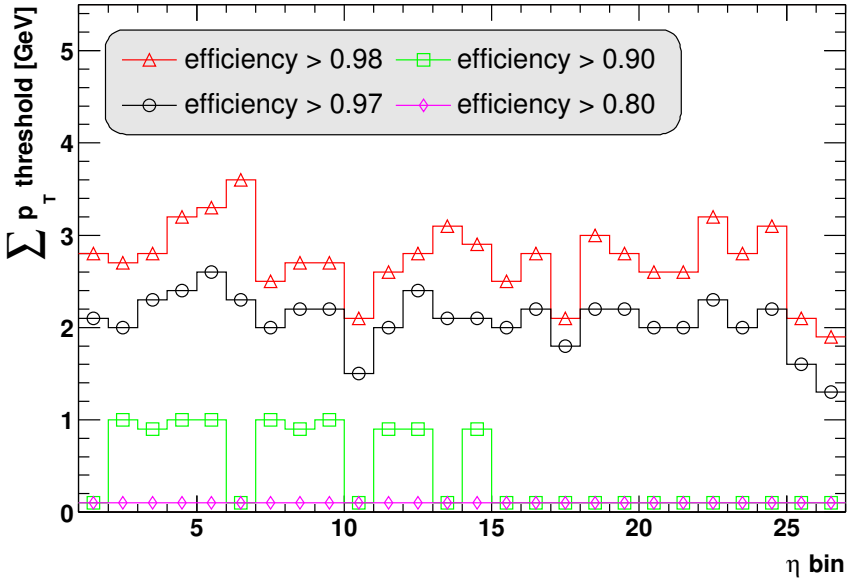


Figure 5.35: *Threshold values of the Level-3 tracker isolation algorithm at low luminosity for cone 6 and nominal efficiencies 0.8, 0.9, 0.97 and 0.98.*

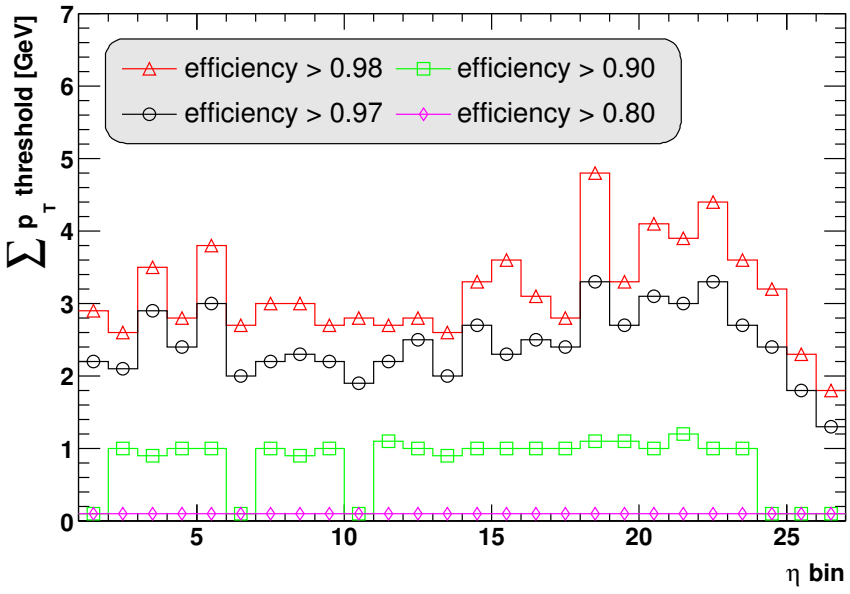


Figure 5.36: *Threshold values of the Level-3 tracker isolation algorithm at high luminosity for cone 6 and nominal efficiencies 0.8, 0.9, 0.97 and 0.98.*

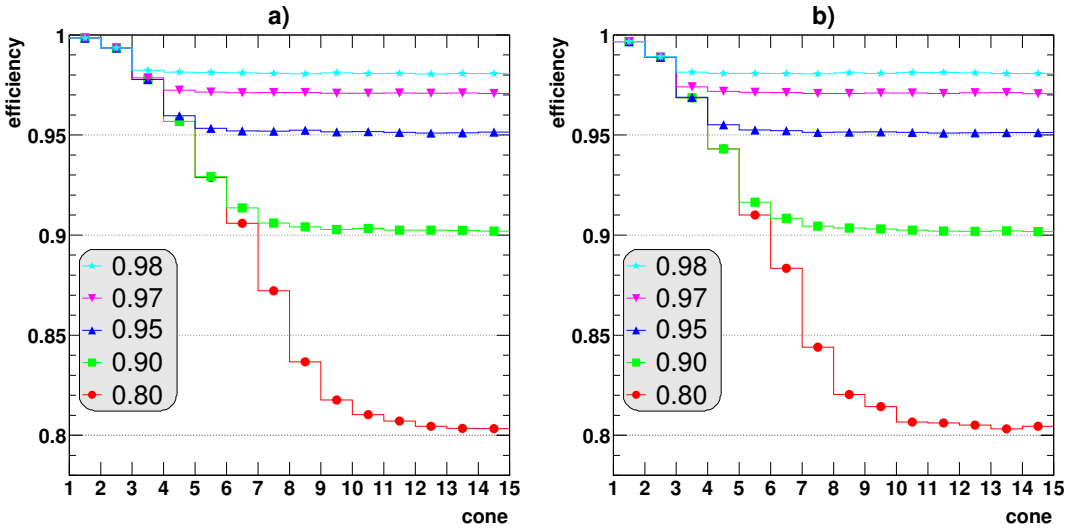


Figure 5.37: *Reference signal efficiency of the Level-3 tracker isolation algorithm for different cone sizes and nominal efficiency values, at low (a) and high (b) luminosity.*

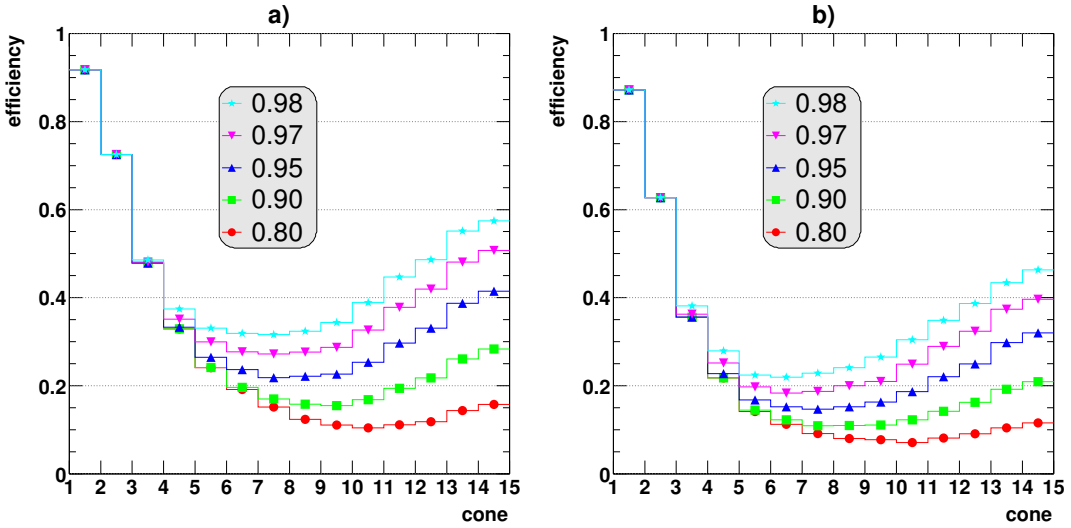


Figure 5.38: *Efficiency of the Level-3 tracker isolation algorithm for the reference background sample for different cone sizes and nominal efficiency values, at low (a) and high (b) luminosity. The minimum of each curve indicates the optimal cone for the corresponding nominal efficiency.*

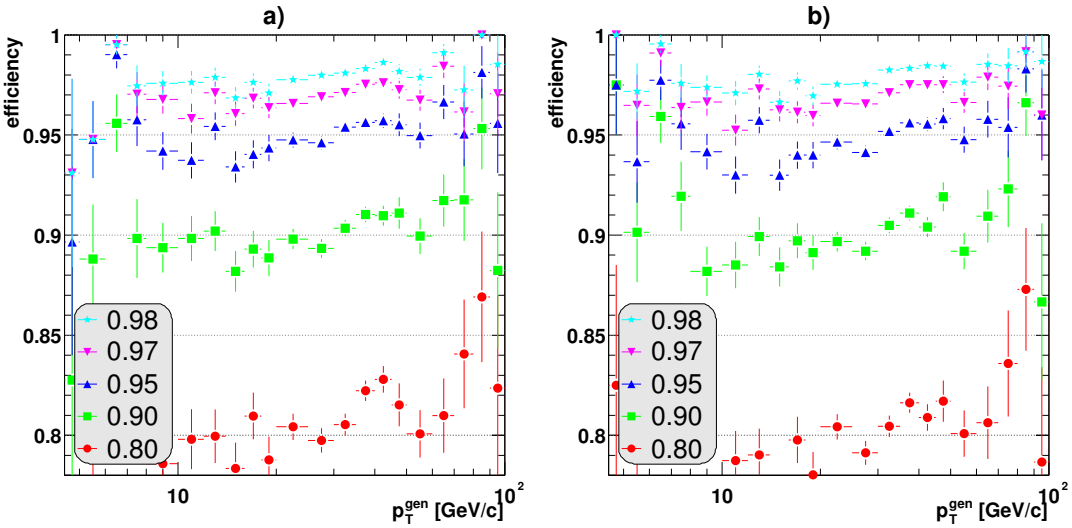


Figure 5.39: *Signal efficiency of the Level-3 tracker isolation algorithm as a function of the p_T of the muon for different nominal efficiency values, at low (a) and high (b) luminosity. The nominal efficiencies are obtained by convoluting these curves with the actual p_T spectra of the reference signal.*

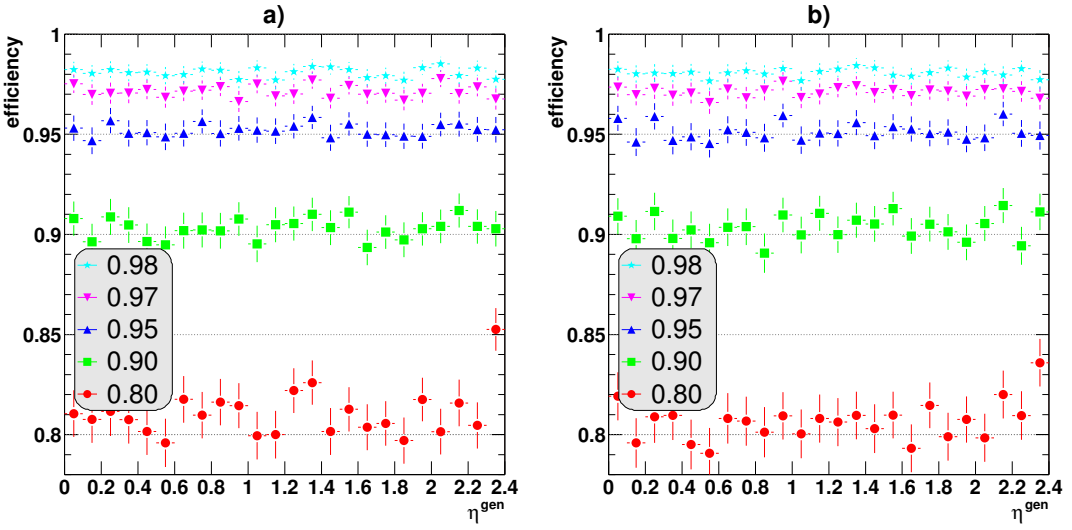


Figure 5.40: *Reference signal efficiency of the Level-3 tracker isolation algorithm as a function of the pseudorapidity of the muon for different nominal efficiency values, at low (a) and high (b) luminosity.*

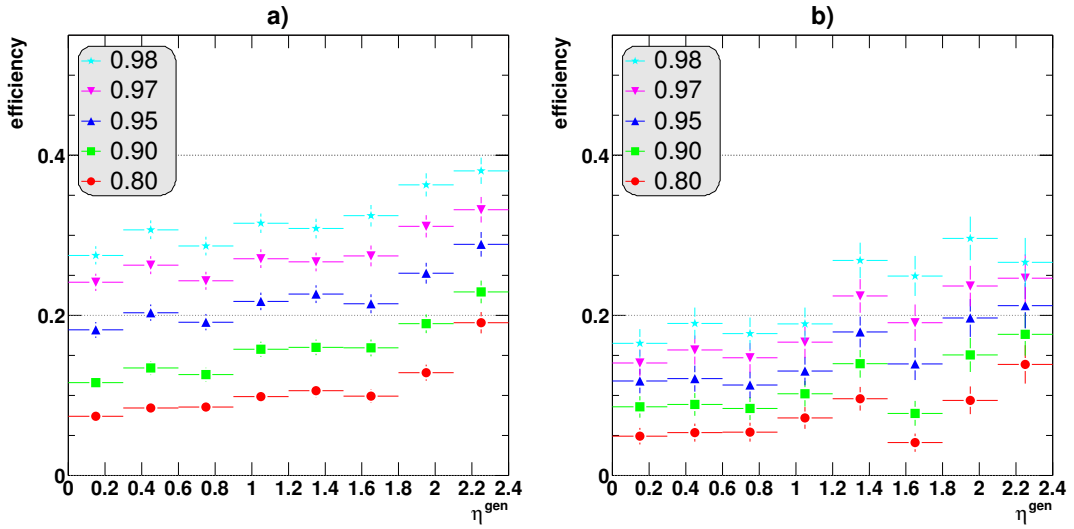


Figure 5.41: *Efficiency of the Level-3 tracker isolation algorithm for the reference background sample as a function of the pseudorapidity of the muon for different nominal efficiency values, at low (a) and high (b) luminosity.*

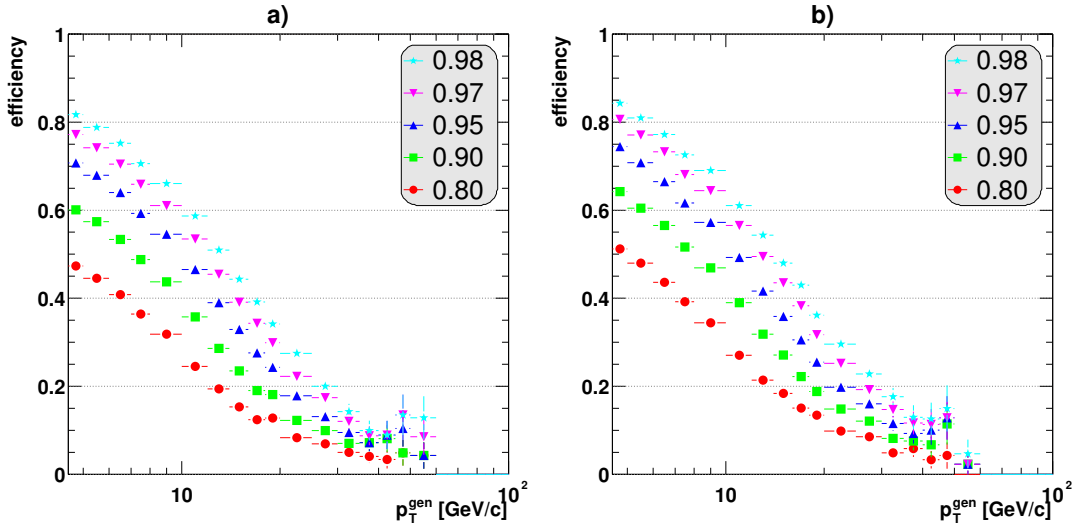


Figure 5.42: *Background efficiency of the Level-3 tracker isolation algorithm for the full sample of minimum bias events as a function of the p_T of the muon and for different nominal efficiency values, at low (a) and high (b) luminosity.*

