Vrije Universiteit Brussel

Faculteit Wetenschappen
Departement Natuurkunde

B-Tagging as a Tool for Charged Higgs Boson Identification in CMS

Steven Lowette

Promotor: Prof. Dr. Stefaan Tavernier
Co-promotor: Prof. Dr. Jorgen D'Hondt

Proefschrift ingediend met het oog op het behalen van
de academische graad Doctor in de Wetenschappen

November 2006

Doctoral examination commission

Chair: Prof. Dr. Catherine De Clercq (VUB)
Supervisor: Prof. Dr. Stefaan Tavernier (VUB)
Co-supervisor: Prof. Dr. Jorgen D'Hondt (VUB)
Prof. Dr. Gigi Rolandi (CERN)
Prof. Dr. Fabio Maltoni (UCL)
Prof. Dr. Ben Craps (VUB)
Prof. Dr. Robert Roosen (VUB)

Waartoe de wereld amuseren?
Ik wens slechts aandacht van een vriend
en zou u graag met gaven eren
van hoger waarde, als mijn tiend
verschuldigd aan een ziel vol dromen
van levend heil, vol poëzie
en edel denken: klare stromen
van eenvoud, licht en harmonie.
Maar 't zij zo. Neem dit onvolkomen
geheel goedgunstig aan en lees
mijn volks, utopisch, komisch, treurig
verhaal of mengeldicht: een kleurig
product dat half doordacht ontrees
aan slapeloosheid, inspiratie,
onrijpheid, dorheid, blijdschap, smart,
verstandelijke observatie
en de gekweldheid van het hart.

Aleksandr Sergejevitsj Poesjkin
Opdracht uit *Jewgeni Onegin*, 1827.

Contents

Introduction

The Standard Model of elementary particles is a relativistic quantum field theory that describes the particles observed in nature and their interactions, except for gravitation. The predictions of this theory coincide with observations in experiments with an astonishing precision, making it the most stringently tested existing scientific theory. Although the Standard Model is found to be self consistent, it exhibits several shortcomings. The mechanism to give mass to the particles is introduced ad-hoc, and requires the existence of a scalar particle in the theory, which is currently the only unobserved Standard Model particle. Moreover, the large difference between the characteristic energy scale of the theory and the much higher scale at which gravitational effects are expected to become important, leads to a naturalness problem, the so-called hierarchy problem, because of the need of extreme fine-tuning in the theory. Additional arguments, like the recent confirmation of neutrino oscillations and dark matter, lead to the consideration of the Standard Model as a low-energy effective theory with new physics entering at higher energy scales than currently accessible in experiment.

In the fall of 2007, the new Large Hadron Collider (LHC) will become operational at the CERN laboratory, near Geneva, Switzerland. This proton–proton collider and its experiments will provide the possibility to probe nature with about an order of magnitude increase in energy (14 TeV) compared to present-day accelerators. Because of this, the LHC will be able to probe an energy domain in which many of the proposed extensions of the Standard Model predict new particles or other observable effects of new physics. Doing so, answers may be found to many of the open questions of physics at both sub-atomic and cosmological scales.

One of the most attractive extensions of the Standard Model, which is able to solve many of its shortcomings, introduces a new so-called supersymmetry that relates fermions and bosons. Theories incorporating supersymmetry require an extension of the Standard Model Higgs sector, which gives in its minimal form rise to three neutral Higgs bosons and two charged ones. In this thesis the observability of sufficiently heavy[1] charged Higgs bosons is studied in the $H^{\pm} \to tb$ decay channel[2] in the Compact Muon Solenoid (CMS) experiment at the LHC. It is shown that the tools for identification of b-quark signatures are crucial experimental techniques for charged Higgs-boson identification and background suppression in this channel. In this context a novel method is designed to measure the b-identification performance on the data in CMS.

[1] The use of the words "light" and "heavy" is strictly speaking incorrect in this context, since there is no gravitational weight involved. Throughout the thesis these jargon words are used nevertheless for objects with respectively small and large mass.

[2] Charge conservation is implicit throughout the thesis.

The thesis is structured as follows. Chapter 1 introduces the theoretical framework of the Standard Model and its supersymmetric extensions. The enlarged Higgs sector in such extensions is described in a general and specific example of two-Higgs-doublet models. The charged Higgs boson, being the main subject of this thesis, is discussed in more detail from the phenomenological and experimental point of view. In Chapter 2 the LHC accelerator is introduced and its physics program is outlined. The theoretical description of proton–proton collisions is treated in Chapter 3. Many aspects of the strong interaction at low and high energy are discussed, and emphasis is put on the phenomenology relevant for charged Higgs-boson signals and backgrounds. In Chapter 4 the experimental setup of the CMS detector and the low-level reconstruction is summarized. Also the software and computing environment, together with the online-selection system are described, elaborating on the way these are used for the analyses in the thesis. In Chapter 5 a detailed survey is given of all the high-level reconstruction tools needed to conduct the analyses described in Chapters 6 and 7. It includes efficient reconstruction methods for charged leptons, jets and b jets, but also a global event-reconstruction method that allows to exploit kinematic constraints in the events.

In Chapter 6 a novel method to measure b-tagging efficiencies on data is presented in detail. Several jet samples are selected in top-quark pair events, and the kinematic properties of these events are exploited with a multivariate technique to enrich the b-jet content of the jet samples. The b-tagging performance can be directly measured on the data in such samples of high b-jet purity. With a detailed estimation of the involved systematic uncertainties, an optimization of the method is performed with respect to the total expected uncertainty. Chapter 7 concerns the analysis of charged Higgs-boson identification in the $H^{\pm} \to tb$ channel. Two search strategies are presented, tagging either three or four b jets. Multivariate likelihood ratios are used to optimize the suppression of both combinatorial and process backgrounds. After evaluation of the systematic uncertainties the visibility of the signal is interpreted in the parameter space of the Higgs sector. The obtained results, finally, are compared to the charged Higgs-boson discovery prospects in other decay channels in Chapter 8, and are discussed in the broader framework of supersymmetric Higgs-boson searches at the LHC.

Chapter 1

The Standard Model and Beyond

The current knowledge of elementary particle physics is the result of many decades of worldwide fundamental scientific research. Theoretical and experimental developments have been evolving side by side, each with their moments of breakthrough and discovery interleaved with periods of struggle or consolidation. Today, elementary particles and their interactions are described by the so-called Standard Model (SM), a field theory which combines special relativity and quantum mechanics. This description agrees extraordinary well with experimental observations, in such a way that the basic principles still hold after about 30 years of extensive testing. As such, the Standard Model is the most precisely tested scientific theory ever.

This chapter[1] discusses the main theoretical framework used for the studies in this thesis. In Section 1.1 the Standard Model is reviewed. Emphasis is put on the conceptual description of interactions and on the theoretical and experimental endeavours to understand the origin of electroweak symmetry breaking. In Section 1.2, the need for new physics beyond the Standard Model is argued. Supersymmetry is introduced as an appealing extension, which manifests itself minimally as a two-Higgs-doublet model. In this framework the charged Higgs boson is predicted, the particle that is the protagonist in this thesis. Finally, some alternative Standard Model extensions are highlighted.

1.1 The Standard Model of Elementary Particles

The theory that encapsulates the current experimental knowledge of elementary particles and their interactions is called the Standard Model [1–7]. It is a relativistic quantum field theory that describes all matter as being made up from spin-1/2 particles, called fermions. Interactions between these fermions happen through the exchange of spin-1 particles, called bosons, which are understood to arise from invariances of the theory under so-called gauge symmetries.

The fundamental fermions observed up to today in experiments are sub-divided in leptons and quarks. These two groups of particles come in three families or generations, that behave identically under interactions. No experimental evidence has been found so far for the existence of a fourth generation [8–11]. The six known lepton flavours are

[1] Throughout this chapter the so-called natural units are employed, in which $c = \hbar = 1$.

the electron e, the muon μ, the tau-lepton τ, and the three corresponding neutrinos ν_{e}, ν_μ and ν_τ. The six known quark flavours are labelled as up u, down d, strange s, charm c, bottom or beauty b and top t. They can be represented as

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

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

with the leptons shown in the first row, the quarks in the second, and where the generations are grouped by column. The quarks in the first row are also referred to as the up-type quarks, the ones in the second row as the down-type quarks. Each of these fermions is in addition accompanied by an anti-particle with opposite quantum numbers. All stable matter observed in the universe is made from the first generation of fermions and no stable anti-matter has been detected so far.

Three types of fundamental interactions between the fermions have been experimentally observed. The electromagnetic interaction is mediated by the photon γ, the weak interaction by the massive W^+, W^- and Z bosons, and the strong interaction is carried by gluons. The fourth known fundamental interaction, gravity, is very difficult to observe at sub-millimetre distances. Since the electromagnetic attraction between an electron and a proton is about 40 orders of magnitude larger than the corresponding gravitational attraction, gravity is negligible compared to the other interactions in today's particle-physics experiments.

In this section the main aspects of the theoretical framework of the Standard Model are summarized, following a bottom-up approach. First, interactions between fermions are introduced as a consequence of the invariance of the theory under local symmetry transformations. Then the symmetries underlying the Standard Model are discussed, the corresponding gauge fields are introduced, and the link is made with the observable gauge bosons. A discussion of the interactions and their energy dependence, is followed by a solution to the problem of an apparently massless theory. To conclude, an overview is given of the experimental status of the electroweak sector of the Standard Model.

1.1.1 Interactions from gauge symmetries

In the Standard Model, a free fermion with mass m is described as a spinor ψ by the Lagrangian

$$\mathcal{L} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi, \tag{1.1}$$

from which the Dirac equation of motion $(i\gamma^\mu\partial_\mu - m)\psi = 0$ can be derived.

Suppose a fermion's wave function changes, under a local phase transformation with rotation parameters $\vec{\epsilon}(x)$ in an internal space represented by the generators $\vec{\tau}$, as

$$\psi' = U\psi = e^{i\vec{\epsilon}(x)\cdot\frac{\vec{\tau}}{2}}\psi, \tag{1.2}$$

such that quantum-mechanical observables, which depend on $|\psi|^2$, remain invariant. The Lagrangian (1.1), however, is in general not invariant under the transforma-

tion (1.2). To retain the symmetry in the theory, a covariant derivative

$$\mathcal{D}_\mu = \partial_\mu - ig\frac{\vec{\tau}}{2}\vec{A}_\mu \tag{1.3}$$

is introduced, which is built from the usual space-time derivative ∂_μ and a new interacting vector field $\vec{A}_\mu$, with g an arbitrary parameter, which will determine the universal interaction strength associated to the field. Substituting the covariant derivative into the Lagrangian (1.1) yields

$$\begin{aligned}\mathcal{L} &= i\bar{\psi}\gamma^\mu\mathcal{D}_\mu\psi - m\bar{\psi}\psi \\ &= i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi - ig\bar{\psi}\gamma^\mu\frac{\vec{\tau}}{2}\vec{A}_\mu\psi,\end{aligned} \tag{1.4}$$

where the last term expresses the coupling between the fermion field and the new vector field. By demanding that

$$\mathcal{D}'_\mu\psi' = U(\mathcal{D}_\mu\psi), \tag{1.5}$$

such that the Lagrangian (1.4) is invariant under (1.2), the transformation relations for the components of the field $\vec{A}_\mu$ are derived to be

$$\frac{\tau^i}{2}A^{i\prime}_\mu = -\frac{i}{g}(\partial_\mu U)U^{-1} + U\frac{\tau^i}{2}A^i_\mu U^{-1}. \tag{1.6}$$

This relation, with U being a transformation matrix in some internal space, is very general. In the case where $U = e^{-i\chi(x)}$ represents a simple $U(1)$ phase transformation, the theory is Abelian and (1.6) simplifies to $A'_\mu = A_\mu - \partial_\mu\chi(x)/g$. A theory with a local non-Abelian phase invariance is called a Yang-Mills theory.

It is found that the requirement of a theory to be invariant under certain symmetry transformations, which are generally called gauge transformations, entails the introduction of associated vector fields, called gauge fields. These fields imply the existence of spin-1 particles, the gauge bosons, that couple to the fermions. In addition, gauge bosons in a Yang-Mills theory also exhibit self-interactions. Although the introduction of interactions follows from the requirement of gauge invariance under symmetry transformations, these actual symmetries are a priori not known. In practise, theories are being built the other way round, with the symmetries underlying the theory empirically proposed in order to explain experimental observations.

1.1.2 Gauge interactions in the Standard Model

To describe the experimental knowledge of the particles and their interactions at the quantum level, three symmetries are found to be necessary and sufficient in the theory. First of all, the Lagrangian exhibits a local $U(1)$ phase invariance. The gauge field required by this Abelian invariance of the theory is called B_μ. A second invariance, under a set of non-Abelian transformations that form an $SU(2)$ group, leads to the introduction of three W^i_μ fields ($i = 1, \ldots, 3$), one for each of the generators of the $SU(2)$ group. The third invariance, also non-Abelian, under a set of transformations that form an $SU(3)$ group, requires the introduction of eight G^a_μ fields ($a = 1, \ldots, 8$).

The covariant derivative, which ensures all three gauge invariances of the theory, takes the form

$$\mathcal{D}_\mu = \partial_\mu - ig_1\frac{Y}{2}B_\mu - ig_2\frac{\tau^i}{2}W^i_\mu - ig_3\frac{\lambda^a}{2}G^a_\mu, \tag{1.7}$$

where the scalar Y and the matrices τ^i and λ^a serve as the generators for respectively the $U(1)$ hypercharge, the $SU(2)$ weak isospin and the $SU(3)$ colour space. The first two terms are singlets in the $SU(2)$ and $SU(3)$ spaces. The third term is a 2×2 matrix in $SU(2)$ and a singlet in the other spaces, and the fourth term is a 3×3 matrix in $SU(3)$ and a singlet in the other spaces.

The full Lagrangian describing fermions and their interactions should arise by writing down the Dirac terms (1.1) for all fermions, and replacing the space-time derivatives by the covariant derivative (1.7). The way the fermions behave under the gauge transformations, however, depends on the charge they carry with respect to the gauge interaction. These charges are not predicted by the theory, and must be determined from experimental observations.

$SU(3)$ **colour** The $SU(3)$ colour charges are only present for the quarks, which appear as triplets under $SU(3)$ transformations, while the leptons are colour singlets. The gluons G^a transform in a colour octet.

$SU(2)$ **weak isospin** In the framework of the $SU(2)$ gauge transformations, it is instructive to project the Dirac fermion spinors ψ into left- and right-handed Weyl spinors ψ_L and ψ_R as

$$\psi = \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix}. \tag{1.8}$$

This way massless fermions are decoupled in the Lagrangian into left-handed and right-handed particles, because

$$\bar{\psi}\gamma^\mu\psi \;\; = \;\; \bar{\psi}_L\gamma^\mu\psi_L + \bar{\psi}_R\gamma^\mu\psi_R. \tag{1.9}$$

The left- or right-handedness of fermions is referred to as their chirality, which coincides with the helicity for massless fermions. For $SU(2)$ gauge transformations, the weak-isospin charge turns out to be different for particles with different chirality. Left-handed fermions transform as isospin doublets within the lepton and quark families, while the right-handed counterparts transform as singlets with zero isospin and hence do not interact with the $SU(2)$ gauge bosons. The W^i gauge bosons themselves appear as a triplet. The relevant quantum number for the particles is the third component T_3 of the weak isospin T.

This chiral nature of the weak isospin transformations has an important consequence. Fermion mass terms in the Lagrangian can be written as

$$m\bar{\psi}\psi = m(\bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R), \tag{1.10}$$

and manifestly violate $SU(2)$ gauge invariance. Therefore, at this point, fermion mass terms must be excluded from the theory.

$U(1)$ **hypercharge** The $U(1)$ hypercharge, necessarily inducing transformations as singlets, is non-zero for all fermions except for the right-handed neutrinos. As a convention the corresponding quantum number Y is chosen $Y = -1$ for left-handed leptons. Since right-handed neutrinos do not couple to any of the introduced interactions, they are sterile, and do not form a part of the theory. Recent observations of neutrino masses, however, require an extension of the presented Standard Model that may include right-handed neutrinos.

To summarize, the interactions between the fermions in the Standard Model are generated by the $U(1)_Y \otimes SU(2)_L \otimes SU(3)_C$ gauge group. Unlike the strong and the $SU(3)$ interaction, the $U(1)$ and $SU(2)$ gauge interactions cannot be identified directly with the electromagnetic and weak interactions respectively. This can for example be understood from the fact that two leptons in the same $SU(2)$ doublet carry a different electric charge. Instead, the observed interactions are a manifestation of the combined electroweak $U(1)_Y \otimes SU(2)_L$ gauge group, where the physical fields A_μ, Z_μ and $W^\pm_\mu$, for respectively the photon, the Z boson and the $\mathrm{W}^\pm$ bosons, arise as combinations of the gauge fields according to

$$\begin{aligned} A_\mu &= \frac{g_2 B_\mu - g_1 Y_L W^3_\mu}{\sqrt{g_2^2 + g_1^2 Y_L^2}} & W^+_\mu &= \frac{-W^1_\mu + iW^2_\mu}{\sqrt{2}} \\ Z_\mu &= \frac{g_1 Y_L B_\mu + g_2 W^3_\mu}{\sqrt{g_2^2 + g_1^2 Y_L^2}} & W^-_\mu &= \frac{-W^1_\mu - iW^2_\mu}{\sqrt{2}}. \end{aligned} \tag{1.11}$$

These combinations, which allows to describe the experimentally observed electromagnetic and weak interactions in terms of the underlying electroweak gauge interaction, implies, among other things, that $Y_R = 2Y_L$ and $Y = 2(Q_{\mathrm{EM}} - T_3)$, with Q_{EM} the observable electromagnetic charge. From these relations and the charge and chirality information, the underlying quantum numbers of the particles can be derived. In Table 1.1 all fermions and their main quantum numbers are summarized. Although the right-handed neutrinos are not considered as part of the Standard Model, they are mentioned for completeness.

Particle	Y	T	T_3	Q_{EM}
$(\nu_{\mathrm{e},\mu,\tau})_L$	-1	$1/2$	$+1/2$	0
$(\mathrm{e},\mu,\tau)_L$	-1	$1/2$	$-1/2$	-1
$(\nu_{\mathrm{e},\mu,\tau})_R$	0	0	0	0
$(\mathrm{e},\mu,\tau)_R$	-2	0	0	-1
$(\mathrm{u},\mathrm{c},\mathrm{t})_L$	$+1/3$	$1/2$	$+1/2$	$+2/3$
$(\mathrm{d},\mathrm{s},\mathrm{b})_L$	$+1/3$	$1/2$	$-1/2$	$-1/3$
$(\mathrm{u},\mathrm{c},\mathrm{t})_R$	$+4/3$	0	0	$+2/3$
$(\mathrm{d},\mathrm{s},\mathrm{b})_R$	$-2/3$	0	0	$-1/3$

Table 1.1: Overview of the main quantum numbers for fermions.

