UNIVERSITÀ DEGLI STUDI DI FIRENZE

DIPARTIMENTO DI FISICA

Tesi di Dottorato di Ricerca in Fisica (FIS/04)
XVIII ciclo di Dottorato

Jet flavour tagging with the CMS experiment

Dott. Andrea Bocci

Relatore
Prof. Raffaello D'Alessandro

Coordinatore
Prof. Alessandro Cuccoli

Firenze, 30 Dicembre 2005

Pitagora disdegnò la scrittura; Platone inventò il dialogo filosofico per ovviare agli inconvenienti del libro, «che non risponde alle domande che gli vengono rivolte» [. . .] *.*

E non ho detto nulla del più alto di tutti i maestri orali, che parlava per parabole e che, una volta, come se non sapesse che la gente stava per lapidare una donna, scrisse delle parole per terra che nessuno ha mai letto.

Jorge Luis Borges

Contents

Introduction

The present theory of elementary interactions, the Standard Model of particle physics, is able to explain most of the observed phenomena to a very high degree of accuracy. However, an oustanding issue is still open: the origin of the mass of elementary particles. The most credited hypothesis is the electroweak symmetry breaking via the Higgs mechanism, which predicts the existence of a new particle, the Higgs boson. Despite all the past efforts, this elusive particle has not been observed yet, due to its extremely low production cross section and unknwown mass: since the latter is a free parameter of the theory, it must be searched over a wide range of values. Chapter 1 gives an introduction to the Standard Model, with the Higgs boson search results and mass limits.

The Large Hadron Collider (LHC) being built at CERN will probe this new frontier of particle physics. It will be a proton-proton collider, reaching a luminosity and a center-of-mass energy more than an order of magnitude greater than the present accelerators, allowing to search for heavy particles produced in processes with very low cross sections. Its characteristics will be exploited by the two multi-purpose experiments that will be installed at the beam intersection points: ATLAS and CMS. Two smaller, dedicated experiments will also be installed: LHCb will carry out high precision measurements of CP violation, while ALICE will investigate heavy ion physics and a new state of matter, the quark-gluon plasma.

During my PhD research I've been involved in the construction of the tracker of the Compact Muon Solenoid experiment, CMS. A general purpose detector, its main feature is the 4 T superconducting solenoid that allows a compact design of the detector with a strong magnetic field. The design priorities fulfilled by the CMS project are a redundant muon system, a good electromagnetic calorimeter and a high quality tracking system. An overview of the CMS experiment is given in Chapter 2, with a deeper description of its tracker in the first part of Chapter 3.

In the first year of my PhD studies I continued the work of my degree thesis, developing and testing, in collaboration with Alenia Spazio - LABEN, the power supply prototypes, their control software, and the overall architecture of the CMS Silicon Strip Tracker power supply scheme. These activities culminated taking part in two Test Beams at the CERN X5 experimental area as cabling and power supply expert for the CMS tracker. This gave me the opportunity to test the power supplies performance and the control software I had developed in a realistic scenario. The second part of Chapter 3 gives a general description of the CMS tracker power supply systems, closing with the Test Beam results.

As detailed in the closing part of Chapter 1, the direct search at LEP-2 showed no conclusive evidence for the existence of the Higgs boson. It established however a 95% confidence level lower limit on its mass of $m_{\mathrm{H}} > 114.4\ \mathrm{GeV}/c^2$. Additionally, Standard Model electroweak observables are influenced by radiative corrections to their tree level expectation values induced by higher order loop diagrams, in which particles mass terms appear. Taking into account the most precise measurements from the LEP and Tevatron experiments it is possible to constraint all the observables with a global fit, with the Higgs boson mass as a free parameter. This yields a one-sided 95% confidence level upper limit of $m_{\mathrm{H}} \leq 186\ \mathrm{GeV}/c^2$. Theoretical constraints also come from the requirement of self-consistency of the Standard Model: if its validity is assumed up to Planck scale $\Lambda \sim 10^{19}$ GeV, the allowed Higgs mass range is between 130 and 190 GeV/c^2.

All these arguments suggest the existence of a light Standard Model Higgs boson, with a mass $m_{\mathrm{H}} \sim 150\ \mathrm{GeV}/c^2$. Although not a proof of its existence, they point to a region of particular interest for the search at future experiments. The discovery of a Higgs boson with masses outside this range, on the other hand, would be a clear indication of the presence of new physics.

In this low mass region, one of the dominant decay process of the Higgs boson is into a $\mathrm{b\bar{b}}$ quark-antiquark pair. These events however will be lost in the abundant QCD background. A much more promising discovery scenario is the associated production of a Higgs boson with a $\mathrm{t\bar{t}}$ pair, $\mathrm{pp} \to \mathrm{t\bar{t}H} \to \mathrm{W^+b\,W^-\bar{b}\,b\bar{b}}$, with four b jets in the final state. In such events an efficient b tagging is clearly necessary to properly reconstruct the two t and H and discriminate the signal from the more abundant QCD interactions.

The inclusive tagging of b jets relies upon the large b hadrons mean lifetime ($\mathrm{c}\tau = 470.1 \pm 2.7\ \mu\mathrm{m}$), their large semileptonic branching ratios, having $\mathrm{Br}(\mathrm{b} \to \ell\bar{\nu}_\ell \ \vee\ \mathrm{b} \to \bar{\mathrm{c}} \to \ell\bar{\nu}_\ell \ \vee\ \mathrm{b} \to \mathrm{c} \to \bar{\ell}\nu_\ell) = 19.3 \pm 0.5\%$ per lepton family, and the large charged track decay multiplicity, with an average of five charged tracks per decay vertex, $\langle n_b^{ch} \rangle = 4.955 \pm 0.062$. Chapter 4 gives a detailed analysis of these b hadron properties, as well as a description of the b tagging algorithms implemented in the CMS reconstruction and analysis software (ORCA) and their performance.

Among these, I have developed the “soft lepton b tagging” algorithm: as a b hadron has roughly a 37% probability of decaying emitting directly ($\mathrm{b} \rightarrow \ell^-$) or via a c quark ($\mathrm{b} \rightarrow \mathrm{c} \rightarrow \ell^+$, $\mathrm{b} \rightarrow \bar{\mathrm{c}} \rightarrow \ell^-$) at least an electron or a muon, it is possible to tag a b jet looking for leptons within it. This method is traditionally called soft lepton b tagging, because it is based on the reconstruction and identification of leptons which are softer than the primary ones, emitted in the decays of a top quark or Z^0, $\mathrm{W}^\pm$ or H heavy boson.

Chapter 5 describes the framework for the soft lepton b tag package I have developed, and the study of the topological and kinematical properties of the reconstructed muons used to increase the purity of the tagged sample, improving the rejection of non-b jets without much affecting the b tagging efficiency. In order to achieve this results I combined these tagging variables with Artificial Neural Networks techniques. Starting with the more standard approach to non-linear discriminant, a research in the dedicated literature pointed me to the developement of more powerful algorithms, such as Neural Networks able to estimante bayesian *a posteriori* probabilities. Chapter 5 closes with a description of the different neural network configuration I have studied, and a comparison of their performance.

Chapter 1

Standard Model of Electroweak Interactions

In this chapter we shall introduce the Standard Model of Particle Physics, the lagrangian description of elementary particle interactions and the basic concepts of the electroweak unification theory.

Starting from the gauge simmetry $\mathrm{SU(2)_L \otimes U(1)_Y}$, we shall define the gauge bosons $\mathrm{W^+}$, $\mathrm{W^-}$ and $\mathrm{Z^0}$ for weak interactions and the photon γ for electromagnetic interactions, and through the spontaneous symmetry breaking mechanism introduce a new particle in the theory, the Higgs boson.

In the last part of the chapter, the generation of boson and fermion masses will be described as couplings to the Higgs field.

This theoretical introduction ends with the review of Standard Model Higgs decay modes and constraints on the value of its mass.

Throughout this work natural units $\hbar = \mathrm{c} = 1$ shall be used, unless otherwise stated. The values of the constants in SI units are [1]:

$$\mathrm{c} = 299\,792\,458 \ \mathrm{m\,s^{-1}},$$

$$\begin{aligned}\hbar \equiv \frac{h}{2\pi} &= 1.054\,571\,68(18) \times 10^{-34}\,\mathrm{J\,s} \\ &= 6.582\,119\,15(56) \times 10^{-22} \ \mathrm{MeV\,s}.\end{aligned}$$

1.1 Elementary Particles

In the Standard Model the elementary particles are leptons and quarks, together with their anti-particles. Ordinary matter is made of leptons and hadrons, the latter divided into mesons (a bound state of one quark and one anti-quark) and baryons (a bound state of three quarks).

Leptons and quarks are fermions, and are classified in three families (generations):

$$\begin{pmatrix} \nu_{\mathrm{e}} \\ \mathrm{e} \end{pmatrix} \begin{pmatrix} \nu_{\mu} \\ \mu \end{pmatrix} \begin{pmatrix} \nu_{\tau} \\ \tau \end{pmatrix}$$

$$\begin{pmatrix} \mathrm{u} \\ \mathrm{d} \end{pmatrix} \begin{pmatrix} \mathrm{c} \\ \mathrm{s} \end{pmatrix} \begin{pmatrix} \mathrm{t} \\ \mathrm{b} \end{pmatrix}$$

All the elements belonging to the three families have been directly observed and no experimental evidence of the existence of a fourth generation is provided.

Elementary particles interact with each other via the three fundamental interactions of Nature, which are mediated by vector bosons. The three forces and the respective force carriers are:

Strong: 8 gluons g

Weak: W^+, W^-, Z^0

Electromagnetic: photon γ

The fourth fundamental interaction, Gravity, is not included in the Standard Model: due to the tiny value of its coupling constant compared to the other three forces, its effects are completely negligible in particle physics experiments.

The weak and electromagnetic forces are described as two aspects arising from the same force, the electroweak interaction.

1.2 Electromagnetic and weak interactions unification

The weak interactions are described in the Standard Model as a theory based on a SU(2) ⊗ U(1) gauge invariant lagrangian. This requires that fermion fields be massless, as any explicit mass term in the lagrangian would break gauge invariance; fermion masses are later recovered by a spontaneous symmetry breaking mechanism.

As a consequence of this requirement helicity is a good quantum number for the fermion fields, and it is possible to describe them defining Dirac spinors with well defined helicity, right-handed (R) or left-handed (L). For the lepton families, L_L and L_R states are defined by Eq. 1.1 and 1.2 respectively:

$$\mathrm{L}_L = \begin{pmatrix} \nu_\ell \\ \ell \end{pmatrix}_L \tag{1.1}$$

$$\mathrm{L}_R = \ (\ \ell \ )_R \tag{1.2}$$

with $\ell = (\mathrm{e}, \mu, \tau)$.

Although neutrinos interact only through weak interactions, and thus only in their negative helicity (left handed) state $(\nu_\ell)_L$, the fact that neutrino masses are non-zero allows the existence of right handed state $(\nu_\ell)_R$ in reference frames different from the one in which the neutrino has been initially emitted. The nature of these right handed particles is still unclear: a credited hypothesis is that neutrinos are Majorana particles, i.e. their own antiparticles. In this case, a right handed neutrino $(\nu_\ell)_R$ would simply be an antineutrino $(\overline{\nu}_\ell)_R$, the helicity being the only different quantum number between the two states. Even if this hypothesis proves to be wrong, "wrong helicitiy" neutrinos $(\nu_\ell)_R$ and $(\overline{\nu}_\ell)_L$ are not affected by electroweak interactions, as we shall see.

For the three quark families the helicity states are similar, as defined in Eq. 1.3 and 1.4:

$$\mathrm{Q}_L = \begin{pmatrix} u \\ d \end{pmatrix}_L \tag{1.3}$$

$$\mathrm{L}_R = \begin{matrix} (\ u \ )_R \\ (\ d \ )_R \end{matrix} \tag{1.4}$$

with $[u, d] = ([\mathrm{u}, \mathrm{d}], [\mathrm{c}, \mathrm{s}], [\mathrm{t}, \mathrm{b}])$.

Weak interactions are described via weak charged currents ("V − A" theory); we can introduce two new quantum numbers to explicitly represent the helicity states within $\mathrm{SU(2)_L} \otimes \mathrm{U(1)_Y}$ gauge theory. They are the *Weak Isospin* T (with third axis projection T_3) and the *Weak Hypercharge* Y, defined by Eq. 1.5:

$$Y = 2(Q_{em} - T_3) \tag{1.5}$$

where Q_{em} is the electric charge in units of positron charge [1] e. In Tab. 1.1 we summarize the values of these quantum numbers. As we can see from Tab. 1.1, if right-handed neutrinos ν_R do exist they would have all electroweak quantum numbers equal to 0, and thus would not directly interact (Eq. 1.8-1.12) with other particles.

---

[1] $e = 1.602\ 176\ 53(14) \times 10^{-19}$ C [1] is the absolute value of the electric charge of the electron, whose charge value is -e.

Fermions	Quantum Numbers			
	Q_{em}	T	T_3	Y
$(\nu_\ell)_L$	0	$\frac{1}{2}$	$+\frac{1}{2}$	-1
$(\ell)_L$	-1	$\frac{1}{2}$	$-\frac{1}{2}$	-1
$(\nu_\ell)_R$	0	0	0	0
$(\ell)_R$	-1	0	0	-2
$(u)_L$	$+\frac{2}{3}$	$\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{3}$
$(d)_L$	$-\frac{1}{3}$	$\frac{1}{2}$	$-\frac{1}{2}$	$+\frac{1}{3}$
$(u)_R$	$+\frac{2}{3}$	0	0	$+\frac{4}{3}$
$(d)_R$	$-\frac{1}{3}$	0	0	$-\frac{2}{3}$

Table 1.1: Quantum numbers of lepton (ℓ=e, μ, τ) and quark ($[u, d]$=[u, d], [c, s], [t, b]) helicity states: electric charge Q_{em} in unit of e, weak isospin T with third axis projection T_3 and weak hypercharge Y.

In terms of the weak isospin and hypercharge quantum numbers, lepton states may then be denoted as

$$\mathrm{L}_L \in (2, -1) \tag{1.6}$$

$$\mathrm{L}_R \in (1, -2) \tag{1.7}$$

where the first number points out the dimension of the representation of the SU(2) group and the second number the value of the hypercharge, connected with the U(1) group representation.

1.3 Electroweak sector

In absence of gauge interactions, the electroweak lagrangian $\mathcal{L}_{EW}^{lepton}$ is

$$\mathcal{L}_{EW}^{lepton} = i \sum_{i=1}^{3} \left(\mathrm{L}_L^{\dagger i} \overline{\sigma}^\mu \partial_\mu \mathrm{L}_L^i + \mathrm{L}_R^{\dagger i} \sigma^\mu \partial_\mu \mathrm{L}_R^i \right) \tag{1.8}$$

with $\sigma^\mu = (1, \vec{\sigma})$ and $\overline{\sigma}^\mu = (1, -\vec{\sigma})$; $\vec{\sigma} = (\sigma^1, \sigma^2, \sigma^3)$ are the Pauli matrices. The index i runs over the three lepton families.

The local $\mathrm{SU(2)_L} \otimes \mathrm{U(1)_Y}$ gauge invariance is obtained by replacing the partial derivative with the covariant derivative

$$\mathcal{D}_\mu = \partial_\mu + i\frac{g'}{2}(Y)Y_\mu + ig\frac{\vec{\tau}}{2} \cdot \vec{W}_\mu \tag{1.9}$$

being Y the hypercharge of the lepton field and $\vec{W}^\mu$ and Y^μ the vector boson fields associated to the $\mathrm{SU(2)_L} \otimes \mathrm{U(1)_Y}$ gauge symmetry with couplings g and g', and $\vec{\tau} = (\tau^1, \tau^2, \tau^3)$ are the generators of the SU(2) group. The kinetic terms of gauge fields

$$\mathcal{L}_{EW}^{boson} = -\frac{1}{4}\vec{W}_{\mu\nu} \cdot \vec{W}^{\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} \tag{1.10}$$

with

$$\vec{W}_{\mu\nu} = \partial_\mu \vec{W}_\nu - \partial_\nu \vec{W}_\mu + g\left(\vec{W}_\mu \times \vec{W}_\nu\right) \tag{1.11}$$

$$B_{\mu\nu} = \partial_\mu Y_\nu - \partial_\nu Y_\mu \tag{1.12}$$

are added to complete the invariant-by-construction $\mathrm{SU(2)_L} \otimes \mathrm{U(1)_Y}$ lagrangian.

Lagrangians in Eq. 1.8 and 1.10 with substitution 1.9 describe massless gauge bosons and fermions. Requiring massless gauge bosons is usual in gauge theories, while a mass term for fermion fields f

$$-m_f\overline{f}f = -m_f\left(\overline{f}_R f_L + \overline{f}_L f_R\right) \tag{1.13}$$

manifestly violates gauge invariance because f_L and f_R are members of a SU(2) doublet and singlet respectively.

1.4 Quark sector

The left-handed components of the quarks are arranged into $Y = +\frac{1}{3}$ isospin doublets

$$\mathrm{Q}_L = \begin{pmatrix} \mathrm{U}_L \\ \mathrm{D}_L \end{pmatrix} \in \left(2, +\tfrac{1}{3}\right) \tag{1.14}$$

and the right-handed into singlets

$$\begin{array}{l} \mathrm{U}_R \in \left(1, +\frac{4}{3}\right) \\ \mathrm{D}_R \in \left(1, -\frac{2}{3}\right) \end{array} \tag{1.15}$$

In analogy with the leptonic case, the quark electroweak lagrangian is

$$\mathcal{L}_{EW}^{quark} = i\sum_{i=1}^{3}\left(\mathrm{Q}_L^{\dagger i}\overline{\sigma}^\mu\mathcal{D}_\mu\mathrm{Q}_L^i + \mathrm{U}_R^{\dagger i}\sigma^\mu\mathcal{D}_\mu\mathrm{U}_R^i + \mathrm{D}_R^{\dagger i}\sigma^\mu\mathcal{D}_\mu\mathrm{D}_R^i\right) \tag{1.16}$$

with $\mathcal{D}^\mu$ from Eq. 1.9.

The Weinberg model is extended to a gauge field theory $\mathrm{SU(3)_c \otimes SU(2)_L \otimes U(1)_Y}$ including the strong interactions between hadrons, called quantum chromodynamics (QCD).

The QCD charge is represented by the colour *red*, *green* or *blue* that any quark carries and exchanges through eight different bi-coloured massless gluons g. The colour interaction is represented by

$$\mathcal{L}_{QCD} = -g_s \sum_{i,j,k} \left(\bar{q}_i^k \gamma^\mu \mathcal{T}_a^{ij} q_j^k \right) G_\mu^a - \frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu} \tag{1.17}$$

where the first term describes the interaction between the gluon fields and quarks, while the second term is the gluon kinetic term, in which

$$G_{\mu\nu}^a = \partial_\mu \mathcal{A}_\mu^a - \partial_\nu \mathcal{A}_\nu^a + g_s f_{abc} \mathcal{A}_\mu^b \mathcal{A}_\nu^c \tag{1.18}$$

In Eq. 1.17 and 1.18 g_s is the QCD coupling constant and γ^μ the Dirac matrices, $\mathcal{T}_{ij}^a$ $(a = 1, ..., 8)$ are the $\mathrm{SU(3)_c}$ colour matrices and f_{abc} the colour structure constants; q_i^k are the Dirac spinors associated with the i-coloured k-type quark fields and $\mathcal{A}_\mu^a(x)$ are the eight Yang-Mills gluon fields. Unlike quantum electrodynamics (QED), in QCD the vector bosons carry colour charge and hence the interaction between two or more gluons is possible.

1.5 Spontaneous symmetry breaking

The spontaneous symmetry breaking of a local SU(2) gauge theory was at first introduced by P.W. Higgs [2, 3, 4, 5]. He theorized a new complex field with mass μ and two components (Higgs doublet) together with an appropriate potential. The interaction term between the Higgs field, expanded around its vacuum expectation value, and the SU(2) gauge fields is the responsible of mass terms of the three gauge field components.

The purpose of this procedure is to dress the weak vector bosons with mass and at the same time to keep the photon, carrier of electromagnetic interaction, massless. The symmetries to be spontaneously broken are $\mathrm{SU(2)_L}$ and $\mathrm{U(1)_Y}$, while the internal symmetry $\mathrm{U(1)_{em}}$ is not broken, since the requirement of the theory is a massless photon.

The choice of the suitable Higgs field was made in 1967 by S. Weinberg [6], who inserted a complex doublet $\phi \in (2, +1)$, as

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tag{1.19}$$

in the $\mathrm{SU(2)_L} \otimes \mathrm{U(1)_Y}$ invariant lagrangian

$$\mathcal{L}_{Higgs} = \left[\left(i\partial_\mu - g\frac{\vec{\tau}}{2}\cdot\vec{W}_\mu - \frac{g'}{2}(+1)B_\mu\right)\phi\right]^\dagger \cdot \left[\left(i\partial^\mu - g\frac{\vec{\tau}}{2}\cdot\vec{W}^\mu - \frac{g'}{2}(+1)B^\mu\right)\phi\right] - V(\phi) \tag{1.20}$$

The Higgs potential is quartic

$$V(\phi) = \mu^2\left(\phi^\dagger\phi\right) + \lambda\left(\phi^\dagger\phi\right)^2 = \mu^2|\phi|^2 + \lambda|\phi|^4 \tag{1.21}$$

with a complex mass term $\mu^2 < 0$ and a positive quartic coupling $\lambda > 0$. No higher order auto-interaction terms than the quartic one $|\phi|^4$ appear in the expression 1.21 of the Higgs potential to guarantee the renormalizability of the theory [7].

The suitable Higgs field vacuum expectation value for keeping the photon mass null is

$$\phi_0 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v \end{pmatrix} \tag{1.22}$$

because it has the right quantum numbers: $T = \frac{1}{2}$, $T_3 = -\frac{1}{2}$, $Y = +1$, especially $\frac{1}{2}Y + T_3 = Q_{em} = 0$ (see Eq. 1.5) to preserve $\mathrm{U(1)}_{em}$ invariance [2].

The mass terms in Eq. 1.20 can be written as

$$\mathcal{L}^{boson}_{mass} = \tfrac{1}{8}g^2v^2\left[\left(W^1_\mu\right)^2 + \left(W^2_\mu\right)^2\right] + \tag{1.23}$$

$$+\tfrac{1}{8}v^2\left(g'B_\mu - gW^3_\mu\right)\left(g'B^\mu - gW^{3\mu}\right) \tag{1.24}$$

where two terms are explicitly expressed.

The first term can be modified by the substitution

$$W^\pm = \frac{W^1 \mp iW^2}{\sqrt{2}} \tag{1.25}$$

and describes a complex field with mass $M_W = \frac{1}{2}gv$.

The second term of Eq. 1.24 is a rotation of the neutral components W^3_μ and B_μ ruled by the Weinberg angle[3] θ_W, defined as $\tan\theta_W = \frac{g'}{g}$:

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix}\begin{pmatrix} B_\mu \\ W^3_\mu \end{pmatrix} \tag{1.26}$$

---

[2] The generator of $\mathrm{U(1)}_{em}$ group is the electric charge $Q = Q_{em} = T_3 + \frac{Y}{2}$ (Eq. 1.5), therefore $Q\phi_0 = 0$ and the local invariance is guaranteed:
$\phi_0 \to \phi_0' = e^{i\alpha(x)Q}\phi_0 = \phi_0 \;\; \forall\alpha(x)$.

[3] The measured value of the weak mixing angle is $\sin^2\theta_W = 0.23120(15)$ [1].

The first field represents the photon field, which remains massless because it does not interact with the Higgs field, while the second is the Z boson associated to the mass term of Eq. 1.24.

In this framework, the masses[4] and couplings are related to g, g' and v:

$$M_W = \frac{1}{2} g v \tag{1.27}$$

$$\frac{M_W}{M_Z} = \cos\theta_W \tag{1.28}$$

$$M_A = 0 \tag{1.29}$$

and the elementary charge is

$$\frac{1}{e^2} = \frac{1}{g^2} + \frac{1}{g'^2} \tag{1.30}$$

The Higgs parameter v is related to the Fermi constant[5]:

$$v^2 = \frac{1}{\sqrt{2} G_F} \simeq (246 \text{ GeV})^2 \tag{1.31}$$

1.6 Fermion masses

As we have seen in Sec. 1.3, gauge invariance requires that no explicit mass terms for the fermion fields be present in the lagrangian. The introduction of the Higgs field however, allows the generatation of the masses of all the fermion fields without introducing symmetry violating terms. This can be done introducing a Yukawa interaction between the fermion and Higgs fields with coupling G_f:

$$\mathcal{L}^{fermion}_{mass} = -G_f[\overline{L}\phi R + \overline{R}\phi^\dagger L] \tag{1.32}$$

In the case of a lepton family, Eq. 1.32 becomes

$$\mathcal{L}^{lepton}_{mass} = -G_\ell[\overline{\mathrm{L}}_L \phi \mathrm{L}_R + \overline{\mathrm{L}}_R \phi^\dagger \mathrm{L}_L] \tag{1.33}$$

---

[4] Measured value of gauge boson masses and widths [1]:
M_W =80.425 ± 0.038 GeV/c^2, Γ_W =2.124 ± 0.041 GeV;
M_Z =91.1876 ± 0.0021 GeV/c^2, Γ_Z =2.4952 ± 0.0023 GeV;
$M_\gamma < 6 \times 10^{-17}$ eV/c^2 @ 95% CL.

[5] The Fermi constant, whose value is $\frac{G_F}{(\hbar c)^3}$=1.166 37(1)×10^{-5} GeV^{-2} [1], was introduced by Fermi to represent the coupling constant assigned to the weak interaction vertex in a theory with punctual interactions, not mediated by vector bosons. It is measured very accurately from the charged current muon decay $\mu^- \rightarrow \mathrm{e}^- \bar{\nu}_\mathrm{e} \nu_\mu$, that occurs through W emission.

Expanding the Higgs field $\phi(x)$ around its vacuum expectation value v (Eq. 1.22) and using the definitions of L_L (Eq. 1.1) and L_R (Eq. 1.2), Eq. 1.33 separates into two contributions:

$$\mathcal{L}^{lepton}_{mass} = -\frac{G_\ell v}{\sqrt{2}}\left(\bar{\ell}_L \ell_R + \bar{\ell}_R \ell_L\right) - \frac{G_\ell}{\sqrt{2}}\left(\bar{\ell}_L \ell_R + \bar{\ell}_R \ell_L\right) h(x) \tag{1.34}$$

with G_ℓ Yukawa coupling constant of the lepton family.

We recognize in Eq. 1.34 the mass term for a fermion of mass m_ℓ :

$$m_\ell = \frac{G_\ell v}{\sqrt{2}} \tag{1.35}$$

With Eq. 1.27, equation 1.34 finally becomes:

$$\mathcal{L}^{lepton}_{mass} = -m_\ell \bar{\ell}\ell - \frac{1}{2} g \frac{m_\ell}{M_W} \bar{\ell}\ell h(x) \tag{1.36}$$

Being G_e, G_μ and G_τ arbitrary parameters, the masses of the charged leptons e, μ or τ cannot be predicted by the theory and their measurements[6] only fix the values of coupling constants.

The second term of Eq. 1.36 represents the vertex of interaction between the Higgs boson and the $\ell^+\ell^-$ pair (or, more in general, any fermion–anti-fermion $f\bar{f}$ pair):

$$h\bar{f}f \longrightarrow -\frac{1}{2} i g \frac{m_f}{M_W} \tag{1.37}$$

The proportionality of the interaction strength to the fermion mass is a prediction of the theory; interactions with lighter fermions are disfavoured, since the vertex coupling depends on the ratio $\frac{m_f}{M_W}$.

1.7 Higgs Decays

As we have seen, in the Standard Model all couplings of the Higgs boson to other particles are given as a function of measured quantities, like fermion and gauge boson masses, and the unknown Higgs mass. Hence, the production cross section and partial decay widths can be calculated in terms of the unknown Higgs mass.

In Fig.1.1 the Standard Model Higgs boson H decay branching ratios are reported as a function of the Higgs boson mass $\mathrm{m_H}$ for a large range of values, from 50 GeV/c^2 to 1 TeV/c^2. All the curves are obtained with the program HDECAY [8], which includes also higher order radiative corrections.

---

[6] Measured value of charged lepton masses [1]: m_e=0.510 998 92(4) MeV/c^2, m_μ=105.658 369(9) MeV/c^2, m_τ=1776.99$^{+0.29}_{-0.26}$ MeV/c^2.

In the low mass region $m_{\mathrm{H}} < 150\ \mathrm{GeV}/c^2$, favoured by current observations (see next section), the dominant decay is into a $\mathrm{b}\bar{\mathrm{b}}$quark-antiquark pair. However, at hadron colliders, this decay will be hidden by the overwhelming non-resonant production of $\mathrm{b}\bar{\mathrm{b}}$ pairs, thus rendering it almost unobservable. Even though the cross section for associate production is two orders of magnitude smaller than that for direct production, as shown in Fig. 1.2, a better chance of detection lies with the associate production $\mathrm{t}\bar{\mathrm{t}}\mathrm{H}$, $\mathrm{H} \rightarrow \mathrm{b}\bar{\mathrm{b}}$. This requires a good discrimination of b jets from the H and t decays from the extremely more abundant light hadron jets produced in QCD interactions.

1.8 Limits on Higgs boson mass

The mass of the Higgs boson is a free parameter for the theory. However, if we assume that the Standard Model is correct in predicting its existence and properties, we can find both theoretical and experimental limits on the Higgs mass m_H:

- **direct observation** at present experiments, or rather the lack thereof, gives a lower limit for the Higgs boson mass;
- **experimental constraints** can be set on the Higgs boson mass to its influence to other well-measured physical quantities;
- **theoretical limits** arise from the the requirement that the Standard Model be valid up to an energy scale Λ, where new physics is supposed to lie.

While no proof of existence, indirect experimental limits can be used along with the theoretical limits as hints for the direct search of the Standard Model Higgs boson. On the other hand, direct observation outside these regions will be a proof of the existence of new physics beyond the Standard Model.

1.8.1 Theoretical limits

If we require the Standard Model to be self-consistent we can limit the allowed mass range of the Higgs boson. The tighter theoretical constraints come from one-loop matching conditions relating the particle couplings to their masses [9, 10]. The allowed upper and lower bounds are shown in Fig. 1.3 as a function of the cutoff parameter Λ at which the Standard Model is replaced by a higher energy theory. The region above the upper curve is forbidden because it would require the quartic coupling of Higgs potential to be divergent, or constantly zero, resulting in a non-interacting theory

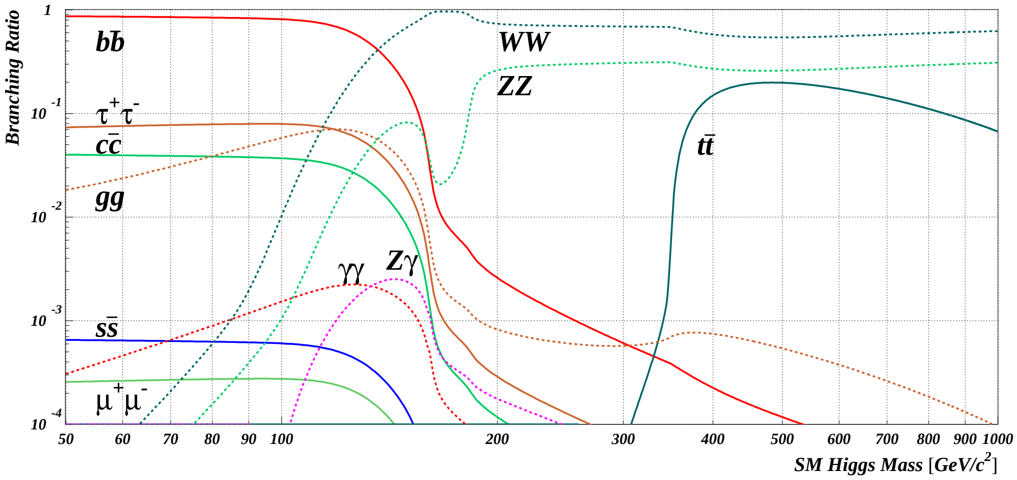


Figure 1.1: Decay branching ratios of Standard Model Higgs boson as a function of the mass. Decays into fermion–anti-fermion pairs are represented by solid lines, decays into gauge boson pairs by dashed lines.

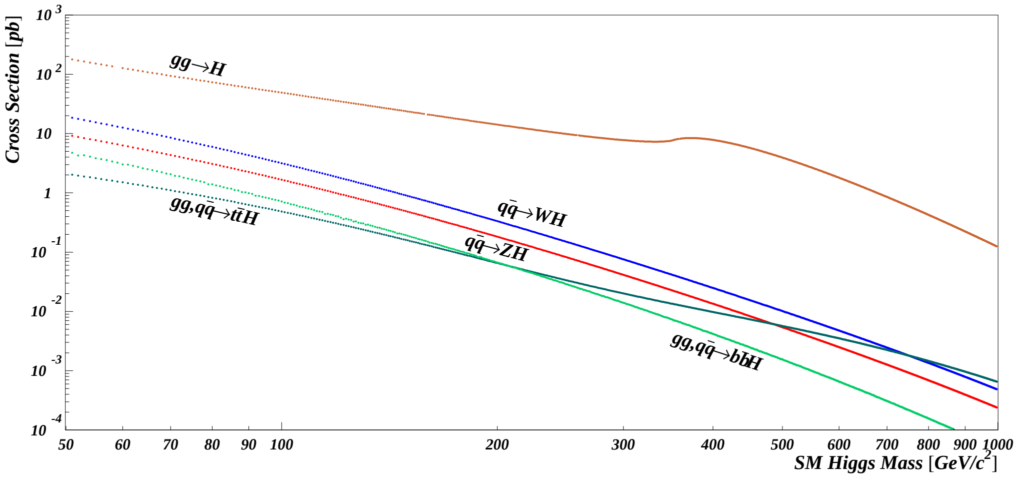


Figure 1.2: Cross sections for direct and associate production of the Higgs boson at the Large Hadron Collider.

(*triviality*). The region below the lower curve is not allowed because the quartic coupling becomes negative and the potential is unbounded from below (*vacuum stability*). The shaded areas reflect the theoretical uncertainties in the determination of the Higgs boson mass limits. If the validity of the Standard Model is assumed up to Planck scale $\Lambda \sim 10^{19}$ GeV, the allowed Higgs mass range is between 130 and 190 GeV/c^2.

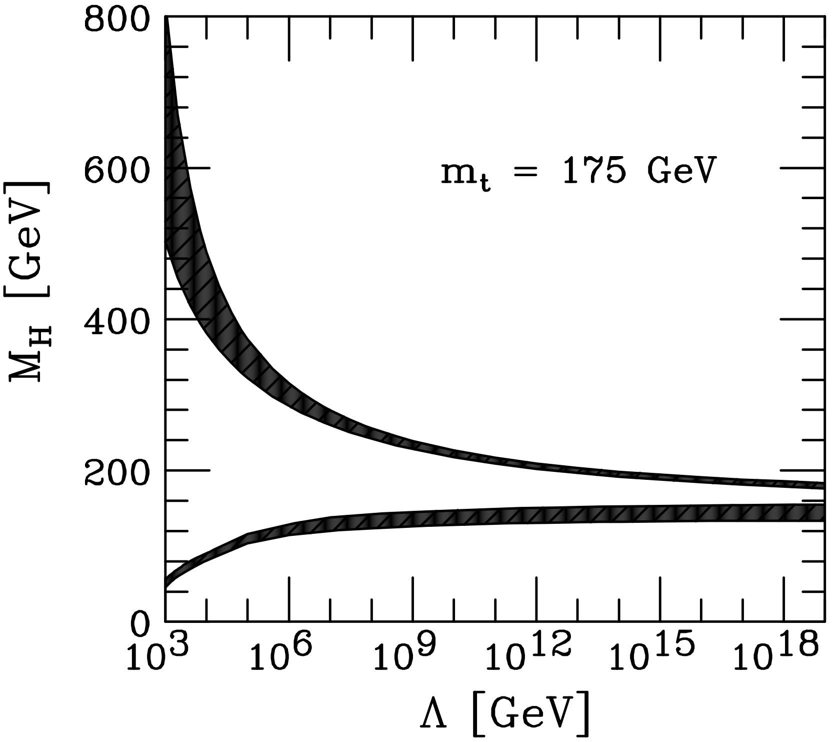


Figure 1.3: Theoretical limits on Standard Model Higgs boson mass. The allowed region, as a function of the energy scale Λ at which the Standard Model breaks down, is between the two curves, obtained assuming $m_{\mathrm{t}} = 175$ GeV/c^2 and $\alpha_s(m_{\mathrm{Z}}) = 0.118$ [10].

1.8.2 Experimental observation

The LEP-2 searches [11] showed no conclusive evidence for the existence of the Higgs boson.

The four experiments ALEPH [12, 13], DELPHI [14, 15, 16, 17], L3 [18, 19, 20] and OPAL [21, 22, 23, 24] collected 2 461 pb^{-1} of e$^+$e$^-$ collision data at centre-of-mass energies $\sqrt{s}$ between 189 and 209 GeV. At LEP [25, 26, 27], the Standard Model Higgs boson is expected to be produced mainly in association with a Z boson, through a process called *Higgsstrahlung*. Small contributions are from the t-channel WW and ZZ fusion processes with Higgs particle production together with a pair of neutrinos or electrons respectively.

According to Fig.1.1, the accessible Higgs particles ($m_{\mathrm{H}} \simeq \sqrt{s} - m_{\mathrm{Z}}$) decay predominantly into $\mathrm{b\bar{b}}$ pairs, hence the final-state topologies are determined by the decay properties of the associated Z boson. The searched final states were:

four-jet $\mathrm{e^+e^-} \rightarrow \mathrm{H}(\rightarrow \mathrm{b\bar{b}})\mathrm{Z}(\rightarrow \mathrm{q\bar{q}})$

missing energy $\mathrm{e^+e^-} \rightarrow \mathrm{H}(\rightarrow \mathrm{b\bar{b}})\mathrm{Z}(\rightarrow \nu_\ell \bar{\nu}_\ell)$ $(\ell = \mathrm{e}, \mu, \tau)$

leptonic $\mathrm{e^+e^-} \rightarrow \mathrm{H}(\rightarrow \mathrm{b\bar{b}})\mathrm{Z}(\rightarrow \ell^+\ell^-)$ $(\ell = \mathrm{e}, \mu)$

tauonic $\mathrm{e^+e^-} \rightarrow \mathrm{H}(\rightarrow \mathrm{b\bar{b}})\mathrm{Z}(\rightarrow \tau^+\tau^-)$ or $\mathrm{e^+e^-} \rightarrow \mathrm{H}(\rightarrow \tau^+\tau^-)\mathrm{Z}(\rightarrow \mathrm{q\bar{q}})$

Data from the four experiments were provided for all the channels and combined together to define a single variable sensitive to the signal-to-background ratio: the log-likelihood test statistics $-2\ln(Q)$ [11]. The measured value of the test statistics as a function of the test-mass m_{H} is shown in Fig.1.4 with the expected curves for the background only and signal+background hypotheses. A broad minimum on the observed curve extends from $m_{\mathrm{H}} = 115$ GeV/c^2 to 118 GeV/c^2 and intersects the expected signal+background curve at m_{H} close to 116 GeV/c^2, 1.74 standard deviations away from the background hypothesis. An optimistic interpretation of this observation is the signal of a Standard Model Higgs boson within this mass range. The signal-like behaviour mainly originates from the four-jet ALEPH data [28].

The data also give a 95% confidence level lower limit on the Higgs boson mass [11]

$$m_{\mathrm{H}} > 114.4 \ \mathrm{GeV}/c^2. \tag{1.38}$$

1.8.3 Indirect searches and Standard Model constraints

Standard Model electroweak observables are influenced by radiative corrections to their tree level expectation values, induced by higher order loop diagrams in which particles mass terms appear. Electroweak measurements have reached such a precision that they are sensitive to these corrections, thus allowing us through a global fit to electroweak data to compare measured quantities, such as the W and top masses m_{W} and m_{t}, with theorethical predictions, and estimate yet unmeasured parameters, i.e. the Higgs boson mass m_{H}.

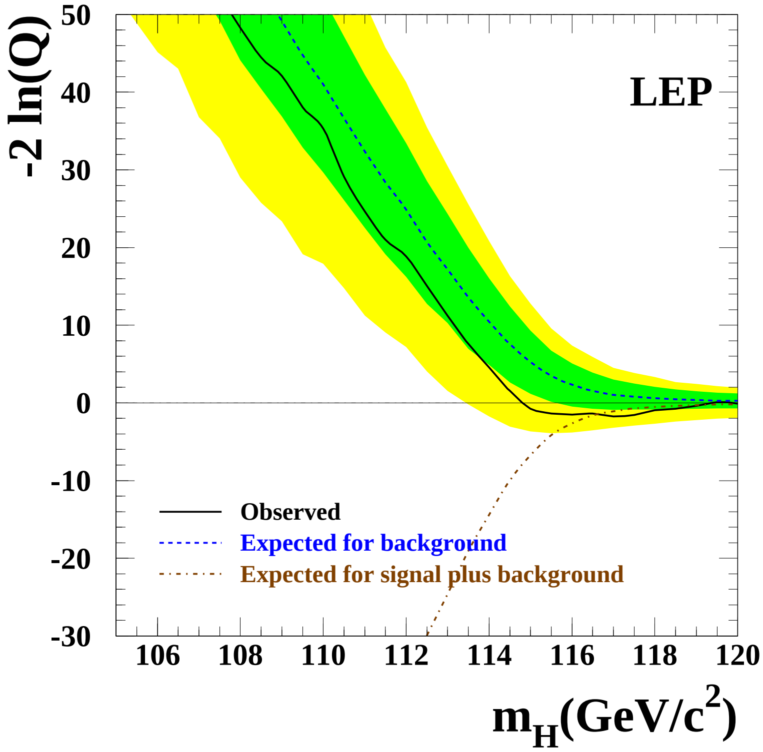


Figure 1.4: Observed and expected behaviour of test-statistics $-2\ln(Q)$ as a function of the test-mass m_{H}, obtained combining the data of the four LEP experiments. The solid line is the observed curve, the dashed (dot-dashed) is the median expectation in the hypothesis of background only (signal+background). The two shaded areas are the 68% and 95% probability bands around the median background expectation.

It is demonstrated that at one loop all electroweak parameters have at most a logarithmic dependence on $m_{\rm H}$ [29, 30, 31], as summarized by the *screening theorem* [32]: electroweak radiative corrections involving Higgs boson loops can be written in the form

$$g^2 \left(\log \frac{m_{\rm H}}{m_{\rm W}} + g^2 \frac{m_{\rm H}^2}{m_{\rm W}^2} \cdots \right) \tag{1.39}$$

with the quadratic term always screened by an additional power of g with respect to the logarithmic term, resulting in a mild contribution to the total correction. Corrections due to top quark loops are stronger, depending at leading order on $(m_{\rm t}/m_{\rm W})^2$; this gave the input for the discovery of the top quark at FermiLab [33].