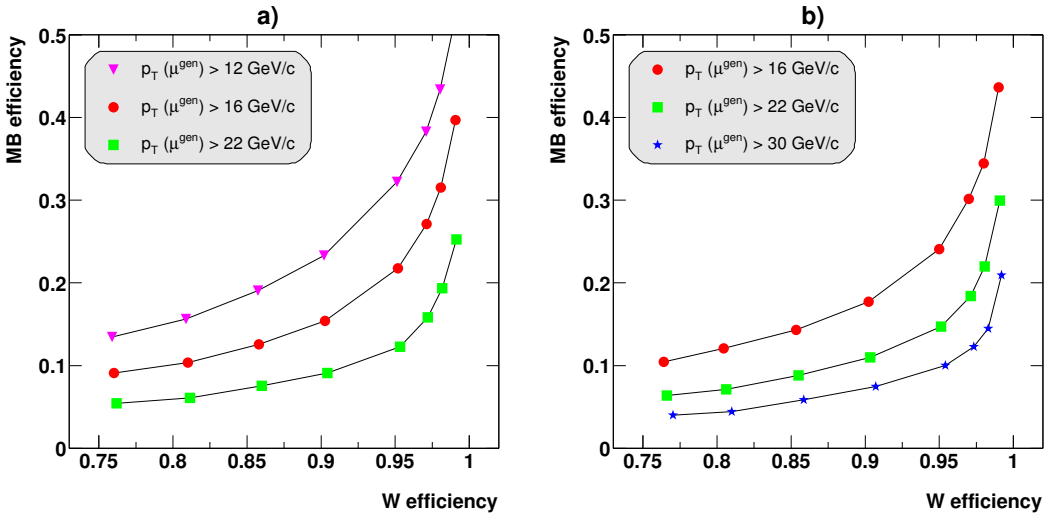


Figure 5.43: *Efficiency of the Level-3 tracker isolation algorithm for minimum bias events as a function of the efficiency for direct $W \rightarrow \mu\nu$ decays at low (a) and high (b) luminosity. The different curves correspond to different cuts on the p_T of the generated muon; the intermediate one corresponds to the definition of the reference signal and background samples.*

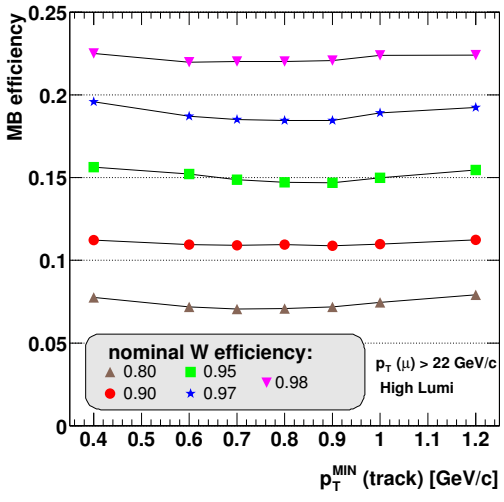


Figure 5.44: *Efficiency of the tracker isolation algorithm for the reference signal as a function of the minimal p_T of the tracks contributing to the cone, for a few different values of nominal efficiency.*

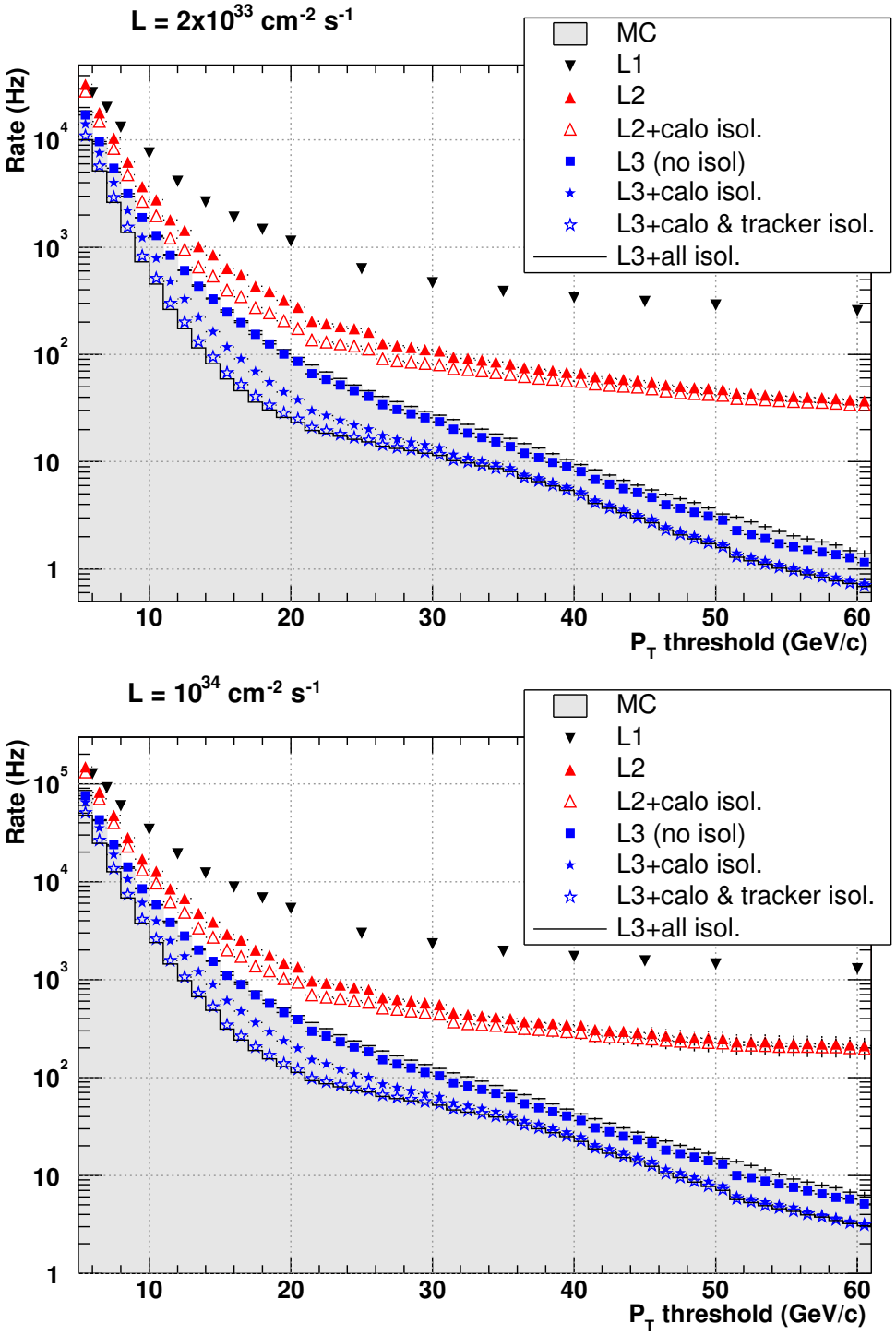


Figure 5.45: *Single-muon rates at low luminosity (top) and high luminosity (bottom) including the isolation selection, as a function of the p_T threshold. The curve labelled "MC" represents the generated rate as a function of the generated p_T of the muons.*

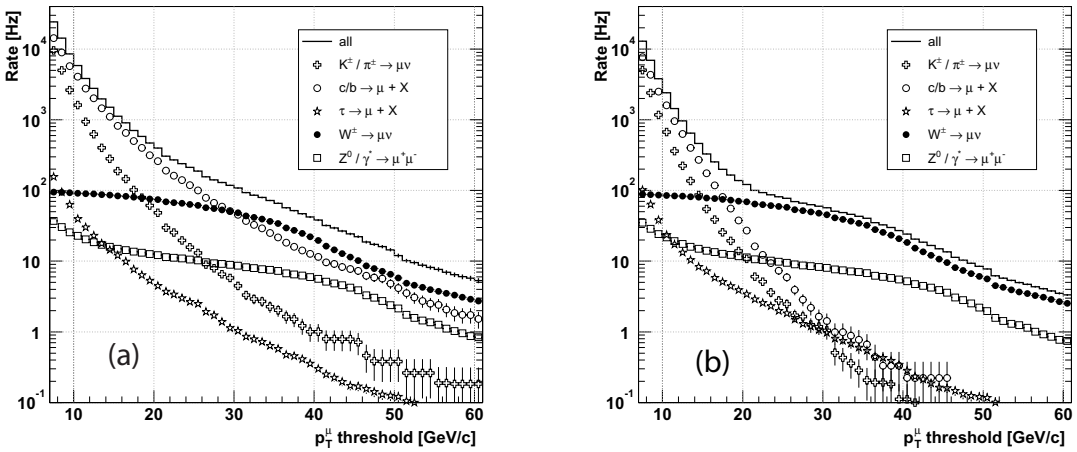


Figure 5.46: *Contributions of the different sources of muons to the Level-3 rate at high luminosity before (a) and after (b) the application of the isolation algorithms [1].*

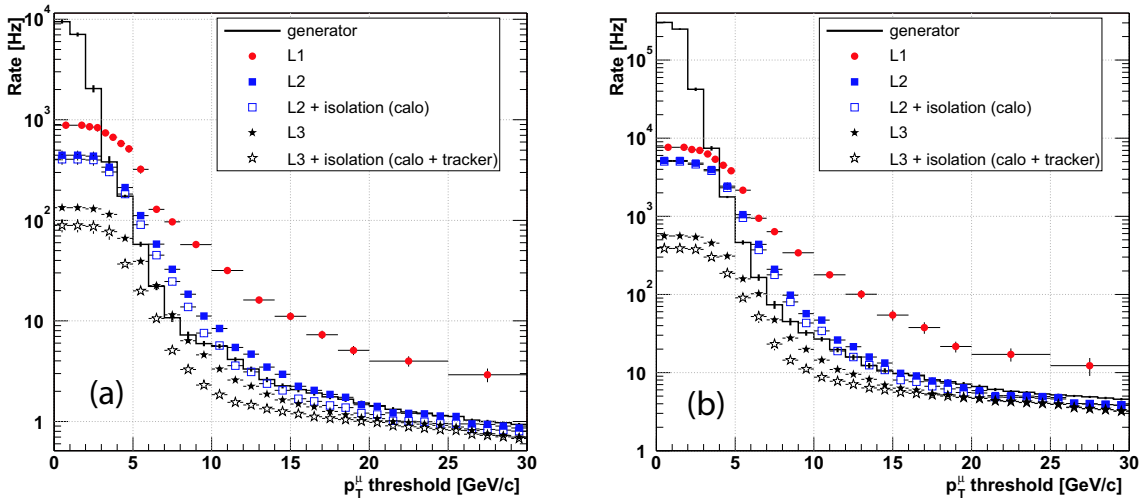


Figure 5.47: *Di-muon trigger rates at low luminosity (a) and high luminosity (b) including the isolation selection, as a function of the symmetric di-muon threshold [1].*

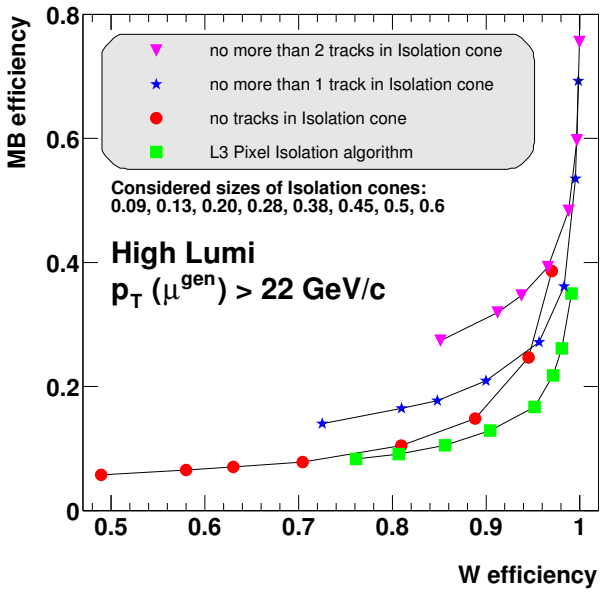


Figure 5.48: *Comparison of the Level-3 pixel isolation algorithm with a simple algorithm based on the requirement of no track, no more than one track, or no more than two tracks in a cone of size ΔR_{MAX} = 0.09–0.6 around the muon.*

Chapter 6

Performance of the Trigger Selection

This chapter presents the performance of the HLT muon trigger algorithms described in the previous chapters. The inclusive muon trigger rates and the choice of p_T thresholds are discussed in Section 6.1. The corresponding efficiency for selecting several benchmark channels is reported in Section 6.2. Finally, the CPU time required to perform this selection is discussed in Section 6.3.

6.1 Bandwidth and Thresholds

The choice of the trigger thresholds is determined by the maximum event rate (bandwidth) that can be accepted at each trigger level. The allocation of the bandwidth at Level-1 has been discussed in Section 2.5.1.4. The basic assumption is that at startup the DAQ system will be able to handle an event rate of up to 50 kHz, which will be increased to 100 kHz when the full LHC design luminosity is reached. Only one third of this bandwidth is allocated, the rest being used as safety margin accounting for all uncertainties in the simulation of the basic physics processes, the CMS detector, and the beam conditions [1]. This bandwidth is then subdivided among different Level-1 objects (muons, electrons and photons, tau jets, jets and combined channels) and for each of these objects between the single- and multiple-object streams. The results are the Level-1 trigger tables shown in Section 2.5.1.4. For the single- and symmetric di-muon streams the thresholds are 14 and 3 GeV/c at low luminosity and 20 and 5 GeV/c at high luminosity, respectively. Except when otherwise stated, these fixed thresholds are always used at Level-1 in the following.

The thresholds for the HLT selection are determined in a similar way, from the request that the final rate to storage be of O(100 Hz). Again, this rate is shared by all channels; about one third will be allocated to the single- and

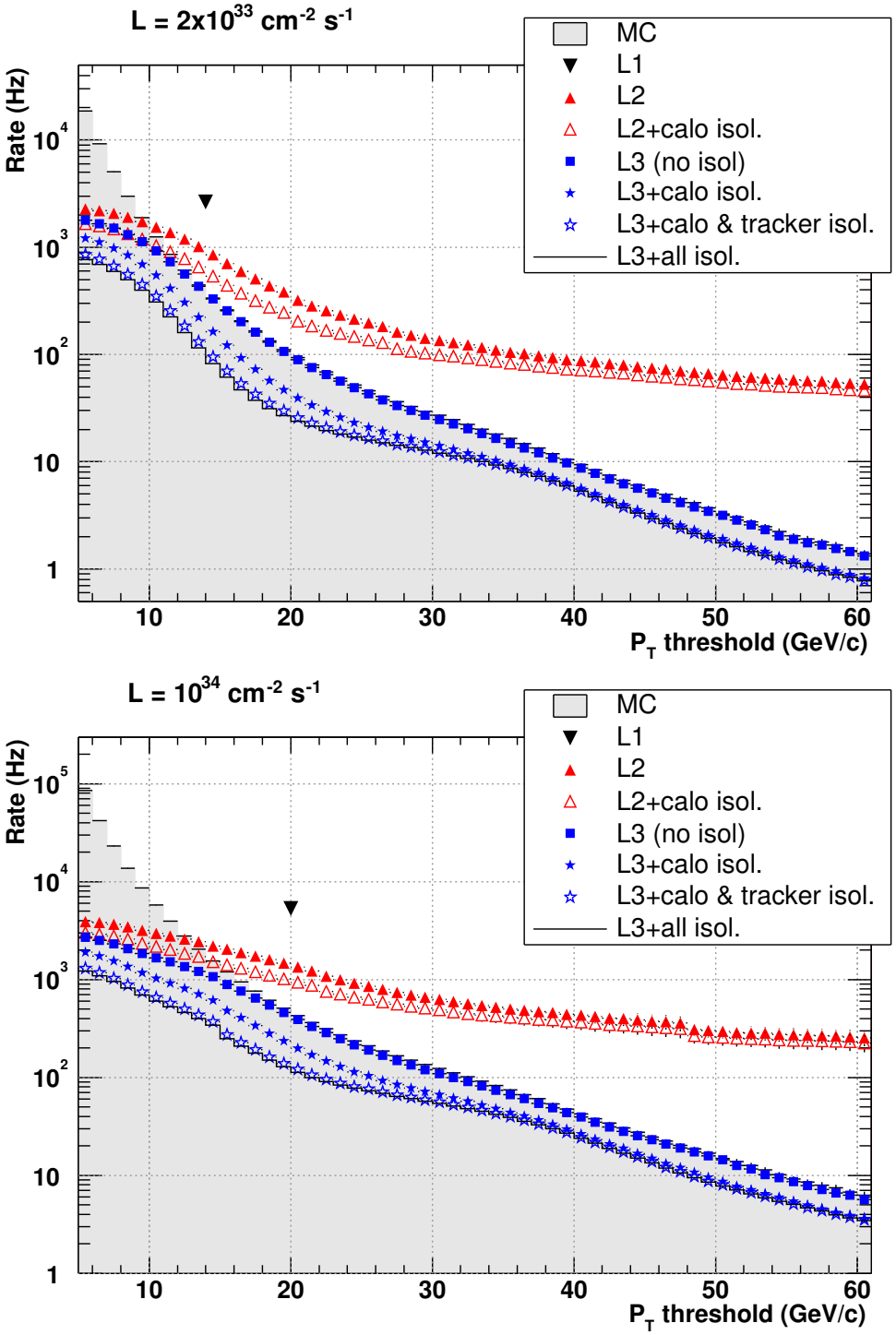


Figure 6.1: *Single muon rates at low luminosity (top) and high luminosity (bottom) for the different HLT steps, as a function of the HLT p_T threshold, assuming a fixed Level-1 single muon trigger threshold of 14 and 20 GeV/c for the two luminosities, respectively. The curve labelled "MC" represents the generated rate as a function of the generated p_T of the muons.*