1.1.3 Renormalization and running coupling constants

The Lagrangian of the Standard Model fully fixes the matter content and the interactions in the theory. Despite its elegance and simplicity, the derivation of predictions from the Lagrangian is a highly non-trivial task. This is simplified by the introduction of Feynman diagrams and rules for their calculation, enabling a diagrammatic approach to calculations of probabilities associated to specific processes. Quantum-mechanical corrections need to be accounted for when performing such calculations, which introduce extra loops and vertices in the Feynman diagrams. By ordering all the diagrams as a function of the number of vertices, a series is formed with increasing powers of the coupling constant. Such an expansion can be used for perturbative calculations, provided the coupling constant is smaller than unity. In such calculations, however, one is confronted with divergences, even in the easiest case of electromagnetism, which is mathematically described by Quantum Electro Dynamics (QED). In the Standard Model, however, these infinities can always be absorbed into the unobservable bare parameters of the Lagrangian by a technique called renormalization [12]. In the following, the so-called renormalization-group approach is described for coupling constants in field theory [1].

Consider a dimensionless physical observable R, dependent on a single energy scale Q, which is calculated as a perturbation series in the coupling $\alpha = g^2/4\pi$. In a quantum field theory the calculation of the series up to a certain finite order requires the introduction of a renormalization scale μ, without physical meaning, to remove so-called ultra-violet divergences at increasing energy scales. Since R is dimensionless, it can only depend on the ratio Q^2/μ^2 and the normalised coupling α. Defining

$$t = \ln\left(\frac{Q^2}{\mu^2}\right) \qquad \text{and} \qquad \beta(\alpha) = \mu^2 \frac{\partial \alpha}{\partial \mu^2}, \tag{1.12}$$

the μ independence of R implies that

$$\mu^2 \frac{\mathrm{d}}{\mathrm{d}\mu^2} R(\frac{Q^2}{\mu^2}, \alpha) = \left[-\frac{\partial}{\partial t} + \beta(\alpha)\frac{\partial}{\partial \alpha}\right] R(e^t, \alpha) = 0. \tag{1.13}$$

Introducing the so-called running coupling $\alpha(Q^2)$ through

$$t = \int_{\alpha(\mu^2)}^{\alpha(Q^2)} \frac{\mathrm{d}x}{\beta(x)}, \tag{1.14}$$

such that the renormalization-group equation

$$Q^2 \frac{\partial \alpha}{\partial Q^2} = \beta(\alpha) \tag{1.15}$$

is satisfied, the expression $R(1, \alpha(Q^2))$ is a solution of (1.13) and it is seen that all scale dependence of R enters only through the running of the coupling constant $\alpha(Q^2)$.

For each interaction, the β function can be calculated as an expansion in the coupling α of the higher-order corrections to the vertices of the theory. Once calculated up to a certain order, the β function predicts how the coupling constant varies with

the probed scale Q^2 through (1.15). The absolute value of the coupling α, however, has to be obtained from experiment. The fundamental parameter of the theory $\alpha(Q'^2)$ can thus be chosen at a reference scale Q'^2.

The calculation of the first-order term in the expansion of the β function leads to

$$\alpha(Q^2) = \frac{\alpha(\mu^2)}{1 + b\,\frac{\alpha(\mu^2)}{4\pi}\ln(\frac{Q^2}{\mu^2})}, \tag{1.16}$$

with b negative in the case of electromagnetism. The resulting increase of the electromagnetic coupling with the probed scale is related to the screening of charges by the vacuum. For both the weak and the strong interaction, the factor b is positive. This follows from the non-Abelian structure of the theory, that induces boson self-interactions. In the calculations of the expansion of the β function, these self-interactions add opposite-sign loop diagrams with bosons to the smaller fermionic-loop contributions. As a consequence the interaction strength increases at small scales, with the interaction generated by the largest symmetry group becoming strongest.

The running of the coupling constant is particularly important for the strong interaction, described in the theory of Quantum Chromo Dynamics (QCD). The decreasing coupling α_s with increasing Q^2 reflects the so-called asymptotic freedom, being the property of partons in QCD interactions to be less affected by higher order effects at large energy scales. The increase of the interaction strength for low Q^2 scales, on the other hand, results in the breakdown of perturbativity, which ensures the confinement of quarks inside colour neutral hadrons (see also Section 3.3.3).

In QCD, the reference scale for the definition of the coupling α_s is typically taken as $\mu = m_{\mathrm{Z}}$, large enough to be in the perturbative domain. Another approach is to introduce the scale Λ_{QCD} directly into the definition (1.14) for the strong coupling α_s, as the integration constant in

$$\ln\frac{Q^2}{\Lambda^2_{\mathrm{QCD}}} = -\int_{\alpha_s(Q^2)}^{\infty}\frac{\mathrm{d}x}{\beta(x)}. \tag{1.17}$$

By this convention Λ_{QCD} is the scale at which the coupling α_s would diverge if extrapolated outside the perturbative domain. The value of Λ_{QCD} is $\approx 200\,\mathrm{MeV}$, depending on the order at which α_s is being evaluated, the number of active quark flavours n_f in the calculation and the renormalization scheme used. As an illustration, the running of the QCD coupling is shown in Figure 1.1 at leading and next-to-leading order, with $\Lambda_{\mathrm{QCD}} = 200\,\mathrm{MeV}$ and $n_f = 5$.

1.1.4 The origin of mass: spontaneous symmetry breaking

The theory built in Section 1.1.2 not only allows to quantitatively explain experimental observations, but also holds predictive power from calculations of perturbative expansions. However, up to now it still contains no mass terms for the fermions, since these violate $SU(2)$ invariance. Also bosonic mass terms of the form $-m^2 A_\mu A^\mu$ for the Z and $\mathrm{W}^\pm$ bosons cannot be added without breaking gauge invariance. The mechanism that has been found to mend these problems is referred to as spontaneous symmetry

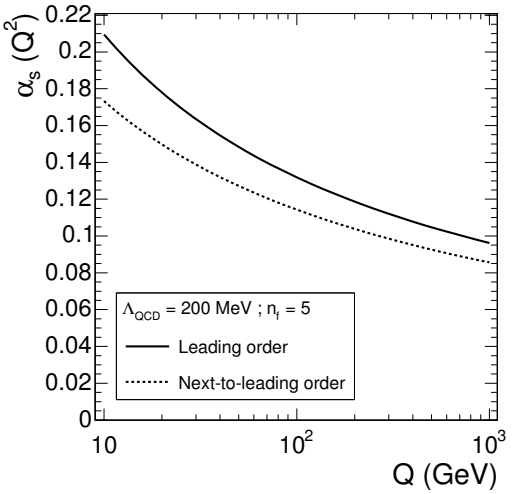


Figure 1.1: Evolution of the QCD coupling constant at leading and next-to-leading order.

breaking, which leaves the Lagrangian of the theory still invariant under the gauge transformations, but which removes the gauge symmetry of the vacuum [13–15].

In the Standard Model this mechanism is referred to as the Higgs mechanism. It entails the introduction of the scalar Higgs field ϕ, that is a doublet in $SU(2)$ space, carries non-zero $U(1)$ hypercharge and is a singlet in $SU(3)$ colour space,

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \tag{1.18}$$

with ϕ^+ and ϕ^0 complex fields. The invariant Lagrangian for this field can be written as

$$\mathcal{L}_\phi = \left(\mathcal{D}^\mu\phi\right)^\dagger \mathcal{D}_\mu\phi - V(\phi) = \left(\mathcal{D}^\mu\phi\right)^\dagger \mathcal{D}_\mu\phi - \mu^2\phi^\dagger\phi - \lambda\left(\phi^\dagger\phi\right)^2, \tag{1.19}$$

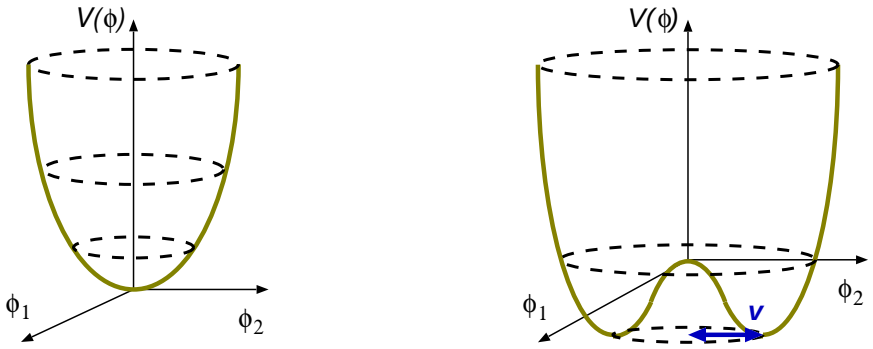
with μ^2 a mass parameter and $\lambda > 0$ the strength of the Higgs field's self interaction. For $\mu^2 > 0$ the potential $V(\phi)$ has a global minimum for $\phi = 0$. When taking $\mu^2 < 0$ on the other hand, the potential has many non-zero degenerate minima for

$$\phi^\dagger\phi = -\frac{\mu^2}{2\lambda} \equiv \frac{v^2}{2}, \tag{1.20}$$

with v the vacuum expectation value of the field ϕ. Writing $\phi^+ = (\phi_1 + i\phi_2)/\sqrt{2}$ and $\phi^0 = (\phi_3 + i\phi_4)/\sqrt{2}$, it can be seen that these minima form a four-dimensional sphere, since $2\phi^\dagger\phi = \phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2 = v^2$. In the two-dimensional case the degenerate vacua lie on a circle, which is visualized in Figure 1.2(a) for an unbroken theory with $\mu^2 > 0$, and in Figure 1.2(b) for spontaneous symmetry breaking, with $\mu^2 < 0$.

The particle content of the theory must be studied in the region of the vacuum. In order to do this, a direction is chosen in $SU(2)$ space by selecting a particular vacuum $\phi_3 = v$ and $\phi_1 = \phi_2 = \phi_4 = 0$, and an expansion is made around the minimum:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}, \tag{1.21}$$



(a) $\mu^2 > 0$ (b) $\mu^2 < 0$

Figure 1.2: Two-dimensional representation of the Higgs potential in the case (a) of an unbroken theory ($\mu^2 > 0$) and in the case (b) of spontaneous symmetry breaking ($\mu^2 < 0$).

with $H(x)$ a real scalar field. Inserting (1.21) and the covariant derivative (1.7) into (1.19), it can be shown that the photon remains massless, while the mass terms $(vg_2/2)^2 W^{+\mu} W^-_\mu$ and $(v\sqrt{g_1^2+g_2^2}/2)^2 Z^\mu Z_\mu/2$ appear. Identification of these terms with the electroweak gauge-boson mass terms allows to deduce the expectation value $v = 246\,\text{GeV}$ of the Higgs field in the vacuum. Apart from the mass terms, several couplings arise between the gauge bosons and the field $H(x)$, which is associated to a new scalar particle, the Higgs boson. This new boson has mass $\sqrt{2\lambda}v$, with λ a free parameter, carries no electric charge, has as weak isospin component $T_3 = -1/2$, and hence as hypercharge $Y = 1$.

With the Higgs field available in an $SU(2)$ doublet, it is straightforward to write an $SU(2)$-invariant interaction of left-handed fermions with the Higgs field. Adding the right-handed counterparts in these interactions allows to obtain fermionic-mass terms in the Lagrangian, after the spontaneous symmetry breaking with (1.21).

In contrast to the bosonic mass terms, the presence of the fermionic couplings to the Higgs field is not determined by a gauge principle. Therefore also the corresponding coupling constants, called Yukawa coupling constants, or equivalently the fermionic masses, are free parameters of the theory. For the bosons, on the other hand, the masses are fixed by the couplings of the gauge interactions and the vacuum expectation value of the Higgs field. This led to the accurate prediction of the $W^\pm$- and Z-boson masses before their discoveries in 1983 [16–19], one of the most impressive achievements of the electroweak theory.

1.1.5 Experimental status of the Standard Model

The Standard Model is a theory that has been constructed to explain the observations in the world of elementary particle physics, but that in the course of time itself gave rise to precise predictions of new particles or phenomena, leading to new experimental discoveries. The most famous recent discoveries of predicted particles have been the observation of W and Z bosons in the UA1 and UA2 experiments in 1983, and the

discovery of the top quark in the CDF and DØ experiments in 1995 [20, 21]. The existence of only one particle is still unconfirmed: the Higgs boson, the key to electroweak symmetry breaking in the Standard Model. The most stringent limit is set by direct searches at LEP in the dominant Higgsstrahlung mode, where the Higgs boson is produced in association with a Z boson and decays into a $\mathrm{b\bar{b}}$ pair. The combination of the results from all four LEP experiments did not show conclusive evidence for a Standard Model Higgs boson, which resulted in a lower limit $m_{\mathrm{H}} > 114\,\mathrm{GeV}$ on the Higgs-boson mass at 95% C.L. [22].

The measurements of Standard Model observables have reached such a precision that they are in many cases sensitive to radiative corrections to the tree-level expectations. Since these corrections are often induced by other virtual particles, such observables can hold information on several parameters of the theory at the same time. The effect induced by the presence of the Higgs boson is a particularly interesting example. Although this particle has not been observed yet, its mass enters in one-loop corrections to electroweak parameters at most through a logarithmic dependence [23]. By combining all experimental knowledge on the electroweak sector of the Standard Model in a global fit, the self-consistency of the theory is tested, and an indirect bound can be derived on the unknown Higgs-boson mass [24, 25]. In Figure 1.3 the observed value of $\Delta\chi^2 = \chi^2 - \chi^2_{\mathrm{min}}$ is shown as a function of the hypothesized Higgs-boson mass m_{H}. The result for the best fit is $m_{\mathrm{H}} = 94^{+36}_{-27}\,\mathrm{GeV}$, with $\chi^2/\mathrm{n.d.f.} = 17.9/13$, corresponding to a 16% probability. Including the theoretical uncertainty, the one-sided 95% C.L. yields the bound $m_{\mathrm{H}} < 166\,\mathrm{GeV}$. Taking also the lower limit from the direct search into account, the bound is raised to $m_{\mathrm{H}} < 199\,\mathrm{GeV}$.

The indirect upper limit on the Higgs-boson mass quoted above is obtained from a global fit which includes both W-boson and top-quark masses. It is instructive to separate these two experimental constraints from all other electroweak data in the fit, since the experiments at the Tevatron are expected to further constrain these masses in the coming years. The current world averages are $m_{\mathrm{W}} = 80.392 \pm 0.029\,\mathrm{GeV}$ [24, 25] and $m_{\mathrm{t}} = 171.4 \pm 2.1\,\mathrm{GeV}$ [26]. In Figure 1.4 the 68% C.L. contours are shown in the $(m_{\mathrm{t}}, m_{\mathrm{W}})$ plane, for the direct measurements of the W-boson and top-quark masses and for the global fit excluding the direct measurement of these masses. Also shown in this figure is the corresponding region for the Standard Model Higgs-boson mass between its lower limit and 1 TeV. It is seen that the most important parameters constraining the Higgs-boson mass in the global electroweak fit are the W-boson and top-quark masses, which is expected due to the large couplings of these particles to the Higgs boson.

Currently, the direct search for the Higgs boson is being continued at the Tevatron proton–anti-proton accelerator. As discussed in Chapter 2, searches with hadron colliders allow to probe a large energy domain with a single centre-of-mass energy, but need sufficient collision rates to overcome the large hadronic backgrounds. Currently the limits obtained are not competitive yet with the direct searches at LEP. It is expected, however, that an exclusion at 95% C.L. will come within reach during the Tevatron Run-II period, first in a mass range around $m_{\mathrm{H}} \sim 160\,\mathrm{GeV}$ where the $\mathrm{H} \to \mathrm{W^+W^-}$ decay dominates, and later also in the $\mathrm{H} \to \mathrm{b\bar{b}}$ mode at the lower mass limit $m_{\mathrm{H}} \sim 115\,\mathrm{GeV}$ set by LEP [27, 28].

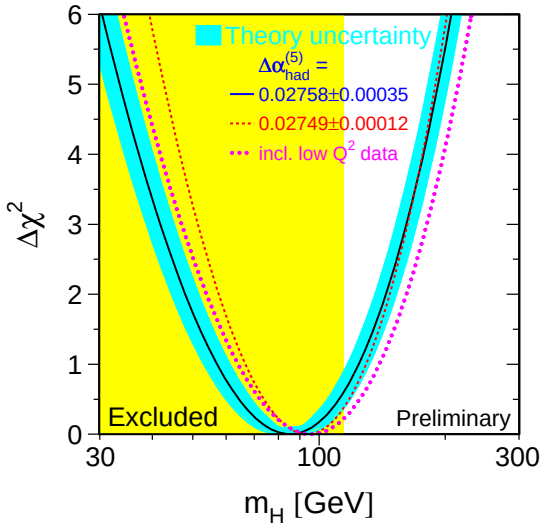


Figure 1.3: $\Delta\chi^2 = \chi^2 - \chi^2_{\text{min}}$ as a function of the hypothesized Higgs-boson mass m_{H} in a global fit of all electroweak data. The vertical band corresponds to the exclusion from direct Higgs-boson searches.

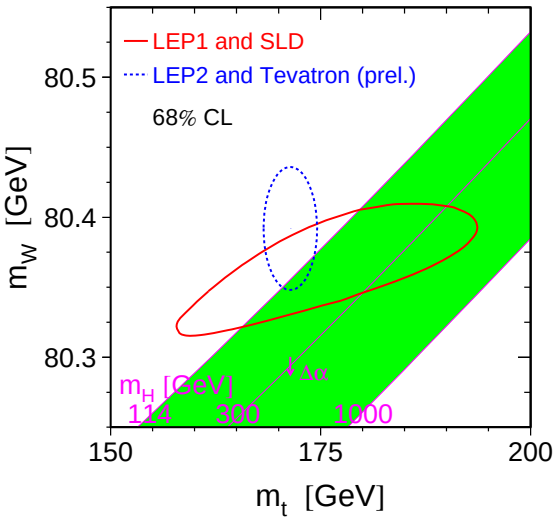


Figure 1.4: Contours at 68% C.L., for the global fit of electroweak data excluding the W-boson and top-quark masses, and for the direct measurement of these masses, in the $(m_{\text{t}}, m_{\text{W}})$ plane.