Using the most accurate and recent measurements listed in [34, 35], several Standard Model fits are made with model predictions calculated with TOPAZ0 [36] and ZFITTER [37, 38, 39], and χ^2 minimization performed by the MINUIT [40] program. Using the most precise measurements of $m_{\rm t}$[7] and $m_{\rm W}$[8], electroweak results from high Q^2 experiments at LEP and Tevatron are fitted to obtain the best constraints on $m_{\rm H}$.

The results of the global fit for $m_{\rm H}$ are reported in Tab. 1.2 and Fig. 1.5.

$m_{\rm H}$	91^{+45}_{-32} GeV/c^2
$\log_{10}(m_{\rm H}/[\text{ GeV}/c^2])$	$1.96^{+0.18}_{-0.19}$
χ^2/d.o.f.	17.8/13 (17%)

Table 1.2: Results of the electroweak fit to all data summarized in [35]. Since the sensitivity to $m_{\rm H}$ is logarithmic, both $m_{\rm H}$ and $\log(m_{\rm H})$ are quoted.

In Fig.1.5 the observed value of $\Delta\chi^2 = \chi^2 - \chi^2_{min}$ as a function of $m_{\rm H}$ is shown. The solid curve corresponds to the result of Tab. 1.2 [35], the blue band represents the uncertainty due to theoretical uncertainties. The one-sided 95% confidence level upper limit on $m_{\rm H}$ from the electroweak fit is

$$m_{\rm H} \leq 186 \text{ GeV}/c^2 \tag{1.40}$$

---

[7] The latest results from the Tevatron CDF and D0 experiments yield a world average for $m_{\rm t} = 172.7 \pm 1.7 \pm 2.4$ GeV/c^2 [41]

[8] Combining Tevatron data with the latest presented results from LEP-2 experiments yield a world average of $m_{\rm W} = 80.410 \pm 0.032$ GeV/c^2, $Gamma_{\rm W} = 2.123 \pm 0.067$ [35]

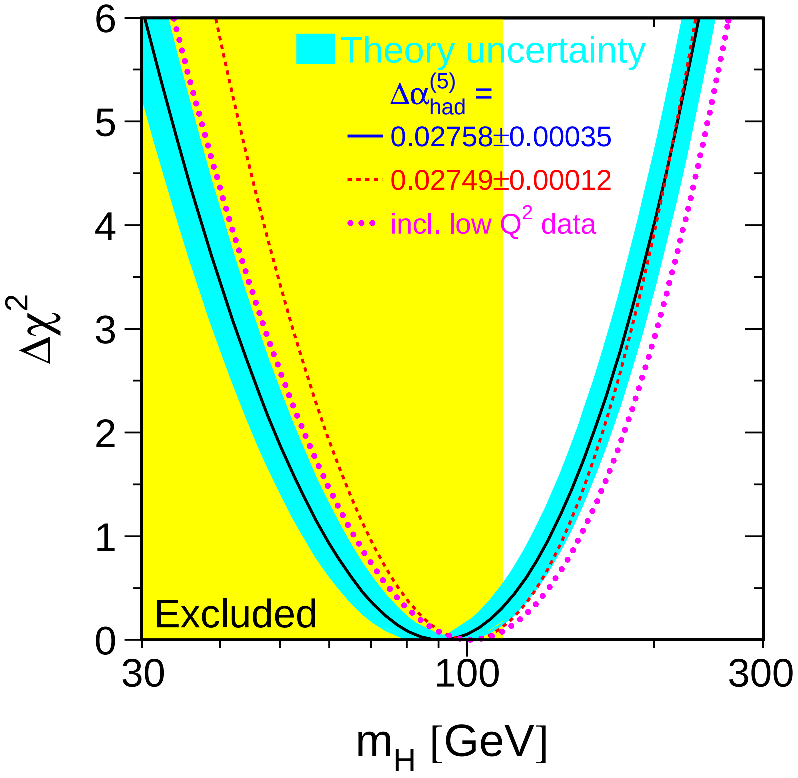


Figure 1.5: Observed value of $\Delta\chi^2 = \chi^2 - \chi^2_{min}$ as a function of Higgs mass $m_{\rm H}$. The line is the result of the electroweak fit using all data [35], the blue band represents the uncertainty due to theoretical uncertainties. The vertical yellow band is the region excluded by direct searches at LEP-2.

and raises to

$$m_{\rm H} \leq 219 \text{ GeV}/c^2 \tag{1.41}$$

taking into account the lower limit from direct search at LEP-2.

The limits 1.40 and 1.41 are valid assuming that the Standard Model is the correct theory of nature. Absence of observation of the Higgs boson within such limits would be a compelling evidence for new physics beyond the Standard Model.

Chapter 2

The CMS experiment

The new frontiers of particle physics are the searches for extremely elusive particles, which are produced in processes with very low cross sections, the femtobarn being the natural unit. With current technologies, leptonic colliders are not able to reach the high energies needed for these searches, which can presently be achieved only with hadron colliders. Despite the production of a lot of low energetic particles resulting in a not clean environment if compared to a leptonic interaction, a proton-proton collider offers the possibility to span over a wider energy spectrum that can be explored simultaneously and permits to reach higher production rates. These are the motivations of the CERN choice for the Large Hadron Collider (LHC) with the related experiments; two of them being multi-purpose experiments, ATLAS [42, 43, 44] and CMS [45, 46], the other two being dedicated experiments, one to heavy ion physics, ALICE [47], and the other to b quark physics and precision measurements of CP violation, LHCb [48, 49]. Fig. 2.1 shows a map of the LHC underground complexes, with the accelerator ring and experimetal areas.

2.1 The Large Hadron Collider

The Large Hadron Collider LHC [50] will be the most powerful hadron collider running in the next years. It is under construction in the already existing LEP [25, 26, 27] tunnel at CERN laboratories in Geneva, Switzerland. The Large Electron-Positron collider, which ceased to operate in the year 2000, was a circular e^+e^- accelerator, situated about 100 m depth underground across the French-Swiss border. The new accelerator LHC will produce collisions between proton beams with $\sqrt{s} = 14$ TeV, the highest value ever reached in accelerator experiments. The LHC tunnel is 26.659 km long, approximatly a circumference, composed with 8 curvilinear sections (2.840 km)

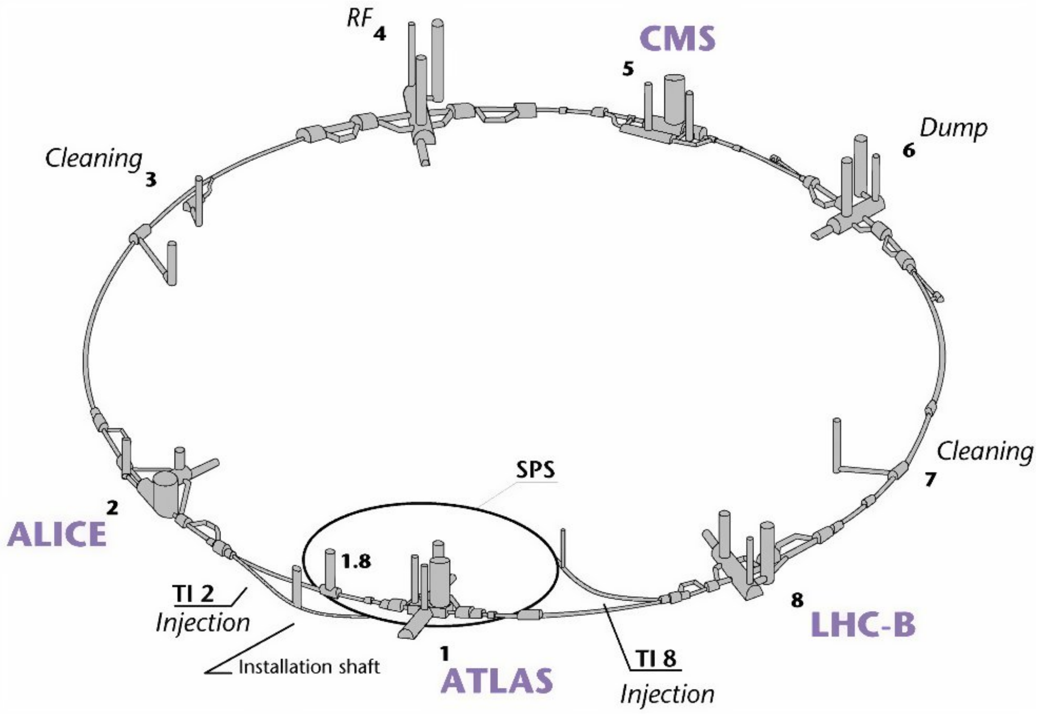


Figure 2.1: Map of LHC and related experiments.

and 8 rectilinear sections, where the beams may be brought to collide.

To reach such a high energy, proton beams will be initially accelerated by existing CERN facilities, upgraded for this purpose: protons will be accelerated and brought up to 50 MeV by a linear accelerator, their energies further raised up to 1.4 GeV by a Booster; the proton beams will be injected into the old Proton Synchrotron circular accelerator (PS), where they will be accelerated to an energy of 25 GeV; they will then be extracted from PS and injected into the bigger Super Proton Synchrotron (SPS), which will introduce 450 GeV proton beams into the LHC ring.

The accelerating power of LHC is limited by the bending magnetic field needed to keep the beams circulating in the tunnel, that is

$$p[\text{ GeV/c}] = 0.3 \cdot B[\text{T}] \cdot \rho[\text{m}] \tag{2.1}$$

with B magnetic field supplied to maintain p momentum particles in a circular orbit with radius ρ. The choice of 7 TeV beam energy is forced by the maximum achievable magnetic fields and depends on the curvature radius $\rho = 4.3$ km of the existing circular tunnel. As the collisions will occur between particles of the same kind, a unique magnetic field is required to

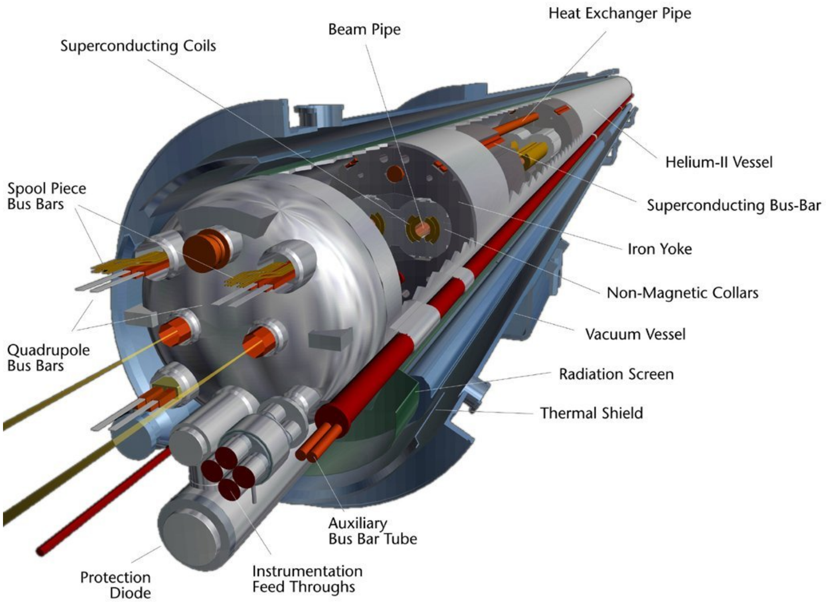


Figure 2.2: The 15 m long LHC cryodipole. The coil inner diameter is 55 mm.

accelerate the proton beams in opposite directions and the two beam pipes will be inserted into a single cryostatic structure with the superconducting magnets and the corresponding coils. A sketch of one of the LHC cryodipole is shown in Fig. 2.2.

In the curvilinear sections 1 232 main dipoles operating at 1.9 K and generating a magnetic field up to 8.33 T will be used to steer the particles into curvilinear trajectories together with 292 quadrupoles, 360 sextupoles and 336 octupoles for beam focusing and stability control. In the linear segments, 400 MHz superconducting radiofrequency cavities will provide the boost and supply energy losses with electric fields ranging from 3 $\mathrm{MV\,m^{-1}}$ at injection to 16 $\mathrm{MV\,m^{-1}}$ at full energy. LHC will also be able to accelerate heavy ions with total energy up to 2.76 A TeV in $\mathrm{Pb-Pb}$ collisions. The main design parameters of LHC are listed in Tab. 2.1.

An important parameter to characterize the performance of a collider is the luminosity $\mathcal{L}$, a quantity completely determined by the colliding beam properties. With a small crossing angle between the beams, composed with gaussian-shaped bunches, LHC luminosity $\mathcal{L}$ is expressed as

$$\mathcal{L} = F\frac{\upsilon n_b N_1 N_2}{4\pi\sigma_x\sigma_y} \tag{2.2}$$

Circumference	26.659 km
Maximum Dipole field	8.33 T
Magnet Temperature	1.9 K

	p − p	$^{82}_{208}$Pb-$^{82}_{208}$Pb
Beam energy at injection	450 GeV	37.4 TeV
Beam energy at collision	7 TeV	574 TeV(2.76 A TeV)
Nominal Luminosity	1×10^{34} cm^{-2}s^{-1}	1×10^{27} cm^{-2}s^{-1}
Number of Bunches	2 808	592
Bunch separation	24.95 ns	99.8 ns
Number of particles per bunch	1.15×10^{11}	7×10^{7}
Total crossing angle	285 μrad	2.8 μrad
Bunch Length (r.m.s.)	7.55 cm	7.94 cm
Transverse beam size at Impact Point	15 μm $\times$ 15 μm	15 μm $\times$ 15 μm¡M-F4¿
Luminosity lifetime	13.9 h	4.3 h
Filling time per ring	4.3 min	10 min
Energy loss per turn	7 keV	
Total radiated power per beam	3.8 kW	
Stored energy per beam	362 MJ	

Table 2.1: Technical parameters of LHC.

where v is the revolution frequency of the n_b bunches, $F = 0.9$ is a correction factor due to non-zero crossing angle (lower than 0.3 mrad), N_1 and N_2 number of protons in the two colliding bunches, σ_x and σ_y the beam profiles in horizontal (bend) and vertical directions at the interaction point. A realistic scenario foresees a three-year initial period at “low luminosity” $\mathcal{L} = 2\times10^{33}$ cm^{-2}s^{-1}, after which the planned “high luminosity” value of 10^{34} cm^{-2}s^{-1} for pp collisions will be gradually reached. To achieve this unprecedented value, the two beams will contain 2 808 closely-spaced bunches filled with an average of 1.15×10^{11} protons each and will collide every 25 ns. The designed luminosity for lead-lead cllisions is about 10^{27} cm^{-2}s^{-1}.

This performance can be obtained also because the two proton beams will be very collimated, with an estimated spread of the beam spot $\sigma_x \simeq \sigma_y \simeq 15$ μm in the plane perpendicular to the beam direction. The uncertainty on the collision point position along the beam axis is about 7.5 cm around the nominal point.

The number of interactions N_i corresponding to the process “i” with cross section σ_i can be written as

$$N_i = \int \sigma_i \mathcal{L} dt = \sigma_i L \tag{2.3}$$

where the integration is performed upon the running time of the machine with luminosity $\mathcal{L}$. The expression L is referred to as *Integral Luminosity* and it is measured in *inverse barn* (b^{-1}), 1 b=10^{-24} cm^2. An integrated luminosity of 20 fb^{-1} per year in the first three years at low luminosity for a total of 60 fb^{-1} should be collected. The second phase at high luminosity will last at least five years for a total amount of 500 fb^{-1} of data.

2.2 The CMS Experiment

The Compact Muon Solenoid experiment, CMS [46], is a general purpose detector which will operate at LHC. The main feature of CMS is the 4 T superconducting solenoid that permits a compact design of the detector with a strong magnetic field. The design priorities fulfilled by the CMS project [45] are a redundant muon system, a good electromagnetic calorimeter and a high quality tracking system.

CMS is built following the typical structure of a general purpose experiment designed for a collider: several cylindrical layers coaxial to the beam direction, referred to as *barrel* layers, closed at both ends by detector disks orthogonal to the beam pipe, the *endcaps*, to ensure good detector hermeticity. Fig. 2.3 shows a schematic view of CMS. Is is evident the cylindrical symmetry of the detector, which has a full length of 21.6 m, a diameter of 15 m and reaches a total weight of 12 500 t.

The natural coordinate frame used to describe the detector geometry is a right-handed cartesian system with the x axis pointing to the centre of LHC ring, the z axis coincident with the CMS cylinder axis and the y axis directed almost upwards[1] along the vertical. The cylindrical symmetry of CMS design and the invariant description of pp physics drive to use a pseudo-angular reference frame, given by the triplet (r, φ, η), with r distance from z axis, φ azimuthal coordinate with respect to x axis and pseudorapidity η defined as $\eta = -ln(tan(\theta/2))$, θ being the angle from the positive z semiaxis.

In this reference frame it is easy to describe the CMS subdetectors, installed radially from inside out as represented in detail in Figs.2.4 and 2.5:

– **Tracker** $\boldsymbol{r < 1.2}$ **m** $\boldsymbol{|\eta| < 2.5}$ Silicon pixel vertex detector plus 198 m^2 active area of Silicon microstrip detectors to reconstruct charged particle tracks and individuate primary and secondary vertices.

– **ECAL** **1.2 m**$\boldsymbol{< r < 1.8}$ **m** $\boldsymbol{|\eta| < 3}$ electromagnetic calorimeter to precisely measure electrons and photons, composed by PbWO_4 scintillating crystals and a forward preshower detector.

---

[1] Since the beams are 1.23% inclined with respect to a plane perpendicular to the direction of the gravity force vector, the y axis is not exactly parallel to the vertical.

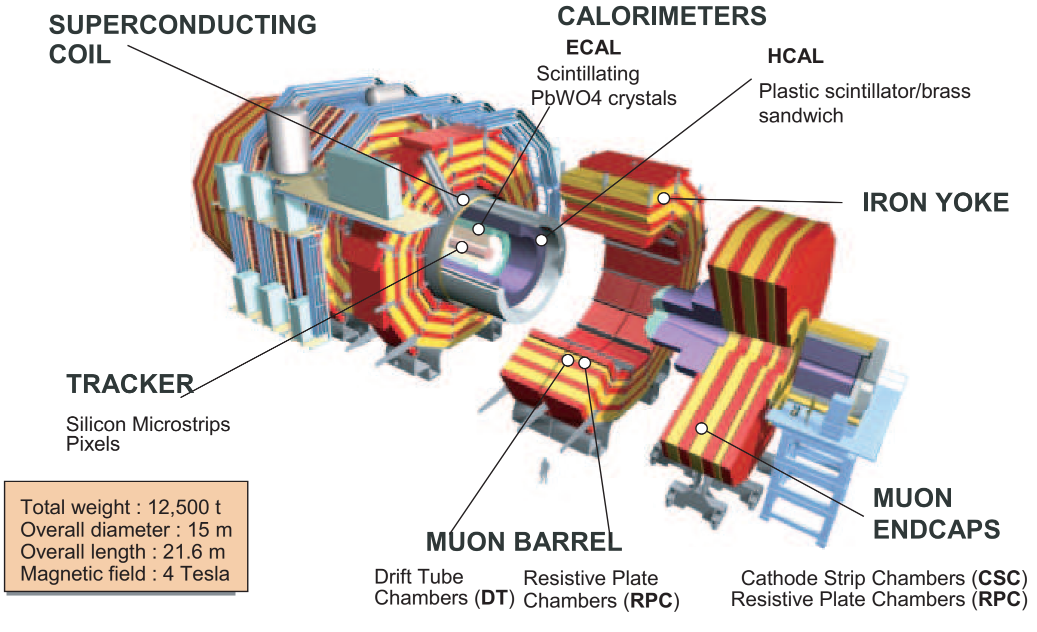


Figure 2.3: Schematic picture of CMS experiment at LHC.

– **HCAL 1.8 m$< r < 2.9$ m $|\eta| < 5$** hadron calorimeter system for jet position and transverse energy measurements, extended in the forward region $3 < |\eta| < 5$ with a very forward calorimeter (HF).

– **Magnet Coil 2.9 m$< r < 3.8$ m $|\eta| < 1.5$** the magnet, large enough to accommodate most of the calorimeters and the inner tracker, with a 4 T longitudinal magnetic field supplied by a superconducting solenoid.

– **Muon System 4.0 m$< r < 7.4$ m $|\eta| < 2.4$** muon chambers merged inside the magnet yoke to detect and reconstruct muon tracks, composed by Drift Tubes (DT) in the barrel and Cathode Strip Chambers (CSC) in the endcaps and complemented overall up to $|\eta| < 2.1$ by Resistive Plate Chambers (RPC).

2.2.1 Magnetic system

The CMS magnet [51] is a 13 m long superconducting solenoid with a diameter of 5.9 m. It provides an inner uniform 4 T magnetic field whose properties, summarized in Tab. 2.2, permit precise measurements of charged particles transverse momentum.

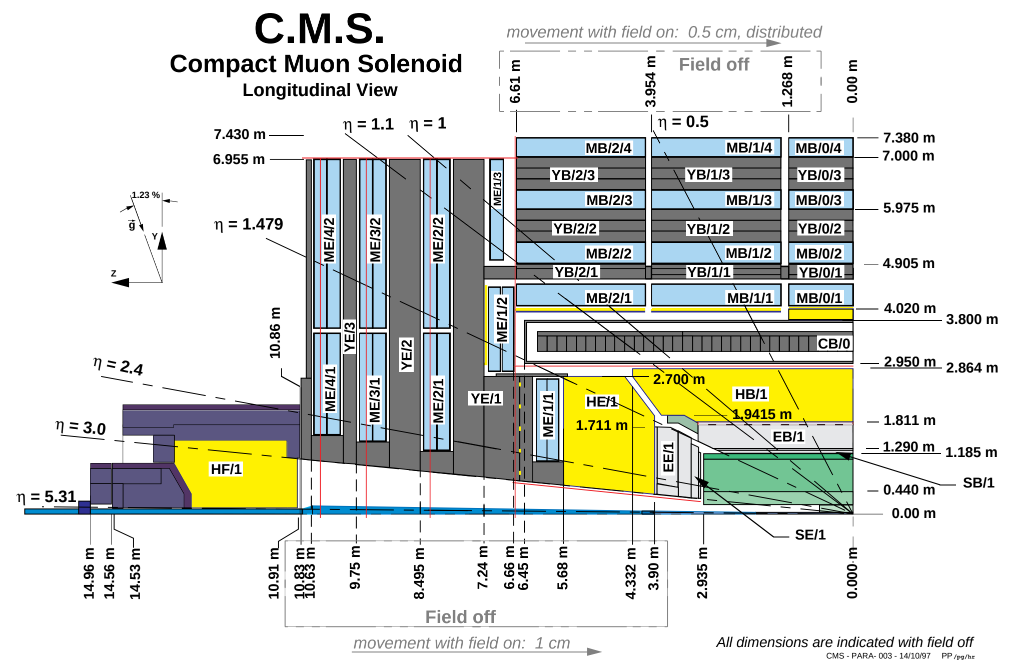


Figure 2.4: Longitudinal view of a quarter of CMS experiment. Detectors and non-sensitive volumes are indicated by two-letter code: the first letter indicates the subdetector (S=Silicon tracker, E=Electromagnetic calorimeter, H=Hadron calorimeter, C=magnet Coil, Y=magnet iron Yoke, M=Muon chambers), the second letter refers to the position (B=Barrel, E=Endcap, F=Forward region).

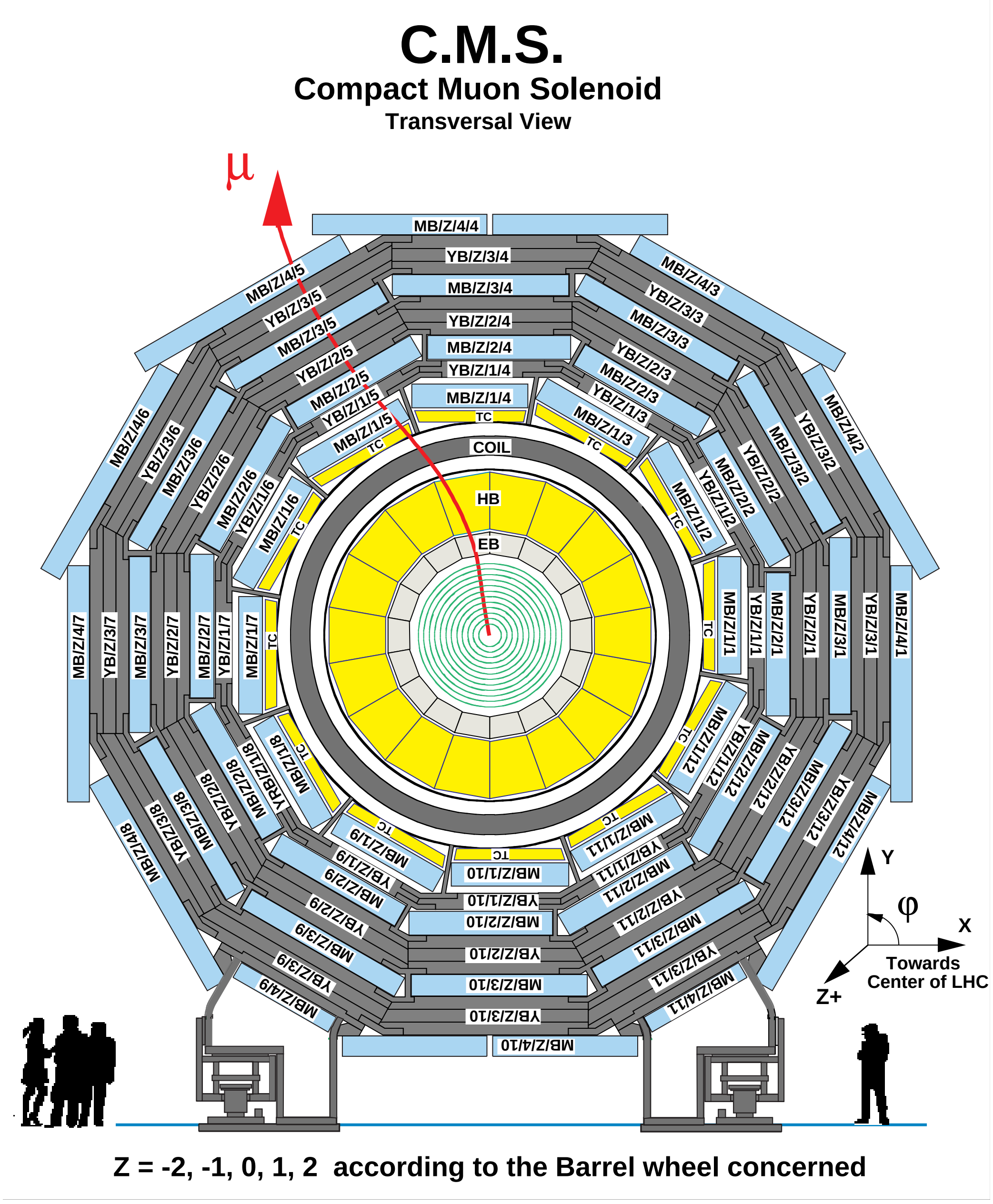


Figure 2.5: Transversal view of the barrel region of CMS. Barrel wheels are identified with the same two-letter code of Fig. 2.4 and numbered along z direction as $z = -2, -1, 0, +1, +2$.

Magnetic induction at impact point	4.0 T
Peak magnetic induction on the conductor	4.6 T
Coil length	12.48 m
Stored energy	2.70×10^9 J
Magnetomotive force	42.24×10^6 At
Magnetic radial pressure	6.47×10^6 Pa
Axial compressive force at mid plane	148×10^6 N

Table 2.2: Main parameters of the CMS magnet.

The conductor consists of three concentric parts: a central flat superconducting cable, a high purity aluminium stabilizer and an external aluminium-alloy to reinforce the sheath. The superconducting cable is a Rutherford type with 40 NiTb strands and is kept cooled by a liquid helium cryogenic system. The magnetic flux is closed in a loop via a 1.8 m thick saturated iron yoke, where the magnetic field is of 1.8 T.

The coil accommodates the tracking system and most of the calorimeters and it is also a supporting structure for the inner part of the apparatus, because it is the main element in term of size, weight and, mostly, structural rigidity. The yoke is instrumented with the two of the four muon chambers, the other two being just inside and outside it, to make full use of the return magnetic field.

2.2.2 Silicon tracker

The silicon tracker [52, 53] is the inner detector of CMS. It is the closest to the interaction point and represents an essential detector to address the multiplicity of LHC physics goals. It extends in the region $|\eta| < 2.4$, $r < 120$ cm, $|z| < 270$ cm and it is completely based on silicon detectors, covering a surface of 198 m^2, the largest ever designed Si detector. As the name suggest, its aim is to reconstruct the tracks and vertices of charged particles in the highly congested LHC environment. The key aspects to solve this pattern recognition problem are a low cell occupancy and a large hit redundancy. To achive this goals it is structured in three distinct layers: an inner silicon pixel detector, composed of 2 or 3 barrel layers and 2 endcap disks per side, and an outer silicon microstrip detector, with 4 layers and 3 disks made of 320 μm-thick sensors, and 6 layers and 9 discs of 500 μm-thick sensors in the outermost part. The low occupancy is obtained by working with high granularity detectors, especially the ones closer to the interaction point as they have to cope with higher particle fluxes, and fast primary charge collection, obtained using thin detectors and overdepleting the silicon bulks. The redundancy is guaranteed by the overall design of the tracker (see Fig. 2.6), which allows many measured points per track within an acceptable material budget, to minimize the adverse effect on the electromagnetic calorimeter performance.

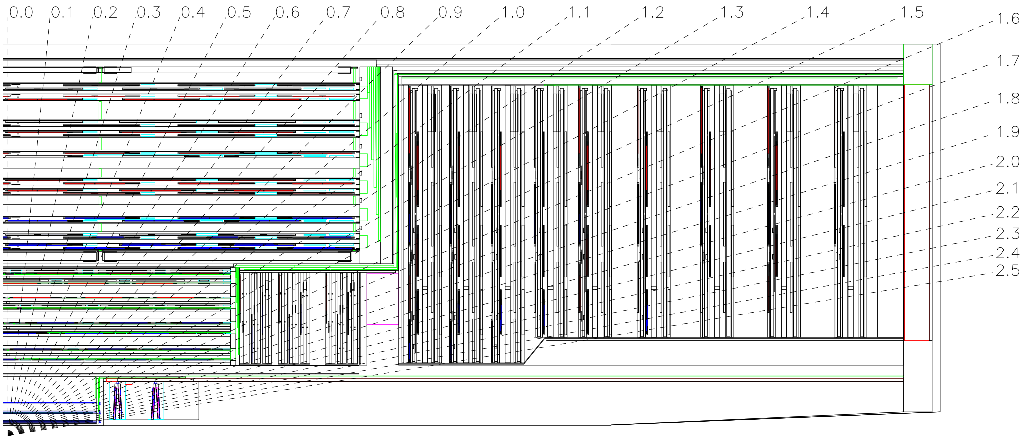


Figure 2.6: Schematic view of a quarter of the CMS silicon tracker comprehensive of the supporting structures, cables and services.

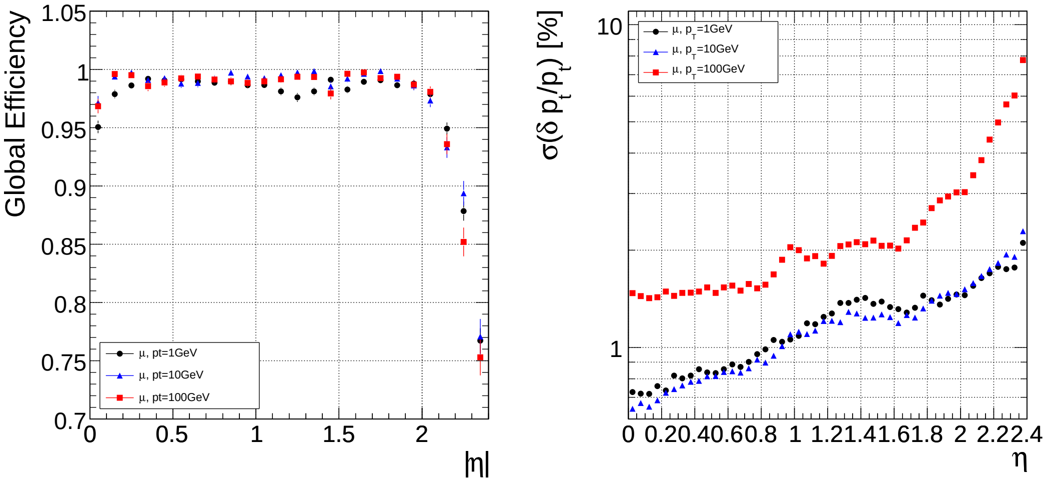
The overall tracker geometry guarantees an average of 12-14 points (*hits*) per track to permit a high tracking efficiency and a low fake rate (10^{-3} or less) of reconstructed tracks not corresponding to any real track.

A direct consequence of high particle density is the radiation damage of the silicon sensors, mainly around the collision area of the proton beams. Another source of radiation in the tracking volume is the high flux due to backscattering of neutrons evaporated from nuclear interactions inside the electromagnetic calorimeter.

To counter the effects of radiation damage, the silicon detectors will be kept cold, working at a temperature of -10^{o} C. This will reduce the adverse effects caused by the radiaton damages cause (increased leakage current, increased noise, worse charge collection efficiency), and freeze out the long term negative annealing. Only during limited maintenance periods the detectors will be “warmed” up to above 0^{o} C.

The physics requirements the CMS tracker has to satisfy are:

- Track reconstruction: Fig.2.7(a) shows the reconstruction efficiency for isolated tracks from the simulation of single muons with different trans-



(a) (b)

Figure 2.7: Reconstruction performance in the CMS tracker for single muons of different transverse momenta (1 GeV/c, 10 GeV/c, 100 GeV/c): (a) global track reconstruction efficiency; (b) transverse momentum resolution.

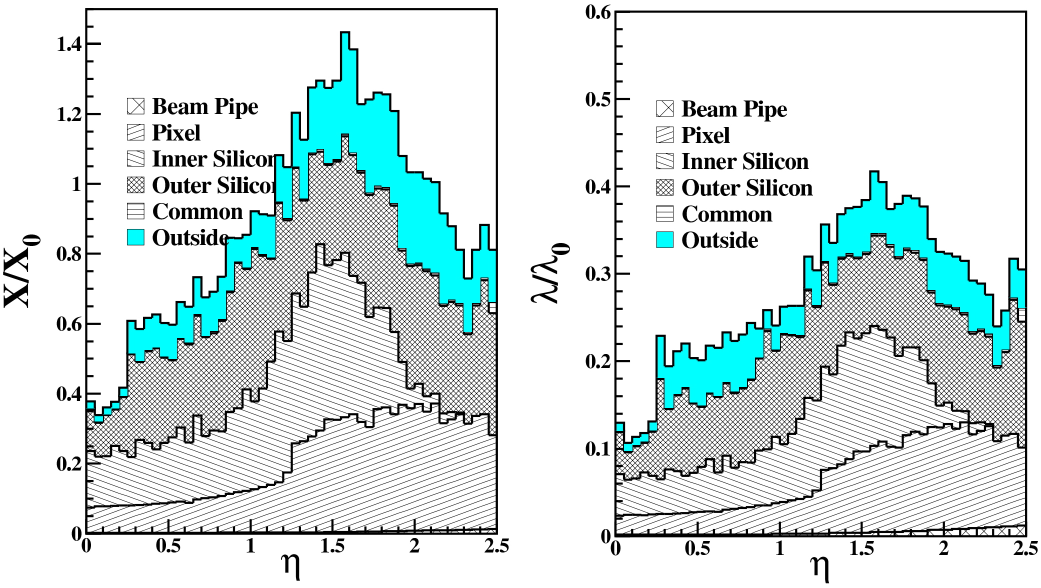


Figure 2.8: Material budget as a function of η for different tracker subunits: material thickness in units of radiation length X_0 (left) and in units of interaction length λ_0 (right).

verse momenta (1 GeV/c, 10 GeV/c and 100 GeV/c); the overall efficiency is expected to be better than 95% in $|\eta| < 2.0$.

- Good transverse momentum resolution: Fig.2.7(b) shows that, from the same simulation samples, the expected resolution $\frac{\sigma(p_{\mathrm{T}})}{p_{\mathrm{T}}}$ is better than 2% for low p_{T} tracks (1 GeV/c and 10 GeV/c) in the whole tracker region; for high p_{T} tracks it is around 2% in the central region ($|\eta| < 1.4$) and worsens up to 8% in the forward region.

- Tagging and reconstruction of b jets, fundamental requirement for new physics studies ($\mathrm{H} \rightarrow \mathrm{b\bar{b}}$), for top quark physics and CP violation measurements.

- Several material budget constraints are imposed on cables, passive materials and active layers, to minimize electron bremsstrahlung and hadronic interactions that would adversely affect tracking performance and the electromagnetic calorimeter energy measurements. In Fig.2.8 both the total radiation length[2] X and nuclear interaction length[3] λ for the tracker material as a function of pseudorapidity are reported. The material budget is higher in the transition region between barrel and endcap ($1 < |\eta| < 2$) due to cables and services that connect the tracker modules to the outside system.

2.2.3 Electromagnetic calorimeter

A high performance electromagnetic calorimeter is a fundamental requirement for any general purpose LHC experiment for precise measurements on electrons and photons. The design of CMS ECAL [54, 55] has been prompted by the possibility to observe the decay of a light Higgs boson into a couple of photons. Since in the region $m_{\mathrm{H}} < 140$ GeV/c^2 the intrinsic Higgs width Γ_{H} is less than 100 MeV, the $\gamma\gamma$ invariant mass resolution is dominated by experimental resolution, which should be of the order of 1% to enhance the significance of a possible signal.

The CMS collaboration has chosen a homogeneous calorimeter composed with finely segmented crystals of lead tungstate ($\mathrm{PbWO_4}$), which is a radiation resistant and chemically inert scintillator suited to work in the LHC high dose environment (from 0.18 Gy/h at $|\eta| = 0$ to 6.5 Gy/h at $|\eta| = 2.6$ at high luminosity). Moreover, the lead tungstate has also a short scintillation

---

[2] The radiation length X_0 is defined as the distance over which a high energy electron loses on average $1 - 1/e$ of its energy within a material.

[3] The nuclear interaction length λ_0 is the mean free path for a hadron before having a nuclear interaction inside a material.

decay time $\tau = 10$ ns that allows to collect 85% of the light in the 25 ns interval between two pp collisions. The small Molière radius[4] of 21.9 mm and radiation length $X_0 = 8.9$ mm permit the shower containment in a limited space resulting in a compact calorimeter design. Figure 2.9 shows a longitudinal view of a quarter of the electromagnetic calorimeter: it is organized in a barrel region $|\eta| < 1.48$ and a forward region to cover the pseudorapidity area up to $|\eta| < 3.0$.

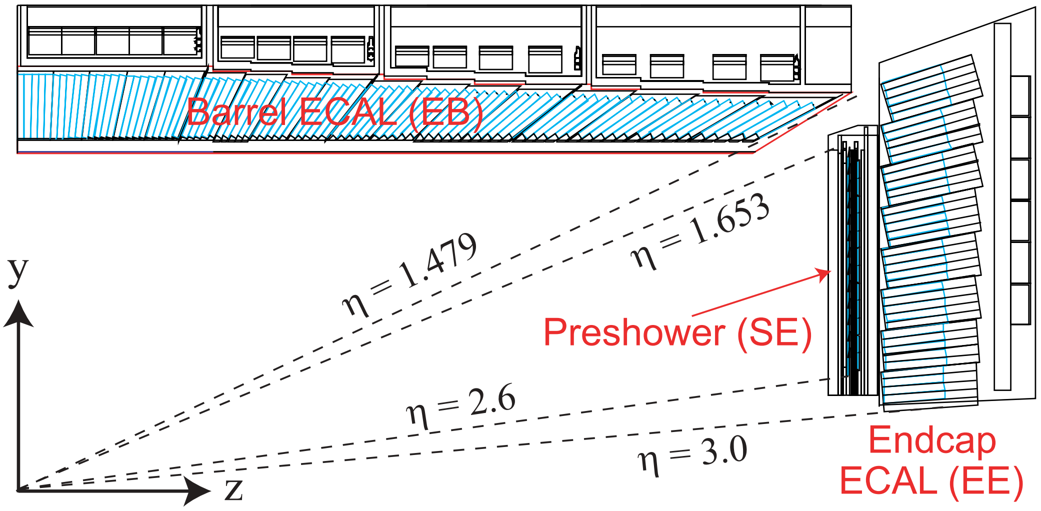


Figure 2.9: Longitudinal view of a quadrant of CMS electromagnetic calorimeter.

It is composed by 61 200 crystals in the barrel region and 21 528 in the endcaps grouped in 36 supermodules. The crystals have trapezoidal shape with squared front faces and are slightly different in the two regions: in the barrel they are 230 mm long with a total radiation length $X = 25.8X_0$ and 22×22 mm^2 front section, equal to the Molière radius. The granularity is $\Delta\eta \times \Delta\varphi = 0.0175 \times 0.0175$, high enough to achieve efficient π^0-γ separation. The collection of light is performed with silicon avalanche photodiodes (APD), which are able to operate inside a high magnetic field and can address the low light-yield of the crystals.

In the endcaps, the crystals have 24.7×24.7 mm^2 square front sections and smaller length (220 mm) and hence a smaller radiation length $X/X_0 = 24.7$, and an increasing granularity at higher $|\eta|$, with a maximum value of $\Delta\eta \times \Delta\varphi = 0.05 \times 0.05$ in the very forward crystals. The higher irradiation

---

[4] The Molière radius is the transversal dimension length scale of an electromagnetic shower evolving within a calorimeter.

levels would also induce too high leakage currents in APDs, therefore the forward crystals are read by vacuum photo-triodes (VPT).

The reduced radiation length and larger granularity is partially balanced by the presence in front of the endcaps, i.e. in the two regions $1.65 < |\eta| < 2.6$, of a preshower with $X = 3X_0$. Each preshower is composed with two lead radiators and two planes of silicon microstrips detectors to increase the π^0 rejection power in the highly irradiated forward regions.

In the range $25 < E[\text{ GeV}] < 500$, of particular interest for the H $\rightarrow \gamma\gamma$ decay, the electromagnetic energy resolution σ_E can be expressed as the squared sum of three independent terms:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{S}{\sqrt{E[\text{ GeV}]}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \tag{2.4}$$

where the first $\frac{S}{\sqrt{E}}$ is referred to as stochastic term and parameterizes the effects of fluctuations in photo-statistics and shower containment. The second term $\frac{N}{E}$ is due to electric noise and pile-up (the shaping time of the preamplifiers is chosen to be 40 ns) and C is a constant term. Fig. 2.10 shows the Test Beam measurements of the energy resolution of barrel crystals, along with fit results for the three parameters of Eq. 2.4.

At the beginning of data taking it is foreseen a staged ECAL without endcaps and only preshowers in the forward regions. This scenario is caused by the longer timescale for construction and crystal calibration, but it seems not to affect too much the resolution on di-jet invariant mass and transverse energy measurements.