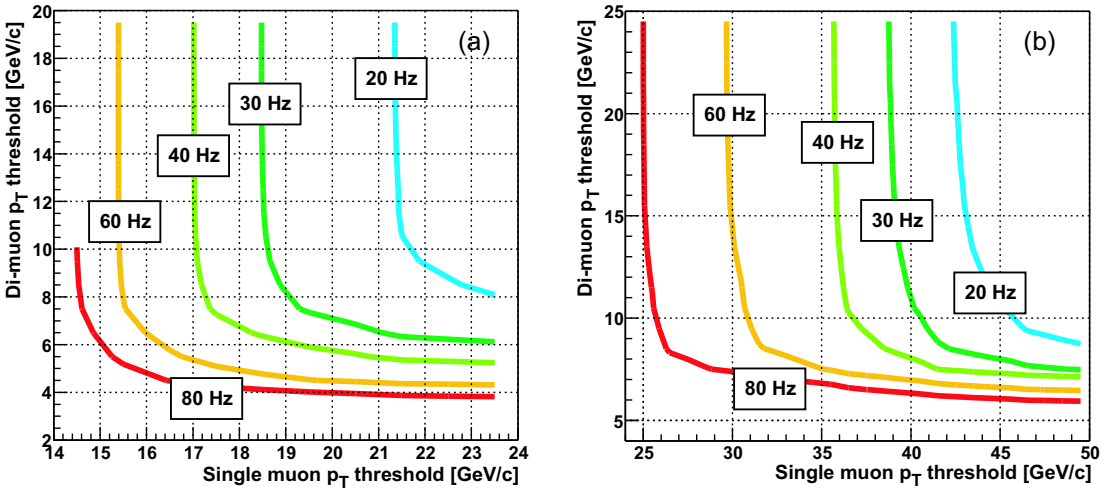


Figure 6.2: *Iso-rate curves for the isolated Level-3 single muon and symmetric di-muon triggers for (a) low luminosity and (b) high luminosity [1].*

di-muon trigger selections. The HLT single-muon trigger rates as a function of the HLT threshold are shown in Fig. 6.1 for the high and low luminosity case. Apart for the use of the fixed Level-1 threshold, these plots are obtained in the same way as those in Fig. 5.45. In particular, the HLT thresholds are defined at the 90% efficiency scale, for consistency with the Level-1 trigger[1]. The isolation algorithms described in Chapter 5 are included, and are set to 97% nominal efficiency. For the selection used in the following, calorimeter isolation is applied at Level-2, and Level-3 rates are given assuming this isolation requirement[2]. The use of pixel and tracker isolation together does not reduce significantly the rate, so that in the following only the tracker isolation algorithm is used at Level-3. This selection, consisting of a p_T threshold at Level-2 and Level-3, plus calorimeter isolation and tracker isolation, will be called "isolated Level-3" in the following.

The allocation of the rate to the single- and di-muon HLT trigger streams is done using the iso-rate plots for the isolated Level-3 trigger, shown in Fig. 6.2. In the case of di-muons, isolation is required for at least one of the two muons. At low luminosity, a rate of about 35 Hz can be obtained with an HLT single muon threshold of 19 GeV/c and a symmetric di-muon threshold of 7 GeV/c. At high luminosity, about the same rate is obtained with a single muon threshold of 38 GeV/c and a symmetric di-muon threshold of 12 GeV/c. However, it is reasonable to assume that when the design luminosity is reached, the HLT rate

[1] Cf. Section 2.5.1.3.

[2] The muon section of the DAQ TDR [1] uses a different convention, and reports Level-3 muon rates for a non-isolated Level-2 selection, corresponding to the curve labelled "L3 (no isol.)" in Fig. 6.1.

to storage will be increased in parallel to the increase in the DAQ bandwidth, though it will still be of O(100 Hz). For this reason, a possible HLT working point at high luminosity is 31 GeV/c for the single muon trigger and 10 GeV/c for the symmetric di-muon trigger, which correspond to a rate of about 60 Hz. These thresholds are used for all results presented in the following, and are summarised in Table 6.1. The corresponding single- and di-muon rates at each trigger level are given in Table 6.2.

Table 6.1: *Thresholds for the single- and symmetric di-muon trigger at low and high luminosity.*

	$\mathcal{L}$ =2 × 10^{33}cm^{-2}s^{-1}		$\mathcal{L}$ =10^{34}cm^{-2}s^{-1}	
Level	Single-muon (GeV/c)	Di-muon (GeV/c)	Single-muon (GeV/c)	Di-muon (GeV/c)
Level-1	14	3	20	5
HLT	19	7	31	10

Table 6.2: *Rates for the single- and di-muon selections obtained with the thresholds of Table 6.1. The rate for each level includes the selection at all preceding levels. Di-muon rates are exclusive of the contribution of the single-muon selection.*

	$\mathcal{L}$ =2 × 10^{33}cm^{-2}s^{-1}		$\mathcal{L}$ =10^{34}cm^{-2}s^{-1}	
Level	Single-muon (Hz)	Di-muon (Hz)	Single-muon (Hz)	Di-muon (Hz)
Level-1	2700	900	6200	1700
Level-2	335	25	700	35
Calorimeter isolation	220	20	590	25
Level-3	45	9	60	10
Tracker isolation	29	6	53	7
Total	35		60	

Of course the available bandwidth is not the only constraint in the trigger selection; it is imperative that the trigger guarantees sufficient efficiency for the interesting physics channels. The performance of the inclusive muon selection in terms of efficiency in selecting a few physics benchmark channels is described in the following section.

In case sufficient efficiency cannot be obtained with this inclusive selection, dedicated streams must be used. Thanks to the software nature of the HLT,

whatever physics selection can be implemented, provided that it does not excessively increase the output rate and the average CPU processing time. Such requirements will also be used to pre-select events in order to facilitate the offline analysis. Examples of dedicated streams are those based on the presence of additional objects in the events (*e.g.* jets or missing energy) or on cuts on the invariant mass or relative charge of multiple objects. In particular, when two muons are reconstructed in one event, it is possible to use the invariant mass of the di-muon system and the relative charge of the two muons. Figure 6.3a shows the inclusive spectrum of isolated Level-3 di-muons, as a function of the invariant mass of the di-muon system multiplied by the opposite of the product of their charges. The contribution of muons coming from the Z/γ^* and $t\bar{t}$ decay chains is shown separately from that of the rest of minimum bias events. The

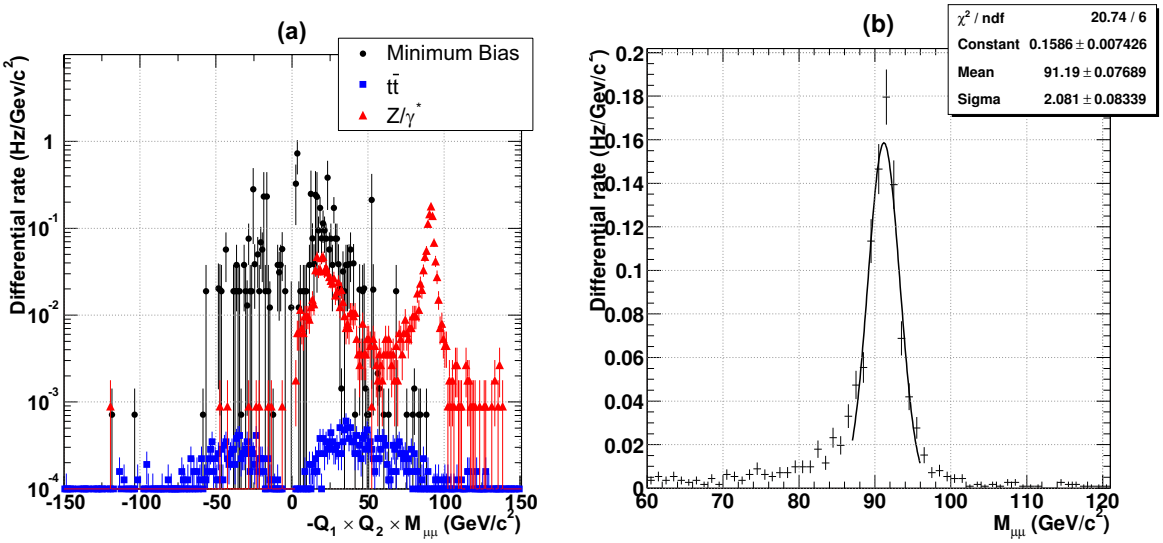


Figure 6.3: *(a) Inclusive differential rate as a function of the invariant mass of di-muons multiplied by the opposite of the product of their charges, after the isolated Level-3 selection. (b) Differential rate as a function of the isolated Level-3 di-muon invariant mass for the Z/γ^* sample, fitted to a Gaussian. Both plots are obtained for the low luminosity case.*

Z resonance peak is reconstructed with good resolution, and allows a clean selection of on-shell Z decays, as shown in Fig. 6.3b. The request that the muons have opposite charge has almost 100% efficiency for isolated Level-3 di-muons from Z/γ^* decays and rejects a good fraction of minimum bias events. It also discards about 36% of the di-muons in the $t\bar{t}$ sample, where the muons can come either from direct W decays or from the fragmentation of jets. Such requirements can be used for dedicated selections, but are not included in the tables and plots of the following sections.

6.2 Efficiency for Benchmark Channels

The efficiency of the inclusive muon trigger selection described in the previous section for several benchmark channels is reported in the following. The samples used and their simulation are described in Chapter 3.

In all cases, efficiencies are normalised to the fraction of events that can be triggered and used in the offline analysis. In particular, at least one muon is required to be within $|\eta| < 2.1$, which is the geometric coverage of the muon trigger (cf. Section 2.5.1.2). In case additional muons are needed to understand the properties of the event, they are required to be within the geometric acceptance of the muon system ($|\eta| < 2.4$), so that they can be reconstructed in the off-line analysis.

The resulting efficiency is a measure of the performance of the trigger; however, from the point of view of the final analysis the event yield is a more useful parameter. The yield for each channel analysed is reported for integrated luminosities of 20 fb^{-1} for the low luminosity case and 100 fb^{-1} for high luminosity, corresponding in both cases to one year of data taking.

6.2.1 W, Z and $t\bar{t}$ Decays

The production of W and Z bosons and of top quark pairs constitutes a background for the muon trigger, since it is a source of high-p_T muons. However, these channels are also an important benchmark for the efficiency of the trigger. For example, the most promising Higgs decay channels for Higgs masses above 130 GeV/c^2 are $H \to ZZ$ and $H \to WW$, with at least one of the two bosons decaying leptonically. It is clear that good efficiency for the selection of muons from W and Z decays is very important.

The production of top quark pairs is a difficult background for many studies like $H \to WW$, since the decay of the two top quarks produces in almost 100% of the cases two W bosons with opposite charge. Good understanding of this process will thus be required.

Moreover, these three processes will also be fundamental for studies of electroweak and QCD physics; deviations of the measured cross sections from the theoretical predictions could provide hints for new physics. As discussed in Section 1.3.2, W and $t\bar{t}$ production will be used for the measurement of m_W and m_t respectively. The latter will also be used for the calibration of the jet energy scale, and leptonic Z decays will be used to calibrate the lepton energy and momentum scale.

The efficiency for selecting $W \to \mu\nu$ decays at each trigger level, relative to the fraction of events where the muon is within the η coverage of the muon trigger (about 50%), is reported in Table 6.3. The selection criteria are those described in the previous section. The efficiency as a function of the Level-1

Table 6.3: *Efficiency for the selection of $W \to \mu\nu$ decays at low and high luminosity, relative to the fraction of events where the muon coming from the W decay is within $|\eta| < 2.1$. The efficiency reported for each level includes the selection at all preceding levels.*

Level	$\mathcal{L} = 2 \times 10^{33}\text{cm}^{-2}\text{s}^{-1}$	$\mathcal{L} = 10^{34}\text{cm}^{-2}\text{s}^{-1}$
Level-1	0.90	0.82
Level-2	0.80	0.58
Calorimeter isol.	0.77	0.56
Level-3	0.72	0.44
Tracker isolation	0.70	0.43

and HLT thresholds is shown in Fig. 6.4 for the low luminosity case. The curve labelled "MC" in these plots and in all following ones represents the efficiency of the selection as a function of the threshold on the real (generated) p_T of the muons, with no other selection criteria. It can be noted that for high thresholds the Level-1 efficiency is above this curve. This is due to the overestimate of the momentum of low-p_T muons. The Level-3 efficiency is below the generator-level efficiency and is very close to it for high thresholds.

The selection of $Z \to \mu\mu$ decays benefits also from the di-muon trigger. The efficiency for the two streams, with the selection criteria described in the previous section, is reported in Table 6.4. The efficiency is normalised to the number of on-shell Z bosons ($85\ \text{GeV}/c^2 < m_Z < 97\ \text{GeV}/c^2$) with at least one decay muon within the η coverage of the muon trigger, and both within the coverage of the muon detector. The efficiency as a function of the Level-1 and HLT thresholds is shown in Fig. 6.5 and Fig. 6.6 for the low luminosity case.

The efficiency for the selection of $t\bar{t}$ events is shown in Table 6.5 relative to the fraction of events with one muon coming from a $W \to \mu\nu$ direct decay within the η coverage of the muon trigger. Other muons can be present as well, *e.g.* from the direct decay of the other W boson or from the decay chains of b or c quarks. The efficiency as a function of the Level-1 and HLT thresholds is shown in Fig. 6.5 and Fig. 6.6 for the low luminosity case.

The total efficiencies of the isolated Level-3 single- and di-muon selection and the corresponding annual yield for W, Z and $t\bar{t}$ decays are shown in Table 6.6.