In this section only the electroweak status of the Standard Model related to the Higgs-boson has been reviewed. Nevertheless, the Standard Model has proved to be also an extraordinary successful theory in many other domains, like for instance in the study of the proton structure, in the description of the quark–anti-quark mixing and related CP violation, and in explaining many of the intrinsic properties of the particles in the theory. A review of these other successes, however, falls outside the scope of this thesis. More information can be found in [29, 30].

1.2 Beyond the Standard Model

Although the Standard Model allows to describe tremendously well all the phenomena that are currently observed in particle physics, it is conceptually incomplete. In this section several of the arguments are described why new physics is to be expected at higher energy scales than probed until now. Next, several extensions of the Standard Model are presented that try to provide elegant solutions to some or all of these problems. Emphasis is put on supersymmetric models, and in particular on the Higgs sector of such models. This Higgs sector is characterized by the presence of charged Higgs bosons, which are the subject of the studies presented in this thesis.

1.2.1 Shortcomings of the Standard Model

A variety of indications exist that more or other fundamental physics remains to be discovered, and that the Standard Model needs to be seen as an effective theory at current energy scales and below. Some major arguments for new physics beyond the Standard Model are summarized below:

Non-observation of the Higgs boson The current non-observation of the Higgs boson leaves the possibility open that the Higgs mechanism is not the key to electroweak symmetry breaking. Unlike the gauge interactions that arise from an invariance principle, no conceptual foundation is present for the rather technical addition of the Higgs field to the theory. Nevertheless, the Higgs mechanism is appealing because of its simplicity.

From the theoretical side upper and lower bounds can be derived on the Higgs-boson mass, requiring the Standard Model to be self-consistent. In Figure 1.5 the upper and lower bounds on the Standard Model Higgs-boson mass are shown as a function of the cut-off scale Λ at which the Standard Model is to be replaced by a higher-energy theory [31]. In the region below the lower curve, called the vacuum-stability bound, the quartic Higgs-boson coupling becomes negative and the potential is unbounded from below. The region above the upper curve, called the triviality bound, is forbidden because it leads to a divergent Higgs-boson self-coupling, a so-called Landau pole, causing a loss of perturbativity. The bound widths reflect the uncertainties in the determination of the Higgs-boson mass limits.

From Newton's constant G, the Planck mass scale $m_{\mathrm{Pl}} \sim 10^{19}\,\mathrm{GeV}$ can be built, which is the scale at which quantum gravitational effects become important. If the validity of the Standard Model is assumed up to the Planck scale, the allowed Higgs-boson mass range is between 130 and 190 GeV. The non-observation of a Higgs boson, or the observation of the Higgs boson outside these bounds, directly implies the existence of non-Standard Model physics at a scale below the Planck scale. Additionally, from the non-observation at the current energy scales of the Higgs boson or physics beyond the Standard Model, the constraint $m_{\mathrm{H}} < \mathcal{O}(1\,\mathrm{TeV})$ can be deduced. Additionally, the same upper bound can be derived from the requirement of WW-scattering amplitudes not to violate unitarity.

The hierarchy problem The typical scale for electroweak physics is found to be of the order of $m_{\mathrm{Z}} \sim 10^2\,\mathrm{GeV}$. The fundamental scale of gravity, however, the Planck mass scale $m_{\mathrm{Pl}} \sim 10^{19}\,\mathrm{GeV}$, is much larger. In the Standard Model, no new physics is expected between these scales, since all fundamental interactions but the gravitational one are already accounted for. This large discrepancy between both scales gives rise to a difficulty referred to as the hierarchy problem.

The observable Higgs boson mass can be written as the sum of the bare mass from the renormalized Lagrangian and corrections from loop diagrams,

$$m_{\mathrm{H}}^2 = m_{\mathrm{H},0}^2 + \Delta m_{\mathrm{H}}^2. \tag{1.22}$$

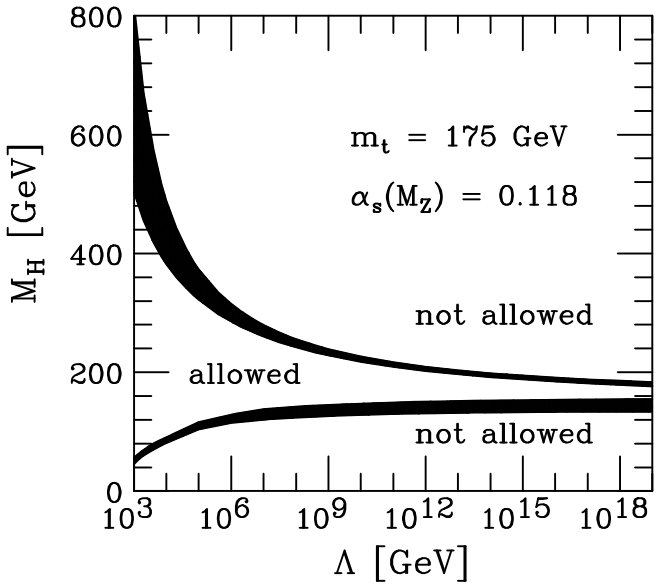


Figure 1.5: Theoretical constraints on the Higgs-boson mass, as a function of the high-energy cut-off Λ of the Standard Model.

The contributions to Δm_{H}^2 from fermion loops, visualized at first order in Figure 1.6(a), lead to a negative quadratic divergence. Boson loops, represented at first order in Figure 1.6(b), lead to a similar quadratic divergence, but with a positive sign. These contributions, which are regularized by the cut-off scale of the theory m_{Pl}, are also proportional to the respective coupling of the particle in the loop to the Higgs field. Therefore, the corrections are dominated by the top quark, and $\Delta m_{\mathrm{H}}^2 \sim -m_{\mathrm{Pl}}^2$, rendering the Higgs-boson mass the highest mass scale in the theory. To obtain the expected Higgs-boson mass $m_{\mathrm{H}}^2 \sim (100\,\mathrm{GeV})^2$, the bare mass $m_{\mathrm{H},0}$ needs to be of the order of the Planck scale, and a cancellation between the square of the bare mass and the corrections is required over more than 30 orders of magnitude. Although such fine-tuning is mathematically no problem, it leaves questions on the naturalness of the theory. In addition, the reason for the electroweak symmetry breaking turns out to be a consequence of the fact that the sum (1.22) leaves a positive number, seemingly by accident. If it were negative, the scalar potential in (1.19) would have $\mu^2 > 0$ and no spontaneous symmetry breaking would occur.

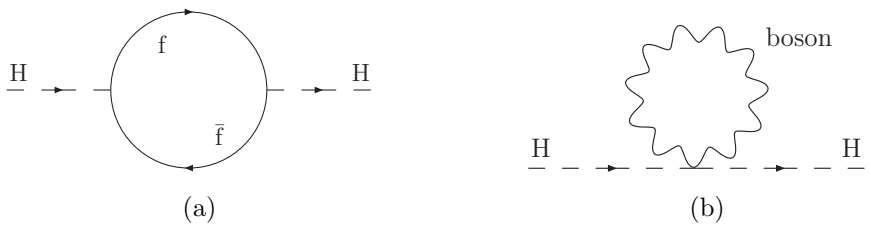


Figure 1.6: Feynmann diagrams of one-loop virtual mass corrections for scalar particles from (a) fermions and (b) bosons.

Unification of the coupling constants The success of the unified description of electromagnetism and the weak interaction has led to the hope that all fundamental interactions can be described by a single symmetry group. This implies that the coupling strengths of the strong, the weak and the electromagnetic interactions should unify at some energy scale. In Section 1.1.3 the dependence of interaction strengths is discussed as a function of the scale at which the interaction happens, leading to the concept of running coupling constants. The expansion of the β function up to first order, leads to the relation (1.16), which expresses the inverse of the coupling strength as a linear function of $\ln Q$ for sufficiently high scales Q. In the Standard Model the coefficients b of this relation are given by [2]

$$\begin{aligned} b_1 &= -\frac{4}{3}n_g - \frac{1}{10}n_{\mathrm{H}} \\ b_2 &= \frac{22}{3} - \frac{4}{3}n_g - \frac{1}{6}n_{\mathrm{H}} \\ b_3 &= 11 - \frac{4}{3}n_g, \end{aligned} \tag{1.23}$$

where n_g is the number of generations of quarks and leptons, and n_{H} is the number of Higgs doublet fields. If the Standard Model theory is expected to be self-consistent, the running couplings, with two-loop corrections taken into account, are proved not to converge to a single value. Therefore, unification of the coupling constants is only expected with new physics at higher energies.

Gravity The Standard Model does not describe gravitational interactions, and as such is incomplete in providing a "theory of everything". Attempts have been made to describe gravity as a quantum field theory, with the interaction mediated by a spin-2 boson, the graviton, associated to the gravitational tensor field. Such extensions of the Standard Model, however, need also an adapted description of the Standard Model itself, because of problems such as non-renormalizability.

Cosmological problems Many astrophysical observations point towards the existence of a significant amount of neutral, non-baryonic matter, referred to as dark matter. The current best measurement from the cosmic microwave background shows the presence of six times more dark than baryonic matter in the universe [32]. In the Standard Model this dark matter cannot be explained. The inclusion of massive right-handed neutrinos in the Standard Model, motivated by the observation of neutrino oscillations, does not provide a way to account for this dark matter either, due to the restrictive neutrino-mass sum constraint $\sum_{i=1}^{3}(m_\nu)_i < 0.2\,\mathrm{eV}$ [33].

Another cosmological problem is the asymmetry between matter and anti-matter experimentally constrained by the non-observation of anti-matter in primordial cosmic rays. Theories of baryogenesis that try to explain this asymmetry, require CP violation at a level not allowed by the Standard Model. Such models also require baryon-number violation, which is severely constrained experimentally by for example the limit on the proton lifetime $\tau_{\mathrm{p}} > 1.9 \times 10^{29}\mathrm{y}$ [30]. In the Standard Model, however, baryon number is exactly conserved.

1.2.2 The 2HDM: an extension of the Standard Model Higgs sector

The many successes of the Standard Model lead to the consideration of this theory as an effective low-energy limit of some more profound theory. Such theories are therefore often constructed by extending the Standard Model or by generalizing it. An example of such an extension is the two-Higgs-doublet model (2HDM) [34]. In such a model the minimally extended Higgs sector consists of two $SU(2)_L$-doublet complex scalar fields ϕ_1 and ϕ_2, represented by

$$\phi_1 = \begin{pmatrix} \phi_1^+ \\ \phi_1^0 \end{pmatrix}, \qquad \phi_2 = \begin{pmatrix} \phi_2^+ \\ \phi_2^0 \end{pmatrix}. \tag{1.24}$$

The Higgs fields at the minimum of the Higgs potential, causing the spontaneous symmetry breaking of $SU(2)_L \otimes U(1)_Y$ down to $U(1)_{\mathrm{EM}}$, are given by

$$\phi_1 = \begin{pmatrix} 0 \\ v_1 \end{pmatrix}, \qquad \phi_2 = \begin{pmatrix} 0 \\ v_2 e^{i\xi} \end{pmatrix}, \tag{1.25}$$

where the squared masses of the resulting physical Higgs bosons need to be positive, and the presence of the phase ξ introduces a source of CP violation in the model when $\sin\xi \neq 0$. In the following the choice $\xi = 0$ is chosen as simplification.

It is convenient to introduce

$$v^2 = v_1^2 + v_2^2 = \frac{4m_{\mathrm{W}}^2}{g_2^2} \qquad \text{and} \qquad \tan\beta = \frac{v_2}{v_1}. \tag{1.26}$$

After electroweak symmetry breaking five physical Higgs bosons remain,

$$\begin{aligned} \mathrm{h} &= \sqrt{2}\left[-(\mathrm{Re}[\phi_1^0] - v_1)\sin\alpha + (\mathrm{Re}[\phi_2^0] - v_2)\cos\alpha\right] \\ \mathrm{H} &= \sqrt{2}\left[(\mathrm{Re}[\phi_1^0] - v_1)\cos\alpha + (\mathrm{Re}[\phi_2^0] - v_2)\sin\alpha\right] \\ \mathrm{A} &= \sqrt{2}\left(-\mathrm{Im}[\phi_1^0]\sin\beta + \mathrm{Im}[\phi_2^0]\cos\beta\right) \\ \mathrm{H}^\pm &= -\phi_1^\pm \sin\beta + \phi_2^\pm \cos\beta, \end{aligned} \tag{1.27}$$

with $m_{\mathrm{H}} \geq m_{\mathrm{h}}$, and the mixing angle α a function of the various couplings in the Higgs potential. Instead of the one free parameter in the Higgs sector of the Standard Model, this model has six free parameters, chosen as the four Higgs-boson masses, the ratio of the vacuum expectation values of the two Higgs doublets denoted as $\tan\beta$, and the Higgs mixing angle α.

The phenomenology of the two-Higgs-doublet model depends in detail on the various couplings of the Higgs bosons to fermions, gauge bosons and Higgs bosons. The Higgs-boson couplings to the gauge bosons follow from gauge invariance and are thus model independent. Some of these couplings can be suppressed if either $\sin(\beta-\alpha)$ or $\cos(\beta-\alpha)$ is very small. The Higgs-boson couplings to the fermions, on the other hand, do not follow directly from an underlying principle, and hence leave freedom in the general model. Two choices for the Higgs-fermion interactions are regularly encountered. In the 2HDM Type-I, the quarks and leptons do not couple to the first Higgs doublet ϕ_1,

but couple only to the second Higgs doublet ϕ_2 in a manner similar to the Standard Model. In the 2HDM Type-II, on the other hand, ϕ_1 is assumed to couple only to the down-type quarks and to the charged leptons, while ϕ_2 couples only to up-type quarks and neutrinos.

In this thesis, the 2HDM Type-II is of particular interest, as it is the required configuration of the Higgs sector in the Minimal Supersymmetric Standard Model (MSSM), introduced in the next Section. In such a model, the phenomenology of charged Higgs-boson production is governed by the fermion couplings,

$$g_{\mathrm{H}^-\mathrm{t}\bar{\mathrm{b}}} = \frac{g_2}{2\sqrt{2}m_{\mathrm{W}}}\left[m_{\mathrm{t}}(1+\gamma_5)\cot\beta + m_{\mathrm{b}}(1-\gamma_5)\tan\beta\right], \tag{1.28}$$

given here for the third-generation quarks. The couplings to up-type and down-type fermions are enhanced respectively at small and large $\tan\beta$. This behaviour is also observed for the A pseudoscalar, and to a lesser extent for the scalars h and H, for which the dependence on the mixing angle α should be taken into account as well.

The study of charged Higgs-boson identification is the main subject in this thesis. Two out of the six parameters of the 2HDM Higgs sector describe the tree-level $\mathrm{H}^\pm$ production and decay: $\tan\beta$ and the mass of the charged Higgs boson $m_{\mathrm{H}^\pm}$. The branching fractions for the decay channels of the charged Higgs boson depend strongly on $m_{\mathrm{H}^\pm}$. In this thesis, these branching fractions are determined with the program `HDECAY` [35], assuming a 2HDM Type-II. As shown in Figure 1.7, for $m_{\mathrm{H}^\pm} < m_{\mathrm{t}} + m_{\mathrm{b}}$, the $\mathrm{H}^\pm \to \tau\nu$ channel dominates. For larger masses, the channel $\mathrm{H}^\pm \to \mathrm{tb}$ studied in this thesis opens up. The suppression of the charged Higgs-boson couplings to down-type fermions for small $\tan\beta$ is reflected in the $\mathrm{H}^\pm \to \tau\nu$ branching fraction. As a consequence heavy charged Higgs bosons decay for almost 100% to top and bottom when $\tan\beta$ is small.

In Section 3.4.3 the phenomenology of charged Higgs-boson production is discussed in the framework of a general two-Higgs-doublet model. The experimental aspects of charged Higgs-boson identification are considered in Chapters 7 and 8.

1.2.3 Supersymmetry and the MSSM

Although two-Higgs-doublet models are interesting because of the extra freedom compared to the Standard Model, they do not solve the problems discussed in Section 1.2.1. Another appealing extension of the Standard Model is supersymmetry, which in the minimal configuration contains the Higgs sector of a 2HDM Type-II. Supersymmetric models naturally provide answers to several of the problems of the Standard Model [36].

A supersymmetry transformation turns a bosonic state into a fermionic state, and vice versa. It can be shown that the operator Q, which generates such transformations, must be an anti-commutating fermionic operator. It follows also that supersymmetry is necessarily a space-time symmetry, in contrast to the gauge symmetries. The operator Q also commutes with the generators of gauge transformations. Therefore particles in the same so-called supermultiplet, having the same mass, must also be in the same representation of the gauge group, and hence must have the same electric charge, weak isospin and colour. An inspection of the Standard Model particles and their quantum numbers shows the impossibility to find superpartners among the known particles.

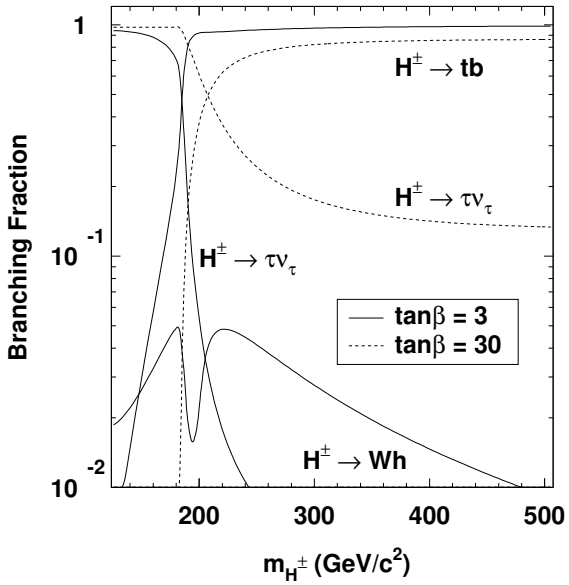


Figure 1.7: Charged Higgs-boson branching fractions as a function of the charged Higgs-boson mass, obtained with `HDECAY` [35].

Therefore supersymmetric models predict at least a doubling of the number of particles compared to the Standard Model.