2.2.4 Hadronic calorimeter

The hadron calorimeter is used together with the electromagnetic one to measure the energy and direction of jets, the transverse energy E_{T} and the imbalance of transverse energy, or missing transverse energy, E_{T}^{miss}. To fulfil these requirements, it has to be thick enough to contain the whole hadron shower and have high hermeticity. Since it is placed inside the magnet, it can not be made with ferromagnetic materials.

The CMS HCAL [56] is a sampling calorimeter with 3.7 mm thick active layers of plastic scintillators alternated with 5 cm thick brass plate absorbers. The signal is readout with wavelength-shift fibres. The granularity $\Delta\eta \times \Delta\varphi = 0.087 \times 0.087$ is fine enough to allow an efficient di-jet separation.

HCAL is subdivided, as it can be seen in Fig. 2.11, into barrel ($|\eta| < 1.4$) and endcap ($1.4 < |\eta| < 3.0$) with an overall thickness from 8.9 to 10

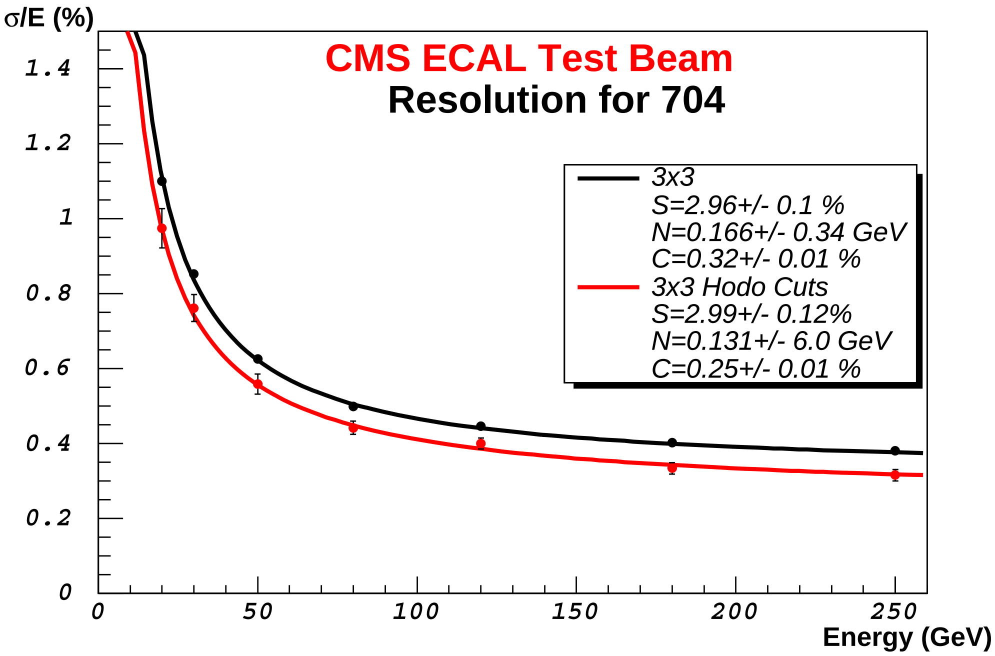


Figure 2.10: Energy resolution, σ_E/E, as a function of energy for crystals in the barrel region. The upper series of points correspond to events taken with a 20×20 mm^2 trigger and reconstructed using the containment correction. The lower series of points correspond to events selected to fall within a 4×4 mm^2 region. Different contributions to the energy resolution are also shown.

interaction lengths λ_0 respectively. Since the barrel part of the calorimeter is not sufficiently thick to contain all the energy of highly energetic showers, an additional "tail-catcher" of scintillators tiles outside the magnet is located.

To improve the pseudorapidity coverage from $|\eta| = 3$ to $|\eta| = 5$, a very forward calorimeter (HF) is placed outside the magnet yoke, ±11 m away along the beam direction from the nominal interaction point. It is a sampling calorimeter with active elements made of radiation hard quartz fibres parallel to the beam direction interleaved into steel plate absorbers. The active elements, whose granularity is $\Delta\eta \times \Delta\varphi = 0.17 \times 0.1745$, are sensitive to Čerenkov light and are readout with photomultiplier tubes. With this configuration the complex of CMS hadron calorimeter has an overall depth of more than $11\lambda_0$ over the full $|\eta| < 5$ coverage.

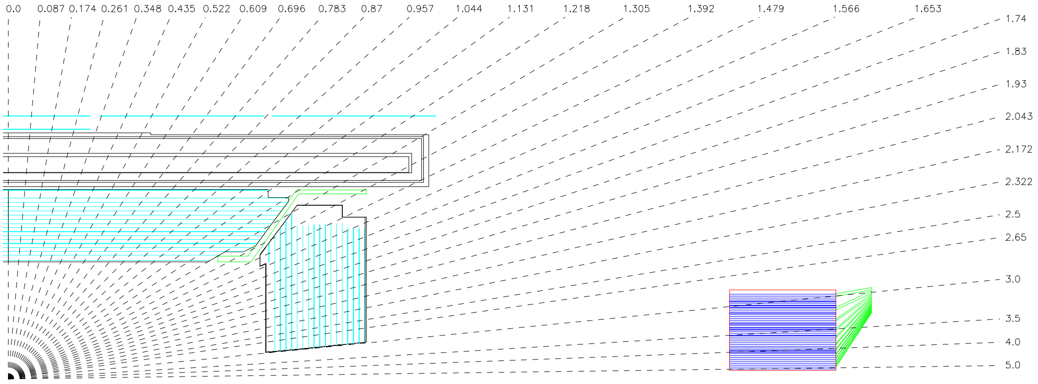


Figure 2.11: Longitudinal view of a quarter of CMS hadron calorimeter, subdivided into barrel and endcap HCAL, placed inside the magnetic coil, the outer barrel "tail-catcher" and the very forward calorimeter HF, sited outside the magnet.

The energy resolution of HCAL is

$$\frac{\sigma_E}{E} = \frac{(140.2 \pm 1.1)\%}{\sqrt{E[\text{ GeV}]}} \oplus (4.7 \pm 0.2)\% \tag{2.5}$$

and is expected to sensibly degrade around $|\eta| = 1.4$, where services and cables will be installed resulting in a higher amount of inactive material.

The performance of the very forward calorimeter

$$\begin{aligned} \left(\tfrac{\sigma_E}{E}\right)^{had} &= \tfrac{182\%}{\sqrt{E[\text{ GeV}]}} \oplus 9\% \\ \left(\tfrac{\sigma_E}{E}\right)^{e} &= \tfrac{138\%}{\sqrt{E[\text{ GeV}]}} \oplus 5\% \end{aligned} \tag{2.6}$$

is sufficient to improve the missing transverse energy resolution to the desired level, both for hadrons and electrons.

2.2.5 The Muon System

The muon system [57] is placed outside the magnet, embedded in the iron return yoke to make the full use of the 1.8 T magnetic return flux. It plays an essential role in the CMS trigger system, because high p_{T} muons are clear signatures of many physics processes. The main goal of this system is to identify muons and measure, when combined with the tracker, their transverse momentum p_{T}.

It is organized into three independent subsystems shown in Fig. 2.12: in the barrel, where the track occupancy is relatively low (< 10 Hz/cm^2), drift tubes (DT) detectors are installed, while in the endcaps cathode strip chambers (CSC) are favoured to work with higher particle rates (> 100 Hz/cm^2) and a larger residual magnetic field within the yoke plates. These two subsystems cover the $|\eta| < 2.4$ region and are arranged in a multi-layer structure to efficiently reject single hits produced by low range particles. In the region $|\eta| < 2.1$ redundancy is provided by resistive plate chambers (RPC), which have a limited spatial resolution, but a faster response and excellent time resolution, less than 3 ns. They are used mainly to unambiguously identify the bunch crossing and to complement the DT+CSC measurement of p_{T} during the trigger period, because RPCs can be finely segmented since they do not demand a costly readout system.

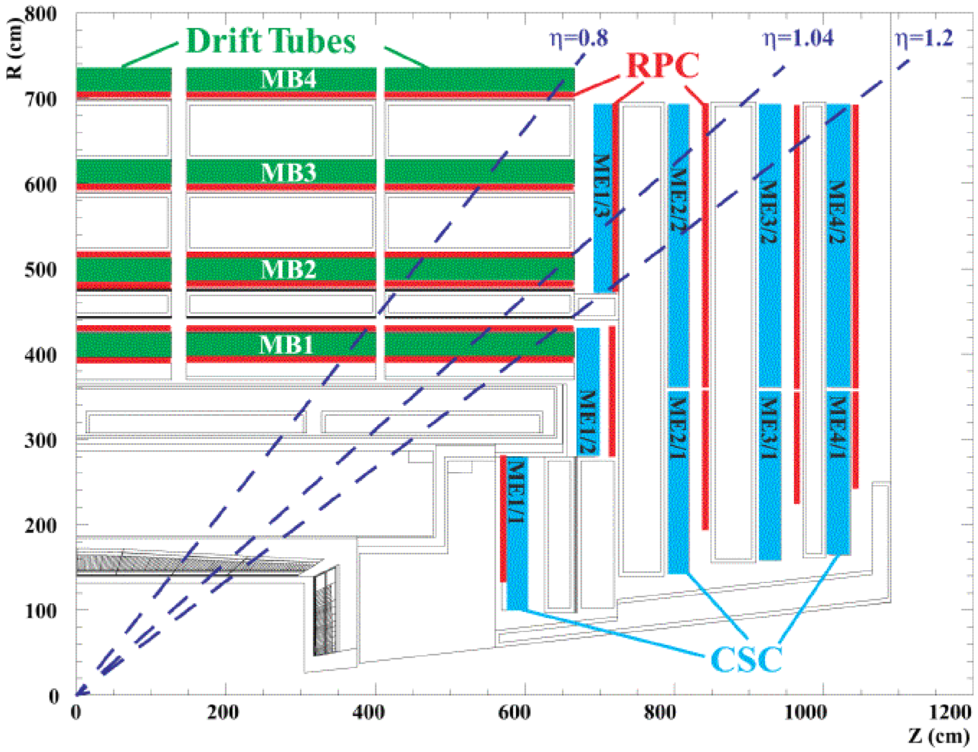


Figure 2.12: Longitudinal view of a quarter of the muon system, subdivided into barrel with drift tubes (DT) and resistive plate chambers (RPC) and endcap with cathode strip chambers (CSC) and RPCs.

Drift Tubes are composed with parallel aluminium plates insulated from perpendicular “I” shaped aluminium cathodes by polycarbonate plastic profile. The anodes are 50 μm diameter stainless steel wires placed between the “I” cathodes. The internal volume is filled with a binary mixture of 80% Ar and 20% CO_2 at atmospheric pressure, because this gas is non-flammable and can be safely used in underground operations in large volumes, as required in CMS. The resolution is about 100 μm both in $r\varphi$ and rz views.

Cathode Strip Chambers are composed with arrays of anode wires between pair of cathode planes, segmented into strips perpendicular to the wires. Gaps are filled with a gas mixture of 30% Ar, 50% CO_2 and 20% CF_4. The interpolation of the signal of neighbouring strips allows a precise spatial measurement of the φ coordinate with 50 μmresolution.

Resistive Plate Chambers are made of planes of a phenolic resin (bakelite) with a bulk resistivity of $10^{10} \ldots 10^{11}\ \Omega\,\mathrm{cm}$, separated from aluminium strips by an insulating film. The gaps are filled with a non-flammable gas mixture of 94.5% freon ($C_2H_2F_4$) and 4.5% isobutane (i-C_4H_{10}), which operates in “avalanche” mode to sustain the high rates.

2.2.6 The Trigger system

At LHC the proton-proton collisions occur at a rate of 40 MHz. However, it is impossible to reconstruct the events at this rate, or even store all the information about every collision: economical and technical difficulties impose a limit of about 100 Hz to the acceptable rate of data reconstructed and permanently stored. Furthermore, due to their tiny cross sections the rate of interesting events is smaller by order of magnitudes than the total interaction rate, hence a trigger system is endowed with the twofold task to reject a factor 4×10^5 of the collisions and to select with high efficiency the interesting physics events within a short time.

The trigger system at CMS is subdivided into two layered entities, the Level 1 and the High Level Trigger: every trigger decision is taken in steps of increasing refinement using more of the available subdetector data.

The Level 1 (L1) trigger is performed by dedicated hardware to reduce to a minimum dead times and take a very fast accept/reject decision, to cut the data rate from the 40 MHz LHC interaction rate down to almost 100 kHz. The events that pass this selection are buffered and submitted to the High Level Trigger Trigger (HLT) system. This is implemented by dedicated software algorithms running on a computer farm, and represents the first step of physical analysis. Using a parallel processing scheme as much as 100 kHz of events can be processed, with an allowed processing time per event estimated up to 500 ms/ev running on a single CPU, and reduced to the data storage rate of 100 Hz.

Level 1 Trigger

The Level 1 trigger selection is based exclusively on calorimeter and muon chamber information, processed with coarse granularity by fast hardware logical circuits [58]. Its aim is to analyise every 40 MHz pp collision and reduce the data rate passed on to the HLT to 100 kHz.

This rate will actually be of only 50 kHz for the low luminosity LHC startup scenario, and will be raised to the designed 100 kHz at full LHC luminosity. Of the available bandwith, only one third, 16 kHz (33.5 kHz) at low (high) luminosity, is allocated to the Level 1 selections, while the rest is kept as a safety margin against miscalculations of the expected rates due to uncertainties in the simulation of physics processes or in the extrapolated values of cross sections. The selection thresholds will be adjusted during the CMS operation to fully exploit the available trigger bandwidth as these parameters are directly measured.

The two elements of the Level 1 trigger, calorimetry and muon system, work in parallel and analyse the data locally, combine the information and produce the output passed to the Data Acquisition system (DAQ). The calorimetric information is synthetized into calorimetric towers, clusters of adjacent signals, by individual Trigger Primitive Generator circuits for ECAL, HCAL and HF. The towers information is fed to the Regional Calorimeter Trigger, which combines them to reconstruct jets, leptons and photons. All these objects have a raw measurement of energy and position and are collected by the Global Calorimeter Trigger (GCT).

Muons are reconstructed independently by the two muon subsystems, RPC and DT+CSC. The reconstructed segments are combined together by the Global Muon Trigger (GMT), which also resolves ambiguities and removes fakes using a map of inactive calorimeter regions provided by the calorimetric trigger.

The information from GCT and the Level 1 muon candidates from GMT are passed to the Global Trigger (GT), where they are combined to provide a first estimate of the missing transverse energy E_{T}^{miss} and determine the regions where the HLT should focus on.

The Level 1 trigger tables, Tabs. 2.3 and 2.4, for 16 kHz and 33.5 kHz output at low and high luminosities are obtained [59] selecting events with one or more reconstructed physics object candidates (jets, leptons, E_{T}^{miss}) above a certain fixed threshold. The tables show the generator-level E_{T} or p_{T} cuts applied in order to achive the 95% efficiency for the reconstructed objects.

An amount of about 1 kHz is left for calibration and monitoring purposes with random triggers to obtain an uniform sample of Minimum Bias events.

Trigger	Threshold [GeV or GeV/c]	Expected Rate [kHz]
Inclusive isolated e/γ	29	3.3
ee/$\gamma\gamma$	17	1.3
Inclusive μ	14	2.7
$\mu\mu$	3	0.9
1 τ-jet	86	2.2
2 τ-jets	59	1.0
1 jet	177	1.0
3 jets OR 4 jets	86, 70	2.0
1 jet AND E_{T}^{miss}	88, 46	2.3
e AND jet	21, 45	0.8
Minimum Bias (calibration)		0.9
Total		16.0

Table 2.3: Level 1 trigger table at low luminosity. Thresholds correspond to generator values giving a 95% efficiency for reconstructed objects [59]. The total rate is lower than the sum of each trigger rate due to events passing more than one criteria.

Trigger	Threshold [GeV or GeV/c]	Expected Rate [kHz]
Inclusive isolated e/γ	34	6.5
ee/$\gamma\gamma$	19	3.3
Inclusive μ	20	6.2
$\mu\mu$	5	1.7
1 τ-jet	101	5.3
2 τ-jets	67	3.6
1 jet	250	1.0
3 jets OR 4 jets	110, 95	2.0
1 jet AND E_{T}^{miss}	113, 70	4.5
e AND jet	25, 52	1.3
μ AND jet	15, 40	0.8
Minimum Bias (calibration)		1.0
Total		33.5

Table 2.4: Level 1 trigger table at high luminosity. Thresholds correspond to generator values giving a 95% efficiency for reconstructed objects [59]. The total rate is lower than the sum of each trigger rate due to events passing more than one criteria.

The trigger time available for the Level 1 decision making is much higher the 25 ns bunch crossing period, as signals are buffered at front-end level into pipelines able to hold hundreds of events. The more stringent limits are given by the amount of data from the silicon tracker and the preshower front-end buffers, which have to be stored during Level 1 processing and subsequently read for events passing the trigger. The decision time for the Level 1 logical circuits is thus limited to about 1 μs, excluding the unavoidable signal propagation delays.

The Level 1 accept signals and raw readout data are then buffered and sent to a computer farm for the HLT processing.

High Level Trigger Trigger

The High Level Trigger [59] selection (HLT) is implemented by analysis software running on a commercial computer farm.

The goal of HLT is to reduce the Level 1 output rate to 100 Hz storage rate with dedicated fast algorithms. The Level 1 measurements of jets, leptons and photons are refined through successive steps, using the full detector information buffered for each event passing the L1 trigger. In addition, the whole tracker information is available. Using the pixel hits together with zero-suppressed microstrip tracker signals, primary vertex reconstruction and track finding are possible, allowing to run online algorithms similar to the offline reconstruction analysis.

The 100 Hz output data to be stored on disk is subdivided into different topologies, listed for the initial period at low luminosity in Tab. 2.5. Some selections follow the Level 1 trigger scheme while others, as the inclusive b jet selection, are dedicated HLT algorithms making use of b tagging techniques and conditional track finding within the silicon tracker.

The estimated HLT selection efficiency for some interesting channels is listed in Tab. 2.6.

Trigger	Threshold [GeV or GeV/c]	Expected Rate [Hz]
Inclusive e	29	33
ee	17	1
Inclusive γ	80	4
$\gamma\gamma$	40, 25	5
Inclusive μ	19	25
$\mu\mu$	7	4
Inclusive τ-jets	86	3
2 τ-jets	59	1
1 jet AND E_{T}^{miss}	180, 123	5
1 jet OR 3 jet OR 4 jet	657, 247, 113	9
e AND jet	19, 52	1
Inclusive b jets	237	5
Calibration and other events (10%)		10
Total		105

Table 2.5: High Level Trigger trigger table at low luminosity. The thresholds correspond to the values of E_{T} or p_{T} with 95% efficiency (90% efficiency for muons) [59].

Channel	Efficiency
H(115 GeV/c^2)$\rightarrow\gamma\gamma$	77%
H(160 GeV/c^2)$\rightarrow$WW*	92%
H$\rightarrow$ZZ$\rightarrow 4\mu$	92%
A/H(200 GeV/c^2)$\rightarrow 2\tau$	45%
SUSY (0.5 TeV/c^2s-particles)	60%
$\mathrm{R_p}$-violation SUSY	20%
W$\rightarrow$eν_{e}	67%
W$\rightarrow\mu\nu_\mu$	69%
$\mathrm{t\bar{t}}\rightarrow\mu + X$	72%

Table 2.6: Efficiency of HLT selection at low luminosity after applying the cuts listed in Tab. 2.5.

Chapter 3

Silicon Strip Tracker

In this chapter I shall give a more detailed description of the CMS Silicon Strip Tracker from the structural point of view, with particular emphasys on the control system and power supply system.

3.1 Detector description

As pictorially shown in Fig. 3.1 the tracker is divided into four subdetectors: the barrel region is divided into Tracker Inner Barrel (TIB), comprising the four inner layers, and Tracker Outer Barrel (TOB), the six outermost layers; at each end the small three disks with $|z|$ between 70 cm and 110 cm, divided in three rings each, are the Tracker Inner Disks (TID), while the bigger nine disks in the $|z| > 120$ cm region, organized in four to seven rings, are the Tracker End-Caps (TEC). Fig. 3.2 shows the longitudinal view of one quarter of the CMS Silicon Strip Tracker (SST).

3.1.1 Modules

The fundamental active element of the SST is a module [60]. Each module is made up of a carbon fiber support structure, a front-end hybrid circuit, and one or two single-sided silicon sensors. Modules are grouped together in each subdetector for building and control purposes: strings of three modules in the TIB, rods of six modules in the TOB, rings in the TID and petals (each one an eighth of a disk) in the TEC.

A detailed description of the physics of silicon strip sensors can be found in [61].

The standard tracker modules are single-sided. Modules in the barrel regions are rectangular in shape, built with strips aligned along the beam di-

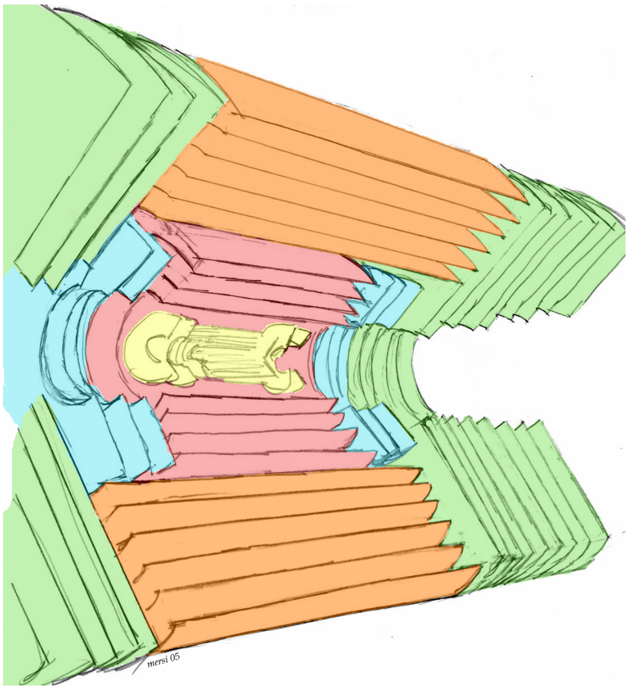


Figure 3.1: Pictorial view of the CMS Tracker. The different colors identify each subdetector: pink (TIB), cyan (TID), orange (TOB) and green (TEC). The pixel detector is yellow.

rection (z axis), and thus have a constant pitch. In the endcaps, on the other hand, modules are wedge-shaped with the strips along the radial direction; thus in each ring they have a constant angular pitch, while the linear pitch varies slightly along the radius.

The single-sided barrel and endcaps detectors provide respectivley only the $r\varphi$ and $z\varphi$ hit coordinates. In order to achieve full three-dimensional position measurements double-sided module are used, a back-to-back sandwich of a standard $r\varphi$ or $z\varphi$ module and a special *stereo* module, built with its sensors and thus the microstrip directions tilted by 100 mrad. Double-sided modules equip the two innermost layers of TIB and TOB and the corresponding endcap rings: the two innermost rings of TID and TEC ($r < 40$ cm) and the fifth TEC ring (60 cm $< r <$ 76 cm). This layout allows the SST to provide 8 to 14 measurements points for high momentum tracks with $|\eta| < 2.5$, about half of which are three-dimensional points.

All over the tracker the strip pitch varies from the inner to the outer layers (from 80 μm to 205 μm) [60]. The number of strips is tuned in order to match the electronics readout modularity of 256 channels: modules have

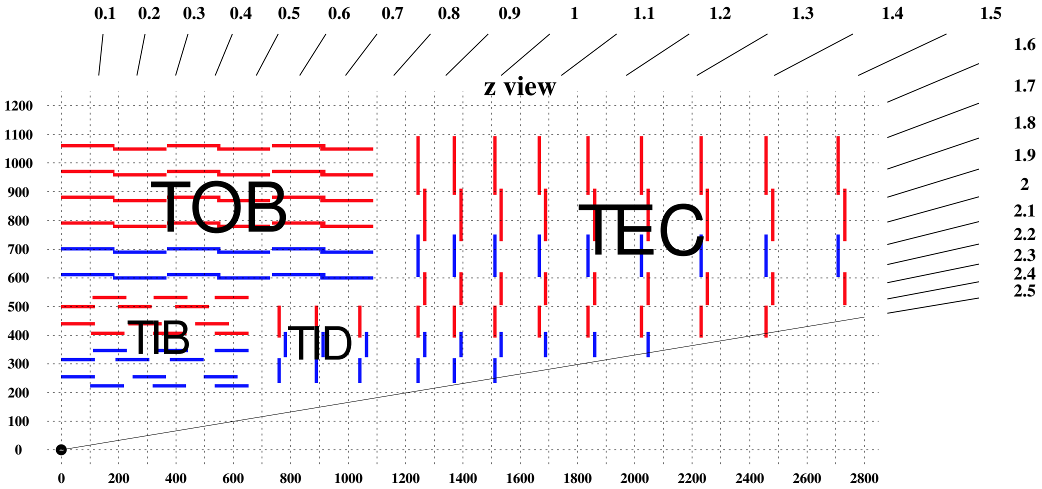


Figure 3.2: Schematic of one rz quadrant of the CMS Silicon Strip Tracker, showing the location of four subdetectors. The horizontal and vertical axes are in mm; the top axis shows the pseudorapidity. The red lines represent single-sided modules; the blue lines represent double-sided, back-to-back modules.

either 512 or 768 strips. The choice of the range for the strip pitches is also driven by two particle separation capability and by the achievable two-hit resolution, whereas the range of strip lengths, between 8.5 and 20.2 cm minimizes occupancy and noise levels.

The innermost layers (TIB, TID, and the first four TEC rings) are instrumented with 320 μm thick, low resistivity (1.5. . . 3.0 kΩ) sensors. The outermost layers (TOB and last three TEC rings) mount two daisy-chained higher resistivity (3.5. . . 8.0 kΩ), 500 μm thick sensors per module. The lower occupancy in the outer layers allows the use of longer strips (183 mm for TOB vs. 117 mm of TIB modules), thus reducing the total number of read-out channel needed. The increased noise due to the higher capacitance is balanced by the higher signal due to the thicker sensor bulk. In spite of the increased thickness, these sensors can be polarized at comparable voltage thanks to the use of higher resistivity silicon crystals, allowed by the lower radiation fluence in the outermost region.

Fig. 3.3 pictures a single-sided inner barrel module, like those mounted in TIB layers 3 and 4. The silicon sensor is connected to the front-end electronics (right) via a *pitch adapter*, made of metal lines deposited on a glass sheet. This allows a single layout of the frond-end for all the sensor

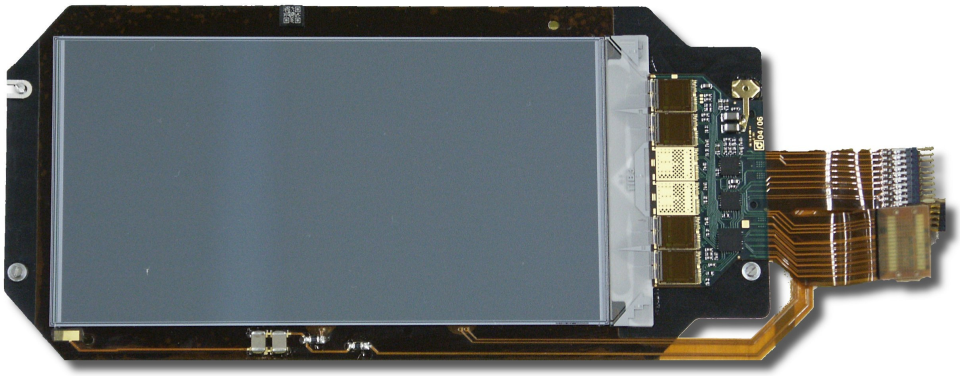


Figure 3.3: Picture of a single-sided TIB module. The greyish rectangole is the silicon sensor; on the right there are (left to right): the pitch adapter (white); the front-end hybrid with four APVs (brown) and two empty pads (golden); the kapton ribbons bringing power and control signals to the module (upper), analog readout data to the optohybrid connector (middle), and high voltage to bias the sensor (lower).

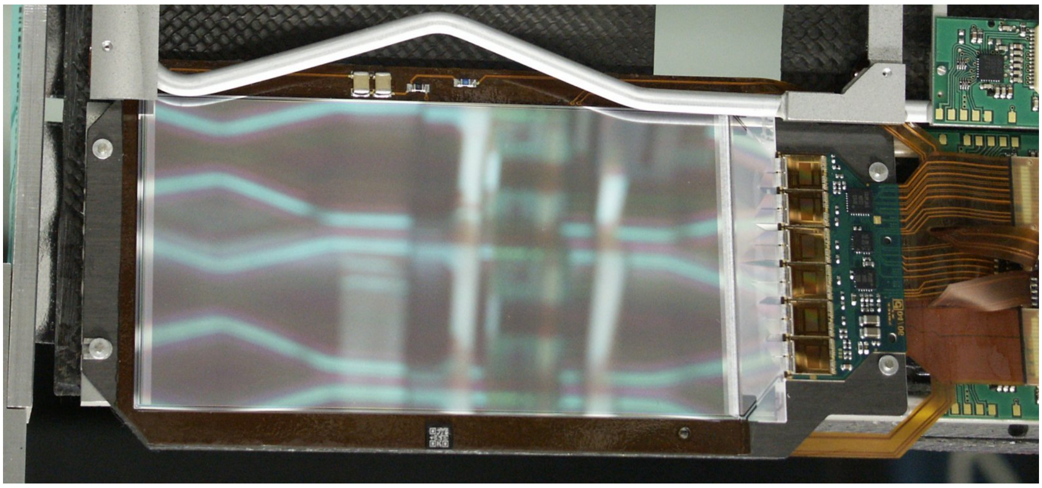


Figure 3.4: Picture of a double-sided TIB module. The $r\varphi$ face is shown, while the *stereo* part is evident only through the doubled number of kapton ribbons. All TIB double-sided modules have 768 strips, hence they the six APVs.

geometries, where only the pitch adapter needs to be changed. Double-sided modules are made of two single-sided ones, the *stereo* module having its sensors tilted by 100 mrad and glued on the back of the $r\varphi$ one. In Fig 3.4 only the $r\varphi$ module is visible, along with the extra kapton ribbons for the *stereo* module. Since TIB double-sided modules have 768 strips while single sided ones have 512, there are evident differences in the pitch adapter and readout electronics. However, its overall layout is the same.

3.1.2 Readout chain

The electronics responsible for the amplification and shaping of the analog signals is the APV25 chip [62, 63], a custom designed ASIC built with radiation hard technology and operating at the LHC frequency of 40 MHz. Each APV25 has 128 analog lines consisting of a preamplifier, an analog pipeline and a shaper. The pipeline allows holding each signal for up to 4.8 μs waiting for a trigger. The APV25 has two main modes of operation: peak mode and deconvolution mode. The first simply uses the signal at the time marked by the trigger; as the signal has a charateristic time of 50 ns, twice than the LHC interaction time of 25 ns, this leads to noticeable pile-up effects. In deconvolution mode the signal is sampled at three successive times, with an interval of 25 ns among them, and a linear combination of the three values is computed. The value and sign of the three coefficient are chosen to suppress the pile-up, and form a signal with a characteristic time of 25 ns. This has the adverse effect of a smaller signal to noise ratio, but allows operation at the 40 MHz LHC interaction rate.

In both modes of operation, upon receiveig a trigger all the 128 signals are multiplexed at 20 MHz into a single analog output; the signals from each pair of APVs is then again multiplexed into a single 40 MHz analog line. Thus, a module with 512 strips has four APVs resulting in two output lines, while a module with 768 strips has six APVs and three output lines. Double-sided modules obviously have double as much readout channels.

These analog outputs are then fed to the Analog OptoHybrids (AOH), which convert them into optical signals and transmit them over optical fibers out of the CMS detector for digitization.

The last step of the signal acquisition is performed by the Front End Drivers (FED). These are organized in VME devices called FED-9U, housed in the CMS Counting Room. Each FED-9U is able to read the signals of up to 192 APVs, digitize them, and perform a first analysis on the gathered data. First, each channel pedestal value and average detector signal (*common mode noise*) are subtracted; then clusters of adjacent strips with a signal incompatible with zero, given the detector noise, are found. Only

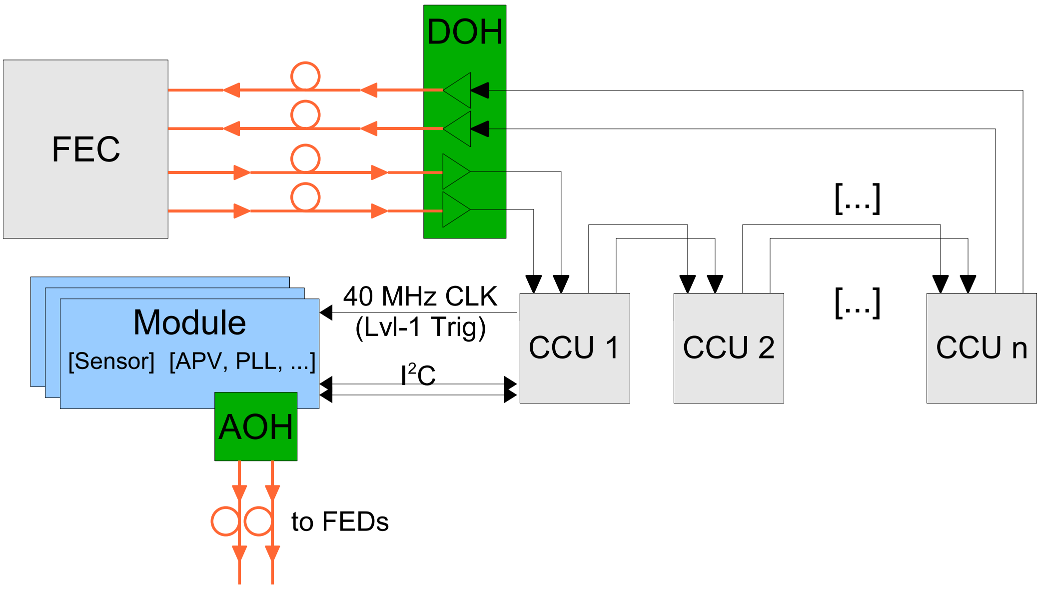


Figure 3.5: Schema of the control ring.

the information relative to these clusters (*zero suppression*) will be transmitted to the CMS Event Filter Farm along with all other detectors data, and eventually saved to disk for offline analysis.

3.1.3 Control electronics

The management of all the tracker electronic devices is performed over 352 *control rings* [64]. Fig. 3.5 shows the schema of one such control ring. This is formed of a Front End Controller (FEC) and many Communications and Control Units (CCU). The FEC is housed in the CMS Counting Room, while the CCUs are deployed over all the tracker. As for the analog readout, the transmission of the control signals to and from the CMS detector is performed over optical fibers; the Digital OptoHybris are responsible for the translation of those signals between optical and electrical form. The overall architecture of the control ring follows a token ring protocol, with the FEC as a master and the CCUs connected in a daisy-chain. Additional safety measures have been taken to insure proper redundancy and eliminate single point of failures.

Each CCU is part of the control ring, and is connected to a group of modules, following the mechanical divisions of each subdetector: one CCU is connected to a TIB string, TOB rod, or part of a TID ring or TEC petal. All the devices in such a group are controlled by the CCU via an I^{2}C bus [65], working at a clock speed of 100 kHz. The FEC thus distributes through each CCU the trigger and 40 MHz clock signals; additionally, it can address each component via its CCU I^2C bus.

3.2 Power supply system

An essential part of every detector is its power supply system. This is especially true for the CMS Silicon Strip Tracker, as a great amount of energy must be fed to the readout and control electronics across over 50 m keeping the induced noise as negligible as possible. The same is true for the bias voltage for the silicon sensors themselves.

During the first year of my PhD studies I worked on the development and testing of the power supply prototypes, their control software, and the overall architecture of the SST power supply scheme. I took part in two Test Beams as cabling and power supply expert for TIB in May 2003 and for TOB in the October of the same year, with the opportunity to test the control software I had developed in a realistic scenario [66].

3.2.1 Overview

Powering 12636 silicon strip modules and their control and readout electronics is no easy feat. Each module requires:

- one high voltage regulator (0-600 V) to bias the silicon sensor; even after 10 years at LHC the sensor's leakage current should be well below 1 mA.
- two low voltage regulators (1.25 V and 2.5 V) to power the front-end electronics; the typical current per module on the 2.5 V line ranges from $I_{2.5} \simeq 510$ mA with four APV25 chips to $I_{2.5} \simeq 750$ mA with six chips; the current on the 1.25 V line varies accordingly, from $I_{1.25} \simeq 260$ mA to $I_{1.25} \simeq 390$ mA.

The devised solution consists of grouping many contiguous and homogeneous detectors together, powering them with a single dedicated Power Supply Unit (PSU), able to supply both the low voltages required by the electronics and the high bias voltage. The control rings are powered with independent power supplies, delivering only the 2.5 V power line.

This solution imposes stringent requirements on the grouping criteria: to be able to use the same bias voltage for all detectors in a group whithout risking a breakdown or insuficient biasing, they must share their characteristics (thickness and resistivity) along all the experiment life; this implies

	APVs	$I_{2.5}$ [A]	$I_{1.25}$ [A]	Power [W]
Avg/group	37	5.2	2.4	16.0
Min/group	12	1.6	0.8	4.9
Max/group	56	7.6	3.6	23.7
Total	72784	10066	4731	31079

Table 3.1: Power consumption of the Silicon Strip Tracker, on the $I_{2.5}$ and $I_{1.25}$ line. Min/group and max/group are the minimum and maximum numbers of APVs in a single group, and the related power comsumption on each line.

that they must be subject to the same radiation dose. This requirement is somewhat alleviated providing two independent high voltage supply lines in each PSU. Other criteria are the maximum number of modules that can be powered by a single PSU, and the desirable fact that powering groups be subgroups of control rings; this effectively allows to power each control ring independently of the others, and resolves issues that might otherwise arise from different ground reference levels inside a single group.

Other problems, related to the length and type of power cables, connection to the modules inside the tracker, and different grounding schemes, have only recently been solved [67].

The adopted grouping configuration consist of 1944 power groups, with different numbers of modules and readout chips per group. Table 3.1 shows the average, minimum and maximum number of APV25 chips in a single power group, along with the measured current drawn and the resulting power consumption. The whole SST requires almost 15 kA, resulting in over 31 kW dissipated inside the detector volume.

Taking into account a safety margin of at leaft 50%, the power requirements imposed on the Power Supply units tested and developed are:

- the low voltage channels must be able to erogate up to 12 A on the 2.5 V line, and 6 A on the 1.25 V line;
- each high voltage channel must be able to erogate up to 10 mA and a voltage in the range 0. . . 600 V;
- both low voltage channels must be able to withstand up to 4 V voltage drop on the power cables.

In order not to negatively affect the tracker detector performance, additional requirements are imposed on the isolation and noise of each PSU:

- less than 10 mV peak-to-peak noise, up to a 20 MHz bandwidth, for the low voltage regulators;

- less than 30 mV peak-to-peak noise, up to a 20 MHz bandwidth, for the high voltage regulators;
- power supply lines electrically isolated from the surroundings, at least 100 Ω at 5 MHz, and better at lower frequencies.

The voltage drop on the power cables is especially noticeable on the low voltage lines, since these convey a high current (over 7 A) over long distances (over 50 m), thus requiring the use of sense cables. These are high impedence wires that bring back to the voltage regulators the actual drop on the loads, thus ensuring that the required values, namely 1.25 V and 2.5 V, are effectiveley present where the sense wires are connected. Taking into account the voltage drops on the cables (up to 2.4 V on both lines), the average power delivered by a PSU rises to 30.5 W, for a total of almost 60 kW.

In addition to the PSUs powering the modules, the digital control optoelectronics requires a distinct powering system, to be able to power on the detector control logic independently of the detectors themselves. This has the added benefit of reducing the noise on the analog readout channels, both from the fast digital electronics and from ground loops picking up environmental noise. Each of the 352 control rings needs one 2.5 V voltage regulator, and consumes about 5 W, depending on the number of CCUs present in the ring. The total power consumption of all the control rings is less than 2 kW, a small fraction of the total.

3.2.2 Architecture

The PSUs are the elementary building blocks of the CMS Silicon Strip Tracker power supply system. Each PSU has an internal microprocessor able to handle the normal operation of the power supplies (powering on/off, high voltage ramping, monitoring) as well as most alarm conditions (overcurrent and overvoltages, external resets, interlocks). Each PSU microprocessor communicates with an Array Controller via a CAN Bus[68] link. These controllers in turn are connected via Supervisor Units to the external systems: the CMS Detector Control System (DCS) and Detector Safety System (DSS). These connections are sketched in Fig. 3.6.

3.2.3 Prototype development

Given the stringent requirememts on output voltages, low noise and good insulation, the development of custom Power Supply Units was mandatory. The CMS Florence Group[1] followed the development of two PSU prototypes

---

[1] See http://hep.fi.infn.it/CMS/.

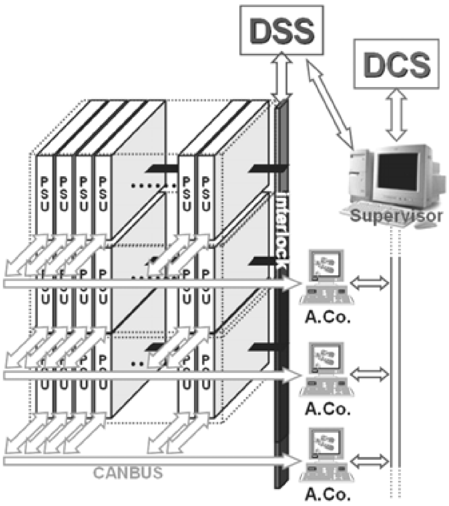


Figure 3.6: Scheme of the power supply system.

from the LABEN and CAEN firms. I have been deeply involved with the LABEN prototype, having developed its firmware as part of my degree thesis [69]. During the first year of my PhD course I kept working on that prototype, developing the supervisor software and human interface to control it. The supervisor software (Fig. 3.7) runs under Microsoft Windows XP on a standard Personal Computer, equipped with CAN Bus PCI expansion cards for the communications with the PSU. I used both a NI-CAN card from National Instruments[2] and two CANnes cards by Trinamic[3]. The use of multiple expansion cards connected to a single PSU allows to run an independent bus monitoring tool to debug the low level CAN bus communication protocol used by the PSU.