6.2.2 Selection of Benchmark Signals

Three benchmark channels have been chosen to demonstrate the performance of the HLT on Higgs discovery channels. As discussed in Section 1.3.1, the channel $H \to ZZ \to 4\mu$ is the "gold-plated" channel for the discovery of the stan-

Table 6.4: *Efficiency for the selection of $Z \rightarrow \mu\mu$ decays ($85 < m_Z < 97\ GeV/c^2$) at low and high luminosity , relative to the fraction of events where one muon is within $|\eta| < 2.1$ and the second is within $|\eta| < 2.4$. The efficiency reported for each level includes the selection at all preceding levels.*

	$\mathcal{L} = 2 \times 10^{33}\text{cm}^{-2}\text{s}^{-1}$		$\mathcal{L} = 10^{34}\text{cm}^{-2}\text{s}^{-1}$	
Level	Single-muon	Di-muon	Single-muon	Di-muon
Level-1	0.99	0.81	0.97	0.79
Level-2	0.98	0.79	0.92	0.76
Calorimeter isol.	0.95	0.79	0.89	0.76
Level-3	0.94	0.77	0.85	0.73
Tracker isolation	0.92	0.77	0.83	0.73
Total	0.97		0.92	

Table 6.5: *Efficiency for the selection of $t\bar{t} \rightarrow W_{\rightarrow \mu\nu} + X$ decays at low and high luminosity, relative to the fraction of events where the muon is within $|\eta| < 2.1$. The efficiency for each level includes the selection at all preceding levels.*

	$\mathcal{L} = 2 \times 10^{33}\text{cm}^{-2}\text{s}^{-1}$		$\mathcal{L} = 10^{34}\text{cm}^{-2}\text{s}^{-1}$	
Level	Single-muon	Di-muon	Single-muon	Di-muon
Level-1	0.94	0.31	0.89	0.28
Level-2	0.87	0.23	0.72	0.18
Calorimeter isol.	0.78	0.22	0.65	0.17
Level-3	0.74	0.19	0.58	0.14
Tracker isolation	0.72	0.18	0.55	0.14
Total	0.75		0.60	

Table 6.6: *Total efficiency of the isolated Level-3 single- and di-muon selection and corresponding annual yield for W, Z and $t\bar{t}$ decays at low and high luminosity. Efficiencies are relative to the selected acceptance, as described in the text.*

	$\mathcal{L} = 2 \times 10^{33}\text{cm}^{-2}\text{s}^{-1}$		$\mathcal{L} = 10^{34}\text{cm}^{-2}\text{s}^{-1}$	
Level	Efficiency	Ev/20 fb^{-1}	Efficiency	Ev/100 fb^{-1}
$W \rightarrow \mu\nu$	0.70	1.4×10^8	0.43	4.0×10^8
$Z \rightarrow \mu\mu$	0.97	1.1×10^7	0.92	5.0×10^7
$t\bar{t} \rightarrow W_{\rightarrow \mu\nu} + X$	0.75	1.6×10^6	0.60	6.5×10^6

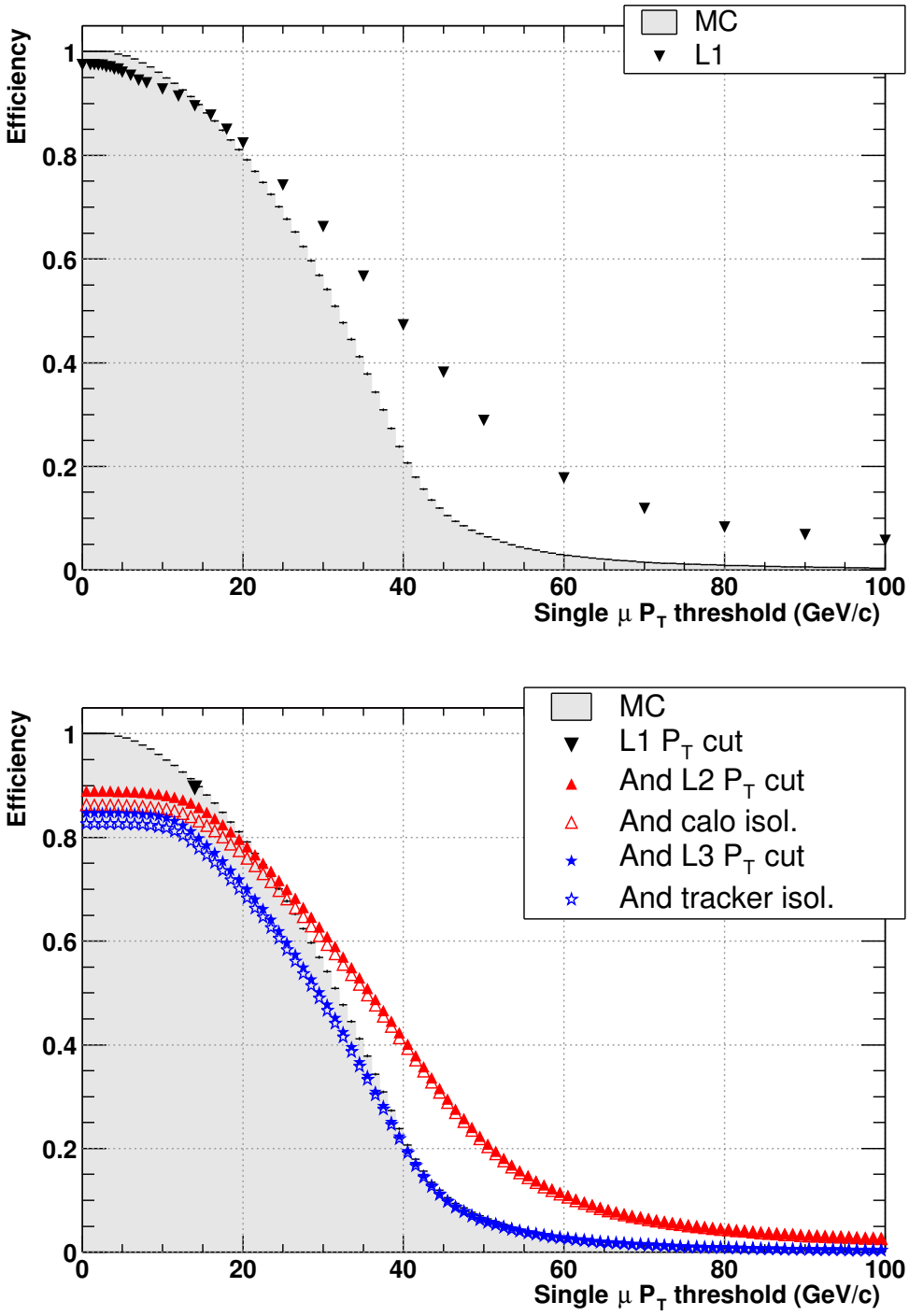


Figure 6.4: *Efficiency for selecting $W \to \mu\nu$ decays at low luminosity as a function of the Level-1 single-muon threshold (top) and of the HLT single-muon threshold for a fixed Level-1 threshold (bottom). The curve labelled "MC" represents the efficiency as a function of the threshold on the generated p_T.*

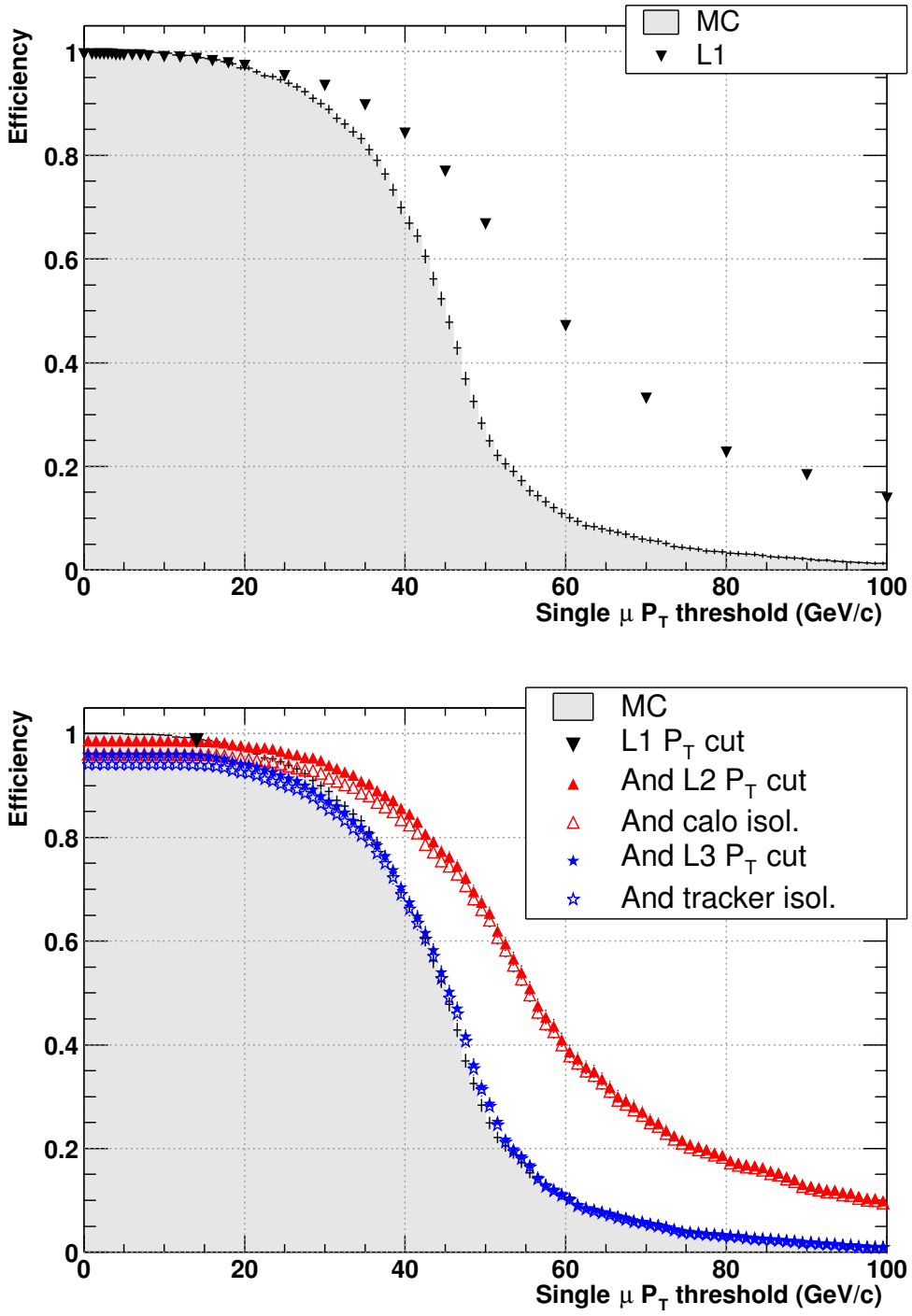


Figure 6.5: *Efficiency for selecting $Z \rightarrow \mu\mu$ ($85 < m_Z < 97$ GeV/c^2) decays at low luminosity as a function of the Level-1 single-muon threshold (top), and of the HLT single-muon threshold (bottom) for a fixed Level-1 threshold. The curve labelled "MC" represents the efficiency as a function of the threshold on the generated p_T.*

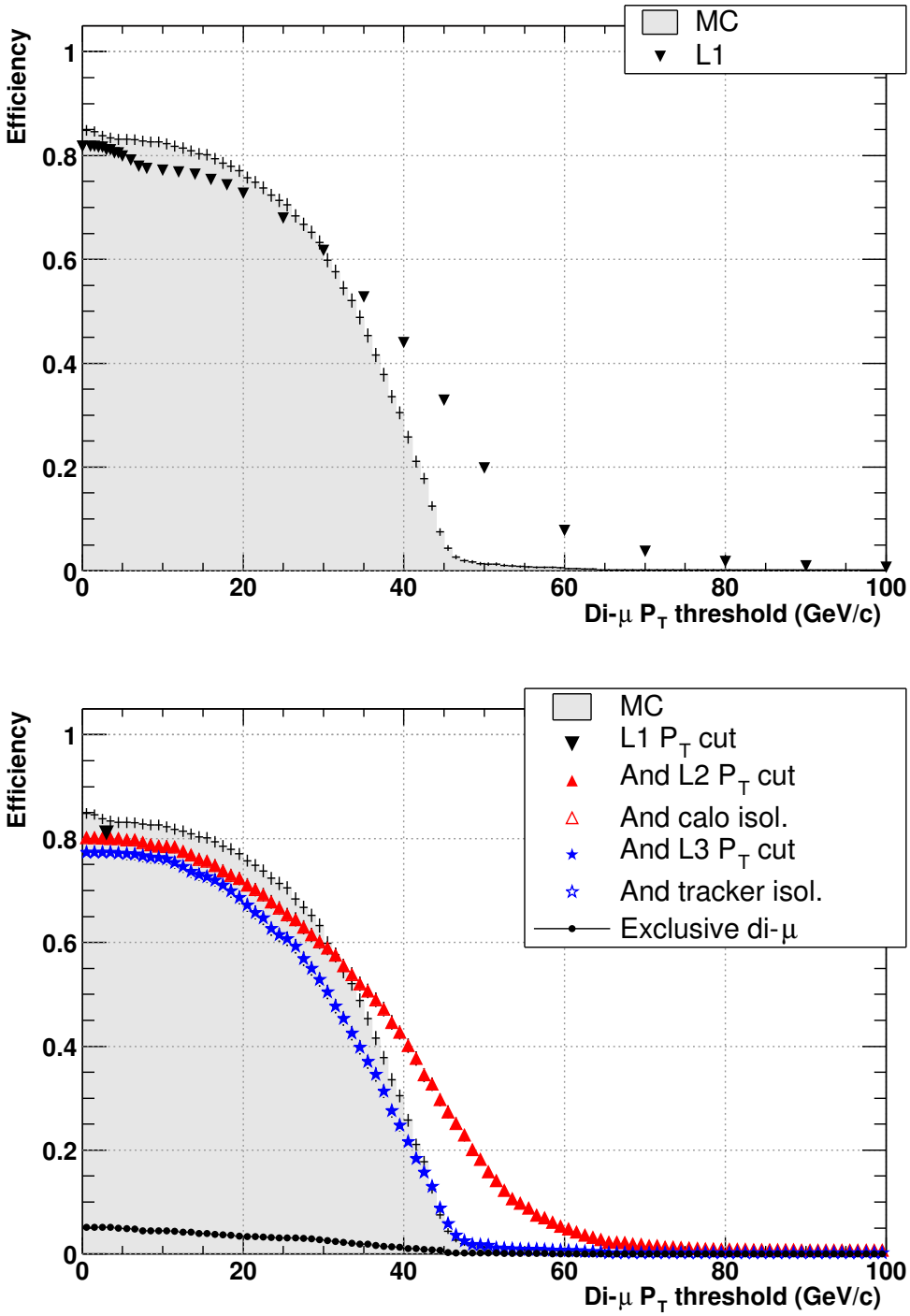


Figure 6.6: *Efficiency for selecting $Z \rightarrow \mu\mu$ decays ($85 < m_Z < 97$ GeV/c^2) at low luminosity as a function of the Level-1 symmetric di-muon threshold (top), and of the HLT symmetric di-muon threshold (bottom) for a fixed Level-1 threshold. The exclusive di-muon contribution refers to events not selected by the single muon trigger after Level-3 and isolation. The curve labelled "MC" represents the efficiency as a function of the threshold on the generated p_T.*