One of the big virtues of supersymmetry is the resolution of the hierarchy problem. It guarantees the exact cancellation of quadratic divergences from fermion loops in the corrections to the Higgs-boson mass, by boson or scalar loops with corresponding superpartners, and vice versa. Since no discoveries have been made yet of supersymmetric partners of Standard Model particles, some of which should be very easy to detect, supersymmetry must be a broken symmetry. To maintain the cancellation to all orders of the fermion and boson loops of the mentioned quadratic divergences, the supersymmetric Lagrangian needs to be invariant under supersymmetry transformations, and the symmetry should be broken in the vacuum state. In this way supersymmetry is hidden at low energies, analogous to the result of the spontaneous symmetry breaking mechanism encountered in Section 1.1.4.

Unfortunately, no simple mechanisms for supersymmetry breaking exist. Therefore this breaking is treated rather phenomenologically, by adding terms to the Lagrangian that violate supersymmetry, but that do not give rise to quadratic divergences in the radiative corrections to scalar masses. It can be shown for the corrections in (1.22) that

$$\Delta m_{\mathrm{H}}^2 \sim m_{\mathrm{soft}}^2 \ln(\Lambda/m_{\mathrm{soft}}), \tag{1.29}$$

where Λ is the upper cut-off scale at which new physics enters, and m_{soft} denotes the largest mass scale associated to these so-called soft-breaking terms in the Lagrangian. The parameters introduced in these terms give rise to the required mass differences. To retain the solution to the hierarchy problem in the presence of the corrections (1.29), m_{soft} is required not to be larger than $\mathcal{O}(1\,\mathrm{TeV})$, which predicts a supersymmetric particle spectrum at the frontier of the present-day experimental reach.

The number of possible soft-breaking terms in the effective potential is limited, but in the minimal supersymmetric extension of the Standard Model, the so-called Minimal Supersymmetric Standard Model (MSSM), the number of additional masses, phases and mixings already adds up to more than hundred. Hence, supersymmetry breaking appears to introduce a large number of free parameters. Several of the new parameters, however, give rise to unacceptable levels of flavour-changing neutral currents, CP violation, lepton-number violation, etc. These potential problems can be avoided by imposing universality constraints on the theory, that are assumed as the result of the considered model of supersymmetry breaking.

The MSSM does not allow to break supersymmetry intrinsically, since none of the physical MSSM fields can develop a non-zero vacuum expectation value without spoiling gauge invariance. Extra fields need therefore to be introduced that form the so-called hidden sector. These fields cause the spontaneous symmetry breaking by interacting very weakly with the usual MSSM fields, which are referred to as the visible sector, and give rise to the soft-breaking terms in the low-energy effective Lagrangian. Several proposals exist for what the mediating interactions between the hidden and visible sector might be. One of these associates the hidden sector to new physics at the Planck scale mediated by the gravitational interaction [37, 38]. In its minimal form, called mSUGRA, the number of free extra parameters is reduced to five. Other types of mediation, for example gauge mediation [39], allow the scale of the hidden sector to be as low as 10^4 GeV.

Many new fields and associated particles are required in the MSSM when demanding supersymmetric invariance. All chiral Standard Model fermions receive scalar superpartners, referred to as squarks and sleptons[2], while the Standard Model gauge bosons get associated to fermionic so-called gauginos and the Higgs scalars are associated with fermionic Higgsinos. After electroweak-symmetry breaking many of these particles possess the same quantum numbers, and mixing of the mass eigenstates will occur, which significantly complicates the interpretation of particle masses. The gauginos and Higgsinos will lead to two physical so-called charginos $\chi^{\pm}_{1,2}$ and four neutralinos $\chi^{0}_{1,2,3,4}$. Also the two scalar partners of each massive lepton and each quark will mix.

A particularly interesting feature of the MSSM is the need for a second Higgs doublet. If only one Higgs doublet is present, the fermionic Higgsino partners generate a gauge anomaly by breaking gauge invariance in triangle loops with fermions. By introducing a second Higgs doublet with opposite hypercharge, these anomalies are cancelled by the respective superpartners. Another reason stems from a technical property of supersymmetric theories, that causes the need for two Higgs doublets to give mass to the up- and down-type fermions separately. The MSSM is thus a 2HDM Type-II, but with some additional powerful constraints on the Higgs sector, which leaves only two free parameters in the description of the Higgs sector. Typically these are chosen as $\tan\beta$ and $m_{\mathrm{H}^{\pm}}$ or m_{A}.

The charged Higgs-boson mass in the MSSM is related to m_{A}, at tree level, by the relation

$$m_{\mathrm{H}^{\pm}} = \sqrt{m^2_{\mathrm{A}} + m^2_{\mathrm{W}^{\pm}}}. \tag{1.30}$$

[2]The addition of the letter "s" to the particle names expresses the scalar nature of the superpartners.

Compared to a 2HDM Type-II, decays of charged Higgs bosons in the MSSM are only affected by supersymmetric decay channels opening up at sufficiently high charged Higgs-boson masses. It has been argued, however, that for a heavy supersymmetric spectrum of $\mathcal{O}(1\,\text{TeV})$ and $m_{\text{H}^\pm} < 500\,\text{GeV}$, the branching fraction for $\text{H}^\pm \to \text{tb}$ is only slightly affected by supersymmetric decay channels opening up [40].

The masses of the two other Higgs bosons h and H are given at the tree level by

$$m^2_{\text{H,h}} = \frac{1}{2}\left[m^2_{\text{A}} + m^2_{\text{Z}} \pm \sqrt{(m^2_{\text{A}} + m^2_{\text{Z}})^2 - 4m^2_{\text{Z}} m^2_{\text{A}} \cos^2 2\beta}\right]. \tag{1.31}$$

As a consequence, $m_{\text{H}^\pm} \geq m_{\text{W}}$, $m_{\text{H}} \geq m_{\text{Z}}$, $m_{\text{A}} \geq m_{\text{h}}$, and $m_{\text{h}} \leq m_{\text{Z}}$. The last inequality, which states that the lightest neutral Higgs-boson mass is bounded from above by the Z-boson mass, seems to rule out the MSSM as an extension to the Standard Model, since no Higgs boson has been observed at LEP, unless an additional suppression of the cross section is introduced in the model. When radiative corrections are added to this tree-level result, however, the bound is relaxed up to $m_{\text{h}} \lesssim 130\,\text{GeV}$, assuming sparticle masses not larger than 1 TeV and a top-quark mass of 175 GeV. Relaxing this requirement, and allowing for extra multiplets in the MSSM the bound can be further raised to 150 GeV. This is a very powerful prediction, since a non-observation of the Higgs boson at the Large Hadron Collider, described in Chapter 2, would rule out all elegant and powerful low-energy supersymmetry models like the MSSM.

Apart from a solution to the hierarchy problem, supersymmetry also addresses several other shortcomings of the Standard Model. Making supersymmetry a local symmetry appears to be the road towards the inclusion of gravitational interactions in relativistic quantum field theories. In addition, it can be shown in the MSSM that the three gauge couplings appear to unify at a scale just below the Planck Scale. Unification of the couplings at high energy scales is a non-trivial accomplishment, since the introduction of extra fermions, bosons or Higgs fields in the theory changes the b coefficients in (1.23) in a highly correlated manner. Conversely, from a fit imposing unification of the couplings, a preferred mass range for supersymmetric particles $\mathcal{O}(1\,\text{TeV})$ is obtained in agreement with other constraints. Supersymmetric theories also allow to introduce additional CP violation, for example in the Higgs sector, and violation of baryon-number conservation.

Supersymmetry finally also provides a candidate particle that can account for the dark matter in the universe. Indeed, the MSSM preserves the R-parity quantum number

$$R = (-1)^{3(B-L)+2S}, \tag{1.32}$$

where B, L and S are respectively the baryon and lepton number and the spin of a particle. For all Standard Model particles and Higgs bosons $R = +1$, while for their superpartners $R = -1$. Hence, if the $\mathbb{Z}_2$ symmetry associated to R-parity is exactly conserved, the lightest supersymmetric particle, called the LSP, is necessarily stable. In the case it is also neutral, the LSP furnishes an attractive candidate constituent of the non-baryonic dark matter.

1.2.4 Limits on charged Higgs bosons in the 2HDM and the MSSM

Some theoretical considerations lead to limits on the parameters relevant for the description of the charged Higgs bosons in the two-Higgs-doublet model and the MSSM. If $\tan\beta$ becomes too small, then the tree-level unitarity of processes involving the Higgs–top-quark Yukawa coupling is violated, rendering any perturbative analysis unreliable. Asking for a Higgs–top-quark coupling in the perturbative region leads to $\tan\beta \gtrsim 0.3$ [41]. A similar argument involving the Higgs–bottom-quark coupling yields $\tan\beta \lesssim 120$. More stringent bounds can be deduced by asking the Higgs–fermion Yukawa couplings to remain finite when running from the electroweak scale up to some large scale Λ, above which new physics is assumed to enter. Assuming no additional physics below the scale $\Lambda = 2 \times 10^{16}\,\mathrm{GeV}$, then the Higgs–fermion Yukawa couplings remain finite at two-loop level in the MSSM for all energies below Λ if $1.5 \lesssim \tan\beta \lesssim 65$ [42].

Many constraints can also be deduced from the current absence of experimental evidence for Higgs bosons. At LEP, pair produced charged Higgs bosons have been searched for assuming that only the decays $\mathrm{H}^\pm \to \tau\nu$ and $\mathrm{H}^\pm \to \mathrm{cs}$ are allowed. Constraints on the charged Higgs-boson mass are derived in the $\tau\nu\tau\nu$, the $\tau\nu$cs and the cscs final states, as a function of the branching fraction $\mathrm{BR}(\mathrm{H}^\pm \to \tau\nu)$. The preliminary combined result from all LEP experiments [43] excludes charged Higgs bosons with $m_{\mathrm{H}^\pm} < 79\,\mathrm{GeV}$ at 95% confidence level (C.L.) for all values of $\mathrm{BR}(\mathrm{H}^\pm \to \tau\nu)$. The maximal exclusion is reached at $\mathrm{BR}(\mathrm{H}^\pm \to \tau\nu) = 1$, where $m_{\mathrm{H}^\pm} > 89.6\,\mathrm{GeV}$ at 95% C.L. Searches in DELPHI for $\mathrm{H}^\pm \to \mathrm{W}^\pm\mathrm{A}$, a decay mode which can be dominant in a 2HDM Type-I at large $\tan\beta$, allowed to place an absolute limit $m_{\mathrm{H}^\pm} > 76.7\,\mathrm{GeV}$ at 95% C.L. for all $\tan\beta$ and $m_{\mathrm{A}} > 12\,\mathrm{GeV}$ [44]. Constraints from LEP data on the MSSM also follow from neutral Higgs boson searches, which leads to exclusion of a low $\tan\beta$ for all charged Higgs-boson masses. These exclusion regions, however, are very sensitive to the MSSM scenario, especially to stop mixing. Details on supersymmetric Higgs searches at LEP can be found in [45], wherein also the possible consequences of CP violation in the MSSM Higgs sector are assessed.

More recent searches are being conducted by the CDF and DØ collaborations in proton–anti-proton collisions at the Tevatron collider. These experiments look for deviations in the top-quark pair cross section due to $\mathrm{t} \to \mathrm{H}^\pm\mathrm{b}$ decays. For small charged Higgs-boson masses, the $\mathrm{H}^\pm \to \mathrm{cs}$ and $\mathrm{H}^\pm \to \tau\nu$ decay modes dominate respectively at small and large $\tan\beta$, while the $\mathrm{H}^\pm \to \mathrm{t}^*\mathrm{b} \to \mathrm{W}^\pm\mathrm{b}\bar{\mathrm{b}}$ becomes important at large $\tan\beta$ for $m_{\mathrm{H}^\pm}$ approaching the top-quark mass. The analyses are performed either in disappearance searches, looking for deficits in for example the lepton+jets cross section, or as an appearance search where the τ+jets cross section gets enhanced. The non-excluded region in the $(m_{\mathrm{H}^\pm}, \tan\beta)$ plane of the 2HDM Type-II from DØ searches [46, 47] spans $0.97 < \tan\beta < 40.9$ at 95% C.L. for $m_{\mathrm{H}^\pm} = 60\,\mathrm{GeV}$. This limit becomes less stringent with increasing $m_{\mathrm{H}^\pm}$. For $m_{\mathrm{H}^\pm} = 124\,\mathrm{GeV}$ the lower limit decreases to $0.3 < \tan\beta$, while for $m_{\mathrm{H}^\pm} = 153\,\mathrm{GeV}$ the upper limit is raised to $\tan\beta < 150$. The CDF experiment has obtained similar exclusion limits in the context of the MSSM, and concludes that $\mathrm{BR}(\mathrm{t} \to \mathrm{H}^\pm\mathrm{b}) < 0.4$ at 95% C.L. within the range $80 < m_{\mathrm{H}^\pm} < 160\,\mathrm{GeV}$ for $\mathrm{BR}(\mathrm{H}^\pm \to \tau\nu) = 1$, and $\mathrm{BR}(\mathrm{t} \to \mathrm{H}^\pm\mathrm{b}) < 0.9$ at 95% C.L.

within the same mass range, and independent of the charged Higgs-boson branching fractions [48].

Several indirect searches for charged Higgs bosons are also pursued, looking for deviations in branching fractions of rare B-meson decays, due to the mediation of a charged Higgs boson. In the Standard Model the branching fraction of the $B^{\pm} \to W^{\pm *} \to \tau^{\pm}\nu$ decay is expected to be $(1.41 \pm 0.33) \times 10^{-4}$ from a full unitarity-constrained fit to experimental data [49]. A preliminary combination [50, 51] of the measurements of the $B^{\pm} \to \tau^{\pm}\nu$ branching fraction in the Belle [52] and BaBar [53] experiments, results in $BR(B^{\pm} \to \tau^{\pm}\nu) = (1.34 \pm 0.48) \times 10^{-4}$. A second sensitive decay mode is the $b \to s\gamma$ decay, which is a flavour-changing neutral current process forbidden at tree level in the Standard Model, but which can occur at the one-loop level via so-called penguin diagrams. The presence of charged Higgs bosons or supersymmetric particles in the penguin loop can alter the Standard Model prediction for the branching fraction $BR(B \to X_s\gamma) = (3.61^{+0.37}_{-0.49}) \times 10^{-4}$ with $E_\gamma > 1.6\,\text{GeV}$ [54]. Measurements from several experiments [55–57] lead to the combined world average $BR(B \to X_s\gamma) = (3.55 \pm 0.24^{+0.09}_{-0.10} \pm 0.3) \times 10^{-4}$ with $E_\gamma > 1.6\,\text{GeV}$ [50, 51].

The presented measurements of B-meson branching fractions put strong limits on the allowed parameter space in general two-Higgs-doublet models. The $b \to s\gamma$ measurements exclude $m_{H^{\pm}} < 340\,\text{GeV}$ for all $\tan\beta$ at 90% C.L. Due to the $B^{\pm} \to \tau^{\pm}\nu$ measurement, this exclusion region extends gradually at large $\tan\beta$, reaching $m_{H^{\pm}} < 800\,\text{GeV}$ for $\tan\beta = 100$. In the MSSM the limits on $m_{H^{\pm}}$ and $\tan\beta$ from the constraints of B-meson decays are strongly model dependent. Depending on the MSSM parameters, the supersymmetric particles can induce large corrections [58] or directly contribute in the loop of the $b \to s\gamma$ decay. The MSSM parameter dependence of the limits is only starting to be explored [59].

To conclude, the mass of the charged Higgs boson in two-Higgs-doublet models is constrained by indirect searches as $m_{H^{\pm}} > 340\,\text{GeV}$ for all $\tan\beta$. In the MSSM this constraint does not hold, depending heavily on the model parameters. Indirect searches in top decays exclude small charged Higgs-boson masses at low and high $\tan\beta$. The best model-independent limit comes from direct searches at the LEP accelerator, resulting in $m_{H^{\pm}} > 79\,\text{GeV}$.

1.2.5 Alternative extensions of the Standard Model

In supersymmetric theories, the origin of electroweak symmetry breaking is tied to the breaking of supersymmetry, since the spontaneous breaking of $SU(2)_L \otimes U(1)_Y$ gauge invariance is generated by the soft supersymmetry-breaking terms in the Higgs potential. The supersymmetry breaking itself, however, remains a mystery. Several alternative ideas exist to mend some shortcomings of the Standard Model in a different way than with supersymmetry.

The introduction of extra space-time dimensions in the theory can help to solve the hierarchy problem by bringing down the Planck scale to the TeV level, and to integrate gravity in the model at the quantum level. In addition, string scenarios, that treat particles as extended objects, require both extra dimensions and supersymmetry. Depending on the physical mechanism invoked to hide the extra dimensions from current observation, there is a large range of possible energy or length scales at which

they may start to appear. The extra dimensions can either be compact and small, or the Standard Model particles can be confined to a so-called brane, and thus cannot probe the full bulk with a larger dimensionality. Examples of extra-dimension scenarios are the universal extra dimensions [60], Arkani-Hamed, Dimopoulos and Dvali (ADD) scenarios with large extra dimensions [61] and Randall and Sundrum (RS) warped extra dimensions [62, 63]. Even in the absence of a completely rigorous theoretical framework, the phenomenological consequences of compactified extra dimensions can be explored. A variety of typical signatures are predicted, like standing waves in the extra compactified dimensions, which are reflected in so-called Kaluza-Klein towers of massive excitations of bosons. Also, new particles like radions and gravitons can be produced at the TeV scale.

Many other exotic scenarios have been proposed as alternatives beyond the Standard Model. Grand Unified Theories (GUT) try to embed the Standard Model gauge groups into one global symmetry group, hence unifying the interactions and reducing the number of parameters in the model. In general new matter fields are needed in such scenarios. The technicolour approach postulates a new large gauge group, involving gauge interactions between new massless technifermions, which condensate into technipions and other technimesons due to a strong QCD-like technicolour interaction. The W and Z bosons acquire mass through interactions with the technipions, providing a dynamical nature to electroweak symmetry breaking. Other dynamical breaking scenarios have been put forward, involving condensates of top quarks. Recently, little Higgs models have been introduced, which contain an alternative method for electroweak symmetry breaking with new global symmetries which are both explicitly and spontaneously broken. Overviews of many of these models and scenarios can be found in [64–68].

Chapter 2

Physics at the Large Hadron Collider

All current experimental results in high-energy particle physics are found to be in excellent agreement with the Standard Model. As discussed in Chapter 1, one particle remains to be discovered to understand the principles of electroweak symmetry breaking. From the theoretical side, on the other hand, strong arguments support the interpretation of this theory as only an effective theory at the energies accessible today. In Chapter 1 several appealing extensions of the Standard Model have been presented. The experimental verification or falsification of the predictions of these theories requires the exploration of physics at the TeV energy scale.