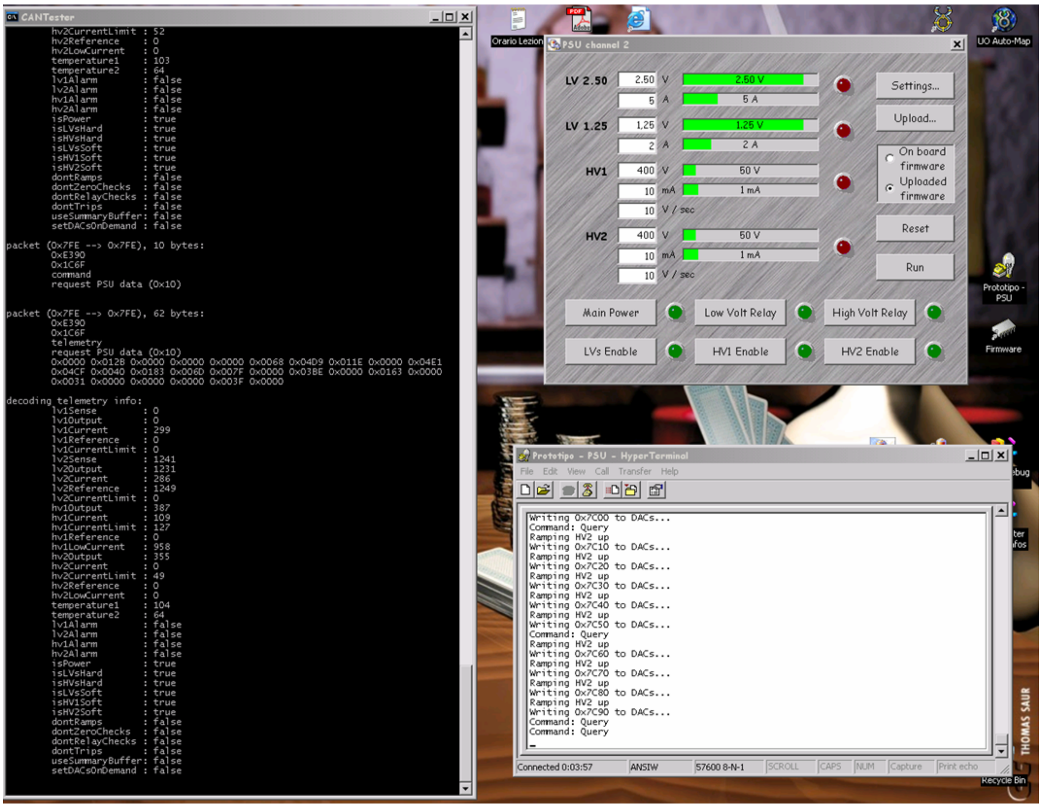
For increased modularity, the supervisor is splitted into two layered parts. The lower level one is a multithreaded dynamic linking library (DLL), written in C++, providing:

- an interface to the different CAN bus expansion cards;
- an implementation of the high level protocol used to communicate with the PSU microcontroller;
- a sharing mechanism and an asynchronous queue system to allow more than one process to interact with the same CAN bus port;

---

[2] National Instruments. URL: http://www.ni.com/ .

[3] TRINAMIC Motion Control GmbH & Co. URL: http://www.trinamic.com/ .



```
        hv2CurrentLimit : 52
        hv2Reference    : 0
        hv2LowCurrent   : 0
        temperature1    : 103
        temperature2    : 64
        lv1Alarm        : false
        lv2Alarm        : false
        hv1Alarm        : false
        hv2Alarm        : false
        isPower         : true
        isLVsHard       : true
        isHVsHard       : true
        isLVsSoft       : true
        isHV1Soft       : true
        isHV2Soft       : true
        dontRamps       : false
        dontZeroChecks  : false
        dontRelayChecks : false
        dontTrips       : false
        useSummaryBuffer: false
        setDACsOnDemand : false

packet (0x7FE --> 0x7FE), 10 bytes:
        0xE390
        0x1C6F
        command
        request PSU data (0x10)


packet (0x7FE --> 0x7FE), 62 bytes:
        0xE390
        0x1C6F
        telemetry
        request PSU data (0x10)
        0x0000 0x012B 0x0000 0x0000 0x0000 0x0068 0x04D9 0x011E 0x0000 0x04E1
        0x04CF 0x0040 0x0183 0x006D 0x007F 0x0000 0x03BE 0x0000 0x0163 0x0000
        0x0031 0x0000 0x0000 0x0000 0x003F 0x0000

decoding telemetry info:
        lv1Sense        : 0
        lv1Output       : 0
        lv1Current      : 299
        lv1Reference    : 0
        lv1CurrentLimit : 0
        lv2Sense        : 1241
        lv2Output       : 1231
        lv2Current      : 286
        lv2Reference    : 1249
        lv2CurrentLimit : 0
        hv1Output       : 387
        hv1Current      : 109
        hv1CurrentLimit : 127
        hv1Reference    : 0
        hv1LowCurrent   : 958
        hv2Output       : 355
        hv2Current      : 0
        hv2CurrentLimit : 49
        hv2Reference    : 0
        hv2LowCurrent   : 0
        temperature1    : 104
        temperature2    : 64
        lv1Alarm        : false
        lv2Alarm        : false
        hv1Alarm        : false
        hv2Alarm        : false
        isPower         : true
        isLVsHard       : true
        isHVsHard       : true
        isLVsSoft       : true
        isHV1Soft       : true
        isHV2Soft       : true
        dontRamps       : false
        dontZeroChecks  : false
        dontRelayChecks : false
        dontTrips       : false
        useSummaryBuffer: false
        setDACsOnDemand : false
```

LV 2.50 2.50 V 2.50 V
5 A 5 A
LV 1.25 1,25 V 1.25 V
2 A 2 A
HV1 400 V 50 V
10 mA 1 mA
10 V / sec
HV2 400 V 50 V
10 mA 1 mA
10 V / sec

○ On board firmware
● Uploaded firmware

```
Writing 0x7C00 to DACs...
Command: Query
Ramping HV2 up
Writing 0x7C10 to DACs...
Ramping HV2 up
Writing 0x7C20 to DACs...
Ramping HV2 up
Writing 0x7C30 to DACs...
Ramping HV2 up
Writing 0x7C40 to DACs...
Ramping HV2 up
Writing 0x7C50 to DACs...
Command: Query
Ramping HV2 up
Writing 0x7C60 to DACs...
Ramping HV2 up
Writing 0x7C70 to DACs...
Ramping HV2 up
Writing 0x7C80 to DACs...
Ramping HV2 up
Writing 0x7C90 to DACs...
Command: Query
Command: Query
_
```

Connected 0:03:57 | ANSIW | 57600 8-N-1 | SCROLL | CAPS | NUM | Capture | Print echo

Figure 3.7: Debug session of the LABEN prototype supervisor software. The command prompt window to the left is running a CAN bus monitor; on the bottom right is the PSU log and debug output, while on the top right is the supervisor interface itself.

- a high level API (Application Programing Interface) used by other applications that load the DLL to control the PSU.

The human interface is implemented as a Visual Basic application, which uses the DLL API to control the PSU. Its main feature is a simple graphical user interface able to control all aspects of the PSU, such as the voltage regulators, voltage and current monitors, power relays and alarms.

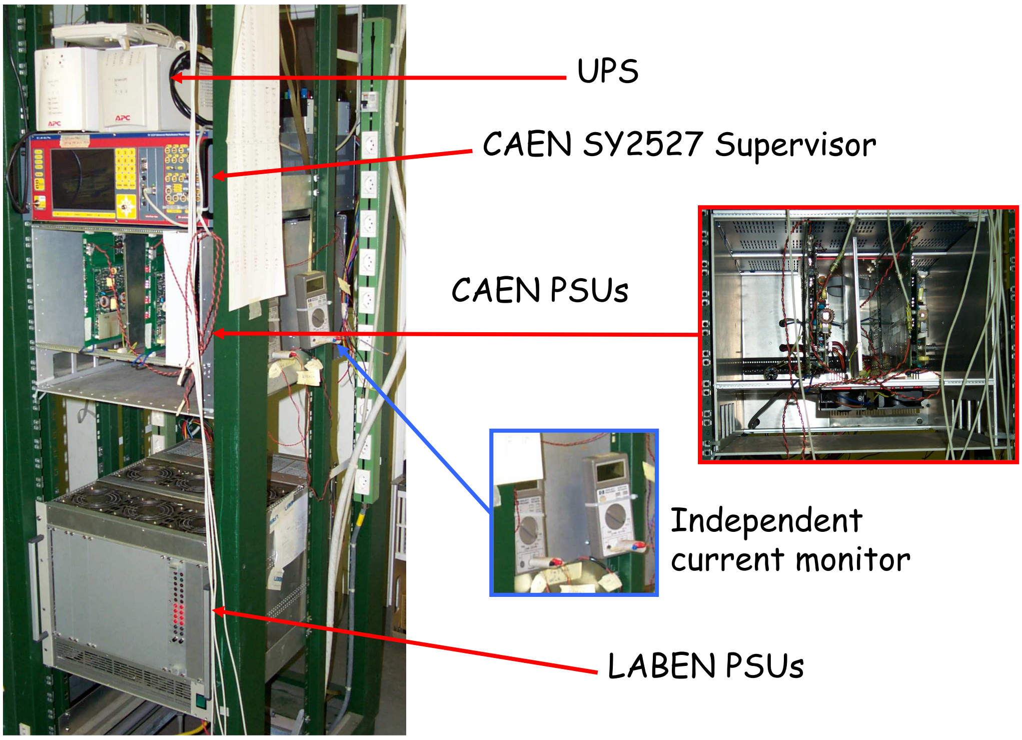


Figure 3.8: Power Supply Units and related paraphernalia, like the independent current monitors and the UPS, for the May 2005 test beam at the CERN X5 beam area.

3.2.4 Test Beam results

The work culminated with the use of both CAEN[4] and LABEN[5] power supply prototypes at the May 2005 Tracker test beam, at the X5 beam area at CERN. Fig. 3.8 shows the setup of the power supply system "rack". The PSU were connected to a first prototype of environmental and electrical monitoring, with temperature and humidiy interlocks, and an UPS system. Both CAEN and LABEN PSU prototypes were used and different power cable setups and grounding schemes were tested.

Fig. 3.9 shows the Landau shaped distributions of the signal-to-noise ratio for a single-sided, 320 μm silicon strip module, with the APV25 operating in peak and decovolution mode. A grounding scheme similar to the CMS final

---

[4] CAEN, Tools for discovery. URL: `http://www.caen.it/` .

[5] Alenia Spazio - LABEN. URL: `http://www.laben.it/` .

one was used; an independent power source for the control ring was provided by a car battery.

For each configuration the most probable value and distribution width parameter are given by the P2 and P3 fit parameters. As can be seen, both power supply prototypes behaved very well, with a very low overall noise and a good signal-to-noise ratio. This was a direct experimental confirmation of the quality of both PSU prototypes, and was expected from the values of noise and isolation measured in laboratory, complying with requirements stated above.

Although the results were good using both prototypes, LABEN decided to end the PSU prototype development shortly afterwards. The CAEN prototype, on the other hand, was fully developed into the final version [67] now being produced and currently in use at the tracker integration centres.

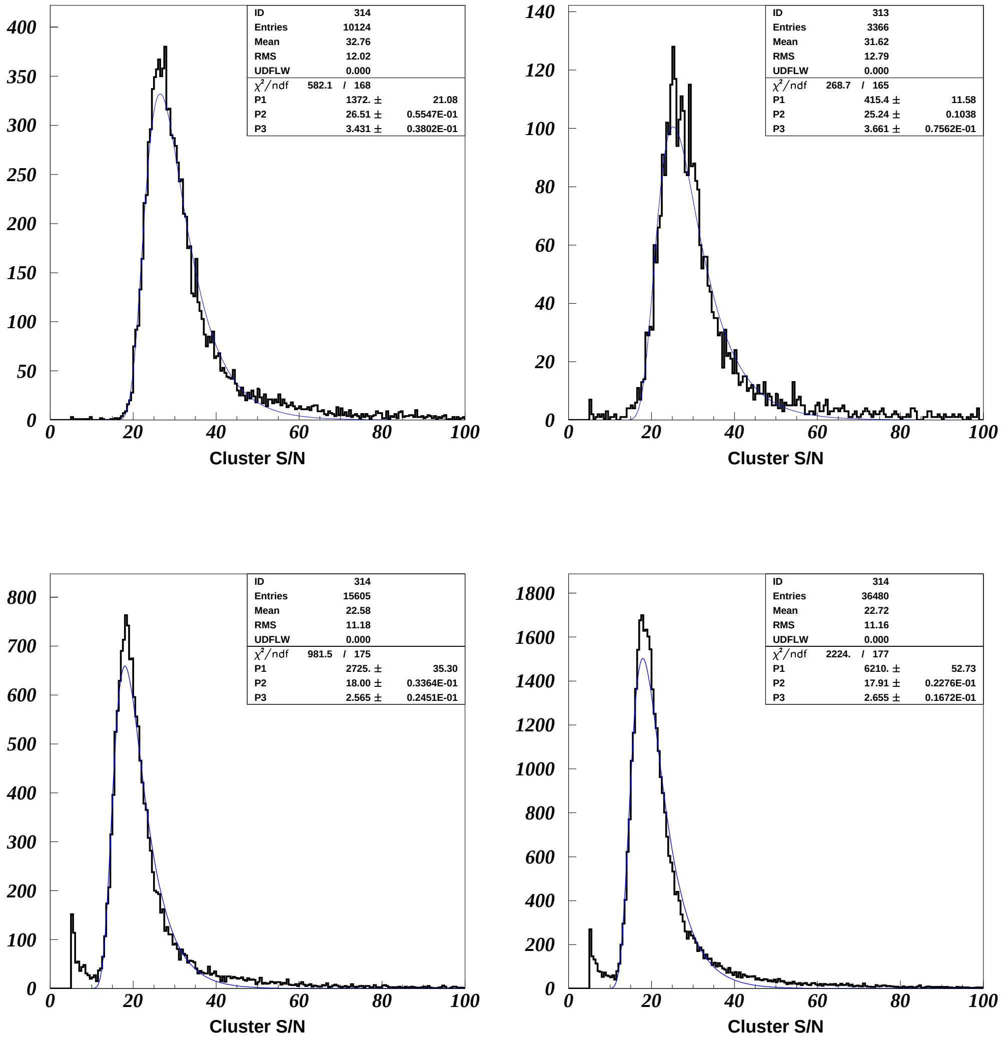


Figure 3.9: Comparison of the performance of a 320 μm single-sided silicon microstrip module using two different PSU prototypes: CAEN (left) and LABEN (right). The APV25 front-end chips are working in peak mode (top) and deconvolution mode (bottom). The measured signal-to-noise and a fit to a Landau distribution are shown; the fit parameters are the area (P1), most probable value (P2) and distribution width parameter (P3).

Chapter 4

Jet b tagging

Many physics channels produce b jets in the final state which need to be separated from the more copious light flavoured jets, such as events containing top quarks, Higgs bosons, or supersymmetric particles. The top quark decays almost exclusively into a W boson and a b quark, and for low masses of the Standard Model Higgs boson ($m_{\mathrm{H}} \leq 135$ GeV/c^2) its dominant decay is into a pair of b quarks, $\mathrm{H} \rightarrow \mathrm{b\bar{b}}$. In this low Higgs mass region a possible discovery channel is the associated production of a Higgs boson with a $\mathrm{t\bar{t}}$ pair

$$\mathrm{pp} \rightarrow \mathrm{t\bar{t}H} \rightarrow \mathrm{W^{+}b\,W^{-}\bar{b}\,b\bar{b}} \tag{4.1}$$

with four b jets in the final state. In such events an efficient b tagging is clearly necessary.

The inclusive tagging of b jets against the much more abundant light flavoured jets background relies upon the large b hadrons mean lifetime ($c\tau = 470.1 \pm 2.7\ \mu$m[70]), their large semileptonic branching ratios, having $\mathrm{Br}(\mathrm{b} \rightarrow \ell\bar{\nu}_{\ell}\ \vee\ \mathrm{b} \rightarrow \bar{\mathrm{c}} \rightarrow \ell\bar{\nu}_{\ell}\ \vee\ \mathrm{b} \rightarrow \mathrm{c} \rightarrow \bar{\ell}\nu_{\ell}) = 19.3 \pm 0.5\%$ (see Sec. 4.1.3), and the large charged track decay multiplicity, with an average of five charged tracks per decay vertex, $\langle n_b^{ch} \rangle = 4.955 \pm 0.062$[34].

Before analyzing in detail the different jet b tagging algorithms used at CMS, I shall give a small review of the more distinguishing properties of b hadrons. The conclusion of this chapter will be a review of the performance of the different b tagging algorithms studied and implemented for the CMS experiment.

4.1 Properties of b hadrons

At high energy experiments, like Tevatron and LHC, b quarks, either from the primary collisions or the decay of heavier particles (t quarks, $\mathrm{W}^{\pm}$, Z^0 or

b hadron species	Fraction
B^0, B^+	$f_d = f_u = 0.399 \pm 0.010$
B^0_s	$f_s = 0.102 \pm 0.014$
b baryons	$f_{\rm baryons} = 0.100 \pm 0.017$

Table 4.1: Fractions of the different b hadron species in an sample produced at high energy[70].

H bosons), hadronize into a heterogeneous "B admixture" of weakly-decaying b mesons (B^+, B^0 and B^0_s) and baryons (Λ^0_b, Ξ_b,Σ_b, Ω_b).

4.1.1 B admixture

The relative production of each meson species and of baryons can be described using the four fractions f_u (fraction of $b \to B^+$), f_d ($b \to B^0$), f_s ($b \to B^0_s$) and $f_{\rm baryons}$ ($b \to$ baryons).

$$f_u + f_d + f_s + f_{\rm baryons} = 1 \tag{4.2}$$

Experimental evidence allows to neglect differences between decays into B^+ and B^0, arising from the different u and d quark mass and electromagnetic interactions, leading to

$$f_u = f_d \tag{4.3}$$

The hadron distributions are assumed to be the same in Z^0 decays, as measured at LEP, and in high p_T[1] jet production at the higher energies available at Tevatron, as measured by the CDF experiment. This assumption has not yet been proved wrong by experimental data.

Combining results from LEP and CDF, subject to the constraints in Eqs. 4.2 and 4.3, yields the results reported in Tab. 4.1.

4.1.2 Inclusive b hadrons lifetime

As b hadrons are produced in a B admixture with similar properties and measured relative abundancies, it is useful to describe the collective properties of such admixture, as well as those of the single mesons and baryons.

We can define the mean lifetime for a B admixture as

$$\tau_b = \sum_i f_i \tau_i \tag{4.4}$$

---

[1] low p_T jets in hadronic colliders can be affected by beam remnants, and thus lead to different relative distributions.

	Experiments	Method	τ_b (ps)	cτ_b (μm)
(a)	LEP, SLD	b vertex	1.572 ± 0.009	471.3 ± 2.7
(b)	LEP	b → ℓ	1.537 ± 0.020	461.8 ± 6
(c)	CDF	J/ψ vertex	$1.533\,^{+0.038}_{-0.034}$	$459.6\,^{+11}_{-10}$
	Average of the above		1.568 ± 0.009	470.1 ± 2.7

Table 4.2: Different measurements of a B admixture mean lifetime: (a) combination of LEP and SLD measurements based on inclusive b vertex tagging; (b) LEP measurements requiring a leptonic b decay; ℓ means either an electron or a muon; (c) CDF measurements based on inclusive b hadron $\mathrm{H}_b \to \mathrm{J}/\psi\ \mathrm{X}$ decays. The world average [70] takes into account correlations among the measurements.

where τ_i is the mean lifetime and f_i the relative abundance of each hadron species in a b hadronic sample. Experimental efficiencies can be different among the hadron species, consequently the measurement of τ_b can yield different results depending on the method used. A summary of the results obtained with different methods is given in Tab. 4.2, along with their world average[70].

The first measurement applies to an inclusive tagging based on the secondary b vertex reconstruction. This is to a good extent independent from the involved b hadron, although an indirect dependency may be introduced by the track multiplicity of the b decay vertex, affecting the vertex reconstruction efficiency. On the other hand, the other two measurements are relative to specific decay channels, and thus can be regarded as independent on the particular b hadron species only as far as the $\mathrm{b} \to \ell$ ($\ell = \mathrm{e}, \mu$) and $\mathrm{b} \to \mathrm{J}/\psi$ branching ratios are; otherwise, the measurements will overestimate the B admixture mean lifetime.

The mean lifetime of each b meson species has been measured using exclusive data samples, from LEP and Tevatron experiments, and Belle and BaBar b factories. The results are shown in Tab. 4.3.

4.1.3 Leptonic branching ratio

The b hadrons decay by and large via the weak decay of the b quark

$$\mathrm{b} \to \mathrm{c}\,\ell^-\,\bar{\nu}_\ell \tag{4.5}$$

$$\mathrm{b} \to \mathrm{c}\,q\,\bar{q}' \tag{4.6}$$

In the limit of zero lepton masses, which is highly justified for a b quark decaying into an electron or a muon, leptonic universality predicts that the

b hadron	τ_b (ps)	cτ_b (μm)
B^0	1.528 ± 0.009	458.1 ± 2.7
B^+	1.643 ± 0.010	492.6 ± 3
B^0_s	1.472 ± 0.045	441.3 ± 13
B^+_c	0.45 ± 0.12	135 ± 36
b baryon	1.210 ± 0.048	362 ± 14

Table 4.3: Mean lifetime for each b meson and for an inclusive sample of b hadrons [70]. The given mean lifetime for B^0_s is an average of the short- and long-lived components, measured on flavour specific decays like $B^0_s \to D_s \ell \nu$ ($\ell =$ e, μ) or $B^0_s \to D_s \pi$; B^+_c measurements have been carried out by CDF and D0, on the semileptonic decay $B^+_c \to J/\psi \ell$ ($\ell =$ e, μ); the baryon mean lifetime comes from the analysis of partially reconstructed final states with a hadron (p, $\bar{p}$, Λ or $\bar{\Lambda}$) and a lepton.

branching ratios for the semileptonic decay (Eq. 4.5), depending only on the $W^- \ell \bar{\nu}_\ell$ couplings, are equal:

$$\mathrm{Br}(b \to c\, e\, \bar{\nu}_e) = \mathrm{Br}(b \to c\, \mu\, \bar{\nu}_\mu). \tag{4.7}$$

This is in good agreement with experimental values for the inclusive branchig ratios of a B admixture decays with an electron or muon in the final state, reported in Tab. 4.4.

The values shown for the $b \to e\, \bar{\nu}_e\, X$ and $b \to \mu\, \bar{\nu}_\mu\, X$ inclusive decays, as well as the lepton flavour blind $b \to \ell\, \bar{\nu}_\ell\, X$ decay, refer to a high p_T sample, as described in Section 4.1.1[2]. The value shown in Tab. 4.4 (d) is a measure from a $B^\pm/B^0$ admixture as produced in B factories at the $\Upsilon(4S)$ peak, assuming the same amount of $B^\pm$ and B^0 produced: $f_\pm \equiv \mathrm{BR}(\Upsilon(4S) \to B^+B^-) = f_{00} \equiv \mathrm{BR}(\Upsilon(4S) \to B^0\bar{B}^0)$.

The b quark will almost always decay into a c quark. An electroweak fit[34] to data from LEP and SLD ($Z^0 \to c\bar{c}$ and $e^+e^- \to c\bar{c}$) yields an average branching ratio per lepton family for c quarks produced in a "clean" environment of

$$\mathrm{Br}(c \to \ell^+ X) = (9.69 \pm 0.31)\%. \tag{4.8}$$

An individual measurement[71] and a world average[1] are availabe for

---

[2] According to [1], they are relative to a somewhat older measurement of the fractions of b hadrons produced: $f_d = f_u = 0.397 \pm 0.010$, $f_s = 0.107 \pm 0.011$, $f_{\mathrm{baryons}} = 0.099 \pm 0.017$, yet compatible with the updated values presented in Tab. 4.1 and [70].

	Deacay:	Branching Ratio
(a)	$\mathrm{b} \to e^- \, \bar{\nu}_e \, \mathrm{X}$	$10.86 \pm 0.35\%$
(b)	$\mathrm{b} \to \mu^- \, \bar{\nu}_\mu \, \mathrm{X}$	$10.95\,^{+0.29}_{-0.25}\%$
(c)	$\mathrm{b} \to \ell^- \, \bar{\nu}_\ell \, \mathrm{X}$	$10.68 \pm 0.22\%$
(d)	$\mathrm{b} \to \ell^- \, \bar{\nu}_\ell \, \mathrm{X}$	$10.95 \pm 0.15\%$

Table 4.4: Inclusive semileptonic decay branching ratios for a B admixture. ℓ means either an elecron or a muon, not the sum of the two. (a), (b) and (c) describe a high p_{T} B admixture[1] as produced at LEP and Tevatron, while (d) is measured from a $\Upsilon(4S)$ B admixture at B factories[70].

the exclusive dacays respectively to electrons and muons:

$$\mathrm{Br}(\mathrm{c} \to \mathrm{e}^+\mathrm{X}) = (10.3 \pm 0.9\,^{+0.9}_{-0.8})\%; \tag{4.9}$$

$$\mathrm{Br}(\mathrm{c} \to \mu^+\mathrm{X}) = (9.6 \pm 0.4)\%. \tag{4.10}$$

In a b quark decay the produced c quark interacts with other partons, resulting in a shorter mean lifetime and a reduced branching ratio for the decay into leptons. The same fit[34] to LEP and SLD data gives a branching ratio for the decay chain

$$\mathrm{Br}(\mathrm{b} \to \mathrm{c} \to \ell^+\mathrm{X}) = (8.02 \pm 0.19)\%. \tag{4.11}$$

An additional contribution to the production of leptons in b quark decays comes from the “wrong sign” chain decay $\mathrm{b} \to \bar{\mathrm{c}} \to \ell^-$, so called because it involves in addition to the usual c quark the production of a $\bar{\mathrm{c}}$ antiquark via $\mathrm{b} \to \mathrm{cW}^-$, $\mathrm{W}^- \to q\bar{\mathrm{c}}$ (Eq. 4.6 with $\bar{q}' = \bar{\mathrm{c}}$). Combining the production probabilities of the different c hadrons with their semileptonic branching fractions it is possible to obtain a value[34] for

$$\mathrm{Br}(\mathrm{b} \to \bar{\mathrm{c}} \to \ell^-\mathrm{X}) = (1.62\,^{+0.44}_{-0.36})\%. \tag{4.12}$$

The combination of all the above results yields an inclusive branching ratio for a b quark decay to produce either directly ($\mathrm{b} \to \ell$) or via a chain decay ($\mathrm{b} \to \mathrm{c} \to \ell$ and $\mathrm{b} \to \bar{\mathrm{c}} \to \ell$) at least one electron or at least one muon of

$$\mathrm{Br}(\mathrm{b} \to \mathrm{e}^\pm\mathrm{X}) = \mathrm{Br}(\mathrm{b} \to \mu^\pm\mathrm{X}) = (19.3 \pm 0.5)\%, \tag{4.13}$$

and to produce at least one of them of $(36.7 \pm 0.9)\%$.

4.2 b tagging algorithms

The two properties of b hadrons highlighted in the previous section can be used to discriminate a jet coming from the hadronization of a b quark from the more abundant and less interesting jets coming from c and lighter quarks.

The high proper lifetime is the physical base for the development of two distinct discriminants: the search for a fully reconstructed secondary decay vertex, which is the basis of the *combined secondary vertex* b tag, and the measurement of the charged tracks impact parameter with respect to the primary vertex, used by the *track probability* and *track counting* b tag algorithms.

The track impact parameter is a very powerful tool, as it can be combined with other topological and kinematical quantities in order to attain better b tagging performance.

The high leptonic branching ratio is naturally exploited by the *soft lepton* b tag I developed, looking for electrons or muons inside the hadronic jets. Such a simple criterion leads to a good rejection of jets stemming from light quarks, which can be further improved with appropriate cuts on the lepton track parameters.

Each b tagging algorithm is characterized by its b tagging efficiency, defined as the ratio of the number of selected to the total of reconstructed b jets, and mistag rate, defined as the ratio of the number of selected non-b jets to the total of the reconstructed ones. The mistag rate, or mistagging efficiency, are usuelly defined separatly for c jets, light (u, d, s) jets, and gluon jets.

Each reconstructed jet is labeled with a “flavour” based on its Monte Carlo parton (quark and gluon) content: the partons inside a cone around the reconstucted jet axis are considered, and the flavour of the most energetic one is assigned to the jet; the cone aperture size in the $\eta\varphi$ plane is $\Delta R = 0.3$.

This definition can assign the wrong label to jets where a very energetic gluon either is irradiated by a heavy quark or splits to a heavy quark-antiquark pairs before hadronization, thus identifying as “gluon” or “light” a jet physically originating from a heavy quark. In the particular case of the soft lepton b tagging a different approach is available, which assignes the jet flavour by reconstructing the full decay chain which leads to the emission of the lepton. A more detailed description of the jet flavour definition is given in Sec. 5.1.

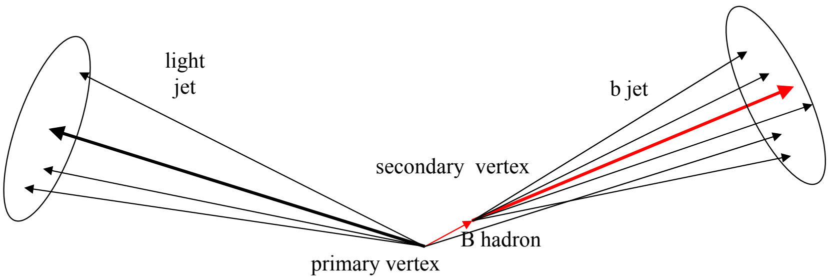


Figure 4.1: Comparison between a light jet (left) and a b jet (right). The b hadron flight path is highlighted.

4.2.1 Track impact parameter

The high lifetime of a b hadron implies that its decay products do not point to the primary interaction vertex, but have a measurable impact parameter.

Figure 4.1 shows the topology of a b jet compared to that of a light jet. The significant flight path of the b hadron causes the tracks of its decay products to have a measurable distance from the primary vertex, or impact parameter[72, 73].

The track impact parameter can be computed either in the plane normal to the beam axis (*transverse impact parameter*) or in three dimensions (*three-dimensional impact parameter*). The first has the advantage of being less sensible to the uncertainty on the primary vertex position, due to the small size of the LHC beam spot in the transverse plane (less than 20 μm × 20 μm). The three-dimensional impact parameter exploits a greater amount of information, but is affected by the larger error on the primary vertex position in the z direction (about 15 cm). The experimental resolution is taken into account using the track impact parameter significance, i.e. the ratio between the track impact parameter and its error.

Both algorithms start from the innermost measurement of the charged track trajectory, extrapolating it backwards to find the point of closest approach to the primary vertex. For the transverse impact parameter the extrapolation is performed analitically, assuming a circular trajectory in the xy plane. For the three-dimensional impact parameter the point of closest approach to the primary vertex is approximated by first finding the point of closest approach of the track to the jet axis, then linearizing the trajectory and propagating it backwards. Fig. 4.2 shows the backward extrapolation and the definition of the signed impact parameters for two tracks, (a) coming from the b decay, (b) falling inside the b jet but coming from the primary vertex. To discriminate this last kind of tracks from the ones actually coming

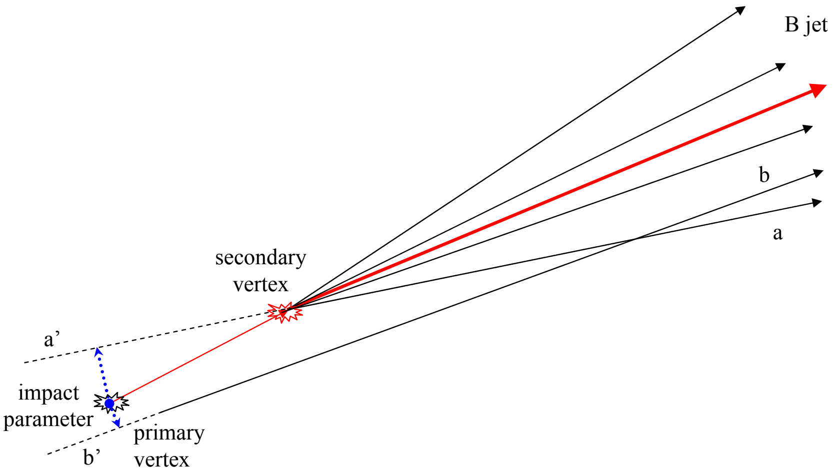


Figure 4.2: Signed impact parameters of tracks inside a b jet. Track (a), coming from the secondary vertex, will have a positive impact parameter, while track (b), born outside of the b jet, will have a negative impact parameter.

from the b decay, the impact parameter is assumed positive for tracks which originate downstream of the b decay, negative for the others: this sign correspond to that of the dot product between the jet axis direction (red thick vectors in Fig. 4.2) and the impact parameter versor (blue dotted vectors in Fig. 4.2). With this prescription, track (a) has a large positive impact parameter, while track (b) has a small negative value.

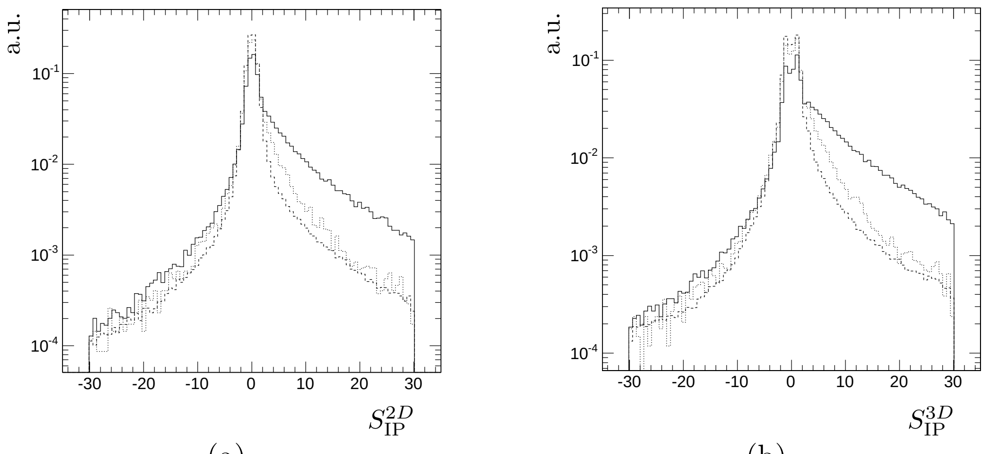
The transverse and three-dimensional impact parameter significance distributions for samples of b, c, and light jets with $80 < E_{\mathrm{T}} < 120$ GeV is shown in Fig. 4.3 [73].

The distribution of transverse impact parameter significance for tracks coming from light jets is strongly peaked around $\bar{x} = 0$, with tails dominated by material interaction effects, such as multiple scattering, which are very difficult to parametrize. Most of those tracks, coming from the primary vertex, have an impact parameter compatible with zero.

For tracks originated in the decay of long-lived particles, the impact parameter is dominated by an exponential decay term, and its distribution shows a marked positive asymmetry.

The three-dimensional impact parameter follows a similar distribution, times a linear term due to the three-dimensional phase space factor (hence the minimum around zero in Fig. 4.3 (b)). It is again evident the asymmetry towards positive value for b jet tracks.

The overall performance of the impact parameter cuts has proved to be



(a) (b)

Figure 4.3: Distributions of the signed impact parameter significance for a sample of b (solid line), c (dotted line) and light (dashed line) jets, from a sample with $80 < E_{\mathrm{T}} < 120$ GeV [73]. (a) transverse impact parameter; (b) three-dimensional impact parameter. The cuts at $S_{\mathrm{IP}} = \pm 30$ are plot artifacts.

better using the three-dimensional impact parameter significance, so this is being used as discriminating factor in all the b tagging algorithms based on the high b hadrons lifetime.

Track counting

The track counting b tagging algorithm[73] is based on the high multiplicity of charged tracks produced in the decay of a b hadron, $< n_b^{ch} >= 4.955 \pm 0.062$ as measured at LEP[34]. It uses the charged track impact parameter significance: the jet is tagged if there are enough tracks with an impact parameter significance exceeding a given cut. Tracks are arranged in decreasing order of impact parameter significance, and the significance of the n-th track is used as discriminant. The number n of tracks requested usually is two or three, depending on the desired trade-off between efficiency and purity of the tagged sample.

Figure 4.4 shows the mistag for light jets versus the b tagging efficiency in a semileptonically decaying $\mathrm{t\bar{t}}$ sample. Black and grey triangles show the performance achieved requiring respectively two and three tracks: as expected, using two tracks yields a better efficiency, while using three tracks gives a better rejection against lighter hadrons.

An additional cut can be introduced on the distance of each track from

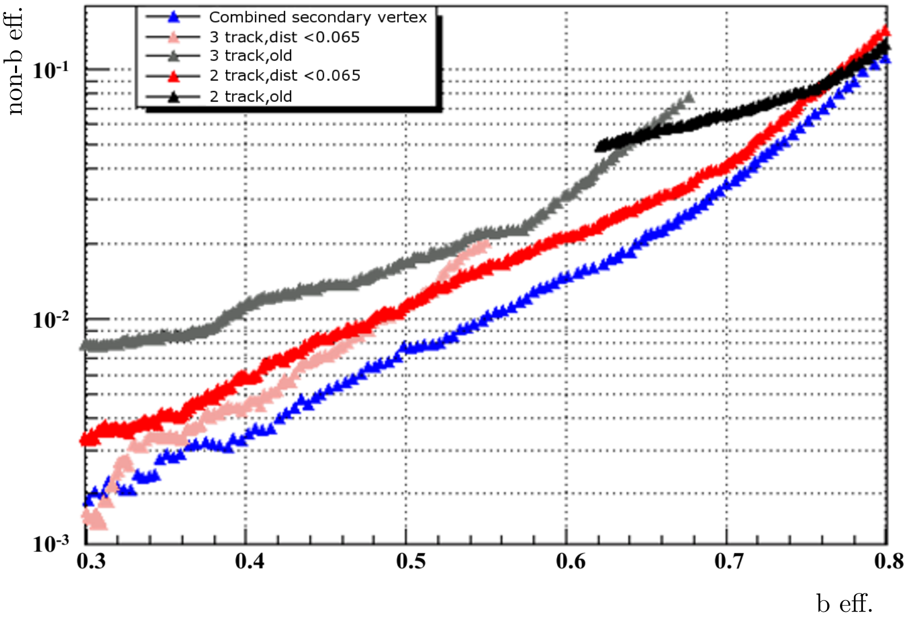


Figure 4.4: Plot of mistag vs. efficiency for the track counting b tag. Black and grey triangles shows the performance of requiring two and three tracks, respectively, and cutting on the impact parameter of the second or third track. The red and light red triangles show the effect of additionally requiring that the used tracks have a distance from the jet axis less than 0.65 mm. The blue triangles are for comparison with the combined secondary vertex algorithm (see Sec. 4.2.2).

the jet axis; rejecting tracks with such distance greater than 0.65 mm greatly improves the suppression of tracks coming from V^0 decays of long-lived neutral hadrons (K^0, Λ^0, etc.), thus improving the overall rejection of light quark jets. As shown in Fig. 4.4, with the introduction of this additional cut the performance of the two tracks algorithm (red triangles) is better than that of the three tracks (light red triangles) down to 50% b tagging efficiency.

Probabilistic approach

As we have seen, the track counting algorithm uses as discriminating variable the impact parameter of only one of the charged tracks. A different approach can be followed, taking into account the impact parameter of *all* the reconstructed charged tracks inside a jet; this method consist in evaluating the probability of the hypothesis that the set of tracks comes from the primary vertex, and using that as b tagging dicriminant. Since the probabilities can be directly calibrated on the impact parameter significance distribution, this approach has the advantage of allowing the detection of the largest range of decay topologies in an inclusive way.

The resolution function $R(|x|)$ describing the distribution of the impact parameter significance x (Fig. 4.3) for tracks coming from the primary vertex can be extracted from the negative part of the distribution itself, since the tracks with negative impact parameter are mainly coming from the primary vertex.

In order to take into account the tails of the resolution function and avoid ad-hoc parametrizations, the resolution function is histogrammed with sufficient precision and the probability computed by numerically integrating the normalized histogram.

Then, for each track we can define a discriminant correlated to the probability that it is coming from the primary vertex as the integral of $R(|x|)$ times the sign of x:

$$\tilde{P}_{tr} = \mathrm{sign}(x) \int_{|x|}^{\infty} R(|t|)dt \qquad (4.14)$$

Following this definition of $\tilde{P}_{tr}$, tracks from light quarks - thus coming from the primary vertex - will show a flat distribution between -1 and 1, as their impact parameter significance will be distributed according to the resolution function $R(|x|)$ itself. Tracks from displaced decays, thus having a large impact parameter significance, will have a positive and small value of $\tilde{P}_{tr}$, as shown in Fig. 4.5.

A discriminant having the required properties to be intepreted as a probability can be recovered renormalizing $\tilde{P}_{tr}$ as:

$$P_{tr}(x) = \begin{cases} \frac{1}{2}\tilde{P}_{tr}(|x|) & \text{if } \tilde{P}_{tr} \geq 0, \\ 1 + \frac{1}{2}\tilde{P}_{tr}(|x|) & \text{if } \tilde{P}_{tr} < 0. \end{cases} \qquad (4.15)$$

By construction, $P_{tr}(x)$ will be flat in $[0 \ldots 1]$ for tracks coming from the primary vertex, while lifetime effects will give a marked asymmetry in favour of small values for tracks coming from the decay of c and b hadrons.

The jet probability estimation P_{jet} for a jet containing N tracks is then performed combining the probability of each track; it is built as the confidence level that any group of N tracks without lifetime would produce the observed value of track probability or any other value equally or more unlikely:

$$L_{jet} = -\sum_{i=0} N \log P_{tr}(x_i) P_{jet} = e^{-L_{jet}} \sum_{j=0}^{N-1} \frac{(L_{jet})^j}{j!} \qquad (4.16)$$

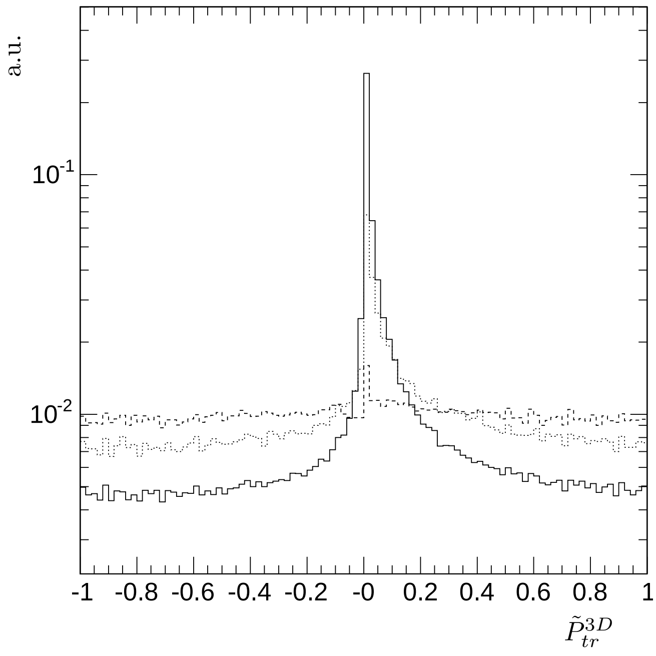


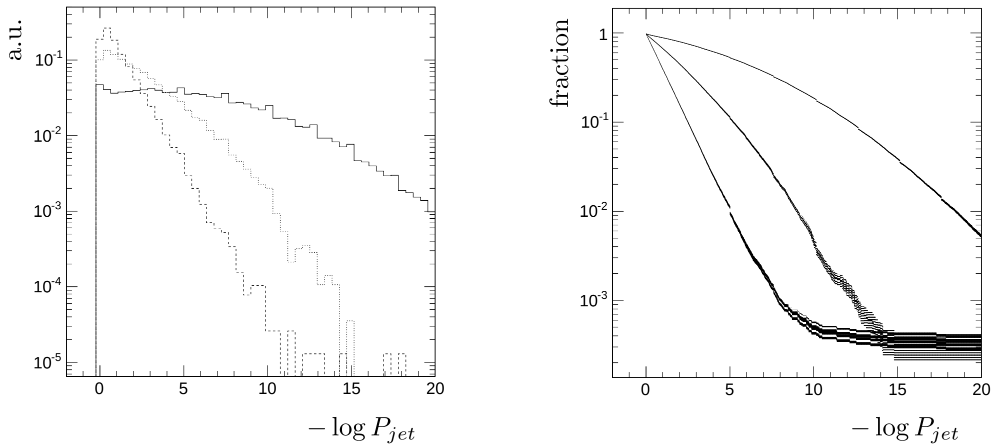
Figure 4.5: Signed track probability $\tilde{P}_{tr}^{3D}$ to come from the primary vertex for b jet tracks (solid line), c jet tracks (dotted line) and light jet ones (dashed line), for jets with $50 < E_{\mathrm{T}} < 120$ GeV and $|\eta| < 1.4$[73]. The three-dimensional impact parameter significance has been used.