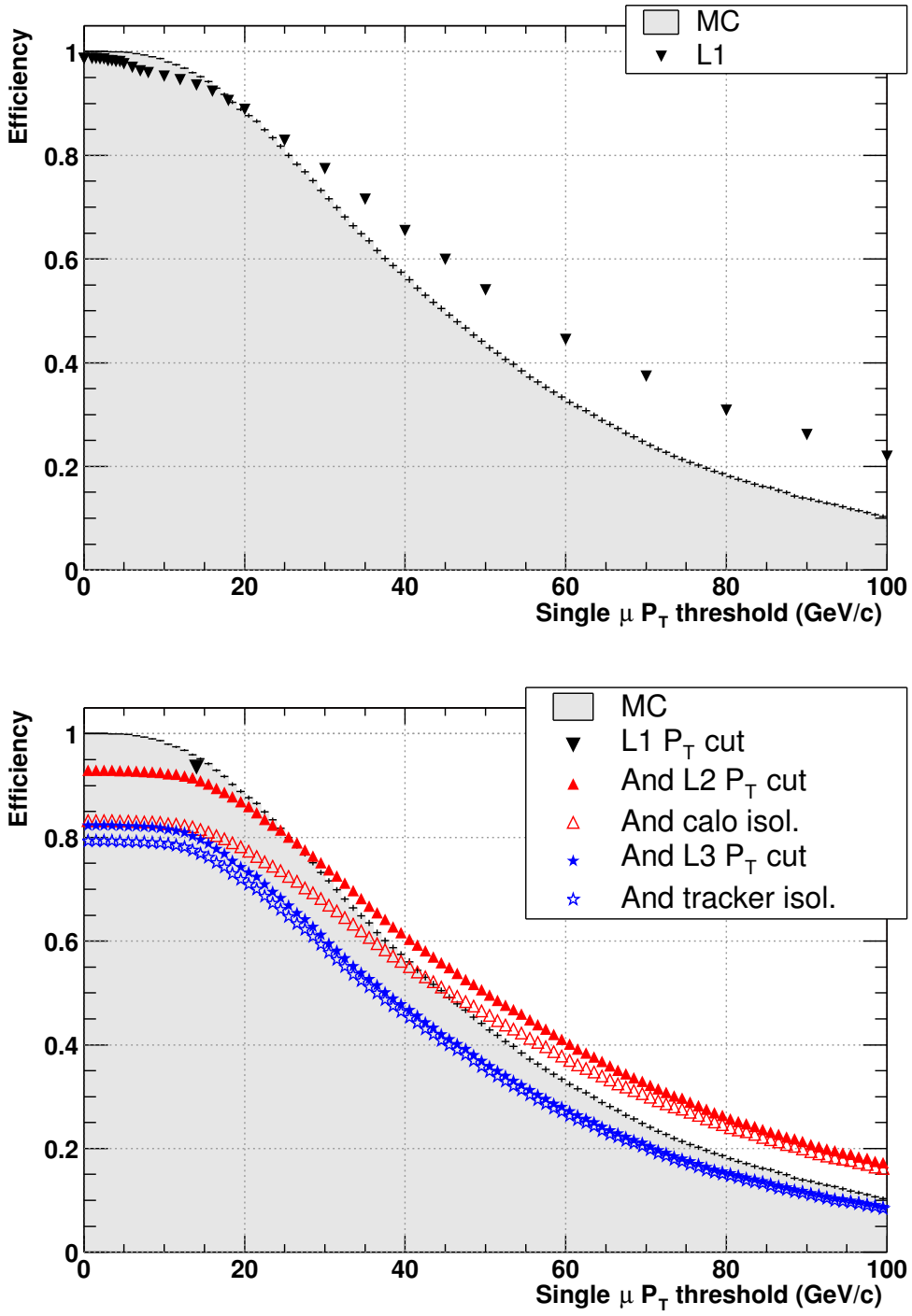


Figure 6.7: *Efficiency for selecting $t\bar{t}\rightarrow W_{\rightarrow\mu\nu} + X$ decays at low luminosity as a function of (a) the Level-1 single-muon threshold, and (b) the HLT single-muon threshold for a fixed Level-1 threshold. The curve labelled "MC" represents the efficiency as a function of the threshold on the generated p_T.*

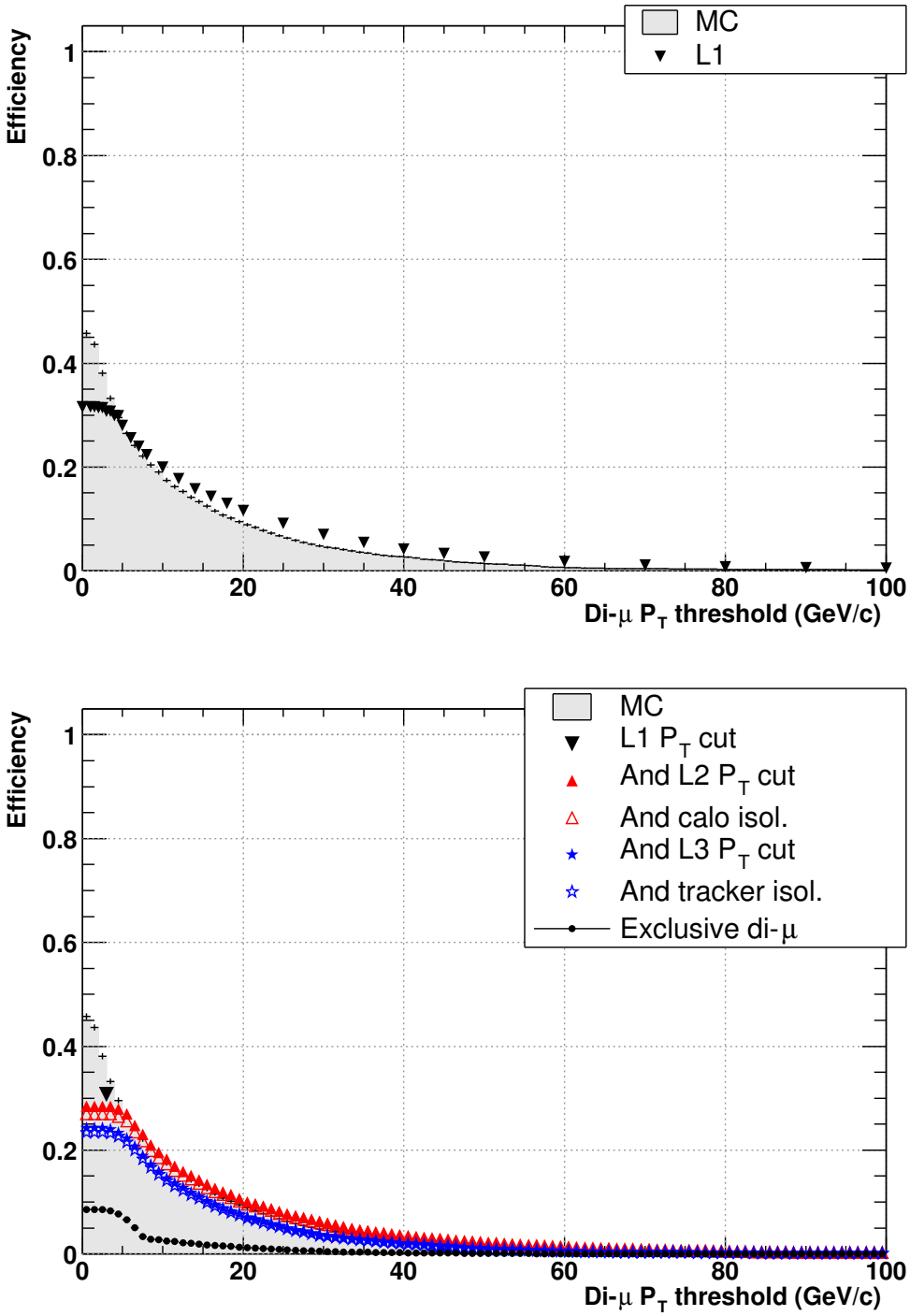


Figure 6.8: *Efficiency for selecting $t\bar{t}\rightarrow W_{\rightarrow\mu\nu} + X$ decays at low luminosity as a function of (a) the Level-1 symmetric di-muon threshold, and (b) the HLT symmetric di-muon threshold for a fixed Level-1 threshold. The exclusive di-muon contribution refers to events not selected by the single muon trigger after Level-3 and isolation. The curve labelled "MC" represents the efficiency as a function of the threshold on the generated p_T.*

dard model Higgs in a wide range of masses. A more difficult channel will be $H \to WW \to 2\mu$. In the field of MSSM Higgs search, the channel $H/A \to \tau\tau_{\to \mu\bar{\nu}\nu}$ was also chosen. The total efficiency of the isolated Level-3 single- and di-muon selection and the corresponding annual yield for these channels is shown in Table 6.7, for different values of the Higgs mass.

6.2.2.1 $H \to ZZ \to 4\mu$

The acceptance for this channel is defined as one muon within the η coverage of the trigger and all four within the coverage of the muon detector, and is in the range 55-75% for the values of m_H considered. The efficiency of the HLT single-muon and di-muon selections described in Section 6.1 is shown in Table 6.7 for five different Higgs masses. Total efficiencies above 98% are obtained in all cases. The effect of varying the HLT thresholds is shown in Fig. 6.9 for the case of $m_H = 130$ GeV/c^2. The isolated Level-3 efficiency for different values of m_H is shown in Fig. 6.11. Both figures refer to the low-luminosity case.

For Higgs masses above 180 GeV/c^2, the two Z bosons and real, and an additional cut on the invariant masses of opposite-charge di-muon systems can be used. However, only a fraction of the events will have all four muons reconstructed in the trigger, mainly due to the limited trigger geometrical acceptance.

6.2.2.2 $H \to WW \to 2\mu 2\nu$

The acceptance for this channel is defined as one muon within the η coverage of the muon trigger and both within the coverage of the muon detector, and is in the range 70-75% for the values of m_H considered. The efficiency of the HLT single-muon and di-muon selections described in Section 6.1 is shown in Table 6.7 for four different Higgs masses. The effect of varying the HLT thresholds is shown in Fig. 6.10 for the case of $m_H = 160$ GeV/c^2. The isolated Level-3 efficiency for different values of m_H is shown in Fig. 6.12. Both figures refer to the low-luminosity case.

The additional requirement that the two Level-3 muons have opposite charge has an efficiency of almost 100% for all considered Higgs masses. Additional background rejection can be obtained with a cut on the angle between the muon momenta in the transverse plane, since the muons tend to be close in direction due to the correlation of the spin of the W bosons [53].

6.2.2.3 $H/A \to \tau\tau \to \mu + X$

In the case of the heavy MSSM Higgs bosons H/A, the decays to two vector bosons, which are the golden modes in the case of the SM Higgs with moderate or large mass, are strongly suppressed. Observation will be possible in the $\tau\tau$ final states. A sample of $H/A \to \tau\tau_{\to \mu\bar{\nu}\nu}$ has been studied for the case of

$\tan\beta = 20$ and $m_A = 200$ GeV/c^2. The acceptance for this channel is defined as one muon (coming from direct τ decay) within the η coverage of the trigger, and is about 80%. The efficiency and yield of the inclusive selection are shown in Table 6.7. The di-muon trigger selects the events where also the second tau decays to a muon; however, the identification will be easier for final states where an electron or a jet are produced instead. Special $\mu - e$ and $\mu - \tau$ jet triggers can be used for this purpose [1], with thresholds lower than those on the corresponding single objects. This will allow to recover part of the large inefficiency of the inclusive selection, which is due to the rather soft spectrum of muons coming from τ decays. Even the inclusive selection, however, allows a significant event yield, thanks to the relatively large cross-section.

Table 6.7: *Total efficiency of the isolated Level-3 single- and di-muon selection and corresponding yield for the Higgs benchmark channels at low and high luminosity. Efficiencies are relative to the selected acceptance, as described in the text.*

	$\mathcal{L} = 2 \times 10^{33}\text{cm}^{-2}\text{s}^{-1}$		$\mathcal{L} = 10^{34}\text{cm}^{-2}\text{s}^{-1}$	
Level	Efficiency	Ev/20 fb^{-1}	Efficiency	Ev/100 fb^{-1}
$H(130)\to ZZ^*\to 4\mu$	0.985	10	0.981	50
$H(150)\to ZZ^*\to 4\mu$	0.987	20	0.987	99
$H(200)\to ZZ\to 4\mu$	0.989	43	0.989	217
$H(300)\to ZZ\to 4\mu$	0.998	29	0.994	147
$H(500)\to ZZ\to 4\mu$	0.998	13	0.995	67
$H(120)\to WW\to 2\mu 2\nu$	0.885	455	0.742	2.6×10^3
$H(140)\to WW\to 2\mu 2\nu$	0.918	1.5×10^3	0.809	6.8×10^3
$H(160)\to WW\to 2\mu 2\nu$	0.922	2.5×10^3	0.824	1.1×10^4
$H(200)\to WW\to 2\mu 2\nu$	0.950	1.5×10^3	0.900	7.2×10^3
$H/A(200)\to\tau\tau_{\to\mu\bar{\nu}\nu}$	0.569	6.1×10^3	0.404	2.2×10^4

6.2.3 Conclusions

Ideally, the trigger selection should be as inclusive as possible: very specific selection criteria bias the sample of accepted events and complicate its understanding in the offline analysis. In the case of the muon trigger, it is therefore preferable to select events with the inclusive single- and di-muon triggers, whenever it is possible. However, the available bandwidth in the DAQ system, as well as the maximum acceptable rate to storage, impose a constraint on the thresholds that can be applied at each level, as discussed in Section 6.1. It

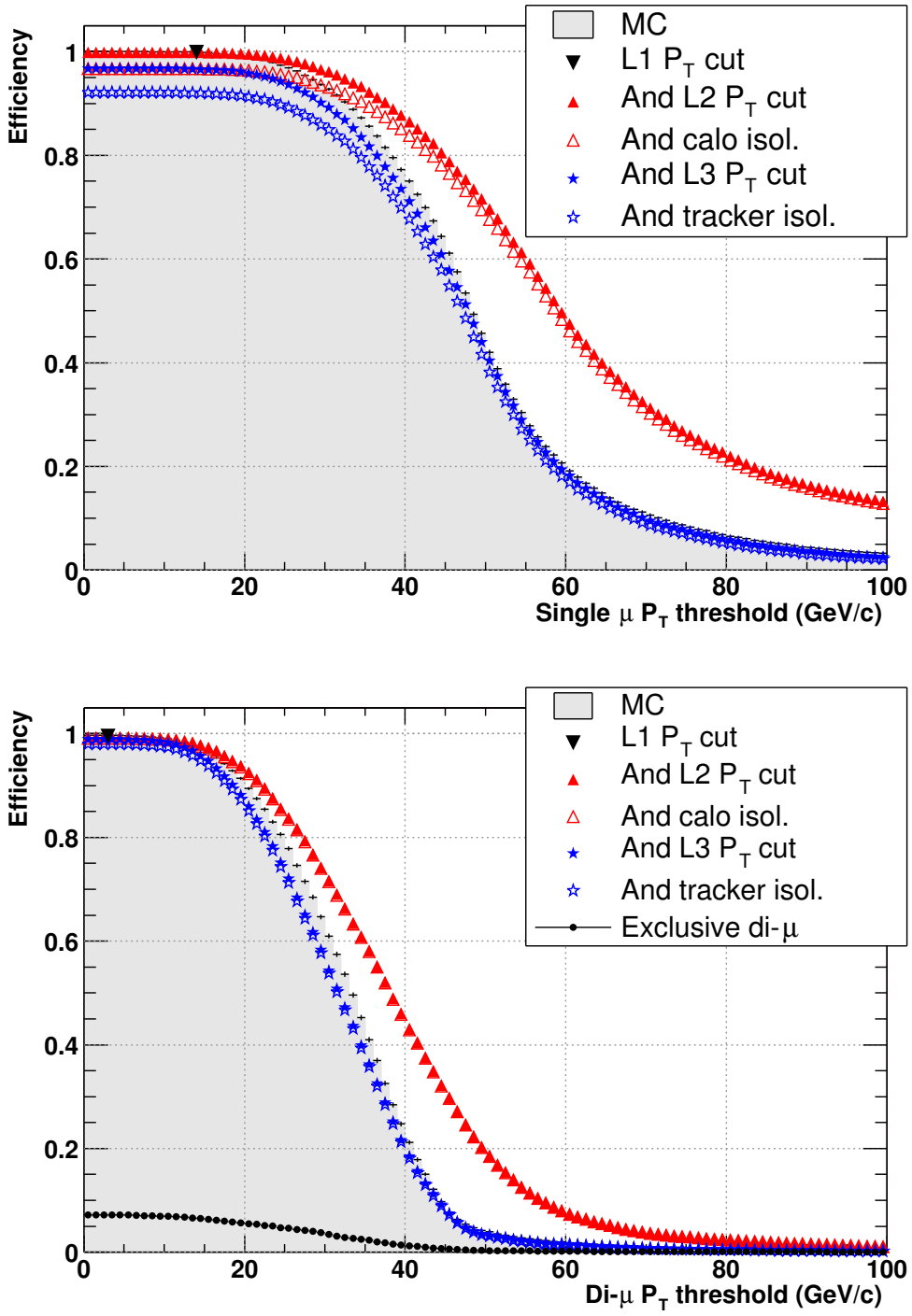


Figure 6.9: *Efficiency for selecting $H \rightarrow ZZ \rightarrow 4\mu$ ($M_H = 130\ GeV/c^2$) decays at low luminosity as a function of the single muon threshold (top) and of the symmetric di-muon p_T threshold (bottom) in the HLT. The exclusive di-muon contribution refers to events not selected by the single muon trigger after Level-3 and isolation. The curve labelled "MC" represents the efficiency as a function of the threshold on the generated p_T.*