In Section 2.1 an overview is given of the properties of recent, present and future colliders that enable the search for new physics at the energy frontier. In Section 2.2 the design and operation of the Large Hadron Collider (LHC) and its experiments are described. In Section 2.3 the general properties of proton collisions are discussed in the context of the LHC. Finally, an outline is given of the physics program envisaged with the LHC accelerator.

2.1 Colliders at the Energy Frontier

A consequence of the quantum-mechanical behaviour of matter and its interactions at the atomic and sub-atomic scale is the experimental need to increase the energy densities in collisions in order to resolve nature at decreasing distances. The de Broglie wavelength $\lambda = h/p$ of a probe particle, which limits the resolved length scale in an interaction, is indeed inversely proportional to the energy of the relativistic probe. Over the past decades many accelerator facilities with increasing centre-of-mass energies have allowed to probe matter and interactions at scales as small as 10^{-18} m.

Searches for rare phenomena and measurements with increasing precision require ever increasing interaction rates. The acceleration in colliders is performed in bunches with a high particle density. The interaction rate is proportional to the luminosity $\mathcal{L}$ of the accelerator, which is given for head-on collisions by

$$\mathcal{L} = \frac{f n_{\mathrm{p}}^2}{4\pi\sigma_x\sigma_y}, \tag{2.1}$$

where f is the bunch collision frequency, n_{p} represents the number of particles in the colliding bunches, and σ_x and σ_y characterize the transverse size of the bunches at the collision point.

In collisions of particles without internal structure, like electrons at the current energy densities, the particles' entire energy is available in the centre-of-mass of the collision to study the interactions. When colliding composite particles, like protons, on the other hand, the collision happens either elastically, as if the particles are point-like, or inelastically, involving a direct interaction between the constituents of the incident particles. The kind of interaction depends on the interaction energy, which sets the length scale probed in the collision. Present day hadron accelerators can reach far into the inelastic regime, where the constituent particles interact directly with each other, such that only a fraction of the available energy of the hadrons is available in the collision. This apparent drawback of a reduced effective centre-of-mass energy, can also be seen as an asset. The wide spectrum of effective collision energies makes hadron colliders ideal discovery machines, while electron–positron colliders serve better for precision measurements at fixed energy scales.

From 1989 up to 2000 the Large Electron–Positron accelerator (LEP) has been functional in a 26.6 km circular tunnel at CERN, near Geneva. It collided electrons onto positrons at centre-of-mass energies ranging from 90 GeV up to 209 GeV, and produced millions of Z and thousands of W bosons. In the ultra-relativistic energy regime in which LEP operated, where

$$\beta = \frac{v}{c} \approx 1 \qquad \text{and} \qquad \gamma \approx \frac{E}{mc^2}, \tag{2.2}$$

the electrons and positrons lose a large amount ΔE of energy due to synchrotron radiation, given by

$$\Delta E = \frac{4\pi\alpha}{3R}\beta^3\gamma^4 \sim \frac{E^4}{Rm^4}, \tag{2.3}$$

for each revolution, where R is the accelerator radius and α is the electromagnetic fine-structure constant. This quartic energy dependence limits the achievable energy at circular electron–positron colliders with a given radius.

An obvious way to overcome the problem of synchrotron radiation at higher energies, is to accelerate and collide protons instead of electrons, which reduces the γ^4 contribution in equation (2.3) by a factor $\sim 2000^4$, due to the increased particle mass. Between 1992 and 1996, at Fermilab, near Chicago, the TEVATRON accelerator has collided protons and anti-protons at 1.8 TeV centre-of-mass energy, and has made the discovery possible in 1995 of the long sought and very massive top quark. The total integrated luminosity

$$L = \int \mathrm{d}t\, \mathcal{L} \tag{2.4}$$

accumulated is about $140\,\mathrm{pb}^{-1}$. From 2001 onwards the TEVATRON has resumed its operation, after an extensive upgrade of both accelerator and detectors, raising the energy to 1.96 TeV and aiming for a rise in luminosity by a factor 20. Up to October 2006, about $1.9\,\mathrm{fb}^{-1}$ of integrated luminosity has been delivered by the accelerator, with instantaneous peak luminosity up to $2.3 \times 10^{32}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ [69].

In parallel to the running of LEP, a next generation hadron collider was planned to be built in Europe. This proton–proton accelerator, the Large Hadron Collider (LHC) will become operational at the end of the summer of 2007. It is not only supposed to raise the energy frontier another order of magnitude compared to the TEVATRON, but also to increase the interaction rate by more than an order of magnitude. Because of the use of the tunnel of the former LEP accelerator, the energy of this machine is limited by the maximal magnetic dipole field that can be achieved to bend the beams in the accelerator. With an energy of $E_{\text{beam}} = B_{\text{dipole}} \times 0.84\,\text{TeV/T}$ for proton beams in the LEP tunnel, the envisaged $E_{\text{beam}} = 7\,\text{TeV}$ requires dipole magnets with a vertical field up to 8.33 T. Such strong magnets have been specially developed for the LHC [70].

At present, advanced research is being conducted on the feasibility of a new type of high-energy and high-luminosity electron–positron accelerator. This international effort aims to build a 0.5 to 1 TeV centre-of-mass linear collider, with superconducting accelerator technology able to reach accelerating fields up to 35 MeV/m. Such a machine would serve very well to complement the discovery potential of the LHC with a tool to perform precision measurements of potentially-discovered new physics [71].

2.2 The Large Hadron Collider

2.2.1 The LHC design and operation

The design of the LHC [70] has been driven by the goal to explore physics at the TeV scale. New phenomena are predicted at this scale, which follow from theoretical attempts to address the shortcomings of the Standard Model, as discussed in Section 1.2. The choice to build a 14 TeV centre-of-mass energy collider within the former LEP tunnel sets the magnetic field requirement of 8.33 T for the 1232 superconducting dipole magnets that will keep the beams on a circular trajectory at nominal energy. In addition about 7000 magnets will be used to clean and focus the beams and correct their trajectories. The acceleration of the beams from their injection energy of 450 GeV up to the nominal energy of 7 TeV will be performed by 8 so-called radio-frequency (RF) cavities, boosting the beams in total by 16 MeV per turn in an electric field of 5.5 MV/m, oscillating at 400 MHz.

Apart form the centre-of-mass energy, also the interaction rate needs to be sufficiently high to carry out the search for new or very rare phenomena. The LHC is designed to ultimately reach a luminosity of $\mathcal{L} = 10^{34}\,\text{cm}^{-2}\text{s}^{-1}$, referred to as high luminosity, after a first few years of operation at the so-called low luminosity of $\mathcal{L} = 2 \times 10^{33}\,\text{cm}^{-2}\text{s}^{-1}$. To reach such high luminosities, bunches with about 10^{11} particles are collided each 25 ns. In addition, the beams are collimated to a transverse size of about 16 μm at the interaction points of the main experiments, to further enhance the collision probability. A small reduction of the luminosity has to be accounted for due to a crossing angle of 285 μrad of the beams, which is needed for the beam separation at nominal energy.

The difficulties in producing sufficient anti-protons to reach the desired luminosity of the LHC forced the choice of a proton–proton collider. Due to the high centre-of-mass energy, interactions involving a few hundred GeV momentum transfer are mostly

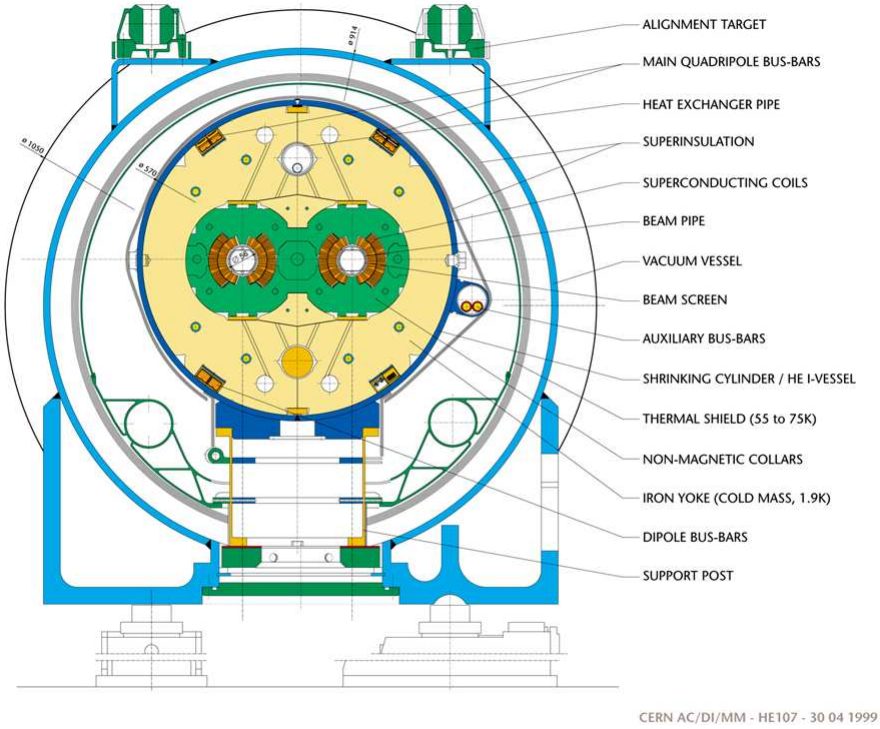


Figure 2.1: Schematic view of the LHC dipole magnet design.

gluon-initiated, such that the choice for proton–proton collisions does not restrict the physics potential of the accelerator. It does however heavily influence the design of the machine, which cannot exploit the opposite charge of the beam particles by using a single beampipe. Instead, the two beams of protons will be accelerated in separate beampipes, that intersect at the experimental interaction points. This results at the same time in a large reduction of beam–beam interactions and collisions outside the experimental regions, which is needed to achieve the high design luminosity of the LHC. The beampipes are embedded in a common cryostat that surrounds the bipolar superconducting dipole magnets, which are the main components of the LHC accelerator. A schematic view of an LHC dipole magnet is shown in Figure 2.1. Superfluid helium at 1.9 K is used to cool the Nb–Ti alloy coil down to superconductive state. The LHC conditions impose many challenges, which needed dedicated research at the technological edge and beyond. Unprecedented constraints need to be met for many components of the magnets and the power and cryogenic services. By the end of October 2006, almost all dipole magnets have been tested at their nominal current of about 12 kA, and over 70% have been installed in the LHC tunnel [72].

Before the protons enter the LHC, they have already passed through several stages of acceleration and beam operations. The protons are delivered to the LHC by the Super Proton Synchrotron (SPS) at an injection energy of 450 GeV. Before entering the SPS, they are first accelerated by a linear accelerator, the Booster and the Proton Synchrotron (PS) up to 26 GeV. The CERN pre-accelerator complex, schematically drawn in Figure 2.2, has already been used for decades for other experimental purposes, and has in the past years extensively been adapted and upgraded for operation at the

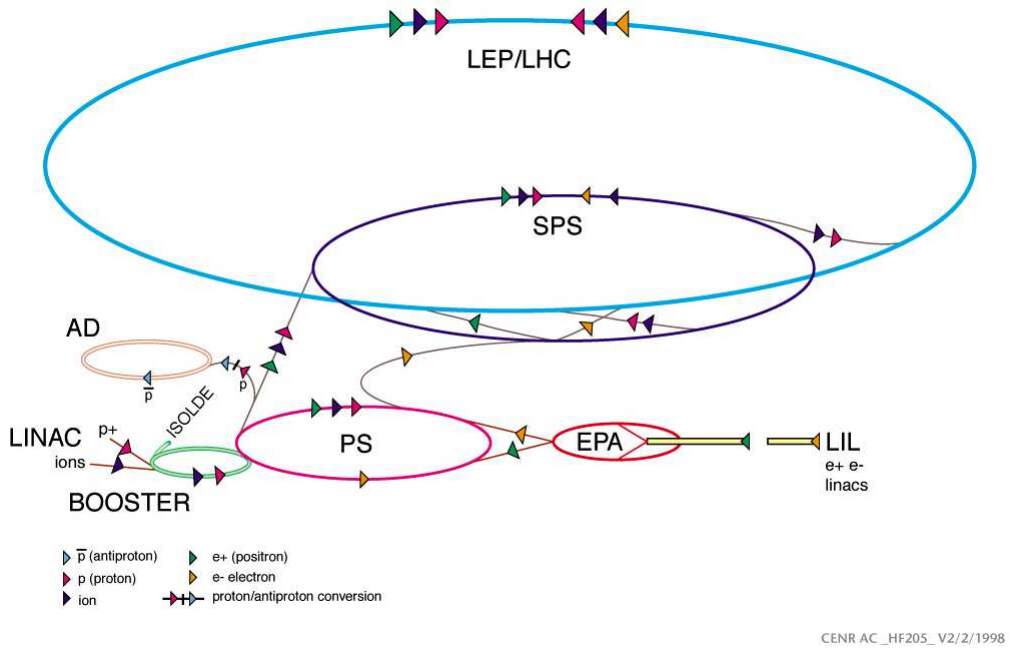


Figure 2.2: Schematic view of the CERN accelerator complex.

LHC. The PS, in particular, will ensure that proton bunches are delivered to the SPS with the correct bunch spacing of 25 ns. At the different stages of injection, the rise time for the injection "kickers" will leave gaps of missing bunches in the beam. In addition, a 3 μs gap in the LHC bunch pattern needs to be foreseen, such that the beam-dump magnets can rise and dump the 350 MJ beams unspoiled in one revolution. As a consequence, 2808 bunches are foreseen to fill the 3564 available 25 ns so-called buckets in the LHC beam structure. The time needed to fill the LHC beams is expected to be about seven minutes, while the beam lifetime will last about ten hours, until the luminosity has degraded too much due to collisions and deformation of the beams.

The LHC will start its operation at the end of the summer of 2007 [73]. Initially the machine will be tested with collisions at the injection energy of 450 GeV, to ensure all power supply tests of the magnets can be completed before the beam energies and hence the dipole magnet fields are ramped up. After that, during a pilot run of about a month, the number of bunches in the beam will gradually be increased, up to 156, and the number of protons per bunch will be raised up to its nominal value for the low-luminosity setting. An integrated luminosity of $10\,\mathrm{pb}^{-1}$ is expected to be collected with a luminosity up to $2 \times 10^{31}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. In 2008 a first physics run will start with 75 ns bunch spacing, useful for synchronization tests, subsequently moving to 25 ns. Depending on the needed machine-development periods and encountered inefficiencies, about $1 - 3\,\mathrm{fb}^{-1}$ of data is foreseen in the first year, with luminosities peaking to low-luminosity conditions. Up to 2010 the LHC should then be able to run at $\mathcal{L} = 2 \times 10^{33}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, acquiring $10 - 30\,\mathrm{fb}^{-1}$ of data. After a foreseen upgrade of the beam dump and collimation systems, the LHC is finally expected to run at its design high luminosity $\mathcal{L} = 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, collecting $100 - 300\,\mathrm{fb}^{-1}$ of integrated luminosity.

In addition to the normal proton–proton operation, the LHC will also perform shorter runs with heavy ions. In these runs completely ionized lead nuclei (Pb^{82+}) will

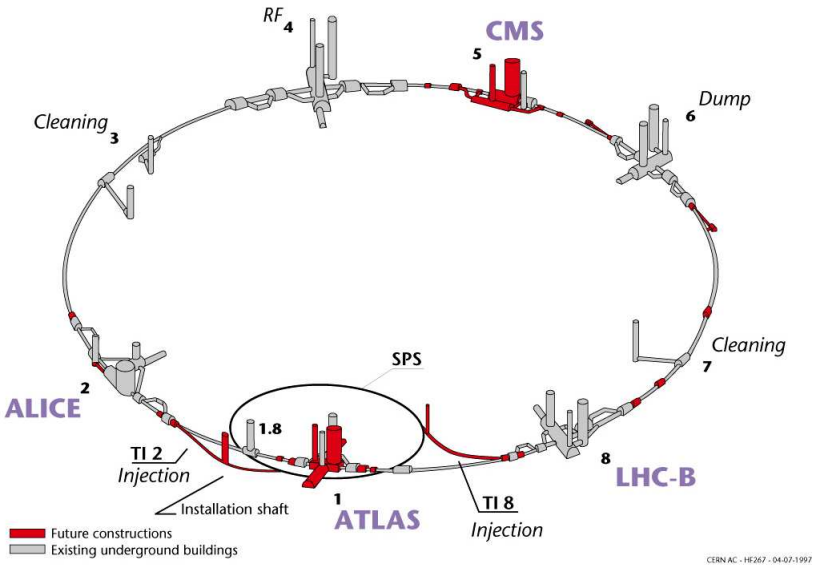


Figure 2.3: Schematic view of the layout of the LHC tunnel, with the location of the main experiments.

be accelerated up to an energy of 2.76 TeV per nucleon with the nominal dipole field strength. At the design luminosity $\mathcal{L} = 10^{27}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, a bunch separation of 100 ns and 7.0×10^7 nuclei per bunch are foreseen. Heavy-ion runs are planned before each yearly machine shutdown, but will initially strongly depend on the machine's performance. A first run with reduced luminosity might occur after the 2008 proton–proton physics run.

2.2.2 Experiments at the LHC

The planned LHC physics program, further discussed in Section 2.3.3, is very rich and diverse. Several detectors will carry out experiments with the colliding protons and/or ions. The location of these detectors in the LHC tunnel is shown in Figure 2.3. Two large general purpose detectors, CMS and ATLAS [74], are being installed at Point 5 and Point 1, respectively. In Point 2 the ALICE [75] experiment will study heavy ion collisions, and in Point 8 hosts the LHCb [76] experiment is dedicated to b physics.

For the CMS and ATLAS detectors, the choice of the magnetic field configuration is crucial for the momentum measurement of charged particles. For a non-interacting charged particle in a homogeneous magnetic field, the relative uncertainty on a transverse-momentum measurement can be written in terms of the uncertainty on the actually measured sagitta f of the particle's bent track,

$$\frac{\sigma_{p_\mathrm{T}}}{p_\mathrm{T}} = \frac{\sigma_f}{f} \propto \sigma_f \frac{p_\mathrm{T}}{BL^2}, \tag{2.5}$$

where B is the strength of the magnetic field and L is the distance between the innermost and outermost measurements of the track. With the precision on the momentum of charged particles inversely proportional to BL^2, the design of a detector has a freedom to balance the magnetic-field strength with respect to the size of the detector.