To prevent tracks with very small probability from spoiling the information from other tracks and drive P_{jet} to zero, a lower limit is introduced: all tracks with $P_{tr} < 5 \cdot 10^{-3}$ are treated as having $P_{tr} = 5 \cdot 10^{-3}$.

The discriminant variable of the algorithm is $-\log P_{jet}$, expected to be low for light quarks jet and high for b jets. Figure 4.6 (a) shows the distributions of $-\log P_{jet}$ for b, c and uds jets. A cut on this variable provides discrimination between b and light flavoured jets; Figure 4.6 (b) shows the fraction of jets of each flavour with $-\log P_{jet}$ above the cut, as a function of the cut.

4.2.2 Secondary vertex properties

The combined b tagging algorithm[74] is based on the reconstruction of the secondary decay vertex of the weakly decaying b hadron. Different topological and kinematical vertex variables are combined together with track impact parameter significances into a discriminating variable to distinguish b quark jets from non-b jets. Secondary vertices are reconstructed in an in-



(a) (b)

Figure 4.6: (a) distribution of the discriminant variable $-\log P_{jet}$ for b (solid line), c (dotted line) and uds jets (dashed line) in $\mathrm{t\bar{t}}$ events, using the three-dimensional impact parameter significance. (b) fraction of jets with $-\log P_{jet}$ above a given cut as a function of the cut[73].

clusive way inside the jet under consideration, using the *Trimmer Kalman Fitter*[75]. This algorithm starts with all tracks and successively rejects outliers which then are used to reconstruct additional vertices. Since this b tagging algorithm uses not only the presence of a secondary vertex, but also topological and kinematical variables related to the vertex, it is desirable to reconstruct the decay vertex as completely as possible, to increase the discriminating power of these variables.

The following cuts are applied to the resulting vertices to select secondary vertex candidates coming from b hadron decays:

- The distance L_t from the primary vertex to the secondary vertex in the transverse plane has to be within 100 μm $< L_t <$ 2.5 cm, and have a significance of at least $L_t/\sigma_{L_t} > 3$.
- The invariant mass of charged particles associated to the vertex must be compatible with a b quark, and thus not exceed 6.5 GeV/c^2.
- The vertex must not be compatible with a light neutral meson (V^0) decay.

Based on the result of the secondary vertex reconstruction and selection, three categories are defined:

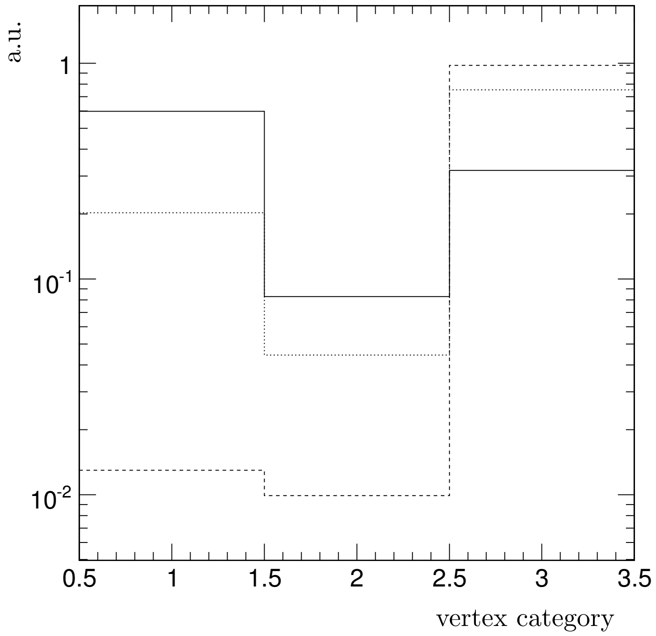


Figure 4.7: Distribution of the vertex category for b jets (solid), c jets (dotted) and light jets (dashed) as obtained from a sample of semi-leptonically decaying $t\bar{t}$ events.

1. *RecoVertex*: At least one secondary vertex candidate has been reconstructed and passed the selection critera. All tracks from all accepted vertices are used for the computation of the vertex related variables if there is more than one accepted secondary vertex.

2. *PseudoVertex*: If no reconstructed secondary vertex candidate has been found, a so-called *PseudoVertex* is created from charged particle tracks not compatible with the event primary vertex, having a signed transverse impact parameter significance greater than two, if at least two such tracks are present in the jet.

3. *NoVertex*: If neither 1. nor 2. above are fulfilled.

The distribution of the vertex categories for the different jet flavours is shown in Fig. 4.7. It can be seen that the presence of a secondary vertex alone is already discriminating between b quark jets and other jets.

The variables used to compute a discriminating variable are:

- the invariant mass of charged particles associated to the secondary vertex;

- the multiplicity of charged particles associated to the secondary vertex;
- only for true reconstructed secondary vertices (*RecoVertex*), the significance of distance in the transverse plane from the primary to the secondary vertex;
- the ratio between the energy of charged particles associated to the secondary vertex and that of all charged particles associated to the jet;
- for each track, the rapidity $y = \frac{1}{2} \cdot \ln\left(\frac{E+p_\parallel}{E-p_\parallel}\right)$ of charged particle tracks associated to the secondary vertex with respect to the jet direction;
- each track impact parameter significance;
- the impact parameter significance of the first track exceeding the charmed threshold: after sorting all the tracks in decreasing order of their impact parameter significance, the invariant mass of the first n tracks is computed, and the impact parameter significance of the n-th track moving the invariant mass above a 1.5 GeV$/c^2$ threshold is used.

The threshold value of 1.5 GeV$/c^2$ is chosen to reflect the typical charm hadron masses, reduced to take into account not reconstructed particles, such as neutral particles, that cannot be detected, and charged particle tracks that may not be reconstructed and accepted.

These variables are combined into a single discriminating variable d using a Likelihood ratio technique. Since for most of the variables c jets and light jets look significantly different, the ratio contains two parts for the discrimination against these two backgrounds. With the quantities

$$\mathcal{L}^{b,c,q} = f^{b,c,q}(\alpha) \times \prod_i f_\alpha^{b,c,q}(x_i) \tag{4.17}$$

the discriminating variable d is defined as:

$$d = f_{BG}(c) \times \frac{\mathcal{L}^b}{\mathcal{L}^b + \mathcal{L}^c} + f_{BG}(q) \times \frac{\mathcal{L}^b}{\mathcal{L}^b + \mathcal{L}^q}. \tag{4.18}$$

where $\alpha = 1, 2, 3$ denotes the vertex category as defined above, x_i are the individual variables listed above, q stands for light (u, d, s quarks, and gluon) jets, $f_{BG}(c, q)$ is the expected bayesian *a priori* probability for the charm and light content in non-b jets ($f_{BG}(c) + f_{BG}(q) = 1$), $f^{b,c,q}(\alpha)$ is the probability for flavour b, c, and q to fall into category α and $f_\alpha^{b,c,q}(x_i)$ is the probability density function for variable x_i for category α and flavour b, c, and q.

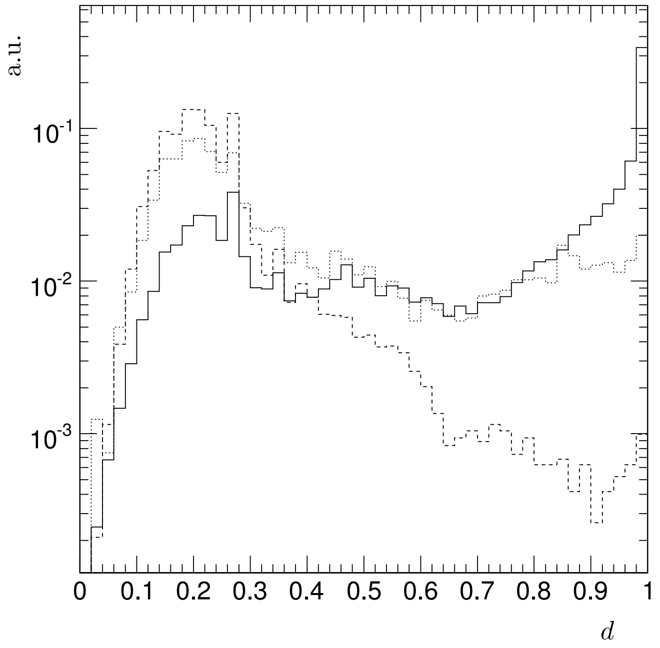


Figure 4.8: Distribution of the discriminating variable d for b jets (solid), c jets (dotted) and light jets (dashed) in the pseudorapidity region $|\eta| < 1.4$ as obtained from a sample of semi-leptonically decaying $t\bar{t}$ events.

The probability density functions $f^{b,c,q}(\alpha)$ and $f_{\alpha}^{b,c,q}(x_i)$, dependent on the transverse jet energy and pseudo-rapidity, are extracted from simulated events, while the prior probabilities $f_{BG}(c, q)$ can be adapted to the specific physics process under study; a default value is $f_{BG}(c) = 0.25$ and $f_{BG}(q) = 0.75$.

The distribution for the discriminator d is shown in Fig. 4.8.

4.2.3 Soft lepton tag

As seen in Section 4.1.3, a b hadron has roughly a 20% probability of decaying emitting directly ($b \to \ell^-$) or indirectly ($b \to c \to \ell^+$, $b \to \bar{c} \to \ell^-$) at least either an electron or a muon.

This can be exploited to tag a b jet looking for leptons within it. This method is traditionally called “soft lepton b tagging”, because it is based on the reconstruction and identification of leptons which are softer than the primary ones emitted in the decay of a top quark or Z^0, $W^{\pm}$ or H heavy boson.

The rejection of c and light jets can be improved without affecting much the b tag efficiency analysing some topological and kinematical properties of the reconstructed lepton with respect to the jet. These tagging variables can be used individually, or combined together to form a linear or nonlinear discriminant. The results in the next Section were obtained using artificial neural networks to combine them.

I shall give a deeper description of the soft lepton b tagging algorithm, its development, and the analysis of different means of combining tagging variables in the next chapter.

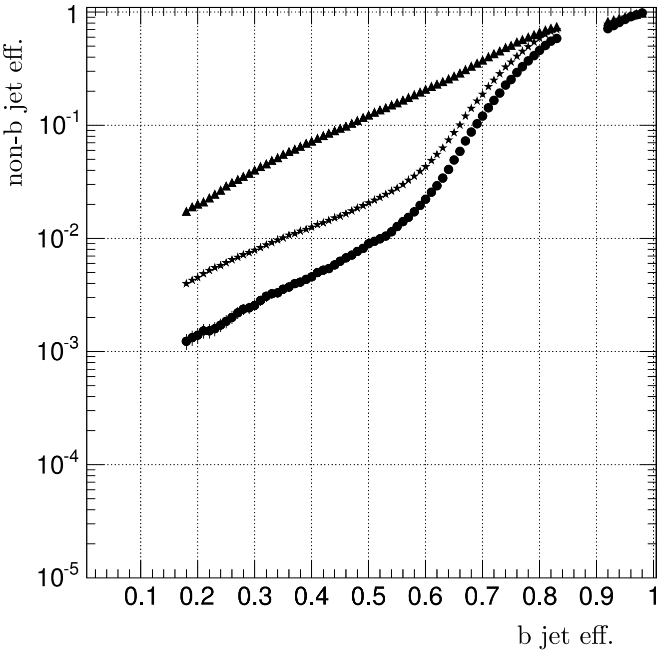
4.3 b tagging performance

I conclude this chapter showing the performance of the b tagging algorithms described in the previous sections.

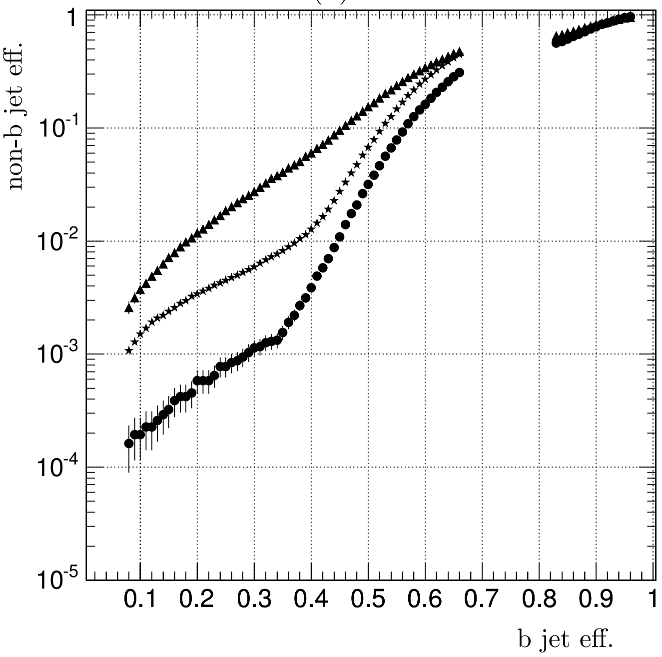
Figures 4.9 through 4.11 show the mistagging efficiency for non-b jets as a function of the b tagging efficiency for the different algorithms. The *combined secondary vertex*, *track probability* and *track counting* algorithms show similar performance. They are very effective tools for b tagging and give a good rejection of light quark jets, with reasonable performance for c jets; jets from gluons fall somewhat in the middle. For example, at a given b tagging efficiency of 50%, the track counting ($n = 2$) and track probability algorithms give a light quark jet mistag rate of about 1%, with the combined secondary vertex better by over a factor three, the mistag being 0.3%; however the mistagging efficiency for c jets is at least an order of magnitude worse, being about 10% for the track based algorithms and almost 7% for the combined secondary vertex.

When a good rejection of background is a stringent requirement, at the expense of a worse efficiency, the soft lepton b tag comes into play. Sacrificing the b tagging efficiency down to 7%, the soft muon tag achieve a mistag rate better than $5 \cdot 10^{-4}$ for light quark jets and than 1% for c jets.

Figures 4.12 and 4.13 show the performance of the combined secondary vertex and soft muon b tagging algorthms as a function of the reconstructed jet transverse energy and pseudorapidity. Keeping their b tag efficiency fixed at the above value (50% and 7%), the non-b jets mistag rate is plotted for jets with transverse energies from 30 to 300 GeV, and pseudorapidity in the full tracker acceptance region $|\eta| < 2.4$.

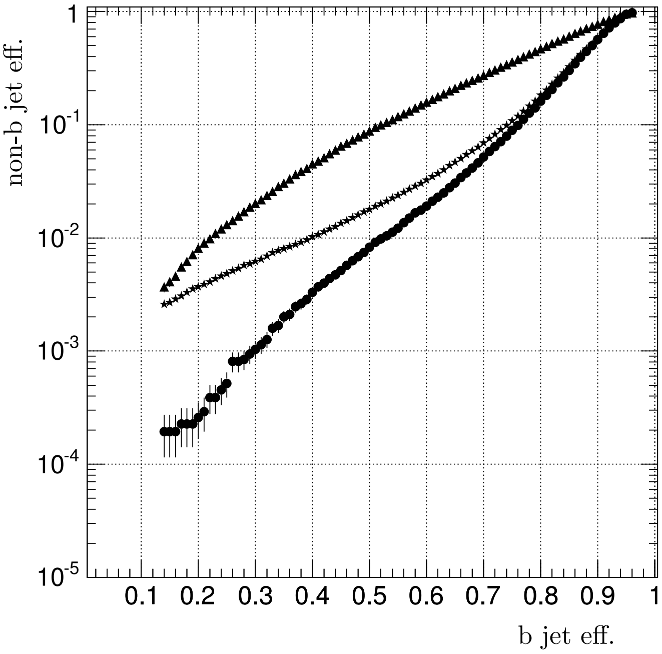


(a)

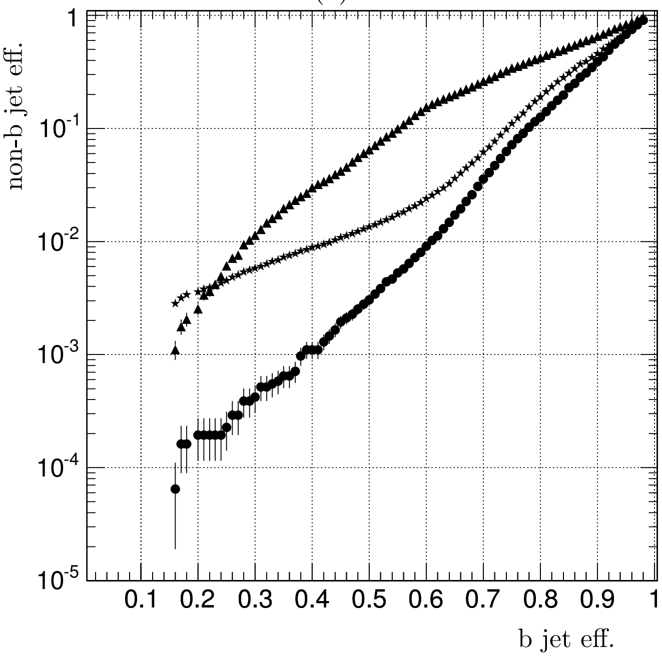


(b)

Figure 4.9: Non-b jet mistagging efficiency versus b jet tagging efficiency for c jets (triangles), uds jets (circles) and g jets (stars) obtained for jets in a QCD sample with transverse jet momenta between 50 GeV/c and 80 GeV/c in the barrel region ($|\eta| < 1.4$) of the detector for the track counting algorithm with (a) $n = 2$ and (b) $n = 3$ tracks.

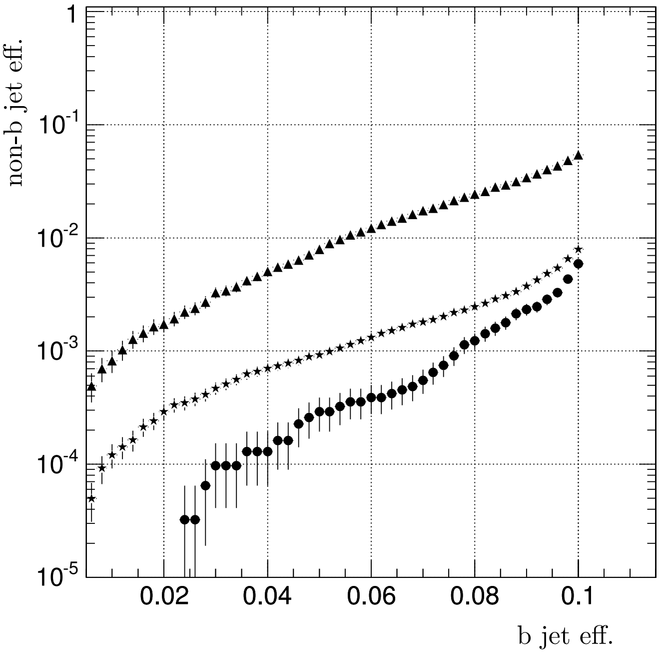


(c)

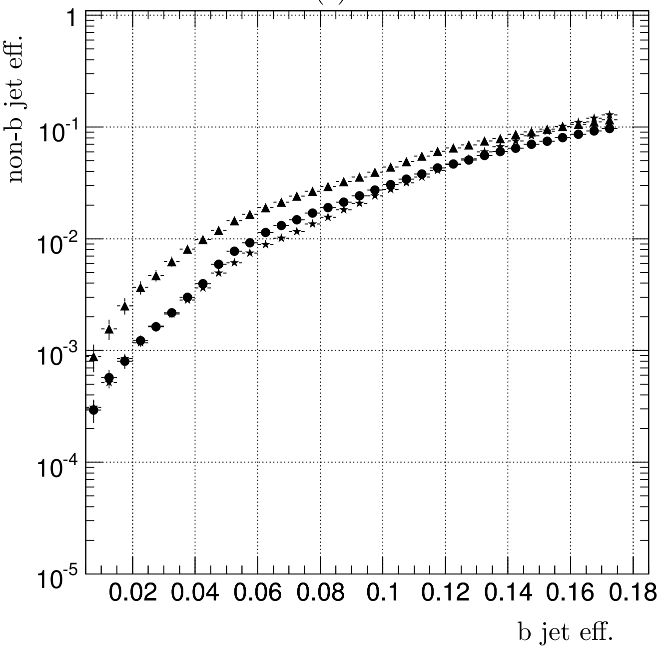


(d)

Figure 4.10: Non-b jet mistagging efficiency versus b jet tagging efficiency for c jets (triangles), uds jets (circles) and g jets (stars) obtained for jets in a QCD sample with transverse jet momenta between 50 GeV/*c* and 80 GeV/*c* in the barrel region ($|\eta| < 1.4$) of the detector for (c) the track probability algorithm and (d) the combined secondary vertex algorithm.

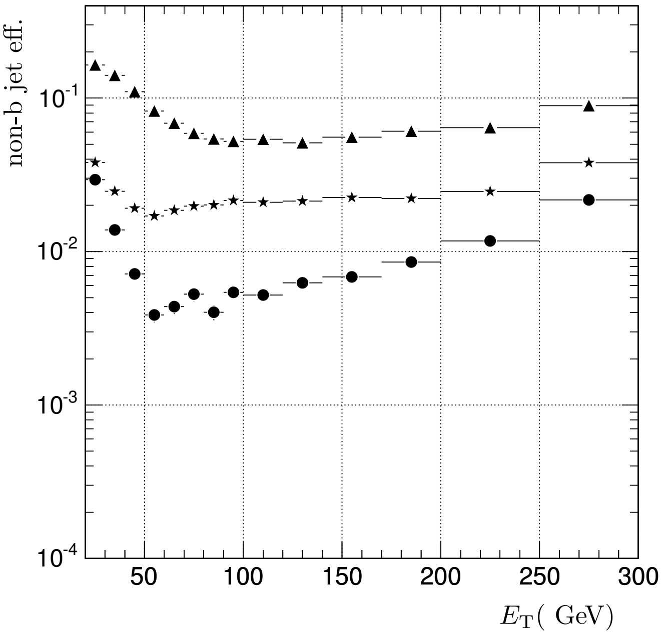


(f)

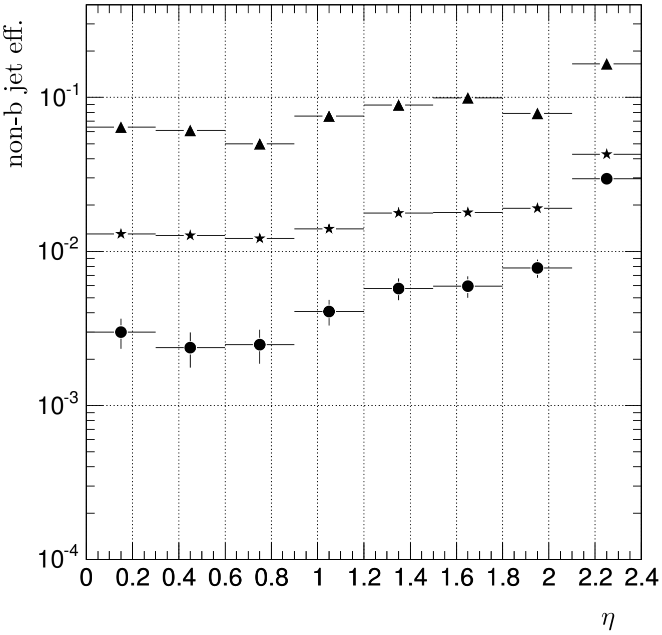


(g)

Figure 4.11: Non-b jet mistagging efficiency versus b jet tagging efficiency for c jets (triangles), uds jets (circles) and g jets (stars) obtained for jets in a QCD sample with transverse jet momenta between 50 GeV/c and 80 GeV/c in the barrel region ($|\eta| < 1.4$) of the detector for (f) the soft muon algorithm and (g) the soft electron algorithm.

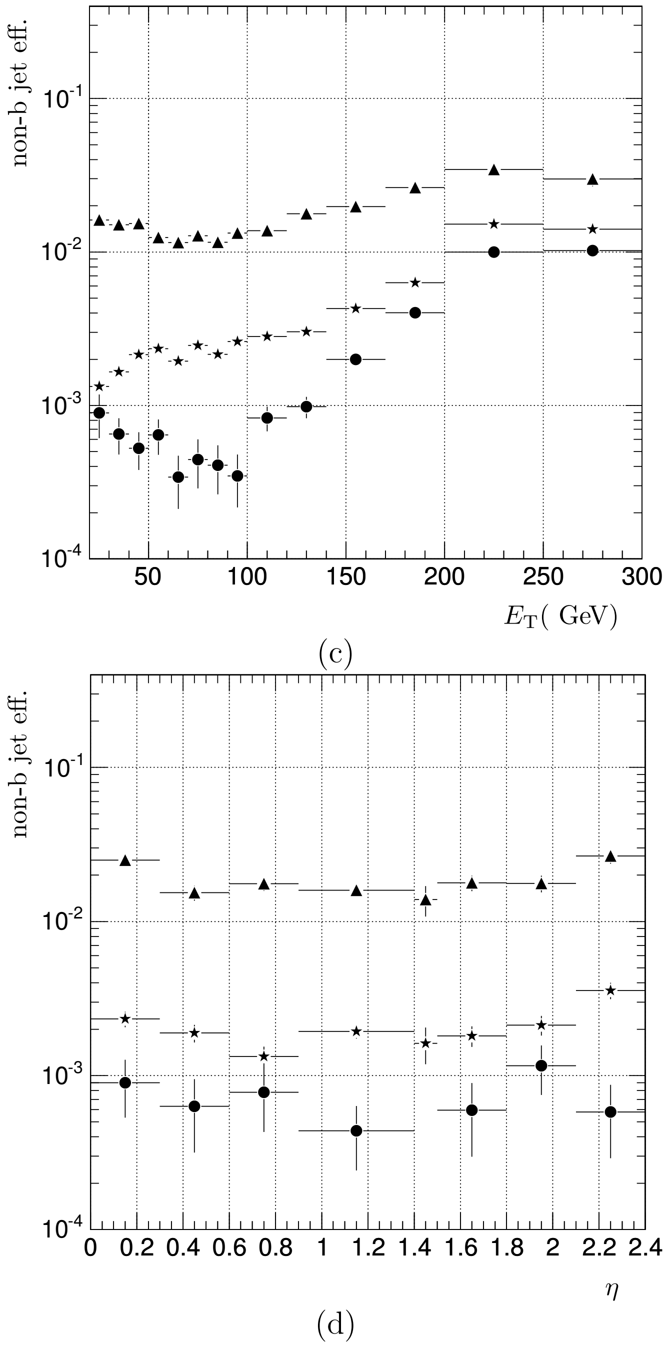


(a)



(b)

Figure 4.12: The non-b jet mistagging efficiencies for a fixed b jet tagging efficiency of 0.5 for the combined secondary vertex algorithm, for c jets (triangles), u, d, s jets (circles) and gluon jets (stars), from a sample of QCD events: (a) mistag rate as a function of the jet transverse energy (in GeV) for jets with $|\eta| < 2.4$, (b) mistag rate as a function of the pseudo-rapidity for jets with transverse energy between 50 GeV and 80 GeV.



(c)

(d)

Figure 4.13: The non-b jet mistagging efficiencies for a fixed b jet tagging efficiency of 0.07 for the soft muon algorithm, for c jets (triangles), u, d, s jets (circles) and gluon jets (stars), from a sample of QCD events: (c) mistag rate as a function of the jet transverse energy (in GeV) for jets with $|\eta| < 2.4$, (d) mistag rate as a function of the pseudo-rapidity for jets with transverse energy between 50 GeV and 80 GeV.

Chapter 5

Soft lepton tag

As seen in the previous Chapter, Section 4.1.3, a b hadron has roughly a 37% probability of decaying emitting directly ($b \rightarrow \ell^-$) or via a c quark ($b \rightarrow c \rightarrow \ell^+$, $b \rightarrow \bar{c} \rightarrow \ell^-$) at least an electron or a muon.

This can be exploited to tag a b jet looking for leptons within it. This method is traditionally called “soft lepton b tagging”, because it is based on the reconstruction and identification of leptons which are softer than the primary ones emitted in the decays of a top quark or Z^0, $W^\pm$ or H heavy boson.

I have developed the framework for the soft lepton b tag package inside the Object-oriented Reconstruction for CMS Analysis (ORCA) software, interfacing it with the existing b tag algorithms (described in the previous Chapter) and developing in detail the soft muon b tag.

In order to improve the performance of this algorithm I used Artificial Neural Networks techniques. Starting with the tools available within the ROOT project, a research in the dedicated literature pointed me to the developent of more powerful algorithms, as the use of Neural Networks able to estimante bayesian *a posteriori* probabilities. The new tools I developed were of course contributed back to the ROOT project.

5.1 Definition of jet flavour

For the development and analysis of any b tagging algorithm a fundamental requirement is the knowledge a priori of a jet flavour. A jet is in itself an object hard to define exactly. Usually they are defined based on the reconstruction algorithm applied to the energy deposits and charged particle tracks reconstructed by the detector: two examples of such algorithms and definitions are the Iterative Cone[76] and Inclusive k_T[77] algorithms.

Given the difficulty in defining the actual constituents of a jet, it is harder still to assign a simulated jet its precise flavour. A common practice is to analyze the Monte Carlo parton content of the jet, i.e. the quarks and gluons that gave origin to the hadronization process. A different approach, available when the jet tagging is performed via leptons, is to use the Monte Carlo lepton decay chain to identify its origin.

5.1.1 Parton content

Flavour assignement through parton content analysis is performed in three steps: the reconstructed jet axis is used to approximate the original parton direction; then the Monte Carlo event description is read, and the partons within a cone with $\Delta R < 0.3$ in the $\eta\varphi$ plane are selected. The flavour of the parton with the highest momentum is assigned to the jet.

This is good “algorithmic” definition, as it assignes to a jet the flavour of its most energetic constituent, and as such the one that influences most its hadronization properties. However it is not always a good “physical” definition, as it misidentifies jets where a very energetic gluon either is irradiated by a heavy quark or splits into a heavy quark-antiquark pair before hadronization. In these cases a jet physically originating from a heavy quark will be identified as a gluon jet, biasing the b tagging effciency measurements.

Jets can be identified as coming from b, c, light quarks (u, d, s), and gluons.

5.1.2 Lepton decay chain

In the particular case of the lepton tagging I’ve developed a different, complementary approach. For reconstructed leptons associated to a Monte Carlo particle the full decay chain can be reconstructed and followed backwards up to the hadronization process. This is done only for leptons correctly associated (electrons to electrons, muons to muons) and allows to classify a jet based on both the lepton direct mother, and the full decay chain:

- if the lepton mother is a b hadron, the jet is classified as a b jet;
- if the lepton mother is a c hadron, the decay chain is taken into account to distinguish between true c and b → c jets, where the c hadron is the decay product of a heavier b hadron; from the viewpoint of the lepton, the distinction is between the $\mathrm{c} \to \ell$ (former) and $\mathrm{b} \to \mathrm{c} \to \ell$ and $\mathrm{b} \to \bar{\mathrm{c}} \to \ell$ (latter) processes;

- if the lepton mother is a light hadron, the jet is considered a light jet, independently of the decay chain origin. This effectively discriminates leptons produced from the decay of light charged mesons, which are abundant in any type of jet, from prompt tagging ones.

Leptons, and thus jets, can be identified as coming from b, b → c, c, and light quarks.

5.2 Event samples

For the development of the various aspects of the soft muon b tag algorithm, we used three distinct simulated data samples; detector response and event pile-up were simulated in a low luminosity ($2 \cdot 10^{33}$ cm^{-2}s^{-1}) scenario.

- **QCD** samples: these events are generated from high-p_T inclusive QCD processes, giving rise to two hard jets, with the possibilty of multiple jets stemming from the parton showering process. The energy of these processes is determined by the transverse momentum $\hat{p}_T$ exchanged between the interacting partons, measured in their own reference frame. Six samples with different $\hat{p}_T$ ranges were used, with different number of events per sample: over 90k with $30 < \hat{p}_T < 50$ GeV/c, almost 200k with $50 < \hat{p}_T < 80$ GeV/c, 280k with $80 < \hat{p}_T < 120$ GeV/c, 100k with $120 < \hat{p}_T < 170$ GeV/c and $170 < \hat{p}_T < 230$ GeV/c, and 50k with $230 < \hat{p}_T < 300$ GeV/c. This choice has been driven both by the number of simulated events available and by the energy range usually interesting for b tagging scenarios. Figure 5.1 shows the reconstructed jet energy and transverse energy spectra for the merged samples. The abundance of low energy jets is due to secondary jets produced in the parton shower process.

  These events are generated with PYTHIA 6.215[78]; the interaction with the CMS detector is performed by OSCAR 2.4.5[79], and the final digitization by ORCA 7.6.1[80, 81].

- **flavour enriched di-jet** sample: these events are similar to the previous ones, but to achieve a good statistics for b and c quark jets, the jet flavour is monitored during the Monte Carlo generation stage. Three samples with different flavours were generated: b-enriched, c-enriched, and light (u, d, s) quark-enriched. This is achieved requiring at least two jets of the corresponding flavour per event; in addition, two jets inside the tracker acceptance $|\eta| < 2.4$ are required. Events not passing these criteria are discarded at generator level.

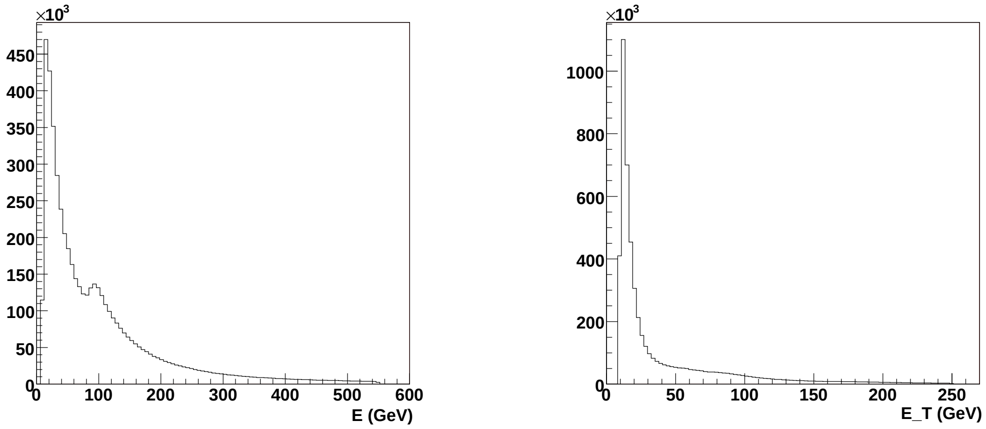


Figure 5.1: Reconstructed energy (left) and transverse energy (right) spectra for a sample of jets produced in hard QCD processes, with $\hat{p}_\mathrm{T}$ ranging from 30 to 300 GeV/c. Different $\hat{p}_\mathrm{T}$ bins have different pupulations, as described in the text.

For each flavour, five samples with different $\hat{p}_\mathrm{T}$ ranges were used: $30 < \hat{p}_\mathrm{T} < 50$ GeV/c, $50 < \hat{p}_\mathrm{T} < 80$ GeV/c, $80 < \hat{p}_\mathrm{T} < 120$ GeV/c, $120 < \hat{p}_\mathrm{T} < 170$ GeV/c, and $\hat{p}_\mathrm{T} > 170$, with roughly 100k events per sample for b and c jets and roughly 250k events per sample for light jets. Figure 5.2 shows the reconstructed calibrated energy and transverse energy spectra for the merged samples. Again, the abundance of low energy jets is due to secondary jets produced in the parton shower process.

These events are generated with PYTHIA 6.215[78]; the interaction with the CMS detector is performed by CMSIM 133[82], and the final digitization by ORCA 7.6.1[80, 81].

- **leptonically and semileptonically decaying** $\mathrm{t\bar{t}}$ samples: these are events where the hard pp interaction produces a pair of $\mathrm{t\bar{t}}$ quark-antiquarks. One (semileptonic decays) or both of them (leptonic decays) are then forced to decay via $\mathrm{t} \to \mathrm{bW}^+$, $\mathrm{W}^+ \to \mathrm{e}^+\nu_\mathrm{e}$ or $\mathrm{W}^+ \to \tau^+\nu_\tau$. These samples are thus very rich in b jets, with comparable amounts of light and gluon jets, and a small fraction of c jets.

  The reconstructed calibrated energy and transverse energy spectra for the two samples merged is shown in Fig. 5.3. Again, the abundance of low energy jets is due to secondary jets produced in the parton shower process.

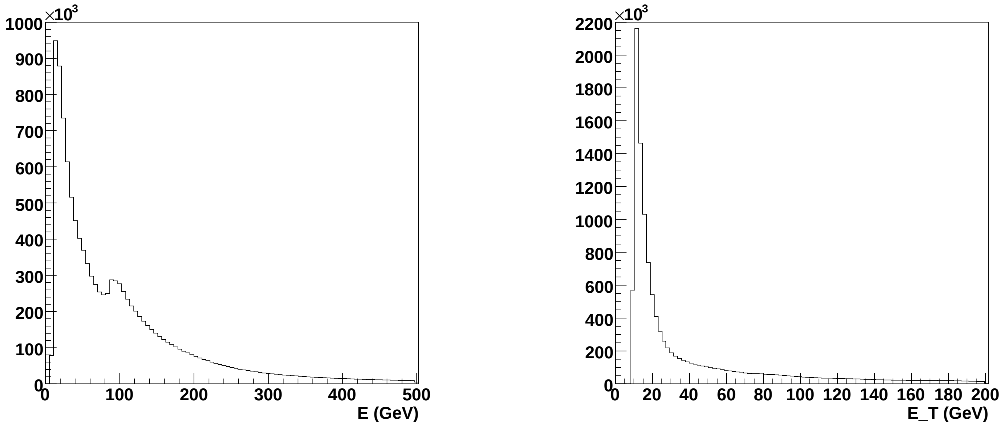


Figure 5.2: Reconstructed energy (left) and transverse energy (right) spectra for a sample of jets produced in flavour-enriched, hard QCD processes, with $\hat{p}_{\mathrm{T}}$ greater than 30 GeV/$c$and no upper limit imposed. Jets of different flavours follow the same energy distributions, so these are not shown individually.

> As for the previous samples, these are generated with PYTHIA 6.215[78]; the interaction with the CMS detector is performed by CMSIM 133[82], and the final digitization by ORCA 7.6.1[80, 81].

5.3 Algorithm description

The rationale behind the soft lepton tag is rather simple: exploit the high semileptonic branching ratio of b hadrons to discriminate them from the overwhelming background of light quark jets due to QCD interactions.

The most naïve approach would be to use the presence of a lepton associated to a jet as a tagging criterion. This in fact works, but the rejection ratio achieved is often not satisfactory. The most important limits to the rejection of light jets are the misidentification of light charged mesons as leptons and the indirect production of leptons in those jets.

The rejection of c and light jets can be improved without affecting much the b tag efficiency analysing some topological and kinematical properties of the reconstructed lepton with respect to the jet.

Thus, for each jet the soft lepton b tagging proceeds in four steps:

- the association of tracks to the jet;

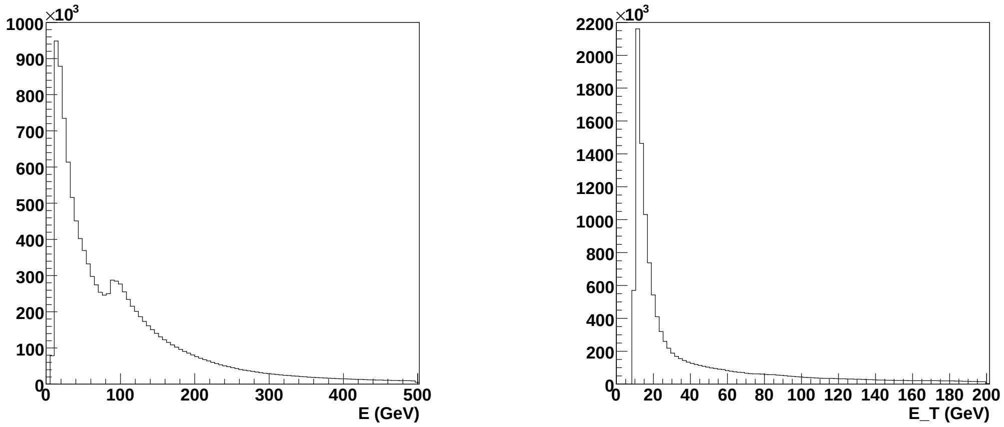


Figure 5.3: Reconstructed energy (left) and transverse energy (right) spectra for a sample of jets produced in the leptonic and semileptonic decays of $t\bar{t}$ quark-antiquark pairs.

- the identification of lepton tracks among them;
- the calculation of the tagging variables;
- the combination of those variables, and the final tagging of the jet.

Jets and tracks have been reconstructed with the standard algorithm available in ORCA. An important issue has been whether to use the default configuration of these algorithms, or to fine tune their parameters to achieve better performance in the cases under examination. As the lepton b tag algorithm aims to be of general applicability, I decided to follow the first approach:

- track reconstruction is done with the Combinatorial Track Finder algorithm[83]. No additional cuts are imposed on the tracks quality and the number of hits in the pixel and silicon strip detectors.
- jets are reconstructed using the Iterative Cone algorithm, applying a Monte Carlo jet energy calibration[76].

The other steps where developed explicitly for the lepton b tag, and will be described in detail in the next Sections.

5.3.1 Algorithm implementation

The software package for the soft lepton b tag had some basic requirements: interface compatibility with the existing b tag framework, reuse, wherever possible, of algorithms and physical objects already available and, while not strictly a requirement, a common implementation for soft electron and muon b tag algorithms was a desirable feature. I wrote a modular implementation: the bulk of the package, taking care of the interfacing to the ORCA b tag framework, looking for reconstructed tracks in each jet, computing tagging variables from lepton and jet information, and building the output b tagged object, would be common to both electron and muon b tag algorithms; it would then defer to two smaller modules the individual lepton identification and the actual tagging of the jet based on the computed variables. The structure of the algorithm is sketched in Fig. 5.4.

This modular structure allows different combinations of tagging variables to be used in different cases. One example is of course the difference between the b tag based on electrons and muons. A different application is the first data taking scenario at CMS, foreseen during the summer of 2007, when the pixel detector will not yet be available[1]. Without it, the reconstruction of the vertex position will have a much larger error, to the point that the track impact parameter described in Sec. 4.2.1 will not be a reliable measurement. A second combination of tagging variables can then be used to achieve acceptable performance in this scenario.