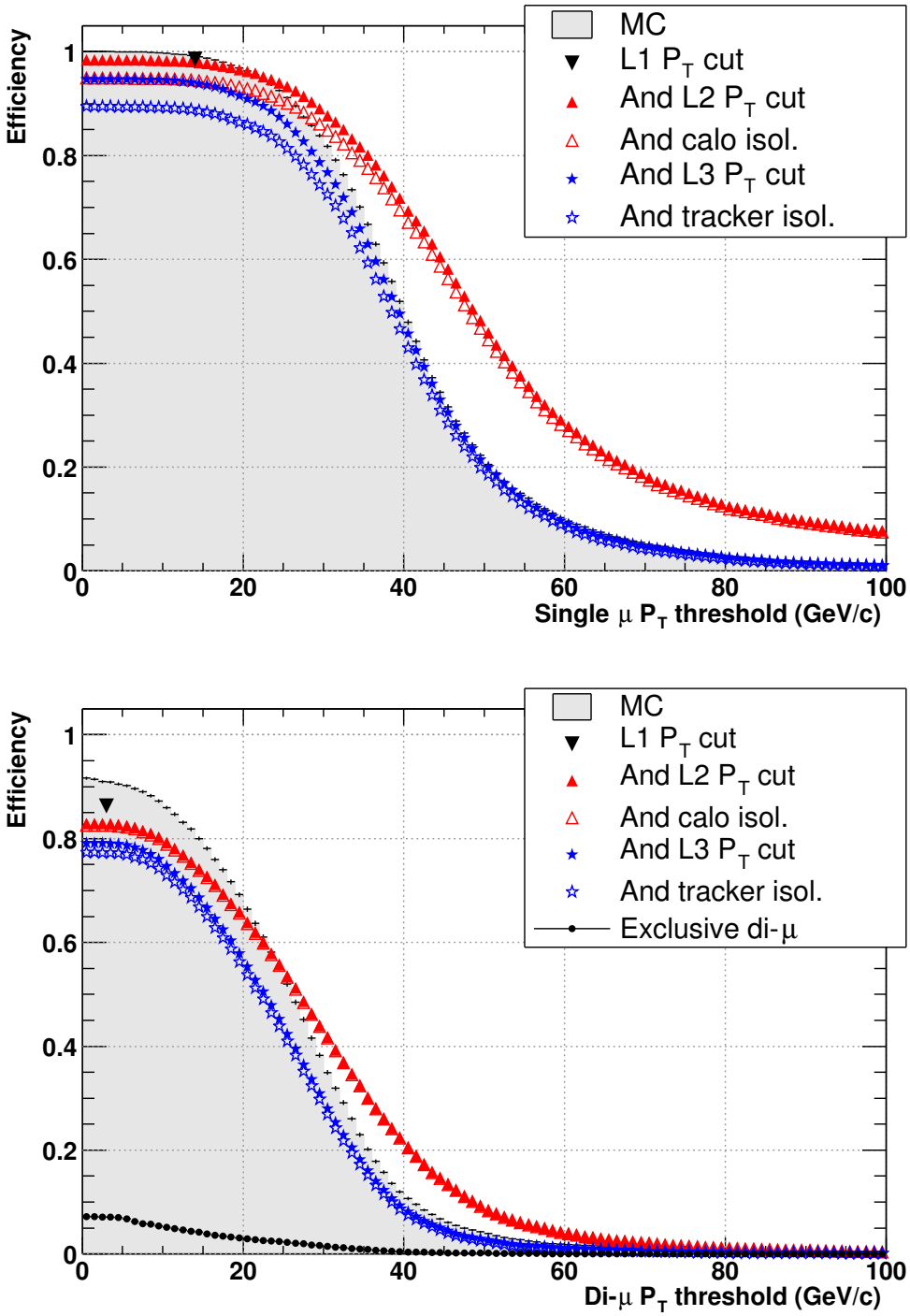


Figure 6.10: *Efficiency for selecting $H \rightarrow WW \rightarrow 2\mu 2\nu$ ($M_H = 160\ GeV/c^2$) decays at low luminosity as a function of the single muon threshold (top) and of the symmetric di-muon p_T threshold (bottom) in the HLT. The exclusive di-muon contribution refers to events not selected by the single muon trigger after Level-3 and isolation. The curve labelled "MC" represents the efficiency as a function of the threshold on the generated p_T.*

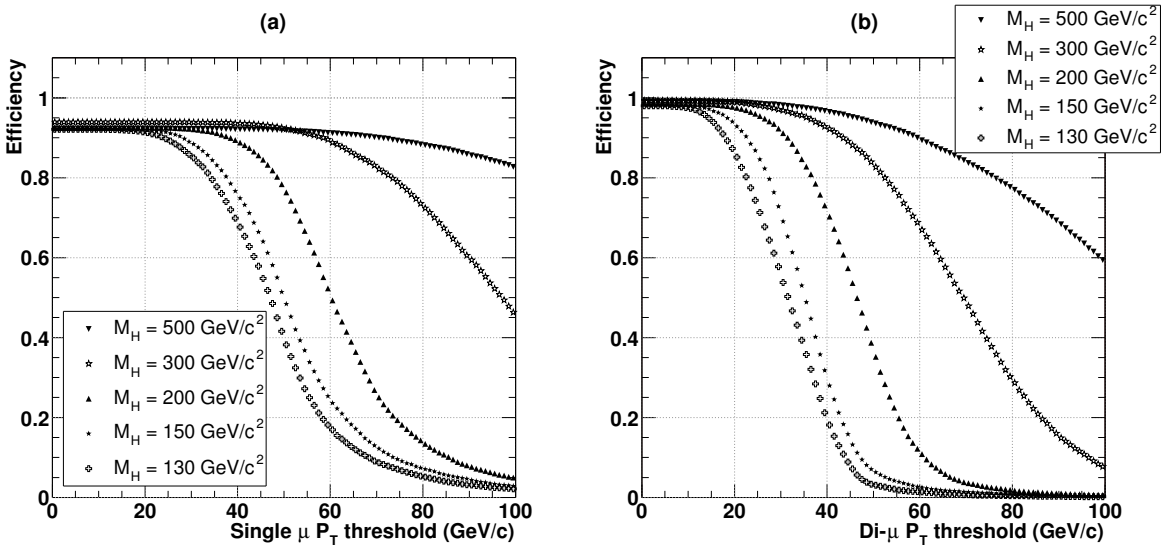


Figure 6.11: *Efficiency for selecting $H \rightarrow ZZ \rightarrow 4\mu$ decays at low luminosity, for different values of the Higgs mass, with (a) the Level-3 isolated single-muon trigger and (b) the Level-3 isolated di-muon trigger.*

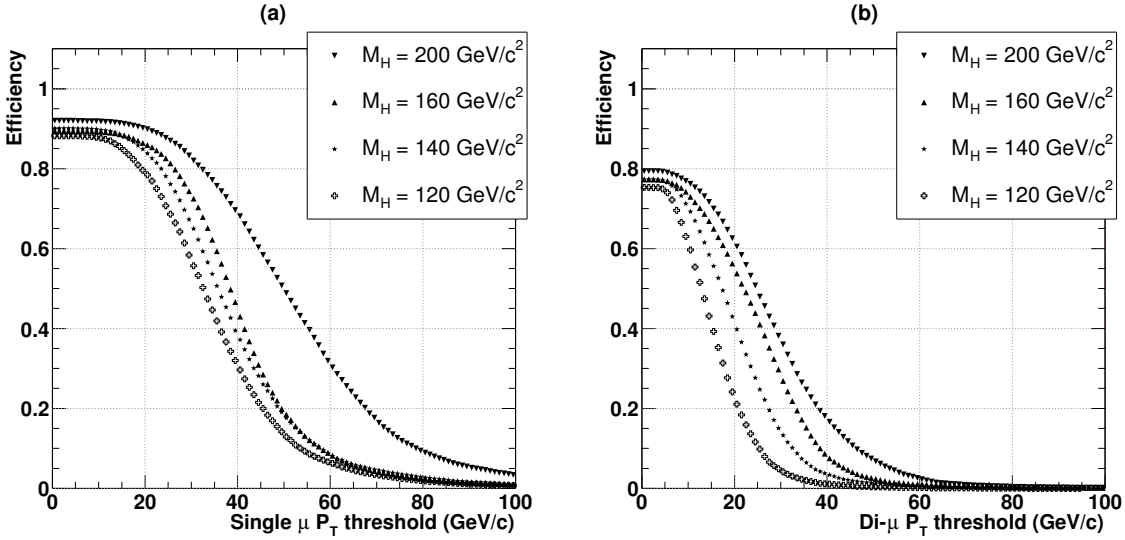


Figure 6.12: *Efficiency for selecting $H \rightarrow WW \rightarrow 2\mu 2\nu$ decays at low luminosity, for different values of the Higgs mass, with (a) the Level-3 isolated single-muon trigger and (b) the Level-3 isolated di-muon trigger.*

is important to verify if the resulting selection criteria still allow sufficient efficiency for the interesting signals. The conclusion from Section 6.2.2 is that the inclusive muon triggers provide very good selection efficiency on several interesting channels, and in particular on the most interesting Standard Model Higgs discovery channels with muonic final states. Moreover, in most of these cases the thresholds for the inclusive selection are located in a range where the efficiencies change slowly with the threshold value. This allows a simple understanding and correction of the trigger inefficiency in the off-line analysis.

Still, more specific selection criteria can be applied, either to flag events in order to pre-select them for the initial offline analysis, or to further reduce the total rate in case of unexpected conditions. More sophisticated and exclusive triggers can also be accommodated, for channels that are not efficiently selected by the inclusive trigger, *e.g.* because of a very soft muon p_T spectrum. An example is the selection of B_s^0 decays [1, App. G2].

6.3 CPU-Time Requirements

The time available for running HLT algorithms in the on-line trigger is limited. The CPU processing time is therefore a relevant parameter to evaluate the feasibility of algorithms to be included in the trigger chain. A detailed study of the CPU time required by the muon HLT chain was performed as part of the validation of the trigger selection.

Timing studies can be subdivided into two main tasks. We call *timing* the determination of the absolute CPU time spent by each algorithm when running in realistic conditions, in order to determine if it fits in the trigger chain. The *profiling* of individual algorithms consists in the determination of the most time consuming operations in order to find possible speed-ups and optimisations. The results of a study of the timing of the muon HLT chain are discussed in Section 6.3.1. A detailed profiling of the Level-2 muon reconstruction, and the options for further improving it, are described in Section 6.3.2.

These two tasks are accomplished with similar tools, though they have to address very different problems. In particular, absolute time measurements are not necessary for algorithm profiling, which also does not require to be performed using a realistic input sample. On the other hand, timing must provide absolute time measurements and, since the reconstruction time depends on the physics content of the events considered, it must be done on a realistic input sample at all trigger levels. A special treatment is of course required when using weighted events such as the inclusive muon samples described in Section 3.

The results of timing measurement are affected by several independent factors. The most evident is the hardware used; in particular the CPU type and

speed. All CPU-time measurements presented in the following were performed on commercial computers equipped with Intel Pentium III processors running at 1 GHz in a controlled environment, *i.e.* without any other data processing or input/output intensive process running, apart from that being measured. The reading and formatting of the input data is another source of uncertainty, since the actual data access mechanisms that will be used in the HLT online farm are not yet implemented and will be different from those used by current reconstruction algorithms. For this reason, whenever possible, the data access time was subtracted on a per-event basis from the absolute timing measurements presented in Section 6.3.1.

Other factors that affect the result are the type of compiler and degree of binary optimisation it produces, the performance of the operating system and that of the underlying software components, which are subject to constant improvements. The version of ORCA used for these results was 6.2.3, running on Linux (kernel ver. 2.2.19). The code was compiled with the `gcc` compiler version 2.95.2 with the optimisation flag `-O2`.

Basic facilities for the timing of algorithms are provided by COBRA, the framework used by the CMS reconstruction and analysis package, ORCA. They are based on system calls that measure both real (wall-clock) time and CPU-time, corrected to account for the fraction of CPU activity actually devoted to the process under study.

6.3.1 Timing of the Muon HLT Chain

The goal of HLT timing is to determine the absolute CPU time spent by the HLT selection when running under realistic conditions. For this task, a program performing the full muon HLT selection chain was implemented, including Level-2 reconstruction, calorimeter isolation, Level-3 reconstruction, and pixel and tracker isolation. A trigger selection was performed at each step, and only selected events were passed on to the following step, as will happen in the real trigger.

The first requirement to have a correct estimation of the CPU time is that the sample used be representative of the inclusive HLT input. For this reason, a subsample with correct proportions of the three minimum bias samples and of the W sample described in Section 3 was used. As shown in Fig. 3.4, these events give the main contribution to muon rates for a wide range of thresholds. Since events in the minimum bias samples are weighted, it is not sufficient to measure the cumulative time spent in each trigger step; the measurement was performed on a per-event basis and each measurement was weighted with the event weight. Special low-overhead software timers were implemented for this task. Per-event timing also allows to obtain the distribution of the time spent per event.

Another factor that influences the timing is the trigger selection and threshold used. For simplicity, only the single muon selection was simulated; however, all muon candidates were analysed in every step and, in case of selection, forwarded to the next step. Reconstruction time depends critically on the threshold used, since low p_T muons require the search of compatible hits in a much wider area. The thresholds chosen were conservatively set to lower values than those described in Section 6.1, *i.e.* to 10 and 18 GeV/c at low and high luminosity, respectively. The same thresholds were applied at Level-1 to select input events for the HLT. Isolation algorithms at Level-2 and Level-3 were applied at 97% nominal efficiency.

The results of this timing procedure are shown in Table 6.8 for both the low and high luminosity cases. For each level, the average time spent per event passing the previous level is shown, together with the measurement obtained after subtracting the time spent in the GEANE propagation routine (cf. Section 4.2), which in most cases accounts for a significant fraction of the CPU time. The total time given in the last row is averaged over the full Level-1 rate, so that it represents the average time spent in the HLT for a Level-1 candidate. It should be noted that higher level algorithms contribute less to this average since they are applied to fewer events, thanks to the rejection obtained by earlier levels. Also, the reason for the fact that the results are similar at low and high luminosity is that the trigger threshold used is different.

Table 6.8: *Average CPU time spent in the muon HLT selection steps. Each row represents the average time spent per event passing the previous trigger level. Also listed is the time measured after subtracting the contribution of the GEANE propagation routine. The totals are averaged over all events passing the Level-1 selection.*

	$\mathcal{L}$ =2 × 10^{33}cm^{-2}s^{-1}		$\mathcal{L}$ =10^{34}cm^{-2}s^{-1}	
HLT Algorithm	Total (ms/ev)	Excl. GEANE (ms/ev)	Total (ms/ev)	Excl. GEANE (ms/ev)
Level-2	640	100	580	100
Calorimeter isolation	100	25	90	40
Level-3	420	200	590	420
Pixel isolation	65	65	320	320
Tracker isolation	190	190	370	370
Total/Level-1 event	710	125	660	150

The distributions of the CPU time spent by the Level-2 and Level-3 reconstruction and by the calorimeter and pixel isolation algorithms are shown in Figs. 6.13 and 6.14 for the low and high luminosity cases. The distribution of

the total time per event passing the Level-1 selection is shown in Fig. 6.15. The component exclusive of the time spent in the GEANE routine is also shown; it is clear that extrapolations account for most of the CPU time and also determine the long tails visible in the plots. The plots are fitted to a log-normal distribution, that in most cases describes well their behaviour.