The CMS experiment opted for a strong solenoidal magnetic field of 4 T and a compact design [77]. The ATLAS collaboration chose a large detector configuration with a moderate central magnetic field of 2 T, complemented with large toroidal magnets in the outer detector [78]. Because of the general purpose of CMS and ATLAS, both detectors consist roughly of four detector layers, cylindrically structured around the beampipe in a so-called barrel, with complementary layers in two endcaps covering the forward direction. From the inside to the outside, a central tracking device, an electromagnetic- and a hadronic-calorimeter system, and a muon spectrometer are placed. In CMS the calorimetry systems fit within the large magnet, while ATLAS hosts only the tracking detector inside the smaller solenoid.

The two experiments have chosen different and complementary detection techniques. The main track detectors of both CMS and ATLAS use silicon pixels and strips. ATLAS, in addition, uses straw tubes in the outer tracker with transition radiation capability. Both experiments are equipped with an electromagnetic calorimeter with excellent energy and position resolution. For CMS lead-tungstate scintillating crystals have been chosen. ATLAS opted for a liquid-argon sampling calorimeter with lead as absorber in an accordion geometry. The liquid-argon technology is also used for the ATLAS hadron calorimeter in the endcaps, while in the barrel an iron absorber is interleaved with plastic scintillator tiles. The CMS hadron calorimeter also uses plastic scintillators with brass as absorbing material. The muon systems of both experiments use comparable detection technologies, with drift tubes in the barrel and cathode-strip chambers in the endcaps for high-precision tracking. Resistive-plate chambers are used for muon triggering, except in the forward region of ATLAS, where thin-gap chambers fulfil this role. The large difference between both muon spectrometers arises from the design of the magnetic field, which makes the ATLAS detector large and open, 22 m in diameter and 46 m long, and the CMS detector compact and dense with 15 m diameter and a length of 21.5 m. The design of the CMS detector and its components is further described in detail in Chapter 4.

The LHC b-physics program will partly be covered by CMS and ATLAS, but for the main part the dedicated LHCb experiment has been conceived [79]. The LHCb detector is a single-arm spectrometer with a geometry matched to the forward emission of b hadrons. It consists of a vertex-locator system within the beam pipe, a large dipole magnet and a detailed tracking system. Ring-Imaging Čerenkov Hodoscopes (RICH) provide charged-particle identification. Further, also calorimeters and a muon spectrometer are present. To reduce the probability of multiple simultaneous proton interactions, the LHC will deliver a smaller luminosity to the LHCb experiment, by means of a weaker focalization of the beams.

Another dedicated experiment, ALICE, aims to study properties of the quark-gluon plasma with collisions of heavy ions [80]. This experiment needs to be able to reconstruct thousands of tracks from a single collision. This will be performed with a large Time-Projection Chamber (TPC) outside a smaller inner silicon tracker. In addition, several techniques for particle identification are employed. Also CMS and ATLAS will take part in the runs with heavy ions.

Two more experiments will be conducted at the LHC. The Totem experiment [81] is independent of, but closely intertwined with the CMS detector. It consists of tracking

detectors integrated in CMS and Roman pots far from the interaction point. Totem aims for a 1% accuracy measurement of the total proton–proton cross section. Also elastic and diffractive processes will be studied in detail, with a pseudorapidity coverage up to 13. The recently conceived LHCf experiment [82], finally, will measure the π^0 production cross section in proton collisions at 14 TeV at pseudorapidities up to infinity. This will serve as valuable input to model cosmic-ray air showers in the ultra-high energy range $\sim 10^{17}$ eV.

2.3 Physics at the Large Hadron Collider

2.3.1 Coordinate conventions in proton collisions

Proton collisions at high energies manifest themselves as inelastic collisions of composite objects, where the actual collisions happen between the constituent quarks and gluons, also called partons. These partons carry a fraction x_1 and x_2 of the total proton momentum. The fractions are statistically distributed according to the so-called parton density functions, which are rigorously introduced in Section 3.1.2. Hence, in general, x_1 and x_2 are not equal and the rest frame of the hard collision will be boosted along the beam line with respect to the lab frame. The reconstruction of the boost of the system on a collision-by-collision basis requires the full reconstruction of the remnants of the colliding protons, which is in practise not possible because of the presence of the beam pipe and instrumentation along the beamline.

Because of the unknown energy balance along the beamline, proton collisions are usually studied in a more convenient coordinate frame. The z axis is placed along the beamline, and the x and y axes hence define the transverse plane, perpendicular to the beam. The spherical coordinates (r, θ, ϕ) are replaced with (r, η, ϕ), where r is the radial distance, ϕ is the azimuthal angle in the transverse (x, y) plane, θ is the polar angle measured from the z axis, and the pseudorapidity η is defined as

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right). \tag{2.6}$$

The advantage of this coordinate frame is the Lorentz invariance of transverse quantities and differences in η under Lorentz boosts along the beamline. As a consequence, a solid "angle" in (η, ϕ) space is also invariant under longitudinal boosts.

2.3.2 Proton collisions at the LHC

In the complex environment of proton interactions, particular classes of collisions can be distinguished, depending on experimental signatures of either soft, elastic or hard, inelastic partonic interactions.

Elastic scattering Elastic collisions preserve the total kinetic energy of the interacting protons. At very low four-momentum transfer electromagnetic Coulomb scattering is most important; at higher momentum transfer hadronic exchange of a colourless pomeron dominates. Only the two slightly-deflected protons at very large pseudorapidity constitute the experimental signature of these collisions.

Single and double diffractive scattering In single- or double-diffractive collisions, respectively one or both of the protons fragment due to the exchange of a pomeron. Because the exchanged pomeron is colour neutral these collisions are characterized by hadronic activity at very large pseudorapidity on one or both sides of the detector, with no activity in the central part of the detector.

Double pomeron or photon exchange In this type of collisions both protons radiate a pomeron or photon that then interacts. The protons leave the scattering unshattered at very high pseudorapidity. These collisions can leave clean and spectacular signatures in the central detector, because of the absence of the coloured proton remnants. Possible creation of new resonances is of particular interest, as hadronic backgrounds can be very efficiently suppressed [83].

Multi-pomeron exchange More complex collisions, involving multiple pomeron exchanges, also give rise to more complex experimental signatures. For these collisions final states are expected involving several regions in pseudorapidity with hadronic activity, separated by inactive rapidity gaps.

Non-diffractive inelastic scattering This class of collisions groups all non-diffractive hadronic interactions. Most of these interactions happen at small four-momentum transfer and are soft compared to the physics processes of interest at the LHC.

The total cross section of proton–proton scattering σ_{T}, extrapolated from previous experiments at lower energies, is expected to be about 110 mb, of which 30 mb is due to elastic scattering, 24 mb comes from diffractive processes and 55 mb involves non-diffractive inelastic interactions [84]. Due to discrepancies between some measurements, the statistical uncertainty on σ_{T} is of the order of 5 mb. The uncertainty from using different models for the extrapolation to LHC energy, on the other hand, is dominant, and reaches about 20 mb. This large uncertainty is a particular nuisance at the LHC, for reasons of multiple simultaneous proton interactions described below. Therefore the total proton–proton cross section needs to be measured at the LHC itself, which is planned to be done by TOTEM in short periods during the first years of LHC operation, running with small luminosity and a reduced number of proton bunches in the beam.

The non-diffractive inelastic collisions are also referred to as minimum-bias interactions[1]. Different definitions of minimum-bias interactions are sometimes used in different contexts. In the summer of 2006, a large computing effort simulated 25 million minimum-bias collisions in the CMS detector. For this, the previous minimum-bias definition has been extended to include the single- and double-diffractive interactions, raising the cross section from 55 mb to 80 mb. For the studies in this thesis, however, this choice makes no physical difference, as diffractive collisions only cause hadronic activity in the very forward region. Therefore, all presented results use the previous definition for minimum-bias interactions, comprising only non-diffractive inelastic scattering processes. For the same reason also elastic proton collisions are ignored.

[1] Originally, the definition of this class of collisions was based on their property to introduce a minimal bias on the performance of the online selection system.

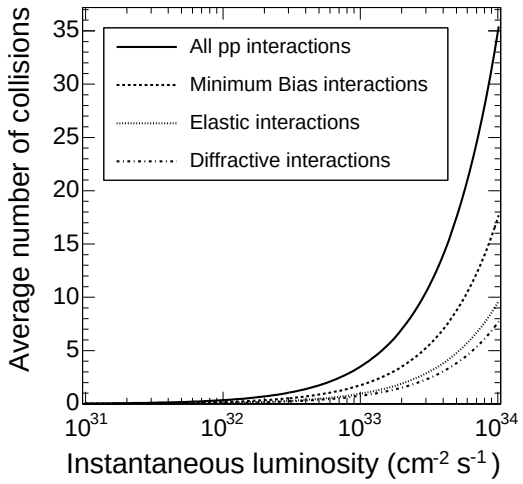


Figure 2.4: Average number of simultaneous collisions in one bunch crossing as a function of the instantaneous luminosity.

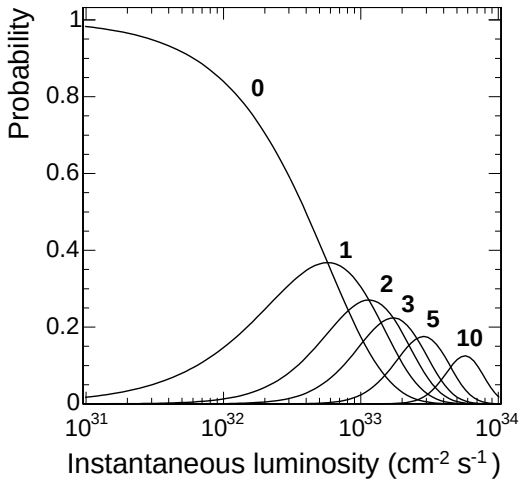


Figure 2.5: Probability for a given number of simultaneous minimum-bias proton collisions in one bunch crossing as a function of the instantaneous luminosity.

Due to the high luminosities that will be reached at the LHC, several proton collisions can occur during the same bunch crossing. In general, only one of these collisions will contain a hard interaction of physics interest, while the other superimposed collisions will be soft minimum-bias collisions. These overlapping minimum bias interactions are called pile-up collisions. The complete set of collisions during one bunch crossing is referred to as the event.

The average number of pile-up collisions per event is directly proportional to the instantaneous luminosity of the accelerator. In Figure 2.4 the average number of proton collisions is shown as a function of the instantaneous luminosity for several types of proton collisions. For low luminosity conditions on average 3.5 minimum-bias collisions are expected per event, increasing to 17.3 at high luminosity. On an event-by-event basis the number of proton interactions contained in the event follows a Poissonian distribution. The corresponding probability for a certain number of minimum-bias proton interactions to happen simultaneously is shown in Figure 2.5 as a function of the instantaneous luminosity.

2.3.3 The LHC physics program

The LHC, being a proton–proton collider, is in general a machine conceived to explore new physics, rather than to perform precision measurements. The large increase in centre-of-mass energy and luminosity compared to previous accelerators opens a window to rare processes and objects of large mass and energy. In Figure 2.6 an overview is given of the cross sections of some major processes as a function of the mass of the produced particle of interest. Also shown are the corresponding rates and number of events per year for the high-luminosity regime of the LHC. The scale in the figure, spanning many orders of magnitude, immediately reveals the formidable experimental

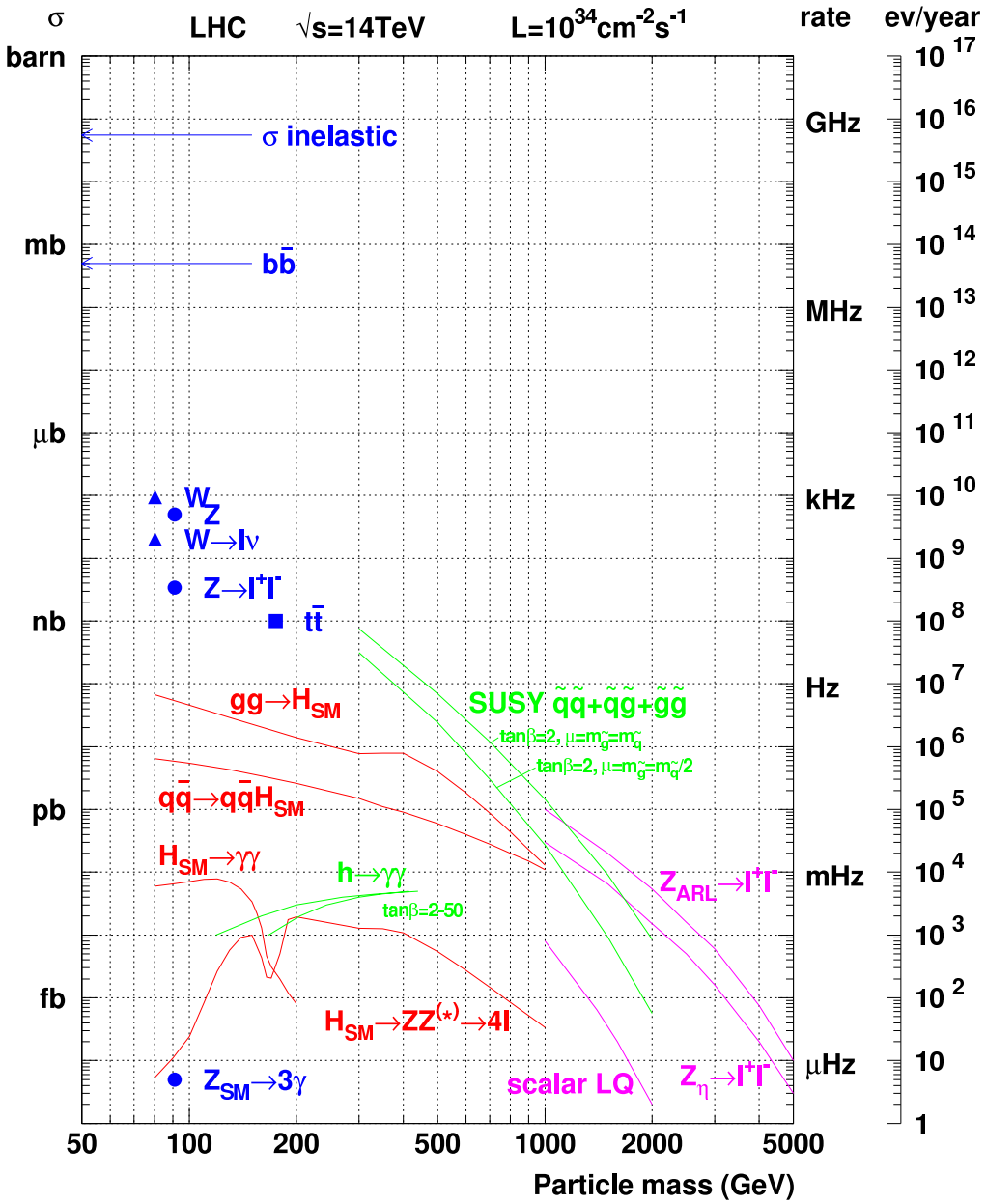


Figure 2.6: Overview of the cross sections of some major processes at the LHC as a function of the mass of the produced particle of interest.

challenge of extracting the signals of new physics from the large total bunch-crossing rate of 40 MHz. Typically the expected cross sections of new-physics processes are of the order of 10 pb or below, which is ten orders of magnitude or more below the total event rate.

The bulk of the physics program that will be conducted at the general purpose CMS and ATLAS detectors can be subdivided in three main categories. A short non-exhaustive overview of these categories is given below. More detailed information on these topics can be found in [85–88].

Higgs-boson searches The identification performance for the Standard Model Higgs boson has been used as a benchmark for the design of both the CMS and ATLAS detectors. This is particularly useful, since a wide range of decay modes and signatures needs to be accounted for, depending on the mass of the Higgs boson. In the low-mass region, just above the limit from LEP, the Higgs boson decays mainly hadronically, which is difficult to distinguish from the large multi-jet backgrounds and is further limited by the poor mass resolution obtained from jets compared to leptons. For masses up to $130\,\mathrm{GeV}/c^2$, however, the two-photon decay channel provides an interesting signature, but is suppressed because the Higgs boson needs to decay through a top-quark loop. For this channel, electromagnetic calorimetry with excellent granularity and energy resolution is needed, such that the small and narrow Higgs-boson mass peak can be resolved on the exponentially decreasing background spectrum. Another signature of interest in this mass range is the production of the Higgs boson together with two forward jets, and the subsequent decay of the Higgs boson into a pair of τ leptons.

In the mass range $130 \lesssim m_\mathrm{H} \lesssim 160\,\mathrm{GeV}/c^2$ the decay into $\mathrm{ZZ}^* \to 4\ell$, with one of the Z bosons off the mass shell, becomes the easiest channel for Higgs-boson detection. The leptons in the final state are electron and muon pairs, which emphasizes the need for good lepton identification. Also for large Higgs-boson masses, $m_\mathrm{H} \gtrsim 180\,\mathrm{GeV}/c^2$, this decay is the major discovery channel. Finally, in the intermediate range, $160 \lesssim m_\mathrm{H} \lesssim 180\,\mathrm{GeV}/c^2$, the decay to a WW pair dominates, and the $2\ell 2\nu$ final state provides the best visibility, again with the leptons being electrons or muons and with the presence of missing energy.

For supersymmetric Higgs bosons, the signatures can deviate significantly from the Standard Model expectations, depending on the parameters of the investigated model. As discussed in Section 1.2.3, the couplings to down-type fermions are enhanced at large $\tan\beta$ in the MSSM, which heavily influences the phenomenology of the Higgs sector. The lightest Higgs boson h decays predominantly in τ and b-quark pairs, except close to its upper mass limit, where it behaves like the Standard Model Higgs boson. Also the H and A bosons decay mainly to b quarks and τ leptons, except for low $\tan\beta$. In addition, they are often produced in association with b quarks. Hence, good tracking capabilities are required to efficiently identify hadronic τ-lepton and b-quark decays. The phenomenology of the charged Higgs boson is covered partly in Section 1.2.2, and is further discussed in Section 3.4.3.

Searches beyond the Standard Model Apart from an extended Higgs sector in the MSSM, many other supersymmetric signals can be searched for at the LHC. Since squarks and gluinos are coloured particles, they are produced via strong interactions with relatively high cross sections. Assuming conservation of R-parity, defined in (1.32), all decay chains of supersymmetric particles must end at the stable LSP, giving rise to signals of significant missing energy. Depending on the sparticle mass hierarchy that follows from the model parameters, also an abundance of leptons and quarks is expected in the final state, especially τ leptons and b quarks. Another possible signature is an excess of same-charged

lepton pairs. Hence, just like Higgs-boson searches, signals from supersymmetry need a good lepton identification, hermetic energy measurements for missing energy determination, and an efficient τ and b identification.