5.3.2 Muon identification

As a dedicated muon identification package[84] had been developed too late to be used in this analysis, I wrote a trivial muon identification tool, based on the matching of reconstructed tracks with reconstructed muons.

This is based on the existing muon reconstruction, but with an identification-like interface: this allows a consistent interface to the rest of the b tag package, and an easy transition to the more advanced muon identification package in the near future.

Muons are reconstructed using the silicon tracker and the muon chambers with the Global Muon Reconstructor algorithm[85] (GMR for short), which entails:

- finding muon candidates inside the muon chambers;

---

[1] A tentative schedule for the installation of the silicon pixel subdetector foresees two barrel layers and the endcap disks installed for the physics run in 2008; the third barrel layer will be installed at the end of the same year.

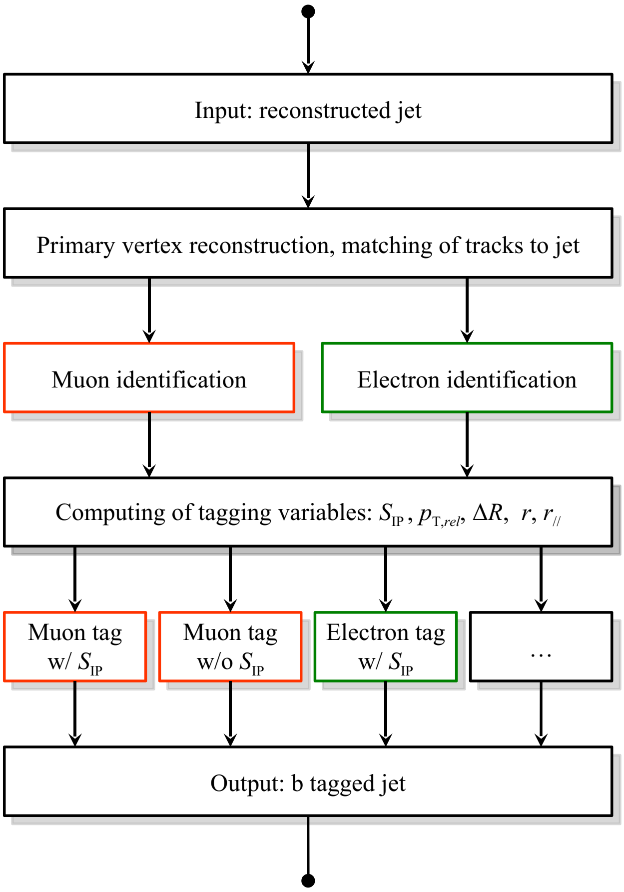


Figure 5.4: Sketch of the soft lepton b tagging algorithm, highlighting the modular structure and the tasks deferred to each module. For variable explanations see Section 5.3.4.

- defining a compatible region in the silicon tracker for each muon candidate;
- reconstructing the tracks inside those regions;
- propagating them outwards from the outer layer of the tracker through ECAL, HCAL and return yokes to the muon chambers, using the GEANE software[86];
- matching tracks and muon candidates;
- refitting matched tracks taking into account the extra hits in the muon system.

Since the GMR muon tracks retain the information about their hits in the tracker, matching them to reconstructed tracks can be done in an efficient way checking for common hits in the silicon detectors. To increase the speed of the association algorithm while keeping it as simple as possible, each reconstructed track is only checked against muon tracks sufficiently near to it, i.e. having $\Delta\eta < 0.1$ and $\Delta\phi < 0.1$. Finally, to take into account the possibility of using different algorithms for the reconstruction of tracks inside jets than those used by the GMR, only 70% of the reconstructed track valid hits are required to be in common with the muon track, thus allowing for slight variations in track-to-hit association within the tracker.

The main drawback of this method is the intrinsically low efficiency of muon reconstruction at low transverse momenta and pseudorapidity. We have investigated the issue, studying the performance of the Global Muon Reconstructor for two samples of isolated muons with transverse momentum in the ranges $1\ \mathrm{GeV}/c < p_\mathrm{T} < 20\ \mathrm{GeV}/c$ and $5\ \mathrm{GeV}/c < p_\mathrm{T} < 100\ \mathrm{GeV}/c$, with pseudorapidity in the full tracker acceptance region $|\eta| < 2.4$.

We found that the muon reconstruction efficiency is very good for isolated muons with a transverse momenum $p_\mathrm{T} > 10\ \mathrm{GeV}/c$ - better than 98% over all the detector acceptance, for the high p_T sample - but drops in the barrel region as the transverse momentum decreases.

This is due to the magnetic field: muons with p_T less than 4 GeV/c hardly reach the muon chambers in the barrel, as can be readily seen in Fig. 5.5 and Tab. 5.1, which show the muon reconstruction efficiency for the low p_T sample in the different η and p_T regions.

In the barrel ($|\eta| < 1.3$) it falls below 90% for $p_\mathrm{T} < 6\ \mathrm{GeV}/c$, below 50% for $p_\mathrm{T} < 4.5\ \mathrm{GeV}/c$, and quickly approaches zero for $p_\mathrm{T} < 3.5\ \mathrm{GeV}/c$. In the region with overlap between barrel and endcap detectors ($1.3 < |\eta| < 1.8$) the efficiency drops similarily, but keeps above 50% down to $p_\mathrm{T} > 3.5\ \mathrm{GeV}/c$,

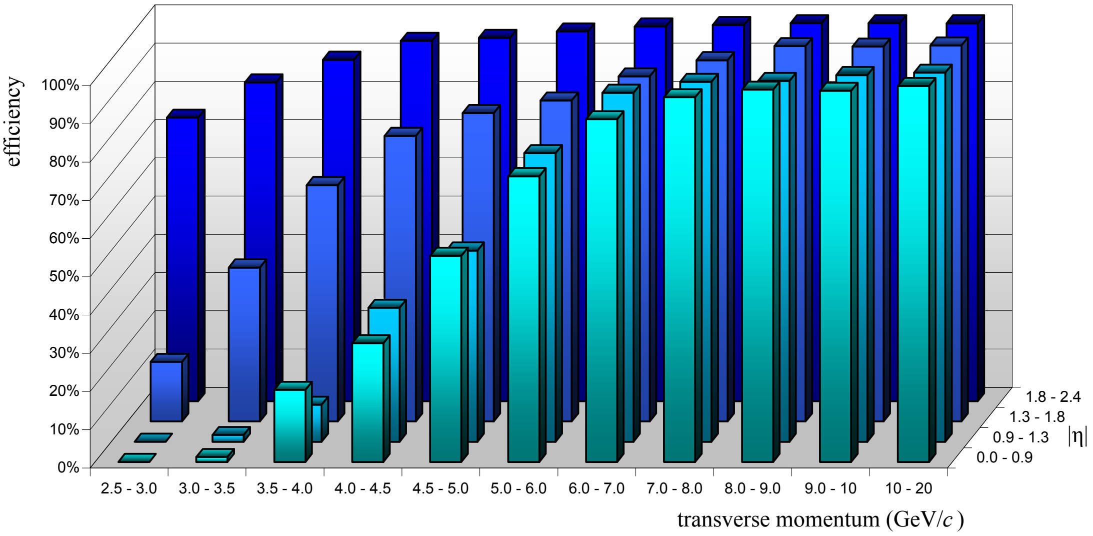


Figure 5.5: Muon reconstruction efficiency for a sample of single muons with flat distributions of transverse momentum 1 GeV/c < p_T < 20 GeV/c and pseudorapidity $|\eta| < 2.4$, reconstructed with ORCA 8.7.4.

p_T (GeV/c) vs. $\lvert\eta\rvert$	0.0 . . . 0.9	0.9 . . . 1.3	1.3 . . . 1.8	1.8 . . . 2.4
2.5 . . . 3	0	0	(15.6 ± 2.1)%	(74.2 ± 2.5)%
3 . . . 3.5	(1.3 ± 0.5)%	(1.8 ± 0.9)%	(40.3 ± 2.9)%	(83.4 ± 2.1)%
3.5 . . . 4	(18.9 ± 1.8)%	(9.7 ± 2.0)%	(61.8 ± 3.0)%	(89.3 ± 1.7)%
4 . . . 4.5	(31.0 ± 2.1)%	(35.0 ± 3.3)%	(74.7 ± 2.7)%	(94.3 ± 1.3)%
4.5 . . . 5	(53.9 ± 2.2)%	(50.0 ± 3.5)%	(80.7 ± 2.5)%	(95.1 ± 1.2)%
5 . . . 6	(74.7 ± 1.4)%	(75.5 ± 2.1)%	(84.0 ± 1.7)%	(96.7 ± 0.7)%
6 . . . 7	(89.6 ± 1.0)%	(91.3 ± 1.4)%	(90.3 ± 1.3)%	(98.0 ± 0.5)%
7 . . . 8	(95.4 ± 0.7)%	(94.1 ± 1.1)%	(94.5 ± 1.0)%	(98.4 ± 0.5)%
8 . . . 9	(97.3 ± 0.5)%	(94.4 ± 1.1)%	(98.2 ± 0.6)%	(98.9 ± 0.4)%
9 . . . 10	(97.0 ± 0.6)%	(95.9 ± 1.0)%	(98.0 ± 0.6)%	(98.9 ± 0.4)%
10 . . . 20	(98.3 ± 0.1)%	(96.5 ± 0.3)%	(98.3 ± 0.2)%	(98.9 ± 0.1)%

Table 5.1: Muon reconstruction efficiency for a sample of single muons with flat distributions of transverse momentum 1 GeV/c < p_T < 20 GeV/c and pseudorapidity $|\eta| < 2.4$, reconstructed with ORCA 8.7.4.

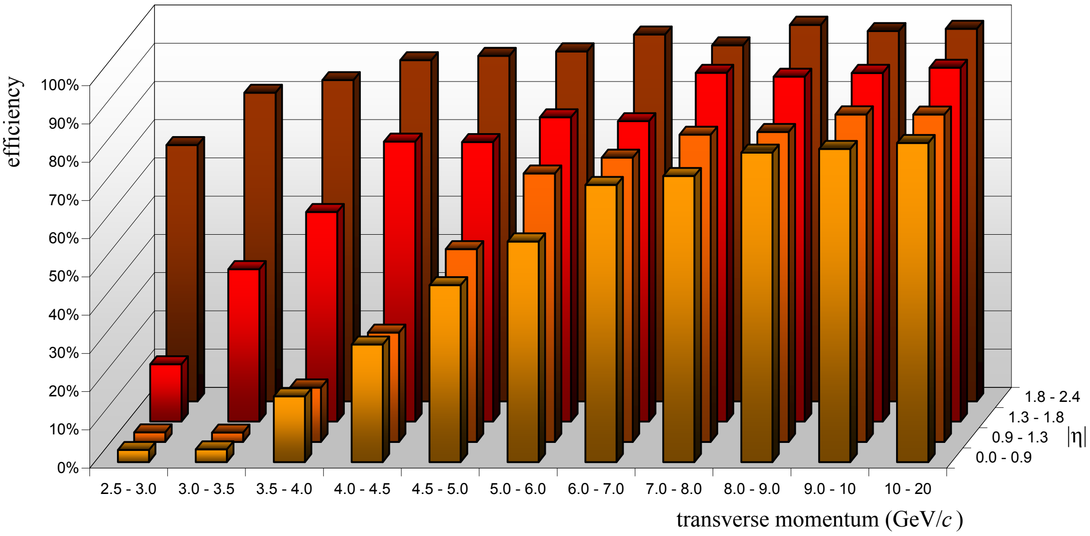


Figure 5.6: Muon reconstruction efficiency for muons in semileptonic dacays of $t\bar{t}$ events, reconstructed with ORCA 8.7.4.

p_{T} (GeV/c) vs. $\lvert\eta\rvert$	$0.0\ldots0.9$	$0.9\ldots1.3$	$1.3\ldots1.8$	$1.8\ldots2.4$
$2.5\ldots3$	(3.2 ± 0.8)%	(2.5 ± 1.1)%	(15.0 ± 2.5)%	(67.1 ± 3.6)%
$3\ldots3.5$	(3.3 ± 0.8)%	(2.4 ± 1.1)%	(39.9 ± 4.0)%	(80.7 ± 3.4)%
$3.5\ldots4$	(17.2 ± 1.9)%	(14.2 ± 2.5)%	(54.8 ± 4.0)%	(84.0 ± 3.3)%
$4\ldots4.5$	(30.7 ± 2.2)%	(28.6 ± 3.5)%	(73.2 ± 3.5)%	(89.2 ± 2.7)%
$4.5\ldots5$	(46.3 ± 2.5)%	(50.4 ± 4.4)%	(73.1 ± 3.5)%	(90.4 ± 2.6)%
$5\ldots6$	(57.7 ± 1.9)%	(70.2 ± 2.9)%	(79.6 ± 2.5)%	(91.5 ± 2.0)%
$6\ldots7$	(72.5 ± 1.8)%	(74.3 ± 2.7)%	(78.6 ± 2.9)%	(96.0 ± 1.4)%
$7\ldots8$	(74.8 ± 1.9)%	(80.3 ± 2.8)%	(91.2 ± 2.2)%	(93.2 ± 2.1)%
$8\ldots9$	(80.9 ± 1.9)%	(81.0 ± 2.8)%	(90.2 ± 2.5)%	(98.5 ± 1.0)%
$9\ldots10$	(81.9 ± 2.0)%	(85.6 ± 2.9)%	(91.2 ± 2.2)%	(96.9 ± 1.5)%
$10\ldots20$	(83.5 ± 0.8)%	(85.6 ± 1.2)%	(92.5 ± 0.9)%	(97.5 ± 0.6)%

Table 5.2: Muon reconstruction efficiency for muons in semileptonic dacays of $t\bar{t}$ events, reconstructed with ORCA 8.7.4.

	μ	e	π	K	other
b jets	84.6%	0.2%	10.9%	2.8%	1.4%
c jets	78.7%	0.2%	14.6%	4.5%	1.9%
QCD jets	44.3%	0.2%	39.0%	11.5%	5.0%
$t\bar{t}$	77.6%	0.3%	16.4%	3.9%	1.8%

Table 5.3: Purity of trivial muon identification, for reconstructed tracks inside jets for different jet flavours and data samples (ORCA 8.7.4). “b jets” and “c jets” refer to tracks inside jets of that specific flavour, from the flavour-enriched di-jet samples. “QCD jets” and “$t\bar{t}$” refer to all the jets from those samples.

and in the full endcap region ($1.8 < |\eta| < 2.4$) it rests above 90% down to $p_T > 4$ GeV/c and roughly above 75% for $p_T > 2.5$ GeV/c, as very forward tracks reach the forward muon chambers even with low p_T.

Due to the complex environment the reconstruction of muons within jets has a worse performance, especially in the barrel region. This can be seen comparing Fig. 5.5 and Tab. 5.1 to Fig. 5.6 and Tab. 5.2, showing the same efficiency data for muons reconstructed in the $t\bar{t}$ semileponic dacay samples. In the overlap and endcap regions the muon reconstruction efficiency drops only slightly, but in the barrel ($|\eta| < 1.3$) it is noticeably worse in the 5 GeV/$c < p_T <$ 10 GeV/c range: just above 80% for $p_T > 8$ GeV/c and 50% for $p_T > 5$ GeV/c.

The second effect that reduces the effectiveness of the soft muon b tagging algorithm is the misidentification of other particles reaching the muon chambers, usually $\pi^\pm$ and $K^\pm$ mesons, as muons. This can be described by the “purity” of the muon selection in different data samples.

Table 5.3 shows the purity of muon identification for tracks inside jets. These tracks were matched to muons reconstructed using the Global Muon Reconstructor algorithm from ORCA version 8.7.4, as explained above. “b jets” and “c jets” refer to tracks inside actual b and c reconstructed jets; the jet flavour has been assigned looking for the most energetic parton in a cone around the jet axis, as described in Section 5.1. These come from flavour-enriched di-jet samples with reconstructed jet energy from 30 GeV to over 200 GeV. “QCD jets” refer to tracks in QCD samples whith no preselection on the jet flavour, thus heavily dominated by light jets. As can be seen, for these jets the purity is rather low: about 50% of tracks identified as muons are actually light mesons (π or K); 5% are heavier particles, mainly protons; only 44.3% are actually muons. The net effect of this misidentification is to artificially increase by over a factor two the observed branching ratio of light jets to reconstructed muons, which in turns adversely effect the muon b tagging purity.

5.3.3 Electron identification

The electron identification algorithms available in ORCA are aimed at the reconstruction of hard, isolated electrons. As such, they are unsuitable for the task at hand, the soft electron b tag.

A dedicated tool for the identification of non-isolated electrons in the barrel region has thus been developed and integrated in the soft lepton b tagging frameork for ORCA.

This tool is based on the extrapolation of reconstructed tracks from the outermost hit in the tracker to the electromagnetic calorimeter (ECAL) and the analysis of the electromagnetic shower in the calorimeter in a region around the extrapolated track.

The following selection cuts are applied to the reconstructed tracks:

- transverse momentum $p_{\mathrm{T}} > 2\ \mathrm{GeV}/c$; lower p_{T} tracks are deflected too much by the magnetic field, and barely reach the electromagnetic calorimeter.

- pseudorapidity of the reconstructed track $|\eta| < 1.2$; within the barrel region.

- at least 8 reconstructed hits in total (pixel and silicon strip detectors); the improved track quality is required to decrease the misidentification rate.

The background to this selection comes from showers looking like an electron shower. The three most common types of background are:

- a charged hadron with a significant energy loss in electromagnetic section of the calorimeter;

- a neutral pion with a charged hadron tracking through its electromagnetic shower cone;

- an electron-positron pair production by a photon which converted in the material before the calorimeter.

The standard clusterization[54] was used on ECAL energy deposits. The extrapolated tracks are matched with ECAL clusters based on the distance

Measurement	Variables
Covariance of cluster energy distribution	$\sigma_{\eta\eta}$, $\sigma_{\eta\varphi}$, $\sigma_{\varphi\varphi}$
Electromagnetic shower shape	$\frac{E_{seed}}{E_{2\times2}}, \frac{E_{seed}}{E_{3\times3}}, \frac{E_{3\times3}-E_{seed}}{E_{5\times5}-E_{seed}}, \frac{E_{2\times2}}{E_{5\times5}}$
ECAL to HCAL cluster energy ratio	$\frac{E_{\mathrm{ECAL}}}{E_{\mathrm{HCAL}}+E_{\mathrm{ECAL}}}$
ECAL cluster energy track momentum ratio	$\frac{E_{\mathrm{ECAL}}-p_{trk}}{E_{\mathrm{ECAL}}+p_{trk}}$

Table 5.4: List of variables used to identify electrons, where E_{seed}, $E_{N\times N}$ and E_{ECAL} respectively are the energy deposit in the most energetic crystal (cluster seed), maximal energy deposit in a square of $N \times N$ ECAL crystals around the seed and total energy of the cluster; p_{trk} is the extrapolated track momentum at ECAL entrance; E_{HCAL} is the sum of energy deposits in HCAL towers next to the ECAL cluster.

between the track impact point on the crystal surface and the threedimensional cluster centre: the track closest to the cluster is matched if such distance is less than $\Delta r(\text{track-cluster}) = \sqrt{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2} < 12$ cm.

The development of electromagnetic and hadronic showers are quite different, so the shower shape information can be used to discriminate electrons and positrons against hadrons: electrons deposit almost all their energy in the electromagnetic section of the calorimeter, while hadrons are typically much more penetrating. In addition, about 90% of the electromagnetic shower is confined to one Molière radius [87]. Fluctuations cause the energy deposition to vary from the average in a correlated fashion among the crystals.

The best discrimination against hadrons can then be achieved taking into account the track momentum, the energy deposits in the electromagnetic (ECAL) and hadronic (HCAL) calorimeters, and the shape of the electromagnetic shower. The relevant variables are listed in Tab.5.4 and shown in Figs. 5.7 through 5.9 for signal electrons, coming from the direct ($\mathrm{b} \to \ell$) or chain ($\mathrm{b} \to \mathrm{c} \to \ell$) decay of b hadrons, and background tracks. To provide a adequate quantity of signal electrons and background tracks the $\mathrm{t\bar{t}}$ and QCD samples described in Sec. 5.2 were used.

The electromagnetic energy fraction of an electron candidate is defined as $E_{\mathrm{ECAL}}/(E_{\mathrm{HCAL}} + E_{\mathrm{ECAL}})$, where E_{HCAL} is the sum of energy deposits in HCAL towers next to the ECAL cluster and E_{ECAL} is the ECAL cluster

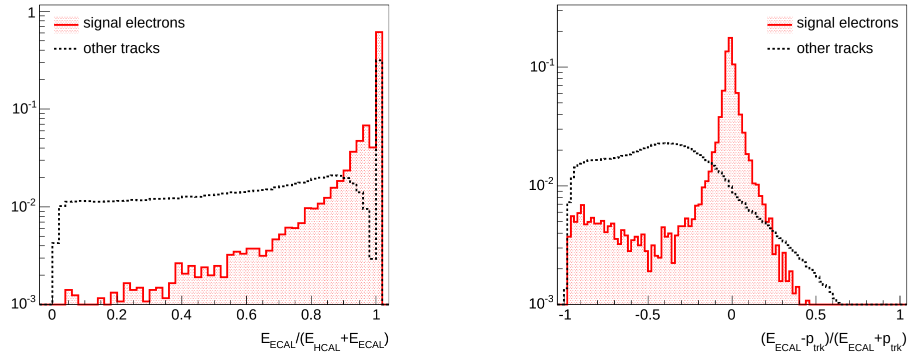


Figure 5.7: Distribution of $E_{\text{HCAL}}/E_{\text{ECAL}}$ (left) and $(E_{\text{ECAL}}-p_{trk})/(E_{\text{ECAL}}+p_{trk})$ (right) for signal electrons and for other charged particle tracks. The variables are described in Tab. 5.4.

energy. A similar ratio can be defined for the electromagnetic energy with respect to the track momentum, $(E_{\text{ECAL}} - p_{trk})/(E_{\text{ECAL}} + p_{trk})$. Fig. 5.7 shows the distribution of these variables for signal electrons and other charged particle tracks from the $\mathrm{t\bar{t}}$ and QCD samples.

The covariance of cluster energy distribution, $\sigma_{\eta\eta}$, $\sigma_{\eta\varphi}$ and $\sigma_{\varphi\varphi}$, describe the correlations between energy deposits in the ECAL crystals. Fig. 5.8 shows the distribution of these variables for signal electrons and other charged particle tracks from the $\mathrm{t\bar{t}}$ and QCD events.

The transverse development of the shower is characterized by four measurements of cluster energy repartition: $E_{seed}/E_{2\times2}$, $E_{seed}/E_{3\times3}$, $(E_{3\times3} - E_{seed})/(E_{5\times5} - E_{seed})$, and $E_{2\times2}/E_{5\times5}$, where E_{seed} and $E_{N\times N}$ are respectively energy deposit in the cluster seed and maximal energy deposit in a square of $N \times N$ ECAL crystals around the seed. Fig. 5.9 shows the distribution of cluster energy repartition for signal electrons and other charged particle tracks from $\mathrm{t\bar{t}}$ and QCD events.

These variables are then combined together with an Artificial Neural Network, whose output value is used as discriminant to separate electromagnetic and hadronic showers. The distributions of the neural network output for signal electrons and other tracks are shown in Fig. 5.10. Efficiency and misidentification rate for the current implementation of the non-isolated electron identification are shown in Fig. 5.11.

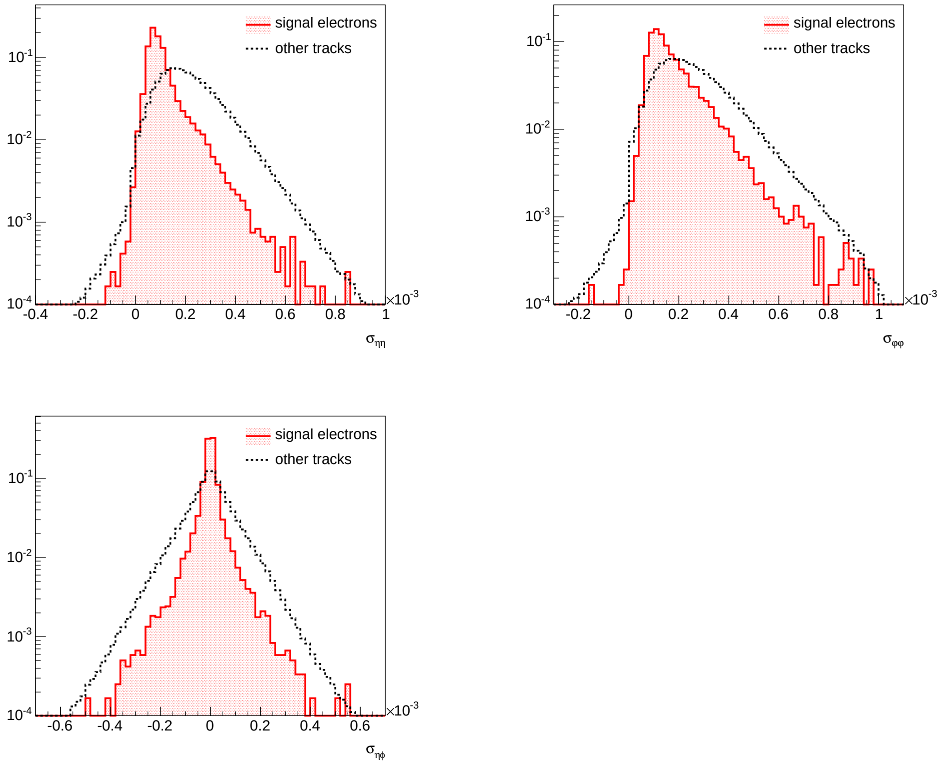


Figure 5.8: Covariance of cluster energy distribution $\sigma_{\eta\eta}$ (top left), $\sigma_{\varphi\varphi}$ (top right), $\sigma_{\eta\varphi}$ (bottom) for signal electrons and for other charged particle tracks. The variables are described in Tab. 5.4.

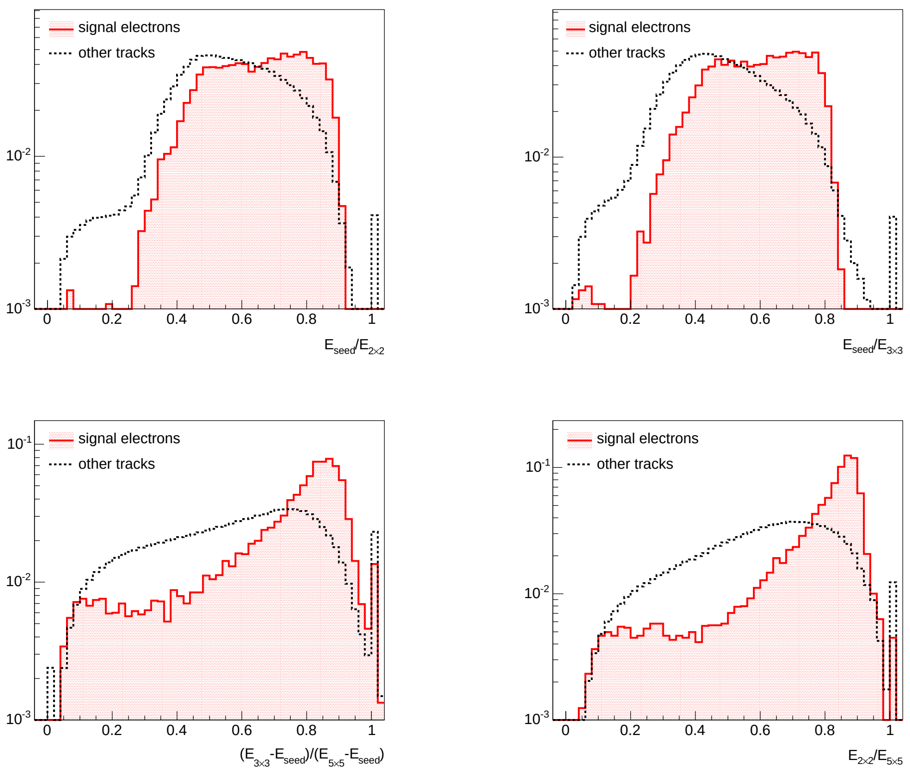


Figure 5.9: Distribution of cluster energy repartition $E_{seed}/E_{2\times2}$ (top left), $E_{seed}/E_{3\times3}$ (top right), $(E_{3\times3}-E_{seed})/(E_{5\times5}-E_{seed})$ (bottom left), $E_{2\times2}/E_{5\times5}$ (bottom right) for signal electrons and for other charged particle tracks. The variables are described in Tab. 5.4.

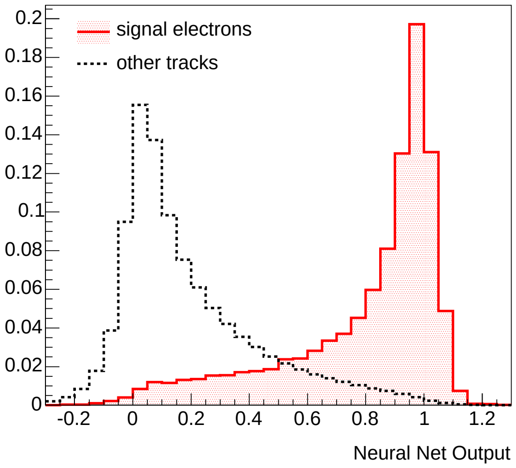


Figure 5.10: Distribution of the neural network output for signal electrons (red solid) and for other charged particle tracks (black dashed).

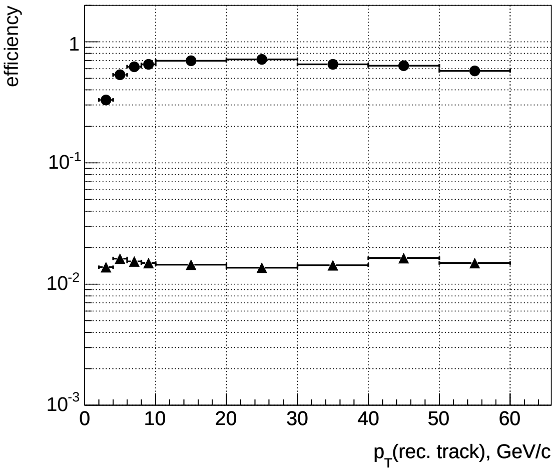


Figure 5.11: Performance of the non-isolated electron identification: efficiency (circles) and misidentification rate (triangles) as a function of the reconstructed track transverse momentum.

5.3.4 Lepton b tagging variables

In order to improve the rejection of non-b jets, the jet axis, used as an approximation of the original parton direction, is refined using the charged particle tracks inside the jet. A weighted average is performed for the pseudoangular coordinates η and φ over all tracks reconstructed inside a jet, except for the track associated to the tagging lepton. The tracks transverse momentum is used as weighting factor.

The following topological and kinematical variables are then computed:

- S_{IP}^{3D}, the lepton three-dimensional signed impact parameter significance;
- $p_{\mathrm{T},rel}$, the transverse momentum, relative to the lepton-excluded refined jet axis;
- r, the ratio between the lepton momentum and the jet energy, and
- $r_{\parallel}$, the ratio between the lepton momentum along the jet axis and the jet energy;
- ΔR, the pseudoangular distance between the lepton track and the lepton-excluded refined jet axis.

In the following Sections I will describe each of them, and show the distributions for reconstructed muons from different data samples: flavour enriched QCD di-jet samples, and the leptonic and semileptonic decay of $\mathrm{t\bar{t}}$ pairs. The distribution of each variable, normalized to unit area, will be shown for each jet flavour and lepton decay chain.

Impact parameter

The three-dimensional signed impact parameter significance S_{IP}^{3D} has been described in detail in Sec. 4.2.1. In the soft lepton b tag algorithm, the S_{IP}^{3D} of the lepton track with respect to the *original* jet axis is used.

Figure 5.12 shows the S_{IP}^{3D} distributions of jets from $\mathrm{t\bar{t}}$ and flavour enriched QCD di-jet samples, for different jet flavours and leptonic decay chains. It has a very good rejection of b against c jets (a, c) and $\mathrm{b} \rightarrow \ell$, $\mathrm{b} \rightarrow \mathrm{c} \rightarrow \ell$ against $\mathrm{c} \rightarrow \ell$ decay chains (b, d). Its performance is worse for the rejection of gluon and light jets, in particular on the QCD sample where part of gluon jets are misidentified b jets. The positive tail of the impact parameter significance for muons in light jets is due to $\pi^{\pm}$ and $\mathrm{K}^{\pm}$ from secondary decays reaching the muon chambers or decaying in flight into muons.

The negative part of the spectrum is due to the primary vertex resolution and errors in the track extrapolation, and as expected is mostly independent from the jet flavour.

Relative transverse momentum

The projection of the lepton momentum on the plane normal to the jet axis $p_{\mathrm{T},rel}$ is sensitive to the available energy in the decay process producing the lepton. It is thus able to discriminate between the decays of heavier and lighter hadrons.

Figure 5.13 shows the distribution of the relative transverse momentum of leptons reconstructed inside jets, for different jet flavours and lepton decay chains. This variable is effective in discriminating the decays of b hadrons from c and lighter hadrons. This difference is more marked looking at the decay chains, where the spectrum of $\mathrm{b} \to \ell$ decays is clearly different from the $\mathrm{b} \to \mathrm{c} \to \ell$ and $\mathrm{c} \to \ell$ ones.

The rejection of light hadrons is good for $p_{\mathrm{T},rel}$ less than 5 GeV/c. At higher momenta the fraction of non-prompt leptons produced via longer decay chains involving light mesons is predominant over the production of prompt leptons, as the multiple decays allow for a wider scattering of the decay products. Again, part of gluon jets are actually misidentified b jets, and thus show a similar spectrum.

The kink at $p_{\mathrm{T},rel} = 2$ GeV/c in the distribution for $q \to \ell$ decays in Fig. 5.13 (d) is due to the mixed population of the sample: for $p_{\mathrm{T},rel} < 2$ GeV/c these are actually indirect decays of b hadrons, while for $p_{\mathrm{T},rel} > 2$ GeV/c they are mostly muons from light and gluon jets.

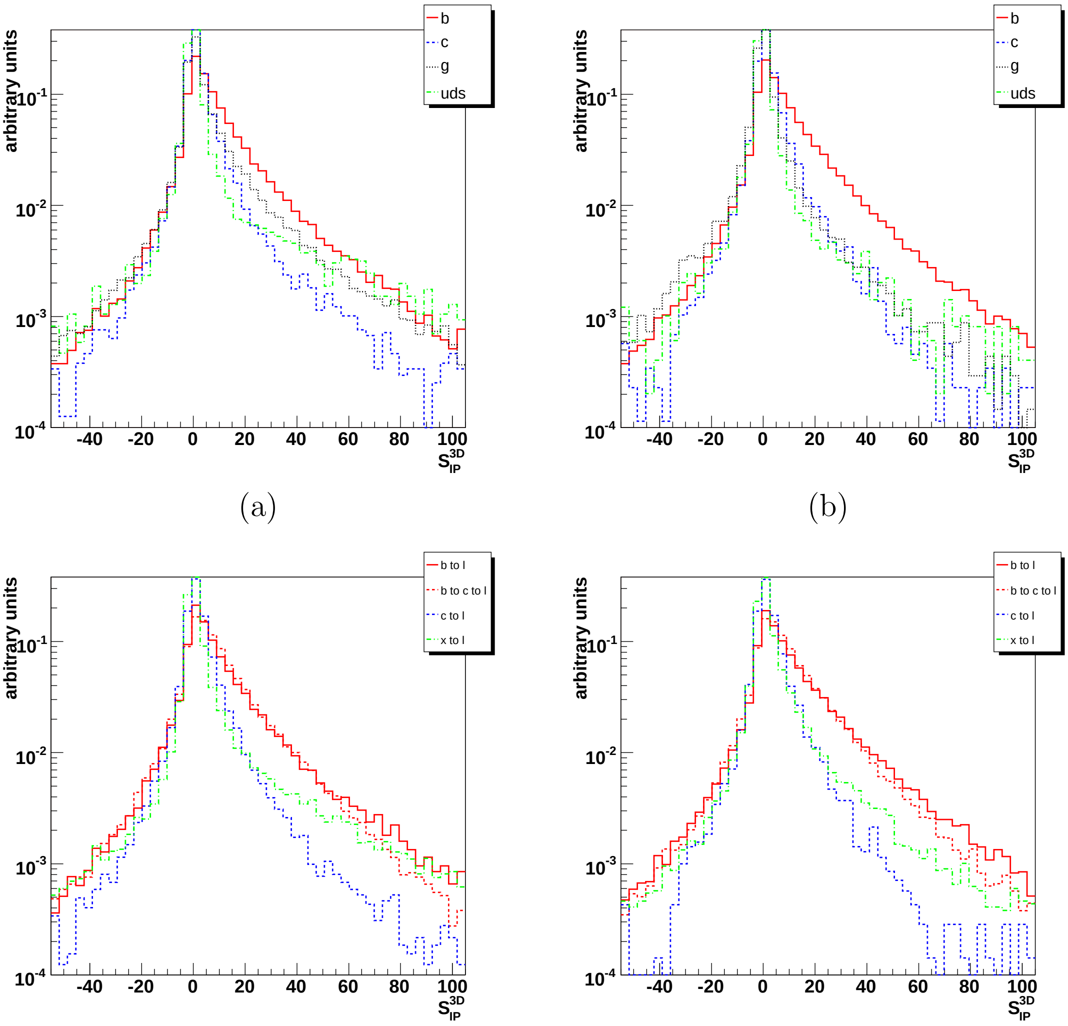
Pseudoangular distance

The pseudoangular distance ΔR in the $\eta\varphi$ plane between the lepton and the lepton-excluded refined jet axis is not effective on its own, but when combined with the above variables helps to discriminate the b from the c and light hadrons.

When applied to jets identified by the lepton decay chain, it helps to discriminate against the decay of c quarks ($\mathrm{b} \to \mathrm{c} \to \ell$, $\mathrm{c} \to \ell$).

Lepton momentum to jet energy ratio

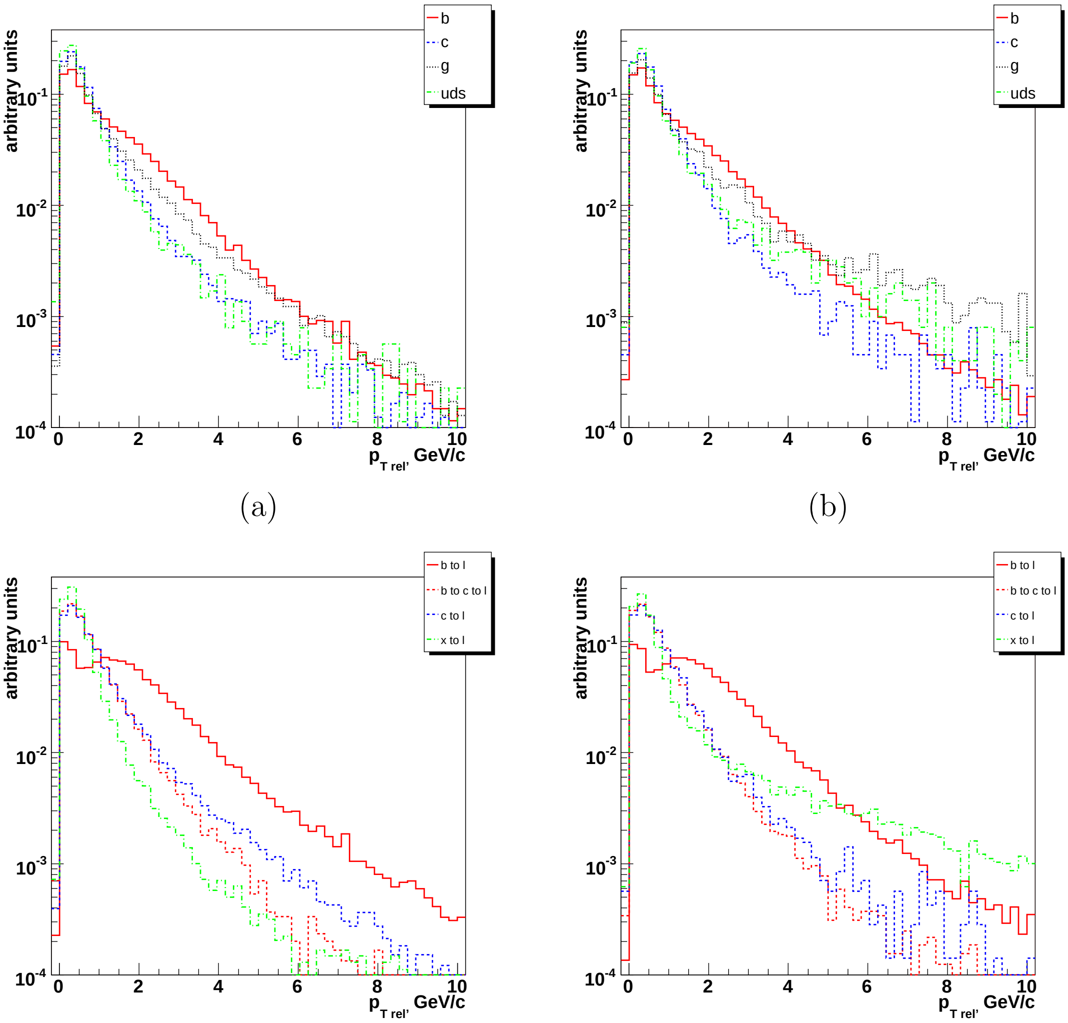
There are two useful definitions for the ratio of the lepton momentum to the jet energy: either using the module of the momentum (r) or the component parallel to the jet axis ($r_{\parallel}$). The two distributions are shown in Figs. 5.15 and 5.16, respectively.