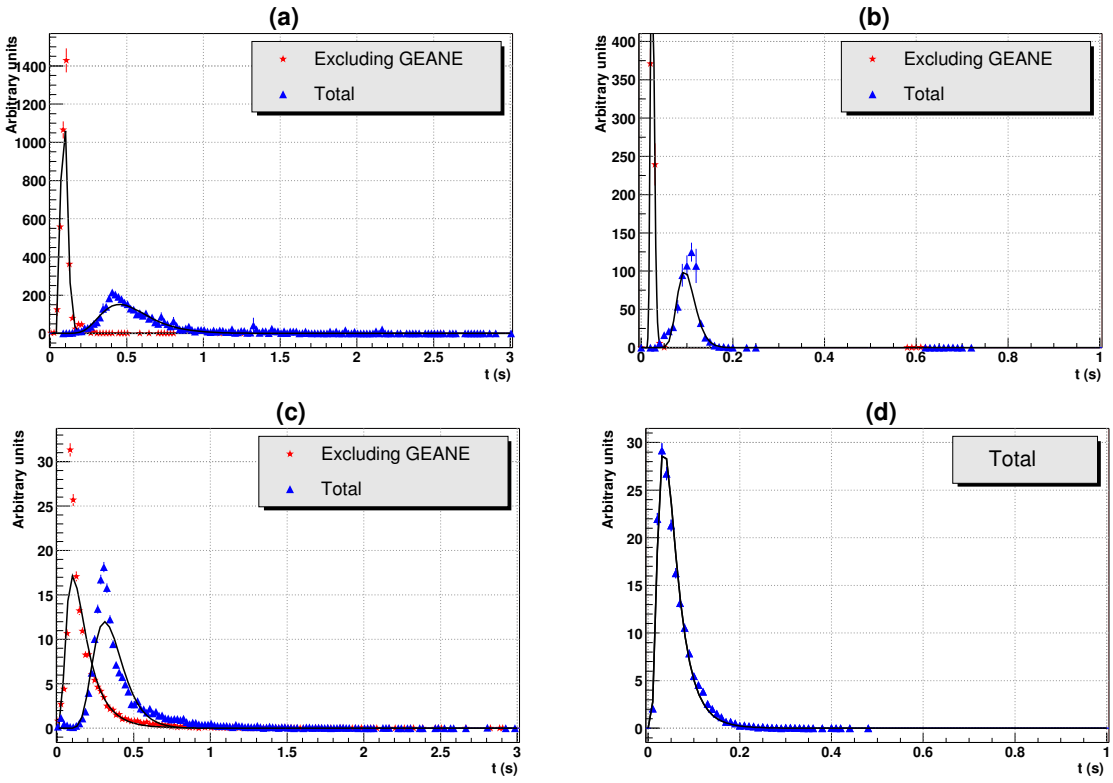


Figure 6.13: *Distribution of the time spent per event in (a) Level-2 reconstruction, (b) calorimeter isolation, (c) Level-3 reconstruction and (d) pixel isolation, at low luminosity. At each step, only events passing the previous steps are included. The solid lines are the result of fits to a log-normal distribution.*

A more difficult exercise is to translate these measurements into requirements on the computing resources needed for the HLT online farm and to determine its feasibility. Several uncertainties affect this estimate, which has recently been discussed in the CMS DAQ TDR [1]. A brief summary is given here, focusing on the situation planned for the LHC startup in the year 2007.

The average time spent by all HLT algorithms (muons, electrons/photons, jets, tau- and b-jet identification) is calculated by weighting the times spent by each algorithm with the corresponding Level-1 rate, and amounts to about 270 ms per Level-1 event. Muon reconstruction gives by far the biggest contribution, accounting for more than 60% of this time. For comparison, the second

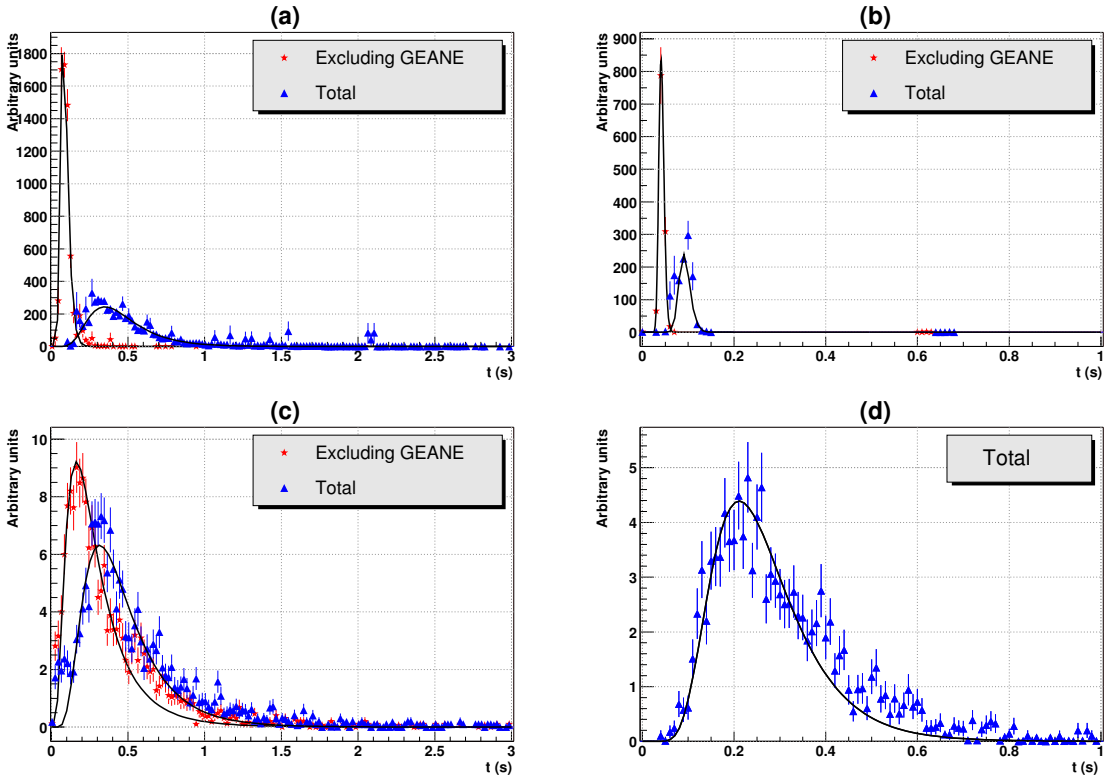


Figure 6.14: *Distribution of the time spent per event in (a) Level-2 reconstruction, (b) calorimeter isolation, (c) Level-3 reconstruction and (d) pixel isolation, at high luminosity. At each step, only events passing the previous steps are included. The solid lines are the result of fits to a log-normal distribution.*

largest contribution is that from electron and photon identification, which accounts for only about 17% of the total. Substantial improvements can therefore be expected from further optimisation of muon reconstruction algorithms.

The HLT is implemented in software, and will benefit from the improvements in computing technology occurring up to the year 2007. An estimate of the future improvement in CPU computing power is difficult, but a figure can be obtained by using the empirical rule known as Moore's law, where the trend is quantified by a doubling of performance every 1.5 years. An overall factor of ~ 8 is therefore to be expected. According to this projection, the current implementation of the HLT chain will require about 40 ms per Level-1 event at LHC startup.

The initial CMS DAQ system will be capable of reading a maximum rate of 50 kHz of events accepted by the Level-1 Trigger. As discussed earlier, only one third of this rate has been allocated for physics selection, the rest being used as safety factor for unforeseen conditions and processes (*e.g.* detector response,

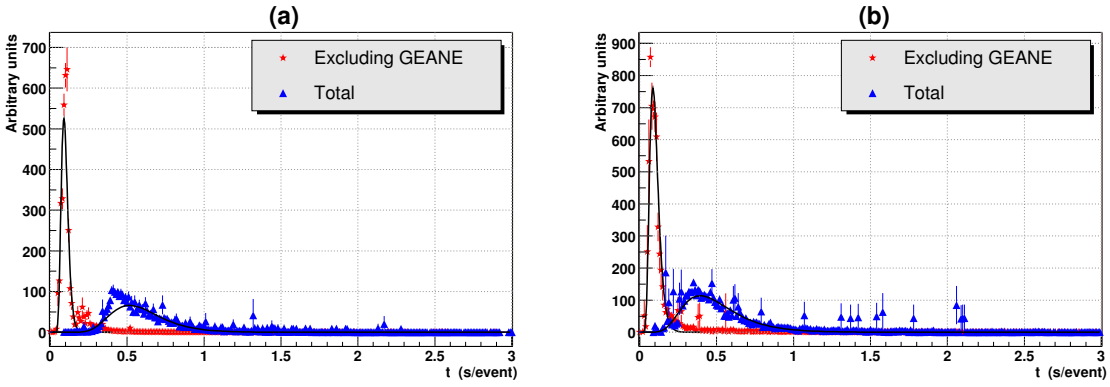


Figure 6.15: *Distribution of the total time spent in the full HLT chain for each Level-1 selected events at (a) low and (b) high luminosity. The solid lines are the result of fits to a log-normal distribution.*

beam halo). However, conservatively assuming that the same average time of 40 ms will be required by the full event rate of 50 kHz, we conclude that the initial HLT farm will require about 2000 CPUs. This is considered to be feasible, and current plans are for a computing farm composed of about 1000 dual-CPU PC boxes.

Of course this estimate is affected by several uncertainties which are difficult to quantify. These are extensively discussed in [1]. The development and improvement of HLT algorithms will continue up to and after the LHC startup, facilitated by the software nature of the HLT. An analysis of the time spent in the Level-2 muon reconstruction is presented in the following section, with the goal of identifying possible improvements.

6.3.2 Profiling of the Level-2 Reconstruction

As shown in the previous section, the muon reconstruction and selection is the most time consuming part of the full HLT. It is particularly important to improve the speed of the muon reconstruction at Level-2, since this is the first HLT level and has to deal with the highest input rate.

Profiling is an iterative process, which consists in identifying the most time consuming processes, improve them, and iterate until the result is satisfactory. A detailed profiling was carried out on the Level-2 and led to the redesign of the most time consuming part, the navigation in the muon detectors, described in Section 4.5, as well as to other minor improvements. The overall speed-up obtained was approximately a factor of eight with respect to the previous prototype of the Level-2 reconstruction. This improvement is already included in the timing measurement discussed in the previous section. The profiling of

the current Level-2 was repeated with the updated code, and the results are shown in the following.

As already mentioned, the absolute time scale is not very important for profiling studies. The most important requirement is that the measurements are performed in consistent conditions. In particular, except for very high optimisation levels, the composition of the input sample is not particularly relevant. A sample of single muons with flat p_T distribution in the range 5-100 GeV/c, digitised without pile-up, was used for the present study, since it also allows easy debugging and inspection of the reconstruction algorithms.

The distribution of the Level-2 time per event of this sample is shown in Fig. 6.16. The contribution exclusive of the time spent in the GEANE propagation routine is also shown; it is evident that calls to GEANE still account for a significant fraction of the total time.

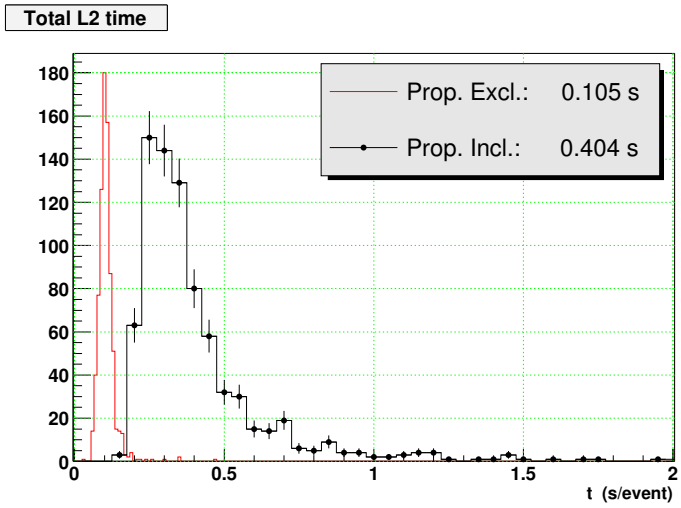


Figure 6.16: *Distribution of time per event spent in the Level-2 reconstruction on a sample of single muons without pile-up. The two histograms correspond to the total time and the time after subtracting the time spent in the GEANE propagation routine.*

The time has then been measured separately for the main blocks of Level-2 reconstruction described in Section 4.4, *i.e.* the seed generation, the RecHit collection and seed refinement, the track fit and the final extrapolation to the vertex and vertex constraint. The result is shown in Fig. 6.17. Most of the time is spent in the seed refinement and track fit. In particular, the time spent in the former seems excessive if compared with the latter. In fact, in the present implementation the seed refinement is algorithmically equal to the track fit, except for the value of χ^2 cuts and for the fact that the initial seed is less precise. A faster, less accurate solution can be used. Several possibilities have

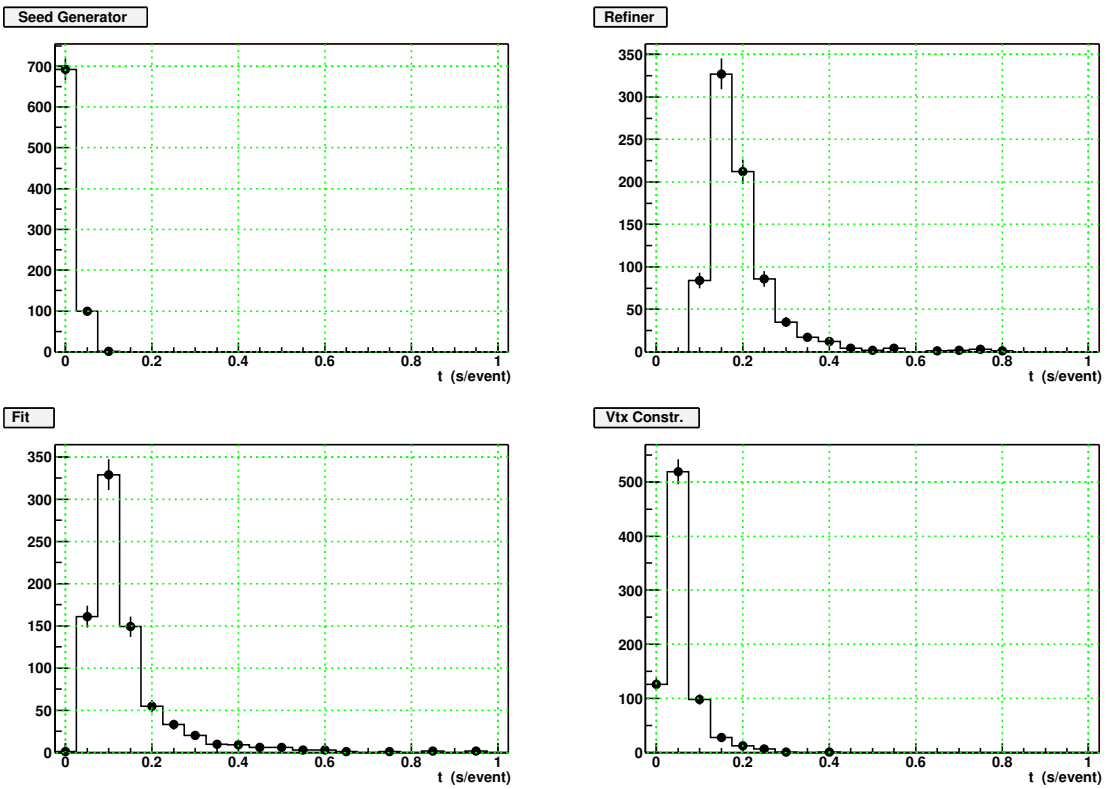


Figure 6.17: *Distribution of the time spent in the different components of the Level-2 reconstruction: seed generation, seed refinement, track fit, extrapolation to the vertex and vertex constraint.*

been proposed, *e.g.* the use of faster and less accurate propagators or the fit of the segments reconstructed in the chambers used to build the corresponding Level-1 candidate. The result will be used as a seed for the track fit as usual.