As discussed in Section 1.2.5, several different scenarios with extra dimensions exist, each with a characteristic phenomenology that can give rise to spectacular signatures. Among the predicted final states are gravitons escaping into the extra dimensions, leading to hard single quarks or photons and large missing energy. Other models produce Drell-Yan-like graviton resonances or heavy Kaluza-Klein Z' bosons, with decays in very energetic electron, muon, photon or quark pairs, and with possible interference effects with the Drell-Yan spectrum. Also many GUT-inspired models predict extra heavy gauge bosons. To study the resulting energetic lepton signatures, sufficiently good track momentum resolution is needed to allow efficient charge determination.

A long list of other exotic scenarios can and will be studied at the LHC. Among these scenarios, technicolour models can be tested by searching for the lightest technimesons. Little Higgs models contain typically a heavy top quark with characteristic decay modes and signatures that can be searched for. If the energy scale of quantum gravity is low enough, certain models even allow for the production of mini black holes, decaying democratically into fundamental particles and leaving a final state with high multiplicity and sphericity.

Standard Model physics Although the LHC machine and the CMS and ATLAS general-purpose experiments have been designed to maximize the discovery potential for Higgs bosons and supersymmetry, a large part of the physics program will consist of Standard Model physics. The main goal at start-up will be the rediscovery of the W and Z bosons and the top quark. Because of their copious production, these particles can then be used as standard candles for calibration purposes. Two examples of such calibration techniques are discussed in this thesis, in Section 5.1.3 and in Chapter 6. Once the detector performance is understood to an acceptable level, searches for physics beyond the Standard Model will look for deviations from the Standard Model expectations. In order to claim possible signals, the Standard Model background will need to be well under control. For this to happen, a large amount of physics commissioning work is required, with for example the measurement and understanding of the pile-up collisions, the tuning of non-perturbative QCD effects, etc. Related to this, various QCD studies will need to be performed. Measurement of the differential jet and $b\bar{b}$ cross sections, for example, will test the strong interaction at yet unexplored energies. Such measurements will also allow to gain further knowledge on the structure of the proton, which is currently extrapolated from the detailed measurements at mainly the HERA collider to the extreme conditions of LHC collisions.

Although much of this physics commissioning work is already to be seen as part of the Standard Model physics program, also a list of precision measurements can be performed that further test the consistency of the Standard Model. Electroweak measurements can be taken into a yet unexplored energy domain. Measurements

of the W and Z total and differential cross sections, and especially the study of multi-boson production and triple-gauge couplings allow to further test the electroweak gauge theory. The top-quark sector of the Standard Model provides a particularly rich environment for a multitude of analyses. Both the top-quark and W-boson mass can be measured with a high precision, which is shown in Section 1.1.5 to be important in constraining the Higgs-boson mass. Also the electroweak production of single top quarks is of interest, since it has not been observed yet. Finally, the top-quark decay consists of several rare decay modes involving flavour-changing neutral currents that are not accessible with current experiments. The high production rate of top quarks, the fact that many precision measurements are possible in the top-quark sector and that top-quark pairs form a main background to many direct searches beyond the Standard Model, makes this part of Standard Model physics from an experimental point of view a particularly interesting window to new physics.

Standard Model physics, Higgs-boson searches and searches beyond the Standard Model do not represent all physics goals of the LHC. Both CMS and ATLAS, along with the experiments described in Section 2.2.2, will study also other aspects of particle physics.

Diffractive physics The Totem, CMS and ATLAS experiments will be able to study diffractive proton collisions. Such collisions, introduced in Section 2.3.2, probe the proton at very small momentum transfer and allow QCD studies and measurements of the proton structure in this soft regime.

b**-Physics** The dedicated LHCb experiment will explore new frontiers in the domain of b-physics, constraining further the quark-mixing CKM matrix by measuring particle–anti-particle lifetime asymmetries and by studying rare decays of b hadrons. The branching fractions of such decays are very sensitive to new physics. Also CMS and ATLAS have a small complementary b-physics program.

Heavy-ion physics The dedicated ALICE experiment, and to a lesser extent also CMS and ATLAS, will observe heavy-ion collisions at unprecedented energies. Such collisions can produce the very energetic and dense state of matter that has been present a split second after the big bang, which is called the quark–gluon plasma. The experiments aim to study the behaviour of this strongly-interacting phase of matter and the transition towards normal free particles.

An overview of the physics program for these domains is beyond the scope of this thesis. More details can be found in [79, 80, 83, 86].

Chapter 3

Strong Interactions in Proton Collisions

Hadronic interactions have been the subject of dedicated theoretical and experimental research since decades [89–91]. The understanding of the physics involved needs elements from electroweak and QCD field theory, as well as from phenomenological models for processes where perturbative calculations are not applicable. In this chapter, an overview is given of the machinery needed to describe and predict the physics of proton collisions, up to the final hadronic objects that are experimentally observable.

Section 3.1 introduces and factorizes the different components that can be distinguished in inelastic proton–proton collisions. In Section 3.2 $2 \to 2$ and $2 \to 3$ parton scattering is discussed on a phenomenological basis. Section 3.3 deals with the formation of the final state through the parton shower and jet formation, with emphasis on heavy flavour. In Section 3.4, finally, several phenomenological aspects are highlighted for a few particular physics channels relevant for this thesis: top-quark pair production and the creation of charged Higgs bosons with associated heavy flavour.

3.1 Inelastic Proton Collisions

In Figure 3.1 a schematic representation is shown of the major processes that can be distinguished in proton–proton collisions. Due to the high energies of the collisions the actual hard scattering (d) takes place between the partons that constitute the protons. These partons are described by momentum distributions (a) that depend on the energy scale at which the proton is probed. The proton remnants, not directly taking part in the hard interaction, give rise to the so-called underlying event (b). Before and after the hard interaction the partons can undergo initial- (c) and final-state (e) showering. If in the final state short-lived resonances are produced, they will also decay in (e). All produced coloured particles will finally fragment and form jets of hadrons (f).

Simulations of proton collisions, often referred to as events, follow roughly the same break up into sub-processes as depicted in Figure 3.1. In this thesis the `PYTHIA` [92] simulation program is used extensively. This general-purpose generator uses Monte-Carlo techniques to simulate the distributions and parametrizations of the models involved in the description of proton collisions. Some parts of the simulation, however, need dedi-

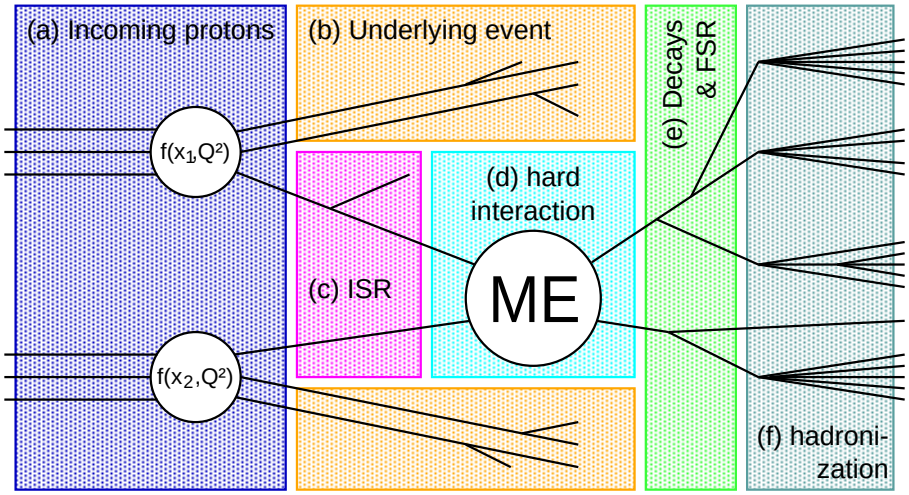


Figure 3.1: Schematic representation of the major processes involved in hadronic interactions.

cated simulation programs or calculations. In particular for the hard parton scattering many generators exist that perform exact calculations for complex processes that are not implemented in `PYTHIA`. Examples used in this thesis are `MadGraph/MadEvent` [93], `CompHEP` [94] and `ALPGEN` [95]. It is possible to interface these dedicated generators with `PYTHIA`, either directly or via intermediate storage of events. Using this feature, the same models for underlying event, parton showering and jet fragmentation have been used throughout this thesis, greatly simplifying the comparison or combination of results obtained with different generators.

3.1.1 From proton to parton scattering

At the LHC protons will collide head on with beam energy of $E_{\text{beam}} = 7\,\text{TeV}$, giving rise to a centre-of-mass energy of $\sqrt{s} = 2E_{\text{beam}} = 14\,\text{TeV}$. Due to this high energy, inelastic proton collisions can be described by the direct interactions between the constituent partons, being quarks and gluons. The two partons entering the hard interaction carry fractions x_1 and x_2 of the proton momentum. The probability distribution function for a parton of type i to acquire a fractional momentum x and virtuality or squared four-momentum Q^2 is called the parton density function $f_i^{\text{p}}(x, Q^2)$ of the proton. Factorizing out the hard partonic $1 + 2 \to n$ interaction, the total hadronic cross section of a hard scattering is calculated by convoluting the probability functions f_1 and f_2 with the cross section of the partonic interaction,

$$\mathrm{d}\sigma_{\text{p+p}\to n} = \int_0^1 \mathrm{d}x_1 \int_0^1 \mathrm{d}x_2\, f_1^{\text{p}}\left(x_1, Q_F^2\right) f_2^{\text{p}}\left(x_2, Q_F^2\right) \mathrm{d}\hat{\sigma}_{1+2\to n}(Q_F^2), \tag{3.1}$$

where Q_F^2 is called the factorization scale. This scale Q_F^2 marks the point of separation between the hard perturbative partonic interaction which can be calculated, and the soft non-perturbative long-distance effects in the proton which are parametrized by the parton density functions. It provides a cut-off to regulate so-called infrared low-energy

divergences in the partonic cross section, associated to soft branching of the incoming partons. This is further discussed in Section 3.3.1.

With the above factorization, the study of proton collisions is split up in the description of the parton density functions, the treatment of the proton remnants or underlying event, and the calculation of the partonic interaction cross section.

3.1.2 Parton density functions

The actual implementation in calculations of the factorization of the partonic cross section is not unique. Any choice for the parton densities that takes into account all long-distance effects is acceptable. Finite short-distance contributions can be chosen to be either absorbed into the parton density functions or accounted for in the partonic cross section. The factorization scheme used for all studies in this thesis is the minimal subtraction scheme $\overline{\text{MS}}$, which defines the parton density functions directly in terms of hadronic matrix elements. These matrix elements may be given in terms of the operators $b_i(xp, k_T)$ and $b_i^\dagger(xp, k_T)$, which respectively annihilate and create a parton of type i, with longitudinal momentum xp along the beam line and transverse momentum k_T perpendicular to it, in a proton p of momentum p,

$$f_i^{\text{p}}(x, Q^2) = \frac{1}{N} \int \frac{\mathrm{d}^2 k_T}{(2\pi)^2} \langle \text{p}(p)| \, b_i^\dagger(xp, k_T) b_i(xp, k_T) \, |\text{p}(p)\rangle \, . \tag{3.2}$$

with N a normalization factor, for instance chosen such that $\int_0^1 \mathrm{d}x f_i^{\text{p}}(x, Q^2) = 1$. The rightmost operator absorbs the parton from the hadronic state of the proton, and the second emits it again. The hence defined parton distributions represent in essence the expectation value of a number operator on the hadronic state, counting the partons of flavour i with momentum fraction x and virtuality Q^2 in the proton.

A powerful consequence of the introduction of parton density functions and the discussion in Section 1.1.3, is the possibility to measure the parton distributions at one scale Q^2, and predict them via a parametrization for any other scale Q'^2, as long as both $\alpha_s(Q^2)$ and $\alpha_s(Q'^2)$ are sufficiently small such that perturbativity is retained. The scaling is governed by the same principles as the parton shower, described by the DGLAP equations which are discussed in Section 3.3.1. This scaling allows not only to predict parton distributions at different scales, but also to perform a global fit on experimental results obtained at different parton virtualities. Such fits are further constrained by QCD sum rules, expressing conservation laws on the constituents of the considered hadron. Global fits allow to test in addition the global consistency of the theoretical picture of hard scattering, based on factorization and the universality of parton distributions. This universality is assumed for instance when extrapolating the proton-structure results from deep-inelastic electron–proton scattering at lower energies to the energy conditions of the LHC proton–proton collider.

The best known examples of global parton-density fits come from the CTEQ [96] and the MRST [97] groups. They both fit to a vast number of experimental results, mainly measurements from the HERA electron–proton collider at DESY [98] and from neutrino–nucleus fixed target experiments [91]. Their approach differs in the parametrization, where the CTEQ fit employs 20 parameters and the MRST analysis

15, resulting both in very good fit results. Nowadays parton density functions are also determined at next-to-leading order, providing the opportunity to obtain uncertainties on the parameters of the fit. By variation of the parameters within their confidence intervals, uncertainties from parton-density fits on predicted observables can be evaluated, taking into account possible correlations [99]. In Section 6.2 this technique is applied to estimate the influence of these uncertainties on a method for the calibration of the performance of b-identification algorithms.

3.1.3 The underlying event

In proton collisions, each inelastically colliding proton will leave behind a coloured remnant that does not take part in the hard part of the interaction. The composite nature of the incoming protons makes it possible for several partons from the same proton to undergo separate scatterings in a single collision, referred to as multiple interactions. The combination of both the multiple interactions and the remnant of the proton constitutes the underlying event.

Over the last years, large progress has been made in the phenomenological studies of the underlying event in jet events by the CDF experiment at the Tevatron [100]. The multiplicity and transverse momentum spectra of charged tracks are studied in regions in space away from jet activity. The average charged multiplicity per unit of pseudorapidity in these regions turns out to be significantly higher with respect to measurements in minimum-bias events. This observation is called the pedestal effect.

The difficulty in modelling the underlying event stems from the need to extrapolate the perturbative-QCD picture, discussed in Section 1.1.3 to the soft regime where the strong coupling constant is large. In `PYTHIA` a sophisticated model implements both multiple interactions and proton-remnant formation. For each proton remnant, depending on the parton involved in the hard interaction, two-, three- or four-quark objects are created such that momentum, flavour, charge and colour are conserved globally. To simulate multiple interactions, the model starts from the basic partonic $2 \to 2$ differential cross section with respect to the momentum transfer p_{T}. This cross section is divergent for $p_{\mathrm{T}} \to 0$, which is regulated in the model by a colour-screening cut-off $p_{\mathrm{T,min}}$. It is assumed that different pairwise interactions take place essentially independently of each of other, generating a Poissonian distribution for the number of interactions. Furthermore, considering the protons as objects extended in space, collisions range from central to peripheral. The Poissonian distribution of interactions still holds for a given impact parameter of the collision, but needs to be convoluted with a distribution function of the matter inside the proton. A convenient function is constructed by superposing two spherical Gaussian distributions, modelling the proton as a parton cloud with a dense core. This choice results in a larger probability for multiple particle interactions when the cores of the protons overlap, corresponding at the same time with an increased likelihood for hard scattering to occur. This provides a possible explanation for the previously mentioned pedestal effect.

The extrapolation to LHC energies of the current experimental knowledge on the underlying event is problematic, since the parameters in the model, especially the cut-off $p_{\mathrm{T,min}}$, are explicitly dependent on the centre-of-mass energy of the protons. In all simulations used in this thesis, the `PYTHIA` parameters controlling the underlying event

are fixed to the values shown in Table 3.1, in the case of a double Gaussian parton-cloud distribution in the protons and a continuous turn-off of the cross section at $p_{\mathrm{T,min}}$ (`MSTP(81)=1` and `MSTP(82)=4`). Other parameters are kept at their default values in `PYTHIA`. This set of parameter values provides at the same time a good description of the minimum-bias data at CDF and at lower energy at UA5 [101]. For the study of underlying-event systematics in Section 6.2, the extrapolation of the colour-screening cut-off $p_{\mathrm{T,min}}$ to LHC energies is varied between 1.9 and 2.9 GeV/c, corresponding to the limits of the 99% confidence interval. Although this variation is large, the uncertainty from the extrapolation to the LHC regime can currently not be accounted for, and therefore care is needed. The ultimate answer for the simulation of the underlying event will eventually be obtained from a parameter tune to the LHC proton collisions at 14 TeV themselves [102].

Underlying event model parameter	`PYTHIA` parameter	Value
Regularization scale $p_{\mathrm{T,min}}$	`PARP(82)`	2.4 GeV/c
Double-Gaussian core radius fraction	`PARP(84)`	0.4
Matter fraction in the core	`PARP(83)`	0.5
Reference energy scale for $p_{\mathrm{T,min}}$	`PARP(89)`	14 000 GeV

Table 3.1: `PYTHIA` underlying event parameters used in the simulations for the analyses in this thesis [101].

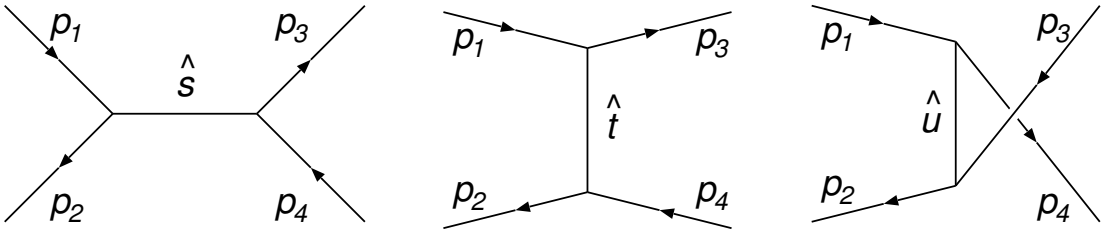
3.2 Partonic Interactions

The partonic cross section $\mathrm{d}\hat{\sigma}$, factorized from the hadronic cross section in (3.1), can itself be decomposed as

$$\mathrm{d}\hat{\sigma}_{1+2\to n} = \frac{1}{F} \times |\mathcal{M}|^2 \times \mathrm{d}\Phi_n, \tag{3.3}$$

where $F = 4\sqrt{(p_1 \cdot p_2)^2 - (m_1 m_2)^2}$ is a flux factor, $\mathrm{d}\Phi_n(p_1 + p_2; p_3, \ldots, p_{n+2})$ is the Lorentz-invariant n-body phase-space element imposing momentum conservation, and $\mathcal{M}(p_1, p_2; p_3, \ldots, p_{n+2})$ is called the matrix element, describing the quantum-mechanical transition amplitude of the considered process, calculated with Feynman rules. In the case of massless incoming partons, the flux factor reduces to $F = 2\hat{s}$ in the partonic centre-of-mass system, with the squared centre-of-mass energy $\hat{s}$ related to s by $\hat{s} = x_1 x_2 s$.