(a) (b)

(c) (d)

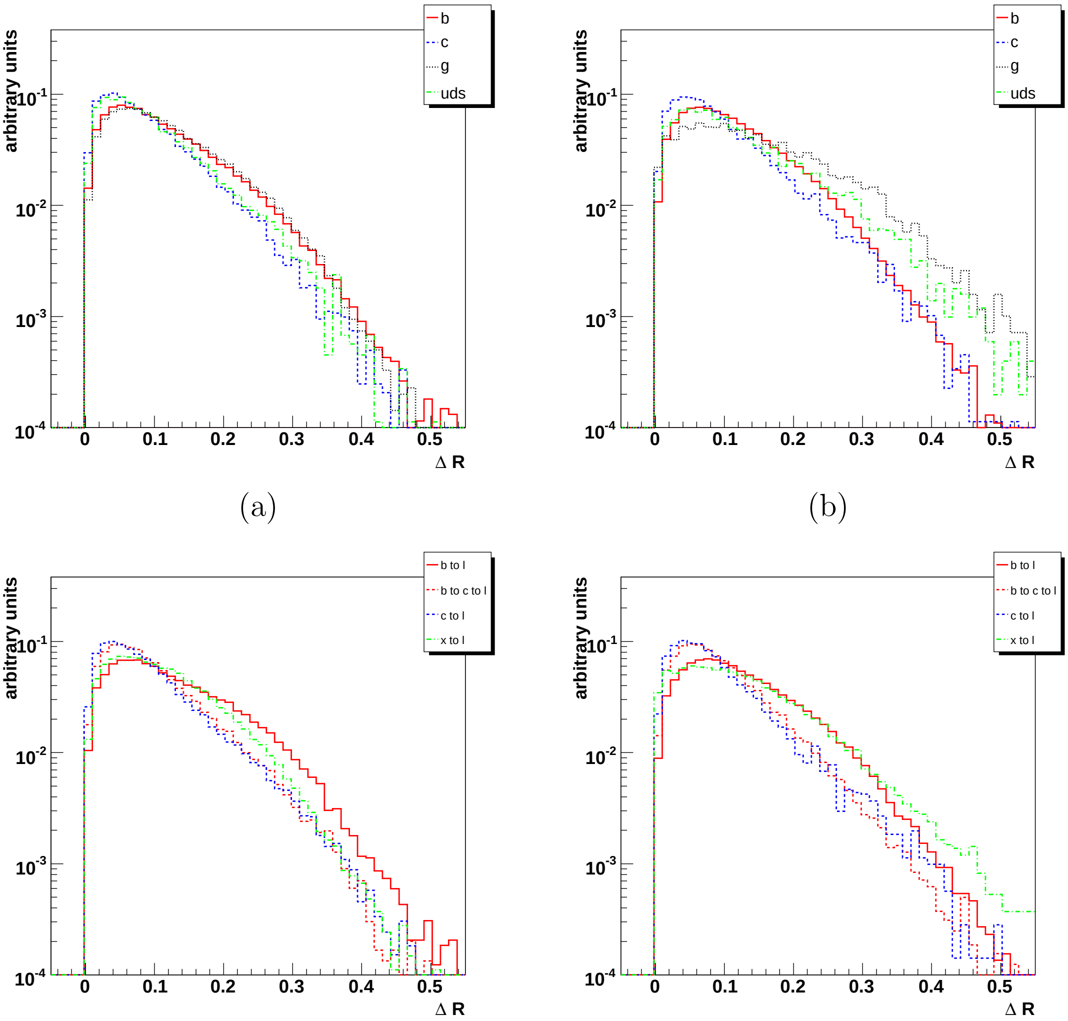
Figure 5.12: Three-dimensional signed impact parameter significance S_{IP}^{3D} for jets in the flavour-enriched di-jet samples (left) and $t\bar{t}$ samples (right) for each jet type, as identified by the parton content (top) and lepton decay chain (bottom) analysis.



(a) (b)

(c) (d)

Figure 5.13: Transverse momentum relative to the lepton-excluded refined jet axis $p_{\mathrm{T},rel}$ for jets in the flavour-enriched di-jet samples (left) and $\mathrm{t\bar{t}}$ samples (right) for each jet type, as identified by the parton content (top) and lepton decay chain (bottom) analysis.



(a) (b)

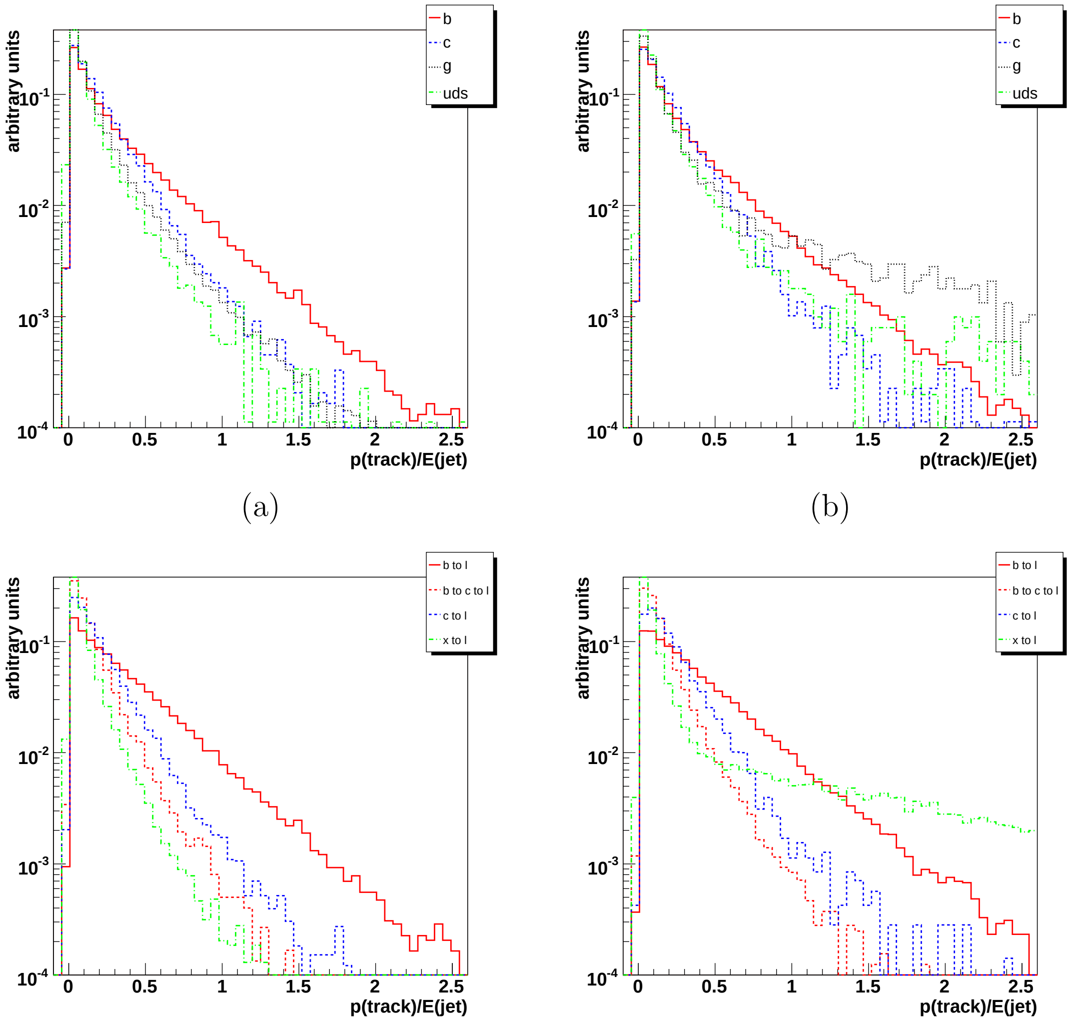
(c) (d)

Figure 5.14: Pseudoangular distance ΔR between the lepton track and the lepton-excluded refined jet axis for jets in the flavour-enriched di-jet samples (left) and $t\bar{t}$ samples (right) for each jet type, as identified by the parton content (top) and lepton decay chain (bottom) analysis.

Both work better in the cleaner environment of the QCD di-jet samples, allowing a good separation of b hadrons from the rest. In $\mathrm{t\bar{t}}$ samples gluon jets overcome the b jets distribution at high r and $r_{\parallel}$; this effect comes from the low end of the jet energy spectrum, and it is possibly due to the reconstruction of muons from τ decays close to the jet direction.

The change of slope in the distributions of r and $r_{\parallel}$ for the $q \rightarrow \ell$ decays in Fig. 5.15 (d) and Fig. 5.16 (d) are again due to the etherogeneous composition of jets labelled $q \rightarrow \ell$: for $r < 0.5$ and $r_{\parallel} < 2$ these are actually indirect decays of b hadrons, while for $r > 0.5$ and $r_{\parallel} > 2$ they are mostly muons from light and gluon jets.

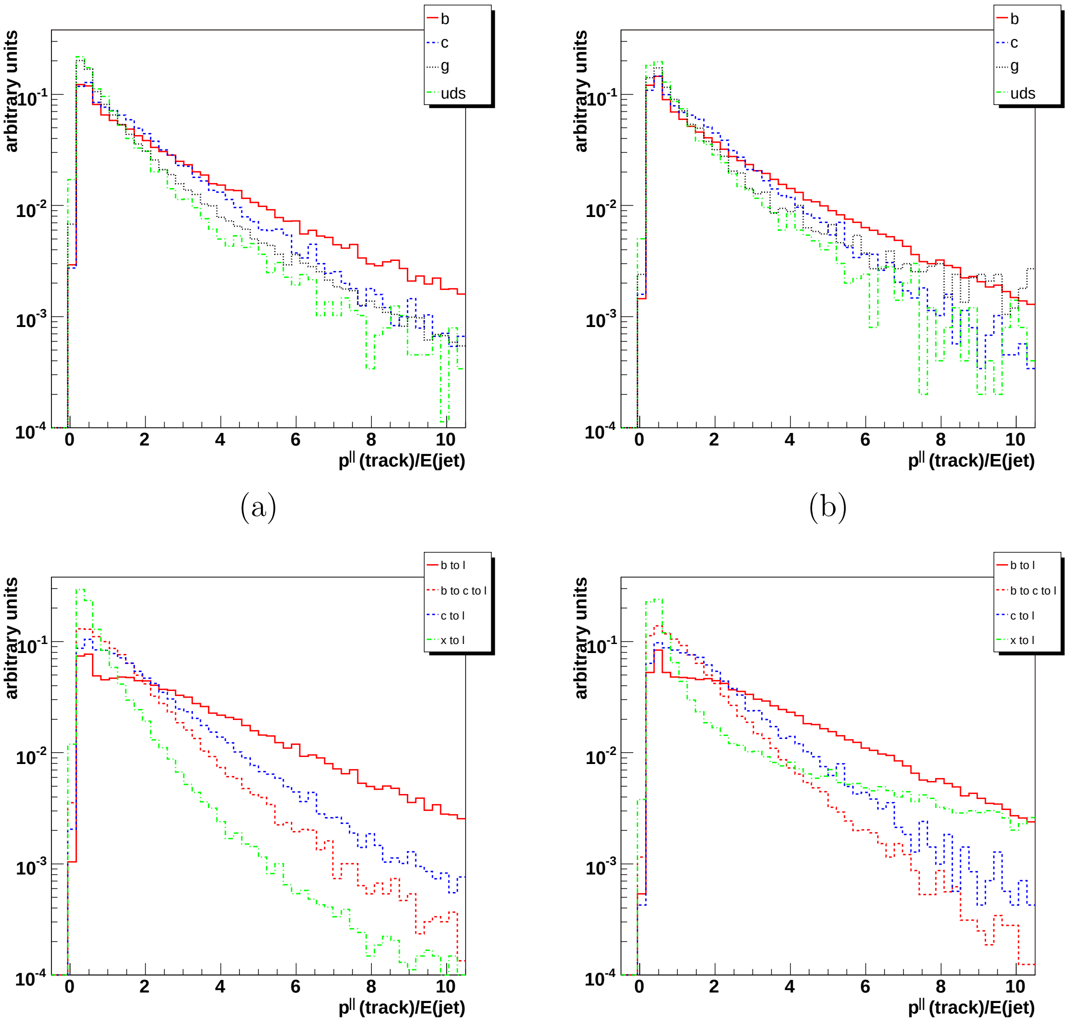
Overall, r and $r_{\parallel}$ distributions have obviously very similar characteristics. The former may be preferred for its independence on ΔR. The second has the advantage of being tightly linked to the lepton rapidity y respect to the jet axis, as $\tanh(y) = r_{\parallel}$.



(a) (b)

(c) (d)

Figure 5.15: Ratio r between the lepton momentum p(track) and the jet energy E(jet) for jets in the flavour-enriched di-jet samples (left) and $t\bar{t}$ samples (right) for each jet type, as identified by the parton content (top) and lepton decay chain (bottom) analysis.



(a)

(b)

(c)

(d)

Figure 5.16: Ratio $r_{\|}$ between the lepton momentum along the lepton-excluded refined jet axis $p_{\|}$(track) and the jet energy E(jet) for jets in the flavour-enriched di-jet samples (left) and $t\bar{t}$ samples (right) for each jet type, as identified by the parton content (top) and lepton decay chain (bottom) analysis.

5.3.5 Combination of tagging variables

Some of the described variables can be directly used to discriminate between b and non-b jets. They are however much more powerful when combined together.

A linear combination of S_{IP}^{3D} and $p_{\mathrm{T},rel}$

$$y = w_0 + w_1 \cdot S_{\mathrm{IP}}^{3D} + w_2 \cdot p_{\mathrm{T},rel} \tag{5.1}$$

gives promising results, but has a big drawback, i.e. the number of free parameters. In order to achieve the best performance, both the coefficients of the linear combination w_i and the cut on the result $y > y_{cut}$ have to be adjusted for each different trade-off between efficiency and purity. Additionally, the different jet flavours are not linearly separable in the $S_{\mathrm{IP}}^{3D}, p_{\mathrm{T},rel}$ plane, and thus a linear combination cannot give the best performance. Nevertheless, the results of this simple study have been used as the reference performance for the development of more powerful and complex techniques.

These problems can be solved building a nonlinear function y of the tagging variables $\mathbf{x} = \{x_i\} = \{S_{\mathrm{IP}}^{3D}; p_{\mathrm{T},rel}; \dots\}$:

$$y = y(\mathbf{x}). \tag{5.2}$$

If the function y is correctly chosen, cutting on its value $y > y_{cut}$ gives improved purity affecting as little as possible the efficiency.

The key point now is the determination of $y(\mathbf{x})$. A possible solution is to choose a sufficiently general parametric form $y = y(\mathbf{x}; \mathbf{w})$, and determine the parameters $\mathbf{w}$ from the data itself. We know from literature[88, 89, 90] that a feed-forward neural network, with at least two layers of weights and a sufficiently large number of sigmoidal hidden units can approximate to an arbitrary precision any continuous mapping between finite-dimensional spaces[91]. This kind of networks, with at least two layers of weights and hidden neuros with sigmoidal activation functions, are called multi layer perceptrons, as they can be seen as a generalization of the first perceptron model which had a single layer of weights and a threshold activation function.

The use of such a neural network is then an obvious choice for modeling the unknown nonlinear mapping from the tagging variables to the discriminant. A feed-forward neural network is usually structured in layers of neurons: the first layer, called input layer, corresponds to the input variables $\mathbf{x}$; each successive layer is made of processing units[2] up to the last, or

---

[2] The names "unit" and "neuron" are used interchangeably, and reflect two point of view: mathematical, and biological. The same is true for the names "connection" and "synapse".

output, layer. The hidden and output neurons j take as input, or activation a_j, a weighted combination $\sum w_{ij} y_i$ of the outputs y_i of neurons from the previous layers plus an optional bias w_{0j}; a nonlinear activation function $g(\cdot)$ is then applied to the activation, giving the output of that neuron $y_j = g(a_j) = g(\sum w_{ij} y_i + w_{0j})$. Following the biological model, the weights w_{ij} are often associated to the connections, or synapses, between the neurons i and j rather than to the neurons themselves.

The activation function can be any real function of its input. The most simple are *threshold* or *step* function

$$g(a) \equiv \theta(a) \equiv \begin{cases} 1 & a \geq 0 \\ 0 & a < 0 \end{cases} \tag{5.3}$$

and the identity, or *linear* activation

$$g(a) \equiv a. \tag{5.4}$$

The most versatile, however, is the *sigmoid* or *logistic* activation function:

$$g(a) \equiv \frac{1}{1 + e^{-a}} \tag{5.5}$$

In the limit of small a it can approximate a linear function, while in the limit for $a \to \pm\infty$ it approximates a threshold function. Since these conditions can be achieved for an arbitrary input pattern adjusting the weights of the synapses leading to and from the neuron, the training of the network with sigmoid activation functions can automatically adjust the shape of its response to the input patterns.

Using a nonlinear function is essential to model complex mappings, while a continuous function allows the use of faster training techniques involving the partial derivatives of the activation functions. This is especially easy to compute for the sigmoid, as

$$g'(a) = \frac{e^{-a}}{(1 + e^{-a})^2} = g(a)(1 - g(a)) \tag{5.6}$$

For each application, the topology of the neural network, i.e. the number of hidden layers, the number of units per layer, and the structure of the connections, must be determined experimentally.

The ability of a network to “learn” scales with its complexity, which is roughly measured by the number of connections, but so does the number of parameters to determine: a feed-forward network has one parameter per neuron (bias) and per synapse (weight). Since the training of a network can be seen as a minimization problem in an P-dimensional space, where P is the number of parameters, the time and number of patterns needed to succesfully train the network can soon exceed the available data and computing resources. An additional drawback of overly complex networks is the risk of overtraining: the network can adapt too well to the sample used for training, modeling the statistical fluctuations rather than the underlying structure.

Regarding the structure of the connections, the most common choice is a full connection between successive layers only: each neuron of layer n has as input the (weighted) outputs of all and only neurons from layer $n-1$. More complex structures are often used to hard code into the network knowledge of the process to be modeled, such as symmetries and invariances in the input variable space.

Training

The training of a neural network consists in presenting to the input neurons patterns x_i for which the expected outcome t_k, or target, is known (*supervised learning*). The corresponding output values y_k are then computed, and the current performance of the network is evaluated by an error function $E_{\mathbf{w}}(\mathbf{y}, \mathbf{t})$. All the network weights $\mathbf{w}$ are then adjusted in order to minimize the error function.

Since the activation functions and their derivatives are nonlinear, the minimization process can be explicitly solved only integrating a system of P nonlinear differential equations. This is usually not feasible and an iterative method is thus adopted, where at each step the weights are updated in order to decrease the energy function. The starting weights for this iterative approach are set to small random values, in order to have all the sigmoid units in the linear region.

The most simple iterative training method is the *gradient descent*: the gradient of the error function with respect to the weights $\frac{\partial E}{\partial w_{jk}}$ is computed, and the weights are adjusted in the opposite direction, usually by an amount proportional to the partial derivative along that axis: $\Delta w_{jk} = -\eta \frac{\partial E}{\partial w_{jk}}$. More powerful methods have been developed, such as the *conjugate gradient* and *quasi-Newton* methods, making implicit or explicit use the Hessian matrix $\frac{\partial^2 E}{\partial w_{jk} \partial w_{lm}}$ or approximations, allowing more efficient updates at each step, and thus a faster training.

All these training algorithms require the knowledge of the derivatives of the error function with respect to the weights. These can be calculated via the chain rule

$$\frac{\partial E}{\partial w_{ji}} = \frac{\partial E}{\partial a_j}\frac{\partial a_j}{\partial w_{ji}} = \delta_j y_i \tag{5.7}$$

$$\delta_j \equiv \frac{\partial E}{\partial a_j} \tag{5.8}$$

For the output units the evaluation of δ_j is straightforward:

$$\delta_j \equiv \frac{\partial E}{\partial a_j} = g'(a_j)\frac{\partial E}{\partial y_j}. \tag{5.9}$$

For the hidden units it can be evaluated applying again the chain rule, and taking into account the topology of the network:

$$\delta_j \equiv \frac{\partial E}{\partial a_j} = \sum_k \frac{\partial E}{\partial a_k}\frac{\partial a_k}{\partial a_j} \tag{5.10}$$

$$= \sum_k \delta_k w_{kj} g'(a_j) \tag{5.11}$$

where the neuron k is in the layer succesive to that of neuron j. Thus, all the derivatives can be computed starting from the output layer and working backwards: for this reason the update rule of feed-forward networks is called *back propagation.*

Whatever the training method adopted, the use of a limited number of training patterns involves the risk of overtraining, a phenomenon in which the network adapts too much to the training sample, modeling its distribution and the particular contributions due to statistical fluctuations more than the underlying mapping between inputs and outputs. This effect is less prominent in smaller networks, which simply don't have the complexity required to follow the fast fluctuations associated to statistical noise. In general, this effect can largely be avoided monitoring the performance of network under training on a different sample, called test sample. To ensure the maximum generalization capabilities the training is stopped as soon as the error, while still decreasing on the training sample, starts to increase on the test sample, a clear symptom of overtraining.

Error functions

Despite the many free parameters involved in building a neural network, its topology only affects its training requirements and approximation quality. The properties of the mapping or classification realized depend instead on the error function used for training. For many error functions the δ_j assume a particularly simple form when an appropriate activation function is used for the output neurons.

The most common pairing of output neuron activation and error functions are a linear activation $g(a) = a$ with a sum-of-square error function:

$$E_{\mathbf{w}}(\mathbf{y}, \mathbf{t}) = \frac{1}{2} \sum_{n} \sum_{k=1}^{c} (y_k^n - t_k^n)^2 \tag{5.12}$$

where k runs over all the output neurons and n over all the training patterns. This combination yields for the δ_j of the output neurons the simple formula

$$\delta_j = g'(a_j) \frac{\partial E}{\partial y_j} = (y_j - t_j). \tag{5.13}$$

This simple form for the error function can actually be inferred from the maximum likelihood principle, applied to the distribution of the network output values under the hypothesis of independent outputs and target values described by a deterministic function of the inputs plus a random Gaussian noise,

$$t_k = f_k(\mathbf{x}) + \phi(0, \sigma_k). \tag{5.14}$$

A sufficiently general network will, after training to the minimum of the error function, reproduce the deterministic function

$$y_k(\mathbf{x}) = f_k(\mathbf{x}). \tag{5.15}$$

Assuming Gaussian errors, an optimal discriminant between b and non-b jets can be approximated by a feed-forward neural network with sigmoid hidden neurons and linear output neurons, trained with a sum-of-square error function over patterns with $\mathbf{x} = \{S_{\mathrm{IP}}^{3D}, p_{\mathrm{T},rel}, \Delta R, r\}$ and target $t = 1$ for b jets and $t = 0$ for non-b jets. The average and range of the network output will depend on the relative fractions of b and non-b jets present in the training sample, but one can expect it to lay roughly in the $[0 \ldots 1]$ range, and to give increasingly better b selection purity for a cut $y > y_{cut}$, as y_{cut} approaches unity.

While such a network can be effectively built and trained to give good performance, its theoretical bases can be improved if we approach b tagging not looking for a discriminant, but as a more general classification problem.

Bayesian classification and cross entropy

In a classification problem the input patterns $\mathbf{x}$ belong to c distinct classes $\mathcal{C}_k$. These classes can overlap in pattern space, so a clean separation between them may not be possible. The problem consists defining the *decision boundaries* in order to classify new paterns, for which the true class membership is not known, minimizing the probability of misclassification. This can be done assigning the pattern $\mathbf{x}$ to the class $\mathcal{C}_k$ that maximizes its membership probabilty $P(\mathbf{x} \in \mathcal{C}_k)$. If the pattern distributions inside each class $p_k(\mathbf{x})$ are known, along with the fraction f_k of patterns belonging to each class, the membership probability for each class can be computed through Bayes theorem of conditional probabilities

$$P(\mathrm{A}|\mathrm{B}) = \frac{P(\mathrm{B}|\mathrm{A}) \cdot P(\mathrm{A})}{P(\mathrm{B})} \tag{5.16}$$

applied to class membership

$$\begin{aligned} P(\mathbf{x} \in \mathcal{C}_k) \equiv P(\mathcal{C}_k|\mathbf{x}) &= \frac{P(\mathbf{x}|\mathcal{C}_k) \cdot P(\mathcal{C}_k)}{P(\mathbf{x})} \\ &= \frac{P(\mathbf{x}|\mathcal{C}_k) \cdot P(\mathcal{C}_k)}{\sum_{j=1}^{N} P(\mathbf{x}|\mathcal{C}_j) \cdot P(\mathcal{C}_j)} \equiv \frac{p_k(\mathbf{x}) \cdot f_k}{\sum_{j=1}^{N} p_j(\mathbf{x}) \cdot f_j} \end{aligned} \tag{5.17}$$

identifying the class membership probability $P(\mathbf{x} \in \mathcal{C}_k)$ with the *a posteriori* probability $P(\mathcal{C}_k|\mathbf{x})$, the class pattern distribution $p_k(\mathbf{x})$ with the conditional probability $P(\mathbf{x}|\mathcal{C}_k)$, and the fraction f_k of patterns in class $\mathcal{C}_k$ with the *a priori* probability $P(\mathcal{C}_k)$ for that class.

The unconditional probability $p(\mathbf{x})$ in Eq. 5.17 plays the role of a normalization, and as such can be dropped when looking for the class with the highest probability; in the general case it can be computed from the conditional probabilities and the prior probabilities for class membership.

The class membership of each pattern for c classes can be described by a 1-of-c coding scheme, with one variable t_k per class $\mathcal{C}_k$ assuming binary values: for a pattern $\mathbf{x}$ belonging the j-th class $\mathcal{C}_j$, $t_k = \delta_{jk}$. The classification problem for a new pattern $\mathbf{x}$ can then be addressed with a feed-forward neural network if its output $y_k(\mathbf{x})$ can approximate the class conditional probabilities $P(\mathcal{C}_k|\mathbf{x})$[91].

Building and training a neural network whose outputs approximate the bayesian class conditional posterior probabilities is possible using the correct error function. This can be derived maximizing the likelihood for the observation for each pattern $\mathbf{x}^n$ of the classification t_k^n, with the requirement that the class conditional probabilities are given by $y_k^n = P(\mathcal{C}_k|\mathbf{x}^n)$:

$$p(\mathbf{t}^n|\mathbf{x}^n) = \prod_{k=1}^{c} (y_k^n)^{t_k^n}. \tag{5.18}$$

Since the patterns are independent of each other, the likelihood for the full sample can be written as

$$\mathcal{L}(\mathbf{t}|\mathbf{x}) = \prod_n \prod_{k=1}^{c} (y_k^n)^{t_k^n}. \tag{5.19}$$

Instead of maximizing the likehood function a better approach is to minimize its negative logarithm:

$$E(\mathbf{y}, \mathbf{t}) = -\log \mathcal{L}(\mathbf{t}|\mathbf{x}) = -\sum_n \sum_{k=1}^{c} t_k^n \ln(y_k^n). \tag{5.20}$$

Since it is independent from the output values y_k^n, it can be convenient to subtract off the minimum of the function, reached for $y_k^n = t_k^n$

$$E_{min} = -\sum_n \sum_{k=1}^{c} t_k^n \log(t_k^n), \tag{5.21}$$

and obtain

$$E(\mathbf{y}, \mathbf{t}) = -\sum_n \sum_{k=1}^{c} t_k^n \log\left(\frac{y_k^n}{t_k^n}\right). \tag{5.22}$$

The function in Eq. 5.20 is known in literature as *cross entropy* of $\mathbf{y}$ and $\mathbf{t}$, $H(\mathbf{y}, \mathbf{t})$, while its minimum in Eq. 5.21 is clearly the entropy of $\mathbf{t}$, $H(\mathbf{t})$. The function in Eq. 5.22 is known as Kullback-Leibler divergence, $KL(\mathbf{t}, \mathbf{y})$; it is non-negative, with a minimum for $y_k^n = t_k^n$, $KL(\mathbf{t}, \mathbf{t}) = 0$.

In order to be interpreted as class conditional probabilities, the output values y_k must lie in the $[0, 1]$ range, $0 \leq y_k \leq 1$, and sum to unity, $\sum_k y_k = 1$. Network training will lead to approximation of these properties. However, they can be directly embedded in the neural network choosing an appropriate activation for the output neurons, the *softmax* function:

$$y_k = \frac{e^{a_k}}{\sum_j e^{a_j}} \tag{5.23}$$

with the sum running over all the output neurons. This function is more complex than the usual activation functions, as it depends on the activations of *all* the output neurons, as required to enforce a proper output normalization. However, coupled with the cross-entropy error function, it again give rise to a simple formula for the δ_k (Eq. 5.8) of the output neurons:

$$\delta_k \equiv \frac{\partial E}{\partial a_k} = \ldots = (y_j - t_j) \tag{5.24}$$

From construction, a sufficiently general network with softmax output units trained to minimize the cross-entropy or Kullback-Leibler error functions will have output values approximating the bayesian class conditional posterior probabilities for the input pattern, $y_k^n \simeq P(\mathcal{C}_k|\mathbf{x^n})$. It can then be used to classify each pattern minimizing the misidentification probability: after training two groups of similar networks on the same sample, the best bayesian network showed better performance than the best classical one. Additionally, as the network outputs are probabilities, they can be easily combined with the output of other b tagging algorithms.

An interesting property of this kind of network is the possibility to reweight the classification parametrization for a better description of different samples: as long as the pattern distributions within each class $p_k(\mathbf{x})$ are unchanged, the network can be adapted to samples with different compositions, i.e. different *a priori* class probabilities f_k, without re-training. Comparing the structure of the softmax activation function (Eq. 5.23) with the definitions of *a posteriori* probabilities (Eq. 5.17), we see that each exponential term e^{a_k} plays the role of the product $p_k(\mathbf{x}) \cdot f_k$. Changes in f_k can be taken into account modifying accordingly the biases w_k of the output neurons:

$$w_k \to w_k + \log\left(\frac{f'_k}{f_k}\right) \tag{5.25}$$

$$p_k(\mathbf{x}) \cdot f_k = e^{a_k} \to e^{a_k} \cdot \frac{f'_k}{f_k} = p_k(\mathbf{x}) \cdot f_k \cdot \frac{f'_k}{f_k} = p_k(\mathbf{x}) \cdot f'_k \tag{5.26}$$

This has been tested training a network on a data sample, and applied to a different sample after renormalizing the output neurons bias. The performance of the normalized network however has been indistinguishable from that of the original one.

5.4 Multi layer perceptron performance

The performance of different neural network configurations has been studied, leading to the use of a rather simple structure: one hidden layer with sixteen nuerons, using sigmoid activation functions, and full connections between consecutive layers. Consequently, the networks have two layers of adaptive weights: from the input neurons to the hidden layer, and from them to the output neurons.

The use of two hidden layers or more neurons per hidden layer has been investigated, but the additional complexity only leads to increased training time without any observable performance gain. On the other hand less hidden units, and especially less than eight, will give rather poor results.

Once the general structure of the networks was estabilished, groups of similar networks were trained with different random starting conditions in order to explore a wider volume of weights space. The fastest training algorithm for this small networks has proven to be the *Broyden-Fletcher-Goldfarb-Shanno* (BFGS) quasi-Newton method[92, 93, 94, 95, 96]. The conjugate gradient *Polak-Ribiere*[97, 98] algorithm, while slower, showed to be less sensitive to overtraining and has been used to double check some of the results (see Fig. 5.17 and 5.23).

All the networks were trained on the flavour-enriched di-jet samples described in Sec. 5.2. 10% of the events (roughly 50k b and c, and 125k light jet events) were used as the training sample, and an other 10% as the test sample.

Network training with BFGS was carried on and monitored for 2000 epochs, saving the weights every 100 epochs. This allowed to recover the weights nearest to the minimum of the error function for the test sample if a network incurred in overtraining. Otherwise the weights at the end of the 2000 epochs training were kept. As shown in Figs. 5.17 and 5.23, for each group the best performing network was kept, i.e. the one having at its minimum the least error on the test sample.

This approach was used to try different combinations of input variables, error functions, and jet flavour definitions. In the next sections I will report the training history and tagging performance of the most representative combinations, listed in Tab. 5.5.

5.4.1 Sum-of-squares error

The first error function used has been the sum-of-squares, as an implementation was readily available within the ROOT framework.

The tagging variables used as input are $p_{\mathrm{T},rel}$, S_{IP}^{3D}, ΔR, r, E_{T}^{jet}, η^{jet};

Input variables	Error function	Jet flavour definition
$p_{\mathrm{T},rel}$, S_{IP}^{3D}, ΔR, r, E_{T}^{jet}, η^{jet}	sum-of-squares	parton content
$p_{\mathrm{T},rel}$, S_{IP}^{3D}, ΔR, r, E_{T}^{jet}, η^{jet}	sum-of-squares	leptonic decay $(\mathrm{b} \to \ell \vee \mathrm{b} \to \mathrm{c} \to \ell)$
$p_{\mathrm{T},rel}$, S_{IP}^{3D}, ΔR, $r_{\parallel}$, E_{T}^{jet}, η^{jet}	cross entropy	leptonic decay (1-of-c)
$p_{\mathrm{T},rel}$, ΔR, $r_{\parallel}$, E_{T}^{jet}, η^{jet}	cross entropy	leptonic decay (1-of-c)

Table 5.5: Different network configurations used in training. For a discussion on the use of the jet flavour definitions see the text.

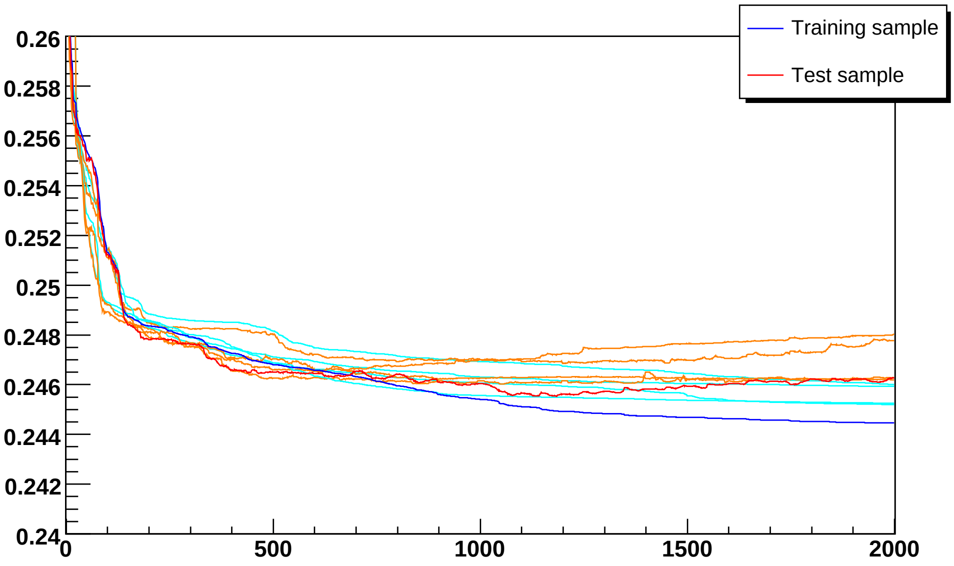
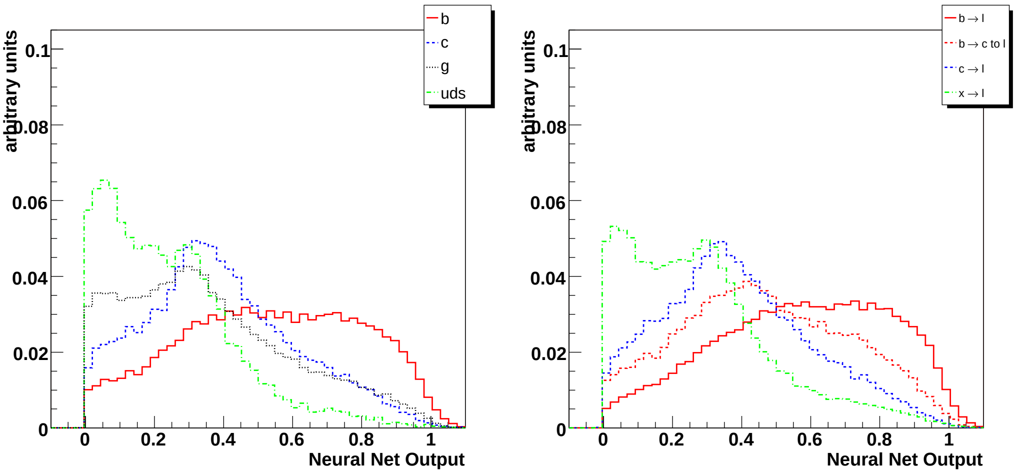


Figure 5.17: Training with the BFGS algorithm of a class of feedforward multi layer perceptrons, with a sum-of-square error function and linear output neurons, showing the error for the training sample (blue and cyan) and test sample (red and orange). Blue and red lines identify the network eventually selected, while cyan and orange refer to other networks.



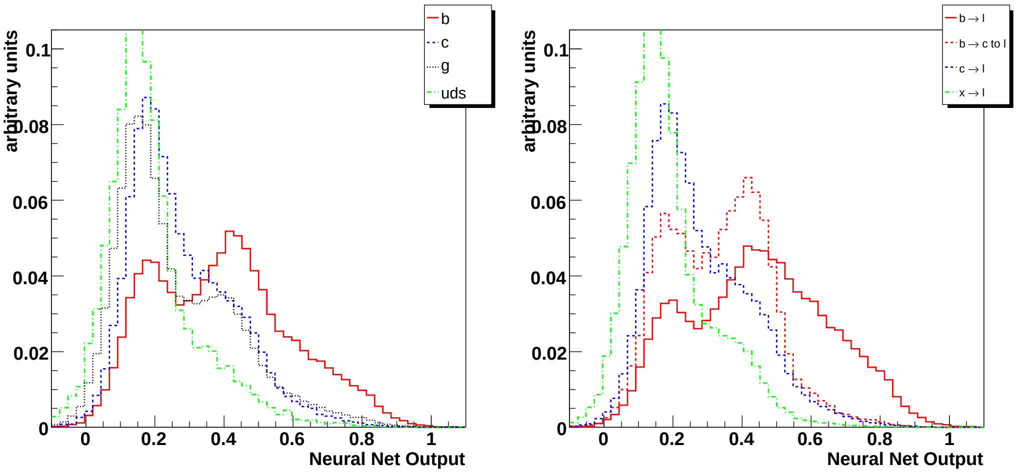
(a) (b)

Figure 5.18: Neural network output for flavour enriched di-jet samples, per (a) jet flavour and (b) muon decay chain. The neural network training was based on jet flavour defined from parton content analysis.

the jet energy and pseudorapidity are included to let the network take into account the detector performance as a function of them. The network has a single linear output neuron, trained to discriminate jets coming from b quarks from all the others: the target variable used for training is set to 1 for b jets, 0 for the rest.

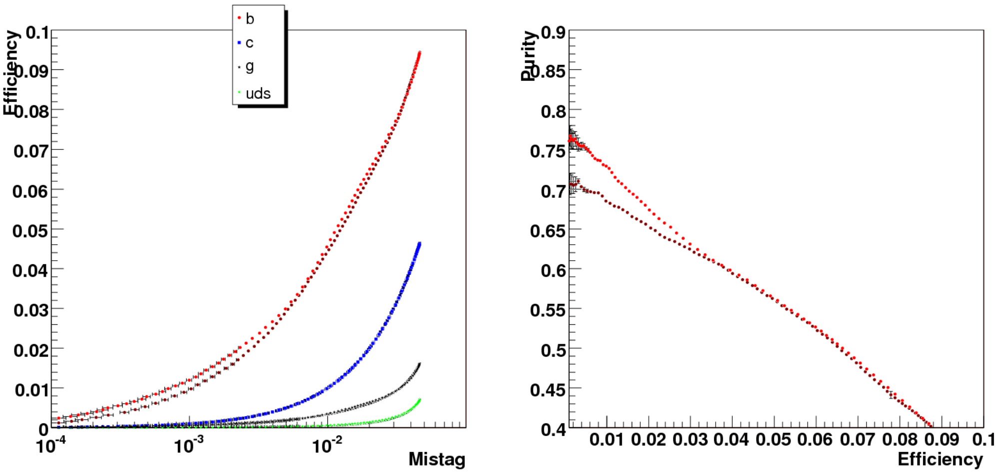
This configuration has been used to decide which jet flavour identification approach to use: whether based on parton content or muon decay chain analysis; in the latter case a jet was considered as coming from a b quark if the muon was either a direct ($b \to \ell$) or chain ($b \to c \to \ell$) decay product of a b hadron. Figures 5.18 and 5.19 show the distribution of network output for each jet flavour for both definitions, normalized to unit area. Although the shapes are quite different, the main characteristics are similar: for light jets the output value is quite low, and for b jets it is higher and more spread out, allowing a good discrimintation between these two classes; the distributions for charm and gluon jets are similar, someway in between the other two, with consequently a worse rejection against those. Overall, the distributions are more sharply peaked for the network trained using the muon decay chain jet identification, hint of a possible better performance.

The direct performance comparison between these network is shown in Fig. 5.20. The left plot shows the b tagging and non-b mistagging efficiency



(a) (b)

Figure 5.19: Neural network output for flavour enriched di-jet samples, per (a) jet flavour and (b) muon decay chain. The neural network training was based on jet flavour defined from muon decay chain analysis.



(a) (b)

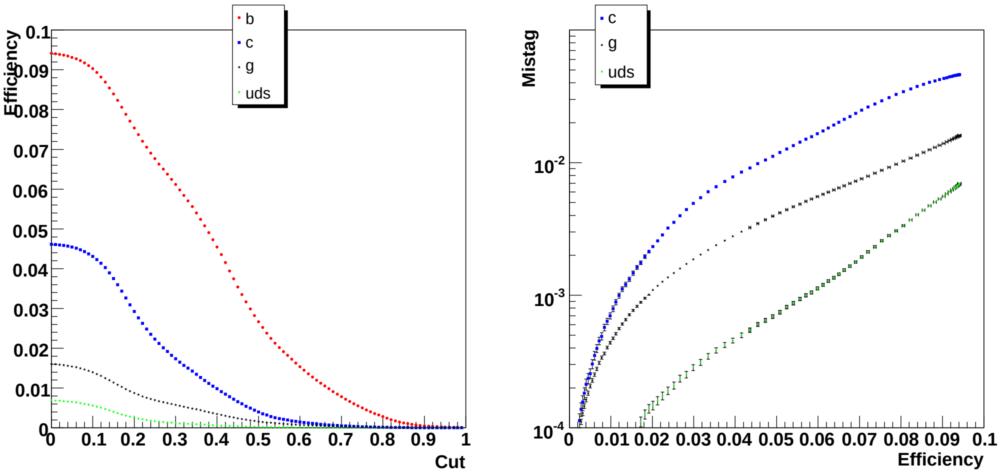
Figure 5.20: Comparison of the performance of similar neural networks, trained with different jet flavour definitions: parton content (dark points) and muon decay chain (light points). (a) b tag and non-b mistag efficiency versus c mistag efficiency (the blue points lie on the $y = x$ line). (b) b tagging purity versus efficiency.

as a function of the mistag rate for c jets - the points describing the c mistagging efficiency obviously lies on $y = x$ line. The plot to the right shows the overall tagging purity versus the b tagging efficiency. The results from the two networks are similar, with the one trained using the muon decay chain definition giving a better performance in the low efficiency region due to the best rejection against c jets.

Figures 5.21 and 5.22 show additional performance details for the network trained using the muon decay chain jet flavour definition.

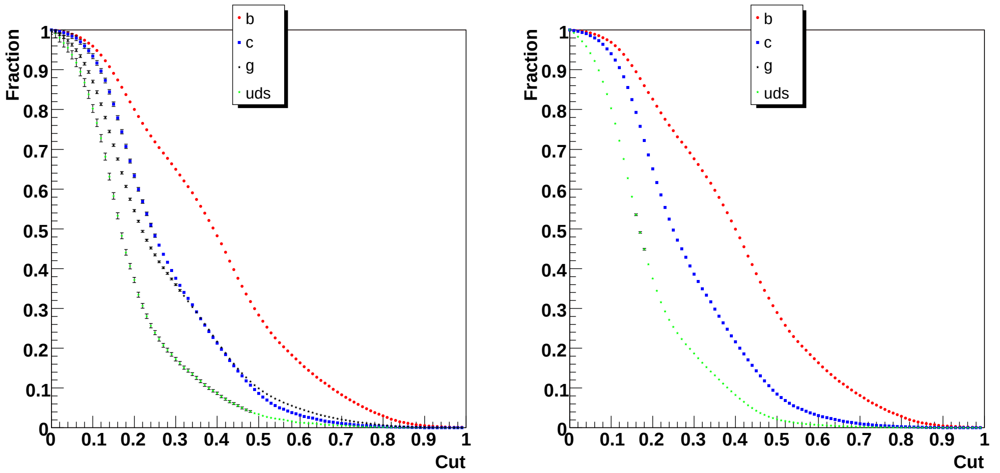
Since the muon decay chain is obviously available only in jets with a reconstructed muons, the overall tagging efficiency can only be computed using the parton content flavour definition. It is shown for the four jet flavours as a function of the applied cut in Fig. 5.21 (a), and Fig. 5.21 (b) as a function of the b tagging efficiency.