The time spent in the extrapolation to the vertex is also relatively long. This is due to the use of GEANE for the extrapolation in the version of the software used for this study. An important improvement can be obtained by substituting it with a propagator based on an analytic calculation in the region inside the calorimeters, where the magnetic field is fairly constant and the amount of material is small.

The main concern regarding the Level-2 timing is anyhow the time spent within the GEANE propagation routine. GEANE itself will have to be substituted in 2003, as part of the program to remove all legacy FORTRAN libraries from ORCA. It is unlikely that an equivalent C++ implementation will be intrinsically faster; however, speed-ups can be be obtained in two ways: by carefully tuning the extrapolation parameters, for example adapting the precision of the calculation and of the estimation of the error matrix to the actual

need of each extrapolation, and by reducing the number of GEANE calls to the minimum necessary. The major part of the latter task was already accomplished with the optimisation of the muon navigation described in Section 4.5. As discussed there, "long" extrapolations between stations and through the iron return yoke were fully optimised; a full optimisation of "short" extrapolations (*e.g.* between chambers within a station) was done in the barrel region only. The fraction of the total extrapolation time per event as a function of the extrapolation length is shown in Fig. 6.18 separately for the barrel and endcap regions. In the latter, the extrapolation time is still dominated by short extrap-

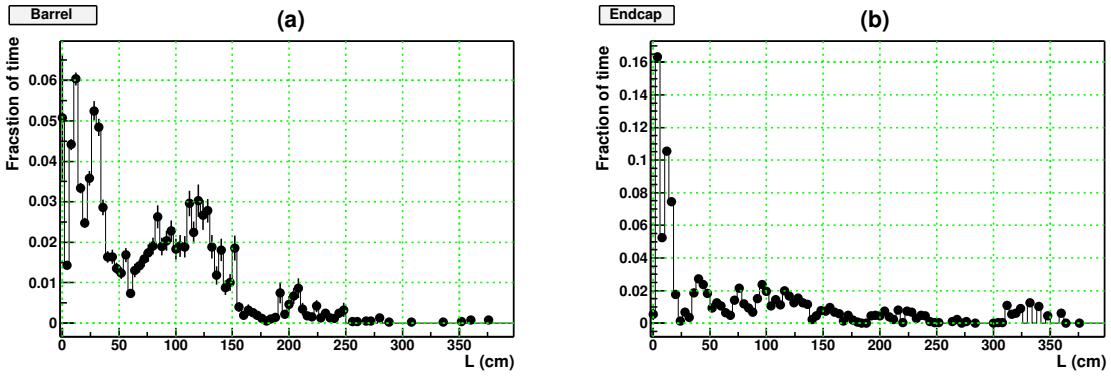


Figure 6.18: *Fraction of the total GEANE extrapolation time spent as a function of the extrapolation length in the Level-2 reconstruction in (a) the barrel and (b) the endcap region.*

olations (less than 20 cm), that account for about 40% of the total time. The same extrapolations are optimised in the barrel and account for about 20% of the total. In principle, the geometry of the endcaps is much simpler; however the extrapolations between layers inside CSC chambers cannot be eliminated. The possible improvement obtained by optimising the number of short extrapolations in the forward region can be quantified to be up to a factor ~ 1.2.

To conclude, the profiling of the Level-2 suggests a number of improvements that can be introduced to speed-up the current reconstruction algorithms, without affecting the quality of the reconstruction:

- The seed refinement, which is the most time consuming single block in the reconstruction, can safely use fast propagators with limited precision and rough treatment of material effects. Further improvements can be obtained with more radical re-implementations that have recently been proposed. Expected speed-ups are of the order of overall factors ~ 1.2 to 1.5;

- Short extrapolations inside the endcap stations can be easily optimised,

as discussed in Section 4.5. The expected improvement is a factor ~ 1.2 in the extrapolation time for muons in the endcap region only;

- Other improvements can come from adapting the extrapolation precision and treatment of material effects to what needed for each extrapolation. The ability to do so is a requirement for the replacement of GEANE that will be implemented in 2003. In particular, it is likely that material effects can be treated in a simplified way every time the extrapolation does not cross iron volumes, like in the case of the very short extrapolation within CSC chambers, those between chambers within stations and between RPC and DT/CSC planes. An estimate of the corresponding speed-up is difficult, but overall speed-up factors higher than ~ 1.3 should be achievable.

Conclusions

The subject of this thesis is the muon High Level Trigger (HLT) system of CMS. In the HLT, muons are reconstructed in two different steps: first, using only the information coming from muon detectors (Level-2); then including information from the central tracking system (Level-3). At both levels, events are selected by requiring the presence of high-transverse-momentum, isolated muons.

The work presented here covers most aspects of the implementation, simulation, validation and analysis of the performance of the muon HLT. A first aspect was the development of the Level-2 muon regional reconstruction, notably the implementation of a highly optimised algorithm for the collection of hits to be used in the track fit. The performance of this algorithm is important not only in terms of efficiency: the reconstruction time is crucial, since the Level-2 has to deal with a very high event rate and has to take a decision within a limited time.

Isolation is a very important selection criterion to reduce the total trigger rate. The Level-1 muon trigger has a built-in isolation algorithm. It was shown that, though being effective for the rejection of high-p_T non-isolated muons, this algorithm does not provide a significant reduction of the Level-1 rate, which is dominated by low-p_T muons. More sophisticated isolation algorithms were implemented for the HLT and the offline analysis. It was shown that they significantly reduce the HLT trigger rate while guaranteeing a high efficiency for muons from heavy object decays.

The study of the trigger performance requires the generation and simulation of large inclusive samples of events with muons in the final state. The generation of these samples is very demanding in terms of computing resources, since it has to include contributions from the entire inelastic proton-proton cross section. A special optimised procedure was developed and used for the production of an inclusive muon sample with good statistical significance up to transverse momenta of ~ 70 GeV/c.

This sample was used to determine the rates at each trigger level as a function of the threshold on p_T. This result is very important since the maximum acceptable rate at each level determines the thresholds that can be used. The resulting thresholds for the inclusive single- and di-muon triggers were discussed.

They were shown to be efficient for a variety of benchmark channels – notably for the selection of W and Z bosons, of $t\bar{t}$ quark pairs and several Higgs discovery channels. For these processes, a selection based on the requirement of high-p_T, isolated muons is sufficient. More exclusive triggers, while possible, do not appear to be necessary.

Finally, the timing of all muon HLT algorithms was studied. It was shown that the CPU time spent in the muon reconstruction and isolation algorithms is acceptable. The possibility of further optimisation of the algorithm was discussed. In particular, a detailed profiling study was carried out for the Level-2 reconstruction algorithm, which has to deal with the highest input rate.

In conclusion, the current prototype of the CMS HLT is a complete, fully functional system. However, this is not intended to be the final implementation. The flexibility of the HLT framework will allow trigger algorithms to evolve in parallel with the off-line reconstruction, adapt to the experimental conditions of LHC and respond to the physics demands of the CMS community.

Bibliography

[1] The CMS Collaboration, *The TriDAS Project, Technical Design Report: Data Acquisition and High-Level Trigger*, CERN/LHCC 02-26, CMS TDR 6.1 (2002).

[2] N. Amapane *et al.*, *Monte Carlo Simulation of Inclusive Single- and Di-Muon Samples*, CMS NOTE 2002/041.

[3] N. Amapane, M. Konecki and M. Zanetti, *TAG based Framework for Muon Analysis*, CMS IN 2001/057.

[4] N. Amapane, M. Konecki and M. Fierro, *High Level Trigger Algorithms for Muon Isolation*, CMS NOTE 2002/040.

[5] M. Guidry, *Gauge Field Theories: an Introduction with Applications*, Wiley Publications, 1991.

[6] T. Hambye and K. Riesselmann, *Matching conditions and Higgs mass upper bounds revisited*, Phys. Rev. D **55** (1997) 7255.

[7] The LEP collaborations, the LEP Electroweak Working Group and the SLD Heavy Flavour Group, *A Combination of Preliminary Electroweak Measurements and Constraints on the Standard Model*, LEPEWWG/2002-01 (2002).

[8] G. Altarelli and M. L. Mangano (editors), *Proceedings of the Workshop on Standard Model Physics (and more) at the LHC* (2000) CERN 2000-004.

[9] The LHC Study Group, *The Large Hadron Collider Conceptual Design*, CERN/AC 95-05 (1995).

[10] H. L. Lai *et al.*, *Improved Parton Distributions from Global Analysis of Recent Deep Inelastic Scattering and Inclusive Jet Data*, Phys. Rev. D **55** (1997) 1280.

[11] The ATLAS Collaboration, *Atlas Detector and Physics Performance, Volume II*, CERN/LHCC 99-15 (1999).

[12] M. Spira and P. M. Zerwas, *Electroweak Symmetry Breaking and Higgs Physics*, CERN-TH/97-379 (1997).

[13] F. Gianotti and M. Pepe-Altarelli, *Precision Physics at LHC*, Nucl. Phys. **B89** (Proc. Suppl.) (2000) 177.

[14] The CMS Collaboration, *Technical Proposal*, CERN/LHCC 94-38, LHCC/P1 (1994).

[15] The CMS Collaboration, *The Magnet Project: Technical Design Report*, CERN/LHCC 97-10, CMS TDR 1 (1997).

[16] The CMS Collaboration, *The Tracker Project: Technical Design Report*, CERN/LHCC 98-6, CMS TDR 5 (1998).

[17] The CMS Collaboration, *Addendum to the CMS Tracker Technical Design Report*, CERN/LHCC 2000-016, CMS TDR 5 Addendum 1 (2000).

[18] The CMS Collaboration, *The Electromagnetic Calorimeter Project: Technical Design Report*, CERN/LHCC 97-33, CMS TDR 4 (1997).

[19] The CMS Collaboration, *The Hadronic Calorimeter: Technical Design Report*, CERN/LHCC 97-31, CMS TDR 2 (1997).

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

[21] F. Gasparini *et al.*, *Bunch Crossing Identification at LHC Using a Meantimer Technique*, Nucl. Inst. & Meth. **A336** (1993) 91.

[22] The CMS Collaboration, *The Trigger Project, Technical Design Report: The Level-1 Trigger*, CERN/LHCC 2000-038, CMS TDR 6.1 (2000).

[23] H. Sakulin, *Design and Simulation of the First Level Global Muon Trigger for the CMS Experiment at CERN*, PhD thesis (2002).

[24] D. Acosta *et al.*, *Results on L2 trigger reconstruction in single and di-muon topologies*, CMS IN 2001/011.

[25] T. Sjöstrand, P. Edén, C. Friberg *et al.*, *High-energy-physics event generation with PYTHIA 6.1*, Comp. Phys. Comm. **135** (2001) 238.

[26] C. Charlot *et al.*, *CMSIM–CMANA CMS Simulation Facilities*, CMS TN/93-63 (1993).
http://cmsdoc.cern.ch/cmsim/cmsim.html

[27] CMS Software and Computing Group, *Object Oriented Reconstruction for CMS Analysis*, CMS IN 1999/001.
http://cmsdoc.cern.ch/orca/

[28] V. Lefebure, *The CMS Production Project home page* (2002).
http://cmsdoc.cern.ch/cms/production/www/html/general/

[29] M. Glück, E. Reya and A. Vogt, *Dynamical Parton Distributions of the Proton and Small-x Physics*, Z. Phys. **C67** (1995) 433.

[30] F. Abe *et al.*, *Measurement of Double Parton Scattering in $p\bar{p}$ Collisions at $\sqrt{s} = 1.8$ TeV*, Phys. Rev. D **41** (1989) 2330.

[31] G. J. Alner *et al.*, *Scaling violation favoring high multiplicity events at 540-GeV cms energy*, Phys. Lett. B **138** (1983) 304.

[32] F. Abe *et al.*, *Measurement of the $B_d^0 - \bar{B}_d^0$ flavor oscillation frequency and study of same side flavor tagging of B mesons in $p\bar{p}$ collisions*, Phys. Rev. **D59** (1999) 32001.

[33] Particle Data Group, *Review of Particle Physics*, Phys. Rev. **D66** (2002).

[34] A. Belkov and S. Shulga, *Status of physics generator of B-decays at Dubna*. Talk given at the CMS B-Physics meeting, April 27, 2001.

[35] H. Sakulin, *Simulation of Single-Muon and Di-Muon Trigger Rates at the CMS Experiment in the Presence of Pile-Up*, CMS NOTE 2002/042.

[36] C. Seez, *Minimum-Bias Pileup Issues in Electron-Photon HLT Studies*, CMS IN 2000/001.

[37] D. Acosta, M. Stoutimore and S. M. Wang, *Simulated Performance of the CSC Track-Finder*, CMS NOTE 2001/033.

[38] M. Huhtinen, *Optimization of the CMS Forward Shielding*, CMS NOTE 2000/068.

[39] G. Bruno and M. Konecki, *Simulation of the Baseline RPC Trigger System for CMS: Efficiency and Output Rates in Single Muon Topology*, CMS NOTE 2001/012.

[40] CN Division Application Software Group, *GEANT3 version 3.21/13 (release 15111999) Detector Description and Simulation Tool*, CERN Program Library Long Writeup W5013 (1993).

[41] C. Zeitnitz and T. A. Gabriel, *The GEANT-CALOR interface and benchmark calculations of ZEUS test calorimeters*, Nucl. Inst. & Meth. **A349** (1994) 106.

[42] V. Innocente, L. Silvestris and D. Stickland, *CMS software architecture, Software framework, services and persistency in high level trigger, reconstruction and analysis*, Comp. Phys. Comm. **140** (2001) 31. http://cobra.web.cern.ch/cobra/

[43] R. E. Kalman, *A New Approach to Linear Filtering and Prediction Problems*, Transaction of the ASME–Journal of Basic Engineering (1960) 35.

[44] P. S. Maybeck, *Stochastic Models, Estimation, and Control*, volume 1, Academic Press, Inc., 1979.

[45] V. Innocente, M. Maire and E. Nagy, *GEANE : average tracking and error propagation package*, CERN Program Library Long Writeup W5013-E.

[46] G. Bruno *et al.*, *Local reconstruction in the muon detectors*, CMS NOTE 2002/043.

[47] N. Amapane, *L2 Navigation and Steering*. Talk given at the CMS PRS/mu meeting, April 30, 2002.

[48] G. Wrochna, *Muon Trigger of the CMS detector for LHC*, CMS NOTE 1997/096.

[49] C. Albajar and G. Wrochna, *Isolated Muon Trigger*, CMS NOTE 2000/067.

[50] A. Fanfani *et al.*, *Monte Carlo simulation for High Level Trigger studies in single and di-muon topologies*, CMS IN 2000/053.

[51] N. Amapane, *QUIET and MIP bits in the GMT*. Talk given at the CMS Trigger meeting, April 24, 2001.

[52] D. Kotlinski, *Track Reconstruction and Primary Vertex Finding using the Pixel Detector Data*, CMS IN 2000/022.

[53] M. Dittmar and H. Dreiner, *How to find a Higgs Boson with a Mass between 155–180 GeV at the LHC*, Phys. Rev. **D55** (1997) 167.

Acknowledgements

First of all, I wish to thank my supervisor, Prof. Alessandra Romero, for her constant support during my doctorate. I am very much indebted to Prof. Michele Arneodo, for several interesting discussions, for his useful advice and for carefully reading and commenting this thesis.

Special thanks go to Marcin Konecki, for the fruitful daily collaboration in the development of isolation algorithms and for many stimulating conversations. It was also a pleasure to work with all the colleagues and friends of the PRS/muon group, and in particular with Silvia Arcelli, Daniele Bonacorsi, Giacomo Bruno, Alessandra Fanfani, Stefano Lacaprara, Norbert Neumeister and Hannes Sakulin.

Finally, I want to thank my family and all my friends, whose kindness and love were invaluable for me.