Since this definition depends heavily on the number and momentum of muons produced inside each jet, a better description of the intrinsic efficiency of the neural network b tagging is given by the fraction of jets of each flavour passing the applied cut. As this is anyway normalized to the number of jets with muons, it can be also computed using the muon decay chain jet flavour definition: Fig. 5.22 shows this normalized efficiency for each jet flavour, for both definitions.



(a) (b)

Figure 5.21: Performance on flavour enriched di-jet sample for a network trained using the muon decay chain jet flavour definition. (a) b tagging and non-b mistagging rates as a function of the neural network cut. (b) non-b mistagging as a function of the b tagging rate.



(a) (b)

Figure 5.22: Algorithmic performance on flavour enriched di-jet sample for a network trained using the muon decay chain jet flavour definition, showing the fraction of jets with a muon passing a given neural network cut for each jet flavour, defined from (a) parton content and (b) muon decay chain.

5.4.2 Cross entropy

The rationale behind the introduction of the cross entropy error function lies in the possibility to develop a neural network able to compute for each jet well defined class membership probabilities, rather than just discriminate b from non-b jets. After ascertaining with sum-of-square networks the possibility to succesfully use of the muon decay chain jet flavour definition for network training, the next logical step has been to develop a class of bayesian neural networks able to identify the flavour of each jet.

The overall sructure of the networks using this approach is the same as for those using the sum-of-square error function described above. The difference lies in the output layer, with four neurons (one per class: $\mathrm{b} \to \ell$, $\mathrm{b} \to \mathrm{c} \to \ell$, $\mathrm{c} \to \ell$, and $q \to \ell$) instead of one, using the softmax activation function instead of a linear one. Once properly trained, these network should give for each jet the probability for it to belong to each class. As there are two distinct outputs for the direct and chain decay of b quarks, the probability $P(\mathrm{b})$ for the jet to come from a b quark is the union of the two, $P(\mathrm{b}) = P(\mathrm{b} \to \ell \vee \mathrm{b} \to \mathrm{c} \to \ell)$. Since these outcomes are incompatible, this is just the sum of the probabilities, $P(\mathrm{b}) = P(\mathrm{b} \to \ell) + P(\mathrm{b} \to \mathrm{c} \to \ell)$. The neural netork computed probability of being a b jet is shown in Fig.5.24 for different flavours of jets and different muon decay chains.

The networks are trained using the cross entropy error function (Eq. 5.20), or rather the Kullback-Leibler divergence (Eq. 5.22), using the BFGS quasi-Newton algorithm. As a consistency check, a separate set of networks has been trained with the slower conjugate gradient Polak-Ribiere algorithm. The comparison between the training development and the errors reached on both the training and test samples is shown in Fig. 5.23. Even though the BFGS algorithm requires slightly more computing power than the Polak-Ribiere per epoch, it is dramatically more efficent in term of training epochs needed: the first has almost converged after less than 500 epochs, and shows only small improvements over the next 1500 epochs; while the latter is becoming stable only after roughly 12000 epochs, and still improves over as many more. Moreover, when comparing the minimum errors achieved, the results reached by the BFGS are slightly better.

Figure 5.25 compares the performance of the second neural network described in the previous section (dark dots) with that of a network using four softmax output neurons and trained with the Kullback-Leibler error function. The left plots shows the b tagging and non-b mistagging efficiency as a function of the c jet mistag, while the right one shows the tagged sample purity versus the b tagging efficiency. The softmax network has a slightly better performance over all the efficiency range.

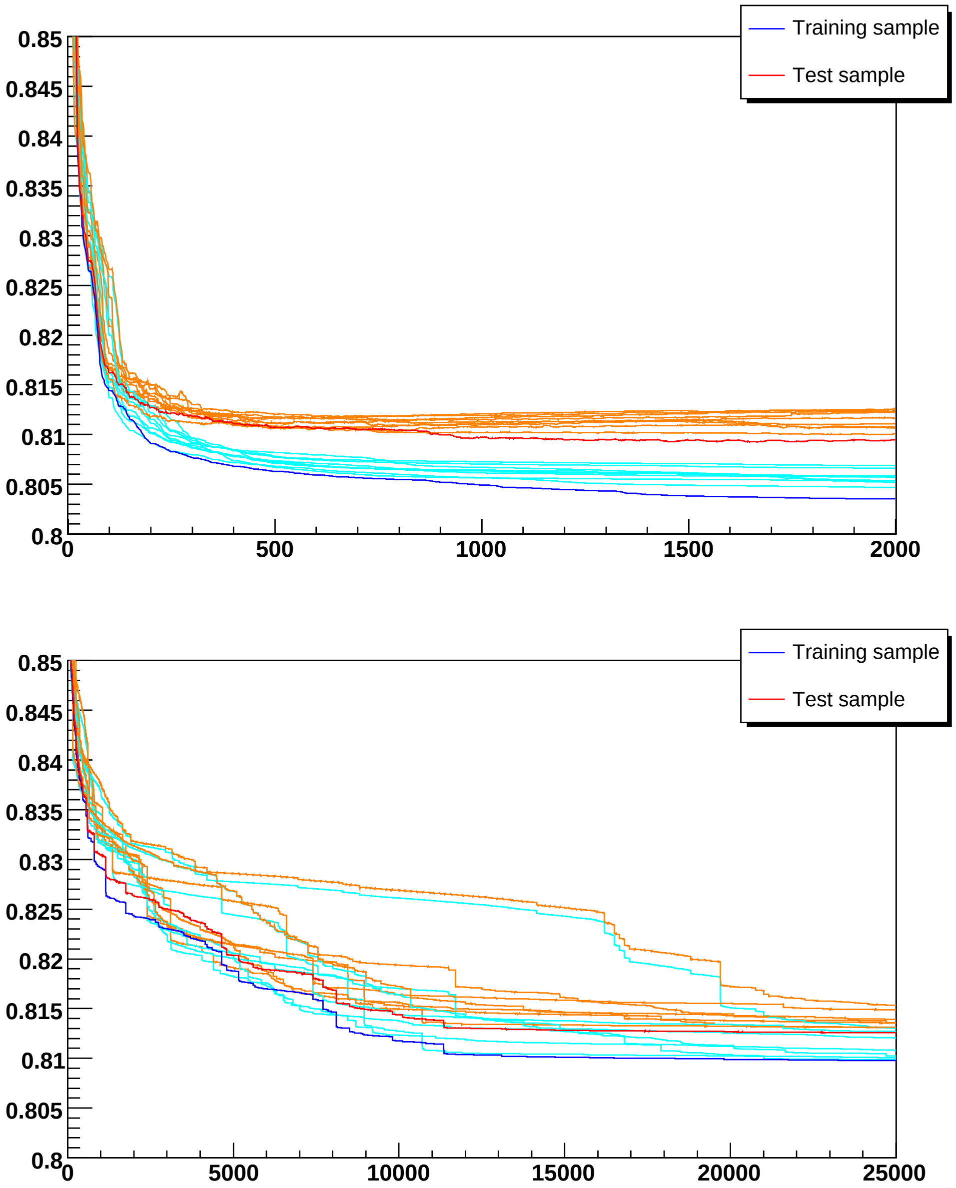
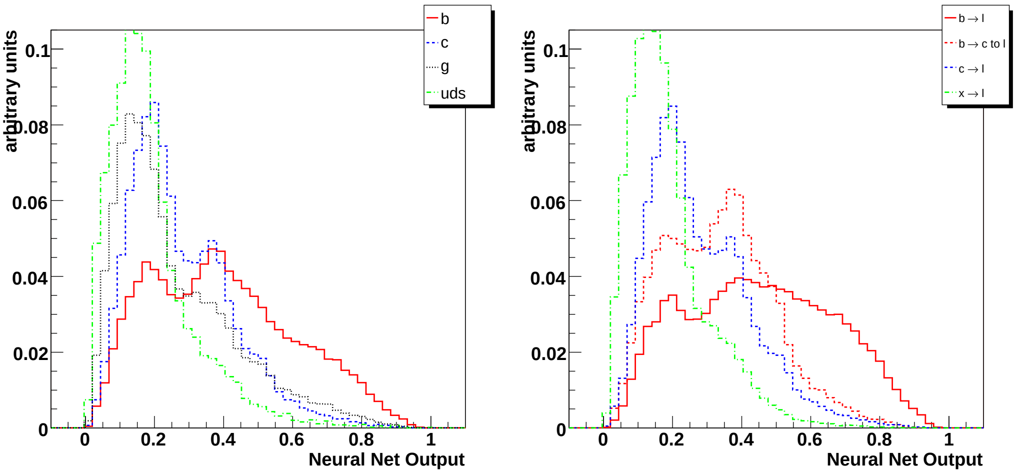


Figure 5.23: Training of a class of feedforward multi layer perceptrons, with a cross entropy error function and softmax output neurons, showing the error for the training sample (blue and cyan) and test sample (red and orange). Blue and red lines identify the network eventually selected, while cyan and orange refer to other networks. (top) training with the BFGS quasi-Newton algorithm. (bottom) training with the conjugate gradient Polak-Ribiere algorithm.

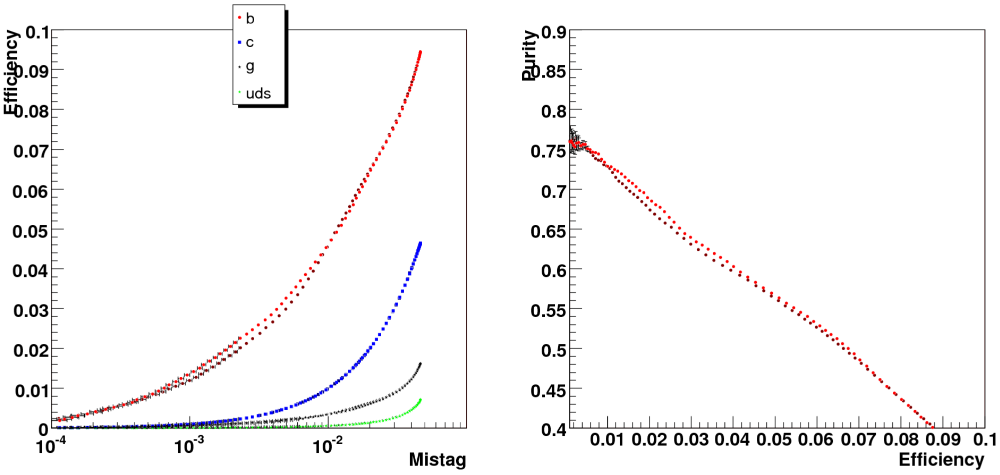


(a)

(b)

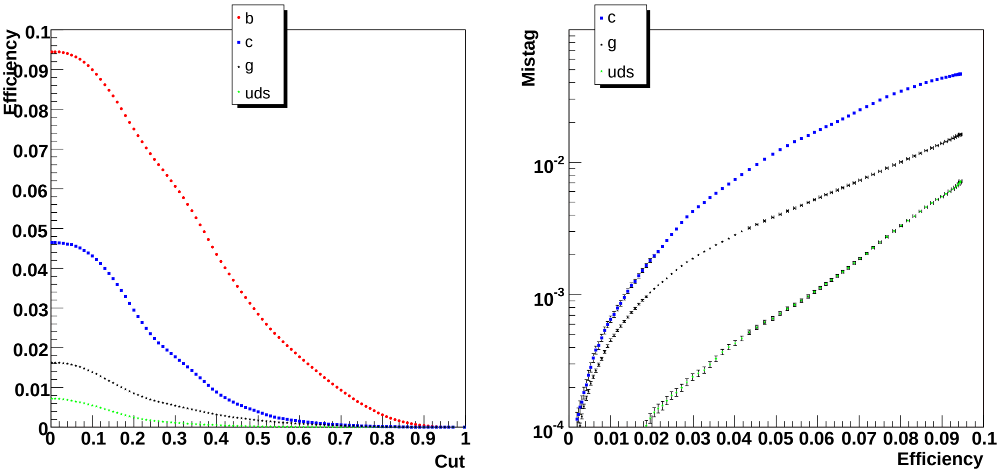
Figure 5.24: Neural network output for b jet probability on flavour enriched di-jet sample for a softmax network trained with the Kullback-Leibler error function, per (a) jet flavour and (b) muon decay chain.

Figures 5.26 and 5.27 show additional performance details for the softmax network. The overall tagging efficiency for the each jet flavour is shown as a function of the applied cut in Fig. 5.21 (a) and as a function of the b tagging efficiency in Fig. 5.21 (b). Fig. 5.22 shows the normalized or algorithmic efficiency for each jet flavour, i.e. the fraction of jets with a muon passing a given neural network cut.



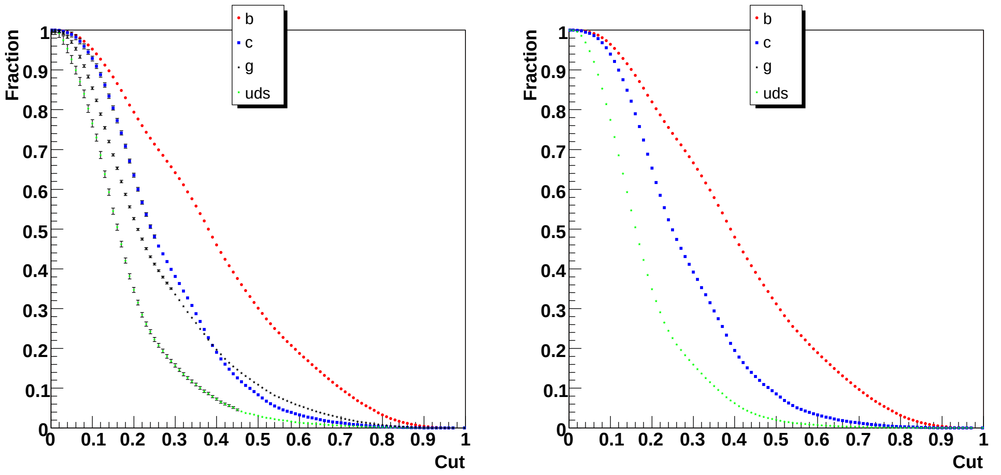
(a) (b)

Figure 5.25: Comparison of the performance of neural networks with different output units, trained with different error functions, with the muon decay chain jet flavour definition: linear output neurons with a sum-of-square error function (dark points) and softmax output neurons with a Kullback-Leibler error function (light points). (a) b tag and non-b mistag efficiency versus c mistag efficiency (the blue points lie on the $y = x$ line). (b) b tagging purity versus efficiency.



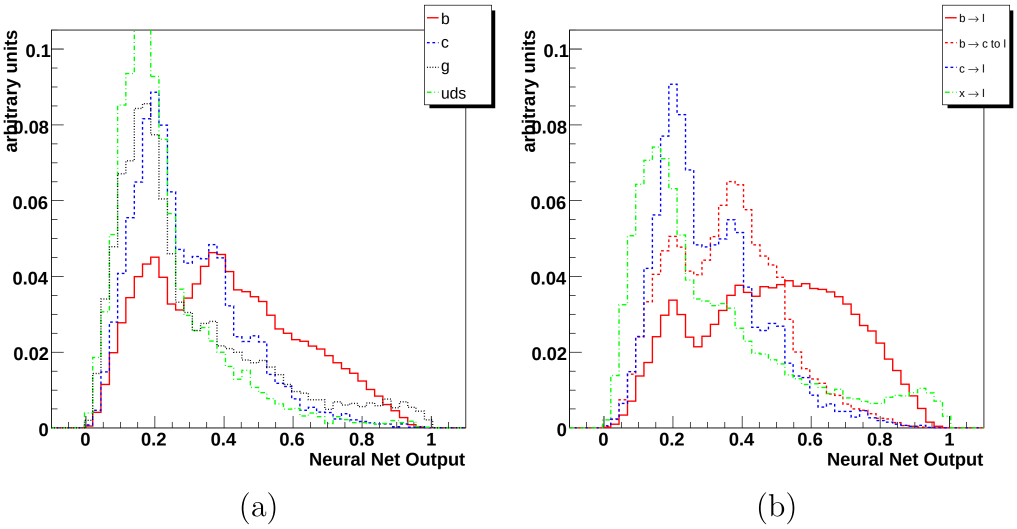
(a) (b)

Figure 5.26: Performance on flavour enriched di-jet sample. (a) b tagging and non-b mistagging rates as a function of the neural network cut. (b) non-b mistagging as a function of the b tagging rate.



(a) (b)

Figure 5.27: Algorithmic performance of the neural network on flavour enriched di-jet sample, showing the fraction of jets with a muon passing a given neural network cut, per (a) jet flavour and (b) muon decay chain.



(a) (b)

Figure 5.28: Neural network output for leptonic and semileptonic $t\bar{t}$ decays, per (a) jet flavour and (b) muon decay chain.

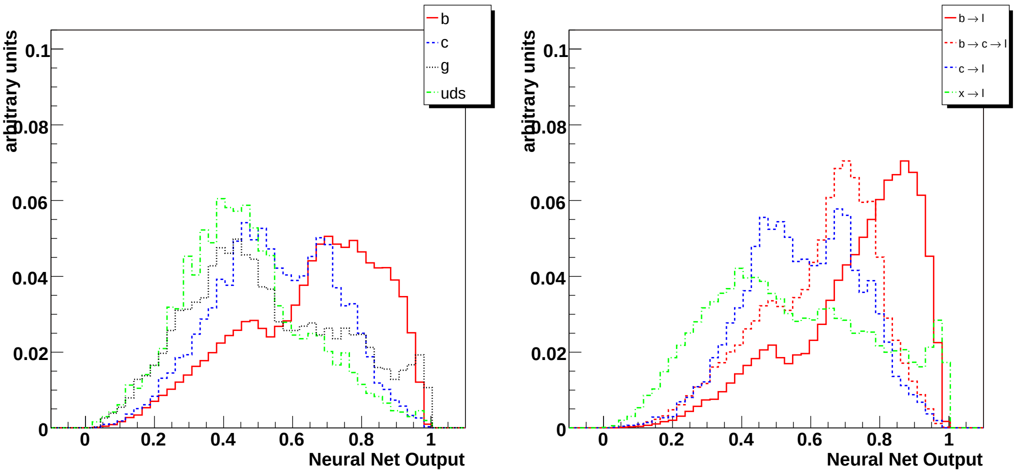
5.4.3 Bayesian weight normalization

One of the reason for studying a bayesian definition of membership probability (Eq. 5.17) is the possibility to adapt such network to samples with different prior probabilities, as long as the event distributions inside each class are unchanged. Applied to b tagging this translates into the possibility to train a network a sample with a particular composition of jet flavours and apply it to a second sample with a completely different composition. In order to test this assumption the softmax network described in the previous section, trained on di-jet samples, has been applied to the leptonic and semileptonic $t\bar{t}$ sample described in Sec. 5.2. Figure 5.28 shows the b jet probability as computed by the neural network for different jet flavours and muon decay chains.

Table 5.6 shows the composition f_k and f'_k of the di-jet and $t\bar{t}$ samples. These fractions describe the prior probabilities for a jet to belong to each class. As the network has been trained on the first sample, its outputs y_k approximate the *a posteriori* probabilities only for prior probabilities f_k. The posterior probabilities for arbitrary priors f'_k can be recovered using Bayes theorem (Eq. 5.16): weighting each output y_k with the ratio $\frac{f'_k}{f_k}$ and normalizing, will give the posterior probabilities $y'_k = \frac{f'_k}{f_k} y_k$ for priors f'_k. As the outputs are normalized, the absolute values of weights are irrelevant, and

Class k	Original fraction f_k	New fraction f'_k	Normalization $\frac{f'_k}{f_k}$
$b \to \ell$	0.192	0.369	1.92
$b \to c \to \ell$	0.126	0.244	1.93
$c \to \ell$	0.114	0.050	0.44
$q \to \ell$	0.568	0.337	0.59

Table 5.6: *A priori* probabilities f_k and f'_k, and their ratio, for the di-jet and $t\bar{t}$ samples.



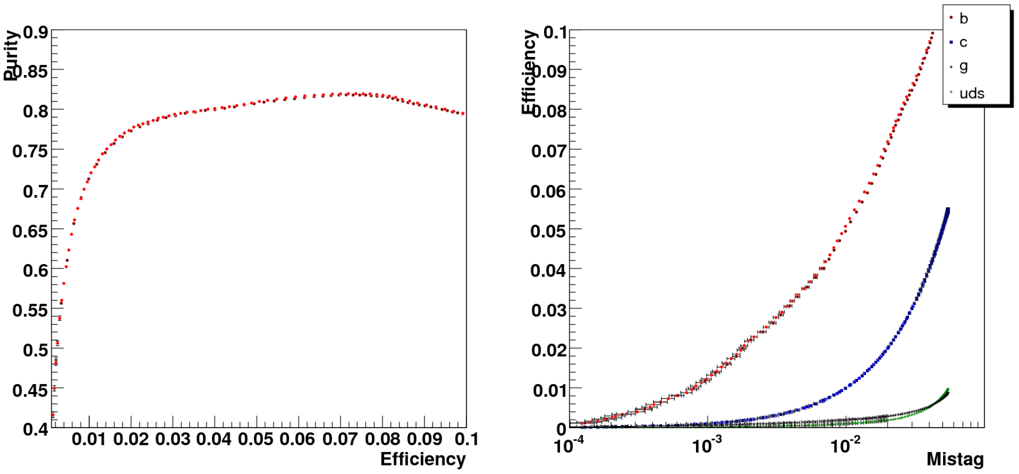
(a) (b)

Figure 5.29: Normalized neural network output for leptonic and semileptonic $t\bar{t}$ decays, per (a) jet flavour and (b) muon decay chain.

only the ratios can affect the performance of the network.

Figure 5.29 shows the b jet probability as computed by the weighted neural network for different jet flavours and muon decay chains. There are evident shifts in the distributions, but these affect more or less in the same manner all the jet flavours. Finally, the performance of the two networks is compard in Fig. 5.30, showing the tagging purity as a function of the efficiency, and this one as a funztion of thec jet mistag rate. In spite of the weight normalization there are no differences in the performance of the two networks.

Further investigation showed this behaviour for a range of weights spanning about an order of magnitude. This can be regarded as a measurement of the noise generated on the output neurons by the in-class statistical



(a) (b)

Figure 5.30: (a) purity of the b tag on leptonic and semileptonic $t\bar{t}$ decays, as a function of the b tag efficiency, and (b) tagging efficiencies as a function of the c mistag rate, for the original (light points) and weighted (dark points) networks.

fluctuations. In order for the weight renormalization to be useful, a better understandig of the in-class pattern distribution will thus be required.

5.5 Summary

The performance of these implementations of the soft lepton tagging algorithm are summarized in Tabs. 5.7 and 5.8.

The former compares the mistag rate of each non-b jet flavour at fixed b tagging efficiencies for the three soft muon b tagging networks described in the previous sections, as obtained from the flavour enriched di-jet training samples. As expected, and as can be seen comparing the mistag rate for the different b tagging efficiencies, the selection purity increases applying more sringent cuts.

The latter shows similar data from the $t\bar{t}$ leptonic and semileptonic decay samples, for the two networks trained using the lepton decay chain jet flavour definition. The characteristics of the leptons and jets in these samples are slightly different from those in the training sample. As a consequence, the b tagging purity does not increase much as more stringent cuts are applied to the neural network output, but keep instead to roughly 80% in the range of

Algorithm	mistag for 8% b eff.			mistag for 5% b eff.			mistag for 3% b eff.		
	charm	light	gluon	charm	light	gluon	charm	light	gluon
Sum-of-squares, parton content	3.42%	0.32%	1.03%	1.25%	0.06%	0.40%	0.52%	0.02%	0.20%
Sum-of-squares, lepton decay	3.38%	0.33%	1.01%	1.16%	0.07%	0.41%	0.50%	0.03%	0.19%
Cross entropy, lepton decay	3.43%	0.33%	1.00%	1.19%	0.07%	0.39%	0.48%	0.02%	0.19%

Table 5.7: Non-b jet mistag rate for three fixed values of b tagging eficiencies, 8%, 5% and 3%, measured on the flavour enriched di-jet samples: the selection purity increases with more stringent cuts. Each algorithm is identified by the error functions and the jet flavour definitions used.

Algorithm	mistag for 10% b eff.			mistag for 5% b eff.			mistag for 3% b eff.		
	charm	light	gluon	charm	light	gluon	charm	light	gluon
Sum-of-squares	4.23%	0.58%	0.58%	1.02%	0.10%	0.18%	0.46%	0.05%	0.13%
Cross entropy	4.29%	0.58%	0.59%	0.97%	0.10%	0.20%	0.36%	0.04%	0.12%

Table 5.8: Non-b jet mistag rate for three fixed values of b tagging eficiencies, 10%, 5% and 3%, measured on the $t\bar{t}$ leptonic and semileptonic decays samples: the selection purity is roughly 80% in all cases. Each algorithm is identified by the error functions used. Both used the lepton decay chain jet flavor definition.

efficiencies shown in the table.

All the selected networks perform quite well on the training sample, and can be succesfully applied to different data samples. The fact that the performace does not change much among them suggests that the algorithms have reached a structural limit, due to the tagging variables distributions inside each jet flavour class. A better understanding of the in-class distributions and a more fine grained jet flavour classification will then be needed to further increase the algorithm performance. This should be even more evident for the application to different data samples from the training ones.

Conclusions

The first part of my work, as presented in this thesis, comprised the development of the software for the communication with and the control of a power supply prototype, in collaboration with Alenia Spazio - LABEN, for the CMS Silicon Strip Tracker. The CMS power supply system architecture foresees intelligent power supplies equipped with radiation hard microcontrollers, able of stand-alone operation and appropriate reactions to alarms, controlled via CAN bus by array controllers, which are interfaced to the CMS slow control system.

The developement work spanned the software run by an array controller, the graphical user interface used to control it, and an efficient communication protocol between the array controller and the power supply microcontrollers over the CAN bus link. In the meantime, I worked on testing the performance of two different power supply prototypes, the one developed by LABEN and one developed by CAEN.

These activities culminated with the use of both power supply prototypes at the May 2005 Tracker test beam, at the X5 beam area at CERN. This gave me the opportunity to measure the power supply performance and test the control software and communication link I had developed in a realistic scenario, where different alarm conditions were simulated. The Test Beam results showed that both power supply prototypes behaved very well, with a very low overall noise and a good signal-to-noise ratio. This was a direct experimental confirmation of the quality of both prototypes, and was expected from the values of noise and isolation measured in laboratory, complying with and exceeding the requirements as set by the CMS collaboration.

In the second part of my work I have studied and developed the soft lepton tagging algorithm for jet b tagging with the CMS detector, based on the high semileptonic branching ratio of b hadrons, $19.3 \pm 0.5\%$ per lepton family. After estabilishing a general framework for the soft lepton b tag algorithm within the CMS reconstruction and analysis software I studied in detail the implementation of b tagging with muons.

A most simple form of soft lepton tagging can be performed just by requiring a reconstructed muon inside a jet cone. To improve the purity of the jet selection that this approach would achieve, the topological and kinematical properties of the lepton are analysed and summarized in four tagging variables: the lepton three-dimensional signed impact parameter significance S_{IP}^{3D}, its transverse momentum, relative to the lepton-excluded refined jet axis $p_{\mathrm{T},rel}$, the ratio between its momentum along such axis and the jet energy $r_{\parallel}$, and the pseudoangular distance in the $\eta\varphi$ plane between the reconstructed lepton track and the lepton-excluded refined jet axis ΔR.

These variables are then combined together with an artificial neural network. I have studied two basically different network architectures, based on different classification methods: the simpler definition of a non-linear discriminant, whose value is used to classify each jet either as a b or non-b jet, and the approximation of the bayesian *a posteriori* probabilty for class membership, each class corresponding to a distinct decay chain leading to the lepton production.

The bayesian approach is more thoretically sound, and was expected to give better results. After training many networks of each kind, the performance of those based on it were only marginally better than that of the more classical networks with a single discriminant. In order to further improve the tagging purity, a better understanding of the in-class distributions of the tagging variables is required, together with a more fine grained jet flavour classification.

All the networks did perform quite well on both the training samples and other ones with different physical characteristics, and the best performing ones have been integrated in the other CMS b tagging algorithms and analysis code.

Bibliography

[1] S. Eidelman *et al.*, *Review of particle physics*, Phys. Lett. **B592**, 1 (2004), `http://pdg.lbl.gov/`.

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

[3] P. W. Higgs, *Spontaneous symmetry breakdown without massless bosons*, Phys. Rev. **145**, 1156 (1966).

[4] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, Phys. Rev. Lett. **13**, 321 (1964).

[5] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global conservation laws and massless particles*, Phys. Rev. Lett. **13**, 585 (1964).

[6] S. Weinberg, *A model of leptons*, Phys. Rev. Lett. **19**, 1264 (1967).

[7] G. 't Hooft, *Renormalization of massless Yang-Mills fields*, Nucl. Phys. **B33**, 173 (1971).

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

[9] J. F. Gunion, H. E. Haber, G. Kane, and S. Dawson, *The Higgs Hunter's Guide* (Addison Wesley, 1990).

[10] K. Riesselmann, *Limitations of a Standard Model Higgs boson* (1997), hep-ph/9711456.

[11] R. Barate *et al.* (LEP Working Group for Higgs boson searches), *Search for the Standard Model Higgs boson at LEP*, Phys. Lett. **B565**, 61 (2003), hep-ex/0306033.

[12] M. De Palma *et al.* (ALEPH), *ALEPH: technical report 1983* (1983), CERN-LEPC-83-2, LEPC-P-1.

[13] *Proposal for the new vertex detector for ALEPH* (1993), CERN-LEPC-93-8, LEPC-P-1-ADD-1.

[14] W. Bartl *et al.* (DELPHI), *DELPHI: technical proposal* (1983), CERN-LEPC-83-3, DELPHI-83-66, LEPC-P-2.

[15] *Proposal for the replacement of the small angle calorimeter of DELPHI* (1992), CERN-LEPC-92-6, LEPC-P-2-ADD-1.

[16] *Proposal for the upgrade of DELPHI in the forward region* (1992), CERN-LEPC-92-13, LEPC-P-2-ADD-2.

[17] *Proposal for the DELPHI very forward tracker* (1993), CERN-LEPC-93-6, LEPC-P-2-ADD-3.

[18] G. von Dardel *et al.* (L3), *L3 Technical Proposal* (1983), CERN-LEPC-83-5, LEPC-P-4.

[19] *Some improvements to the L3 detector for LEP 200* (1994), CERN-LEPC-94-5, LEPC-P-4-ADD-2-INS.

[20] *A very small angle tagger for L3* (1995), CERN-LEPC-95-1, LEPC-P-4-ADD-3-INS.

[21] *OPAL: The OPAL Detector Technical Proposal* (1983), CERN-LEPC-83-4, LEPC-P-3.

[22] *Proposal for a new silicon microvertex detector for OPAL* (1993), CERN-LEPC-93-13, LEPC-P-3-ADD-1.

[23] *Proposal for a new silicon microvertex detector for OPAL* (1993), CERN-LEPC-93-13, LEPC-P-3-ADD-2.

[24] *Proposal to install scintillating tiles in the OPAL endcap* (1994), CERN-LEPC-94-09, LEPC-P-3-Add-3.

[25] L. I. S. Group, *LEP Design Report Vol.1: The LEP Injector Chain* (1983), CERN-LEP-TH-83-29, CERN-PS-DL-83-31, CERN-SPS-83-26, LAL-RT-83-09.

[26] *LEP Design Report Vol.2: The LEP Main Ring* (1984), CERN-LEP/84-01.

[27] *LEP Design Report, Vol.3: LEP2* (1996), CERN-AC-96-01-LEP-2.

[28] R. Barate *et al.* (ALEPH), *Observation of an excess in the search for the Standard Model Higgs boson at ALEPH*, Phys. Lett. **B495**, 1 (2000), hep-ex/0011045.

[29] T. Appelquist and C. W. Bernard, *Strongly interacting Higgs bosons*, Phys. Rev. **D22**, 200 (1980).

[30] A. C. Longhitano, *Low-energy impact of a heavy Higgs boson sector*, Nucl. Phys. **B188**, 118 (1981).

[31] T. Appelquist and G.-H. Wu, *The Electroweak chiral Lagrangian and CP violating effects in technicolor theories*, Phys. Rev. **D51**, 240 (1995), hep-ph/9406416.

[32] M. Veltman, *Second threshold in weak interactions*, Acta Phys. Polon. **B8**, 475 (1977).

[33] F. Abe *et al.* (CDF), *Evidence for top quark production in* $\bar{p}p$ *collisions at* $\sqrt{s}=1.8$ TeV, Phys. Rev. Lett. **73**, 225 (1994), hep-ex/9405005.

[34] The ALEPH, DELPHI, L3, OPAL, SLD Collaborations, the LEP Electroweak Working Group, the SLD Electroweak and Heavy Flavour Groups, *Precision Electroweak measurements on the Z resonance* (2005), hep-ex/0509008.

[35] The ALEPH, DELPHI, L3, OPAL, SLD Collaborations, the LEP Electroweak Working Group, the SLD Electroweak and Heavy Flavour Groups, *A combination of preliminary electroweak measurements and constraints on the Standard Model* (2005), hep-ex/0511027.

[36] G. Montagna, O. Nicrosini, F. Piccinini, and G. Passarino, *TOPAZ0 4.0: A new version of a computer program for evaluation of de-convoluted and realistic observables at LEP 1 and LEP 2*, Comput. Phys. Commun. **117**, 278 (1999), hep-ph/9804211.

[37] D. Y. Bardin *et al.*, *ZFITTER: An analytical program for fermion pair production in* e^+e^- *annihilation*, CERN-TH 6443/92 (1992), hep-ph/9412201.

[38] D. Y. Bardin *et al.*, *ZFITTER v.6.21: A semi-analytical program for fermion pair production in* e^+e^- *annihilation*, Comput. Phys. Commun. **133**, 229 (2001), hep-ph/9908433.

[39] A. B. Arbuzov *et al.*, *ZFITTER: A semi-analytical program for fermion pair production in* e^+e^- *annihilation, from version 6.21 to version 6.42* (2005), hep-ph/0507146.

[40] F. James and M. Roos, *'MINUIT' a system for function minimization and analysis of the parameter errors and correlations*, Comput. Phys. Commun. **10**, 343 (1975).

[41] *Combination of CDF and D0 results on the top-quark mass* (2005), hep-ex/0507091.

[42] *ATLAS: Technical proposal for a general-purpose p-p experiment at the Large Hadron Collider at CERN* (1994), CERN-LHCC-94-43.

[43] *ATLAS: Detector and physics performance technical design report. Volume 1* (1999), CERN-LHCC-99-14.

[44] *ATLAS detector and physics performance. Technical design report. Vol. 2* (1999), CERN-LHCC-99-15.

[45] M. Della Negra *et al.* (CMS), *CMS: The Compact Muon Solenoid: Letter of intent for a general purpose detector at the LHC* (1992), CERN-LHCC-92-3.

[46] *CMS, the Compact Muon Solenoid: Technical proposal* (1994), CERN-LHCC-94-38.

[47] *ALICE: Technical proposal for a large ion collider experiment at the CERN LHC* (1995), CERN-LHCC-95-71.

[48] *LHCb technical proposal* (1998), CERN-LHCC-98-4.

[49] *LHCb Technical Design Report: Reoptimized detector design and performance* (2003), CERN-LHCC-2003-030.

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

[51] *CMS, the Magnet Project: Technical Design Report* (1997), CERN-LHCC-97-10.

[52] *CMS: The Tracker Project Technical Design Report* (1998), CERN-LHCC-98-06.

[53] *Addendum to the CMS Tracker TDR* (2000), CERN-LHCC-2000-016.

[54] *CMS: The Electromagnetic Calorimeter. Technical Design Report* (1997), CERN-LHCC-97-33.

[55] *Addendum to the CMS ECAL Technical Design Report: Changes to the CMS ECAL electronics* (2002), CERN-LHCC-2002-027.

[56] *CMS: The Hadron Calorimeter Technical Design Report* (1997), CERN-LHCC-97-31.

[57] *CMS: The Muon Project. Technical Design Report* (1997), CERN-LHCC-97-32.

[58] *CMS: The TriDAS project. Technical Design Report, vol. 1: The trigger system* (2000), CERN-LHCC-2000-038.

[59] *CMS: The TriDAS project. Technical Design Report, Vol. 2: Data acquisition and high-level trigger* (2002), CERN-LHCC-2002-026.

[60] L. Borrello, A. Messineo, E. Focardi, and A. Macchiolo, *Sensor Design for the CMS Silicon Strip Tracker* (2003), CMS NOTE-2003/020.

[61] M. Lenzi, Ph.D. thesis, Università di Firenze, Dipartimento di Fisica (1999), available online at `http://hep.fi.infn.it/CMS/tesi_mic.pdf`.

[62] M. J. French *et al.*, *Design and results from the APV25, a deep sub-micron front-end chip for the CMS Tracker*, Nucl. Instr. and Meth. **A466**, 359 (2001).

[63] L. Jones, *APV25-S1 User's Guide Version 2.2* (2001), available online.

[64] F. Drouhin *et al.*, *The control system for the CMS Tracker front end*, IEEE Trans. Nucl. Sci. **49**, 846 (2002).

[65] *I^2C Bus Specification, version 2.1*, Philips Semiconductors (2000), document order number: 9398 393 40011.

[66] K. Klein (CMS Tracker), *The CMS Silicon Strip Tracker: System tests and test beam results*, in *Astroparticle, particle and space physics, detectors and medical physics applications. Proceedings, 8th Conference, ICATPP 2003, Como, Italy, October 6-10* (2003), pp. 686–692, CMS CR-2003/052, physics/0312132.

[67] S. Paoletti, A. Bocci, R. D'Alessandro, and G. Parrini, *The Powering Scheme of the CMS Silicon Strip Tracker*, LECC 2004 (2004), CERN-2004-010, CERN-LHCC-2004-030.

[68] *Road vehicles – Controller area network (CAN)*, International Organization for Standardization (2003), ISO 11898:2003.

[69] A. Bocci, *Progetto di sistema di alimentazione per rivelatori al silicio nella sperimentazione su particelle elementari a LHC* (2002), available online at `http://hep.fi.infn.it/CMS/tesi_bocci.pdf`.

[70] Heavy Flavor Averaging Group (HFAG), *Averages of b-hadron properties as of winter 2005* (2005), hep-ex/0505100.

[71] G. Abbiendi *et al.* (OPAL), *Measurement of the semileptonic branching ratio of charm hadrons produced in* $\mathrm{Z}^0 \to \mathrm{c}\bar{\mathrm{c}}$ *decays*, Eur. Phys. J. **C8**, 573 (1999), hep-ex/9810016.

[72] G. Segneri and F. Palla, *Lifetime based b-tagging with CMS* (2002), CMS NOTE-2002/046.

[73] A. Rizzi, F. Palla, and G. Segneri, *Track impact parameter based b tagging with CMS* (2005), CMS AN-2005/041.

[74] C. Weiser, *A combined secondary vertex based b tagging algorithm in CMS* (2005), CMS AN-2005/042.

[75] T. Speer, K. Prokofiev, R. Fruhwirth, W. Waltenberger, and P. Vanlaer, *Vertex Fitting in the CMS Tracker* (2005), CMS AN-2005/022.

[76] V. Gavrilov *et al.*, *Jet Reconstruction with Pileup Subtraction* (2003), CMS RN-2003/004.

[77] S. D. Ellis and D. E. Soper, *Successive combination jet algorithm for hadron collisions*, Phys. Rev. **D48**, 3160 (1993), hep-ph/9305266.

[78] T. Sjöstrand, L. Lonnblad, and S. Mrenna, *PYTHIA 6.2: Physics and manual* (2001), hep-ph/0108264.

[79] S. Abdoulline *et al.*, *An Object-Oriented Simulation Program for CMS*, in *Proceedings of Nuclear Science Symposium (NSS), IEEE 2004* (2004), available online, webpage at `http://cmsdoc.cern.ch/oscar/`.

[80] V. Innocente and D. Stickland, *The design, implementation and deployment of a functional prototype OO reconstruction software for CMS. The ORCA project* (Padova, Italy, 2000).

[81] *Object-oriented Reconstruction for CMS Analysis* (2005), available online, website at `http://cmsdoc.cern.ch/orca`.

[82] CMS detector simulation software group, *CMS Simulation Package CMSIM, Users' Guide and Reference Manual* (2002), available online.

[83] S. Cucciarelli (CMS), *Track and Vertex Reconstruction with the CMS Detector at LHC* (2005), CMS CR-2005/021.

[84] E. James *et al.*, *Muon Identification in CMS* (2005), CMS AN-2005/047.

[85] I. Belotelov and N. Neumeister, *Performance of the CMS offline muon reconstruction software* (2005), CMS AN-2005/010.

[86] V. Innocente, M. Maire, and E. Nagy, *GEANE: Average tracking and error propagation package*, in *Amsterdam 1991, Proceedings, MC91: Detector and event simulation in high energy physics* (1991), pp. 58–78.

[87] T. Ferbel, *Experimental Techniques in High-Energy Nuclear and Particle Physics* (1991).

[88] K. Funahashi, *On the approximate realization of continuous mappings by neural networks*, Neural Networks **2**, 183 (1989).

[89] R. Hecht-Nielsen, *Theory of back-propagation neural networks*, in *Proceedings of the International Joint Conference on Neural Networks* (San Diego, 1989), vol. 1, pp. 593–605.

[90] K. Hornik *et al.*, *Multilayer feedforward networks are universal approximators*, Neural Networks **2**, 359 (1989).

[91] C. M. Bishop, *Neural Networks for Patter Recognition* (Oxford University Press, 1995).

[92] C. G. Broyden, *The Convergence of a Class of Double-Rank Minimization Algorithms 2, The New Algorithm*, J. of the Inst. for Math. and Applications **6**, 222 (1970).

[93] R. Fletcher, *A New Approach to Variable-Metric Algorithms*, Computer Journal **13**, 317 (1970).

[94] D. Goldfarb, *A Family of Variable-Metric Algorithms Derived by Variational Means*, Mathematics of Computation pp. 23–26 (24).

[95] D. F. Shanno, *Conditioning of quasi-Newton methods for function minimization*, Mathematics of Computation **24**, 647 (1970).

[96] E. Polak, *Computational methods in optimization: a unified approach* (New York: Academic Press, 1971).

[97] E. Polak and G. Ribiere, Rev. Fr. Inform. Rech. Oper. **16-R1**, 35 (1969).

[98] W. H. Press *et al.*, *Numerical Recipes in C: the Art of Scientific Computing* (1992).