In the following, several phenomenological aspects of partonic interactions are discussed in the framework of $2 \to 2$ and $2 \to 3$ interactions, with emphasis on heavy-quark production.



(a) s-channel production (b) t-channel production (c) u-channel production

Figure 3.2: s-, t- and u-channel production in $2 \to 2$ scattering.

3.2.1 Phenomenology of $2 \to 2$ parton scattering

The kinematics of the two initial- and the two final-state partons in the $1 + 2 \to 3 + 4$ scattering are conveniently described using the Mandelstam variables

$$\begin{aligned} \hat{s} &= (p_1 + p_2)^2 = (p_3 + p_4)^2, \\ \hat{t} &= (p_1 - p_3)^2 = (p_2 - p_4)^2, \\ \hat{u} &= (p_1 - p_4)^2 = (p_2 - p_3)^2. \end{aligned} \tag{3.4}$$

Energy conservation implies that $\hat{s} + \hat{t} + \hat{u} = m_1^2 + m_2^2 + m_3^2 + m_4^2$, leaving only two independent variables. The partonic cross section (3.3) can be rewritten for two final-state particles as a function of $\hat{s}$ and $\hat{t}$ as

$$\mathrm{d}\hat{\sigma} = \frac{1}{64\pi\hat{s}} \frac{1}{|\mathbf{p}^\star_{1,2}|} |\mathcal{M}|^2 \, \mathrm{d}\hat{t}, \tag{3.5}$$

where $\mathbf{p}^\star_{1,2}$ represents the momentum of the incoming partons in the centre-of-mass frame. For massless incoming partons $|\mathbf{p}^\star_{1,2}| = \hat{s}/4$.

Three types of scattering can be distinguished at leading order: resonance formation, schematically represented in Figure 3.2(a), and scattering with exchange of a virtual particle, shown in Figures 3.2(b) and 3.2(c). These processes are typically referred to as s-, t- and u-channel production, since the four-momentum exchange equals $\hat{s}$, $\hat{t}$ and $\hat{u}$ respectively, as defined in (3.4).

As an example, the squared matrix element for heavy-quark pair $\mathrm{Q\bar{Q}}$ production is discussed at leading order. The exchanged quark in the t- and u-channel contributions will be off-shell by at least the heavy-quark mass m_Q. Heavy-quark pair production is therefore controlled by the coupling α_s evaluated at the heavy-quark mass scale, making perturbative calculations only useful for bottom and top quarks.

The contribution from the s-channel massless quark annihilation is given by

$$\begin{aligned} \overline{\sum} \left|\mathcal{M}_{\mathrm{q\bar{q}}\to\mathrm{Q\bar{Q}}}\right|^2 &= \frac{4(4\pi)^2\alpha_s^2}{9} \left[\frac{\hat{t}^2 + \hat{u}^2}{\hat{s}^2} + \frac{2m_\mathrm{Q}^2(2\hat{s} - m_\mathrm{Q}^2)}{\hat{s}^2}\right] \\ &= \frac{(4\pi)^2\alpha_s^2}{9} \left[(1 + \cos^2\theta^\star)^2 + (1 - \cos^2\theta^\star)^2 + \frac{8m_\mathrm{Q}^2}{\hat{s}} \sin^2\theta^\star\right], \end{aligned} \tag{3.6}$$

where $\overline{\sum}$ averages and sums over respectively all initial- and final-state colour and polarization degrees of freedom, and where the second expression is given as a function

of the scatter angle $\theta^\star$ in the centre-of-mass frame. The result is completely analogous to the corresponding $q\bar{q} \to \gamma^* \to Q\bar{Q}$ process in QED, replacing appropriately the averaging and the coupling constant. The second form of (3.6) reveals the behaviour due to the fermionic nature of the quarks. The first two terms describe the transition where $q/\bar{q}$ and $Q/\bar{Q}$ have parallel or anti-parallel spins respectively, favoured for $\theta^\star \to 0$ and $\theta^\star \to \pi$ respectively. The third term corresponds to a spin-0 final state involving a spin flip, which is only possible for massive particles.

The relation (3.6) is not Lorentz invariant, because the angle $\theta^\star$ is defined in the centre-of-mass frame and changes under Lorentz boosts. When expressed as a function of the Lorentz-invariant transverse mass $m_\mathrm{T} = \sqrt{m_\mathrm{Q}^2 + p_\mathrm{T}^2}$ of the outgoing particles and of the difference $\Delta y = y_3 - y_4$, where the rapidity y is defined as

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right), \tag{3.7}$$

the matrix element (3.6) becomes

$$\overline{\sum} \left| \mathcal{M}_{q\bar{q} \to Q\bar{Q}} \right|^2 = \frac{4(4\pi)^2 \alpha_s^2}{9} \left(\frac{1}{\cosh(\Delta y) + 1} \right) \left(\cosh(\Delta y) + \frac{m_\mathrm{Q}^2}{m_\mathrm{T}^2} \right). \tag{3.8}$$

The amplitude for the heavy-quark pair final state also receives contributions from three diagrams initiated by gluons. A heavy quark can be exchanged in a t- or u-channel process, or a gluon can appear in the s-channel. Evaluation of these contributions must be performed simultaneously in order to keep the final result gauge invariant, showing the need for the gluon self-coupling in the theory of QCD. The matrix element as a function of m_T and Δy is given by

$$\overline{\sum} \left| \mathcal{M}_{gg \to Q\bar{Q}} \right|^2 = \frac{(4\pi)^2 \alpha_s^2}{24} \left(\frac{8\cosh(\Delta y) - 1}{\cosh(\Delta y) + 1} \right) \left(\cosh(\Delta y) + 2\frac{m_\mathrm{Q}^2}{m_\mathrm{T}^2} - 2\frac{m_\mathrm{Q}^4}{m_\mathrm{T}^4} \right). \tag{3.9}$$

It is interesting to investigate the behaviour of the quark-initiated and gluon-initiated production at large Δy,

$$\overline{\sum} \left| \mathcal{M}_{q\bar{q} \to Q\bar{Q}} \right|^2 \overset{\Delta y \gg 0}{\sim} \text{constant} \qquad \text{and} \qquad \overline{\sum} \left| \mathcal{M}_{gg \to Q\bar{Q}} \right|^2 \overset{\Delta y \gg 0}{\sim} e^{\Delta y}. \tag{3.10}$$

Heavy quarks produced by quark annihilation will hence be closer in rapidity than those produced by gluon–gluon fusion. The influence of the transverse momentum p_T on both matrix elements is less pronounced. The dependence of the gluon- and quark-initiated matrix elements on the rapidity difference Δy between the heavy quarks and on their p_T is shown in Figures 3.3 and 3.4 respectively.

The interpretation of these dependencies in terms of observable quantities is not straightforward, as the matrix element still needs to be convoluted with the parton distribution functions. Using the `PYTHIA` Monte-Carlo generator the resulting distributions have been generated for quark- and gluon-initiated collisions separately, assuming top-quark pair production under LHC conditions. In Figure 3.5 the p_T distribution of the top quarks is shown for both types of initial partons. The deviations prove to be small. For the Δy distribution, on the other hand, the matrix-element dependency

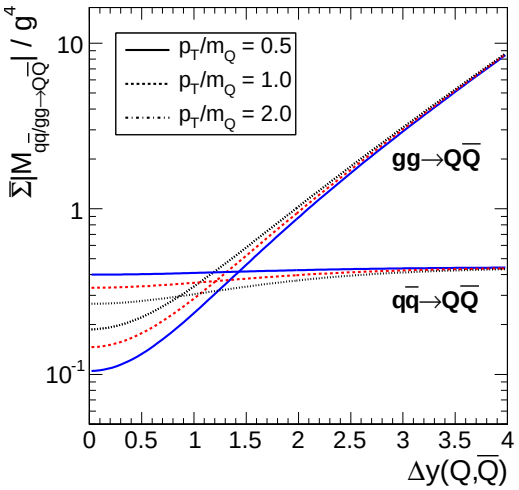


Figure 3.3: Dependence of the leading-order $gg/q\bar{q} \to Q\bar{Q}$ matrix element on Δy between the heavy quarks.

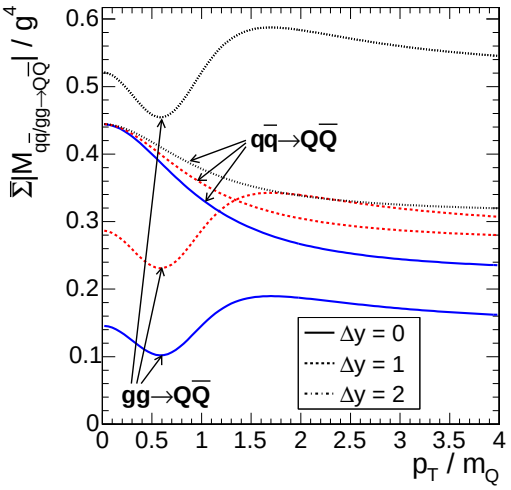


Figure 3.4: Dependence of the leading-order $gg/q\bar{q} \to Q\bar{Q}$ matrix element on the p_T of the heavy quarks.

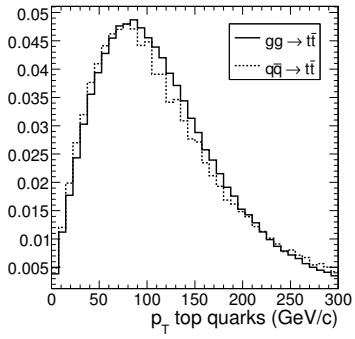


Figure 3.5: Distribution of p_T for top-quark pairs from quark annihilation and gluon fusion.

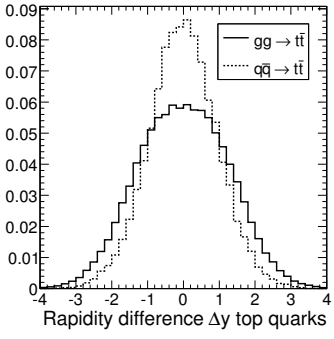


Figure 3.6: Distribution of Δy for top-quark pairs from quark annihilation and gluon fusion.

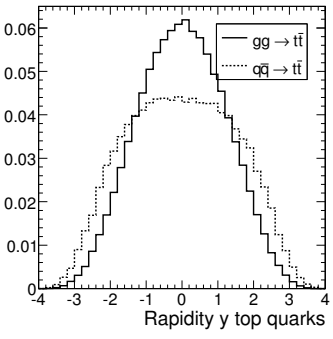


Figure 3.7: Distribution of y for top-quark pairs from quark annihilation and gluon fusion.

is still slightly present after taking into account the parton-density functions. In Figure 3.6 the aforementioned preference of gluon-initiated events for a larger rapidity difference is indeed observed. This does not imply, however, the top quarks from gluon fusion to have larger rapidities compared to quark-annihilation production. On the contrary, as can be seen from the top-quark rapidity distribution in Figure 3.7. Quark-initiated collisions prove to produce top quarks at slightly higher rapidities. In Section 3.4.1 it is shown that top quarks are mainly produced at the Tevatron through quark annihilation, while the gluon fusion dominates at the LHC. Hence top-quark pairs are more centrally produced at the LHC than at the Tevatron.

3.2.2 Phenomenology of $2 \to 3$ parton scattering

The phenomenology of partonic interactions leading to three-particle final states depends heavily on the process under investigation. In this section the exclusive production of a quark pair with an extra final-state gluon is taken as an example to introduce some general concepts, like the renormalization scheme, and to build the bridge towards the parton shower in Section 3.3.1.

The differential cross section for the production of a quark pair and a gluon from electron–positron annihilation is given by

$$\frac{1}{\hat{\sigma}} \frac{\mathrm{d}^2\hat{\sigma}}{\mathrm{d}z_1 \mathrm{d}z_2} \propto \alpha_s \frac{z_1^2 + z_2^2}{(1-z_1)(1-z_2)} \tag{3.11}$$

with

$$0 \leq z_1, z_2 \leq 1 \qquad \text{and} \qquad z_1 + z_2 \geq 1, \tag{3.12}$$

where $z_1 = 2E_\mathrm{q}/\sqrt{\hat{s}}$ and $z_2 = 2E_{\bar{\mathrm{q}}}/\sqrt{\hat{s}}$ are the energy fractions of the final-state quark and anti-quark. Since $1 - z_1 = z_2 E_\mathrm{g}(1 - \cos\theta_{2\mathrm{g}})/\sqrt{\hat{s}}$ and $1 - z_2 = z_1 E_\mathrm{g}(1 - \cos\theta_{1\mathrm{g}})/\sqrt{\hat{s}}$, where E_g is the gluon energy and $\theta_{i\mathrm{g}}$ $(i = 1, 2)$ denote the angles between the gluon and the quarks, it is seen that the cross section is divergent in phase-space regions where the gluon is collinear ($\theta_{i\mathrm{g}} \to 0$) or where the gluon is soft ($E_\mathrm{g}/\sqrt{\hat{s}} \to 0$). For parton scattering the relation (3.11) stays valid provided only gluon branching from final-state partons is considered, and if an additional factor α_s^2 is taken into account.

These singularities are not physical, but rather indicate a breakdown of the perturbative approach in the confined region of phase-space. In this regime the non-perturbative fragmentation of quarks into so-called jets of hadrons takes place, as discussed in Section 3.3.3. When the emission angle of the gluon from a quark becomes small, the hadronic final state that arises from fragmentation of both partons will be contained in the same jet, transforming the event in a $2 \to 2$ type event from the experimental point of view. Hence, an experimental definition of a jet implies a separation between the two-jet and three-jet event types and provides a natural cut-off to regulate the encountered singularities in $2 \to 3$ parton scattering.

The exclusive $2 \to 3$ process proves to be suppressed by a factor α_s compared to the leading exclusive $2 \to 2$ production. Therefore, when considering the total or inclusive quark-pair cross section, which is the sum of all possible exclusive n-parton final states, the $2 \to 3$ process is seen as a correction to the dominant $2 \to 2$ process. Unlike in the exclusive cases, the calculation of the inclusive cross section needs to be robust with respect to the choice of the jet definition. Therefore this jet definition cannot be used to regularize the divergences. Using dimensional regularization, on the other hand, it can be shown that the singularities in the leading-order $2 \to 3$ production are exactly cancelled by the lowest-order virtual corrections to the $2 \to 2$ process for massless quarks [89]. In general, at higher orders or when massive outgoing quarks are considered, such a cancellation is not exact. The divergences need to be subtracted order by order and absorbed into the QCD Lagrangian, to ensure the result to be finite in each order of α_s separately. This renormalization procedure necessarily introduces an unphysical renormalization scale Q_R^2 into the results of calculations. The renormalization scale and the factorization scale, introduced in Section 3.1.1, are in practise often identified to one single scale Q^2.

Physical quantities in QCD are ideally calculated as the sum of all orders in α_s of a perturbation series. Since such a sum needs to be invariant under changes of the unphysical scale Q^2, a variation of Q^2 in a calculation up to order $\alpha_s^n(Q^2)$ leads to corrections of $\mathcal{O}(\alpha_s^{n+1}(Q^2))$. Equivalently, the change in the expansion variable $\alpha_s^n(Q^2)$ due to a variation of the scale Q^2, described by (1.15), needs to be exactly compensated by the scale dependence of the corresponding coefficient in the perturbative expansion. Moreover, by using (1.15), the scale dependent part of the coefficient of the term in $\alpha_s^{n+1}(Q^2)$ can be determined from the results at orders $\leq n$. Using these properties of the scale dependence, a theoretical uncertainty can be attributed to QCD predictions. Typically this uncertainty is obtained by varying both factorization and renormalization scales between $Q^2/4$ and $4Q^2$, although this is not justified in all cases (e.g. $\mathrm{b\bar{b}}$ production). A decreasing scale dependence for increasing calculation order is the only reassuring signal for a well converging perturbation series.

3.3 Hadronic Final-State Formation

The property of asymptotic freedom of the strong interaction is reflected in the confinement of quarks in colour-neutral hadrons. As a consequence bare quarks cannot be observed directly. Quarks and gluons manifest themselves experimentally as jets of particles. After their production in the hard interaction or originating from the proton remnants, quarks and gluons undergo parton branching from their initial energy down to scales where the coupling of the strong interaction becomes too large and perturbative calculations break down. The following non-perturbative process of jet formation through fragmentation and hadronization at low energy scales is therefore to be described with phenomenological models.

3.3.1 The parton shower

Three types of QCD parton branchings are allowed: a quark can radiate a gluon ($\mathrm{q} \to \mathrm{qg}$), a gluon can split into a quark pair ($\mathrm{g} \to \mathrm{q\bar{q}}$), or a gluon can itself split into two gluons due to its colour charge ($\mathrm{g} \to \mathrm{gg}$). In the limit where the branching happens at small angles, the relation between the matrix element for $n+1$ and n final-state partons is given by

$$|\mathcal{M}_{n+1}|^2 = \frac{16\pi\alpha_s}{t}\hat{P}(z)\,|\mathcal{M}_n|^2\,, \tag{3.13}$$

where t is the virtuality m^2 of the initial parton before branching and where $\hat{P}(z)$ is called the splitting function for the considered branching with z the relative energy taken by one of the radiated partons. The splitting functions for gluon radiation from a quark line, for gluon splitting into quarks, and for gluon splitting into gluons are

respectively

$$\begin{aligned}
\hat{P}_{\mathrm{q}\to\mathrm{qg}}(z) &= \frac{4}{3}\left[\frac{1+z^2}{1-z}\right], \\
\hat{P}_{\mathrm{g}\to\mathrm{q\bar{q}}}(z) &= \frac{n_f}{2}\left[z^2+(1-z)^2\right], \\
\hat{P}_{\mathrm{g}\to\mathrm{gg}}(z) &= 3\left[\frac{1-z}{z}+\frac{z}{1-z}+z(1-z)\right],
\end{aligned} \tag{3.14}$$

where n_f is the kinematically allowed number of quark flavours. It is immediately seen that a singularity arises when the gluon radiation becomes soft ($z \to 0$), or equivalently when gluons are emitted at small angles. This so-called collinear divergence is not physical but rather expresses the breakdown of perturbativity at decreasing energies due to the rising coupling α_s.

When multiple consecutive parton branchings are considered from an initial scale Q^2, forming a so-called parton shower, the evolution with Q^2 of the probability distributions $f_{\mathrm{b}}(x, Q^2)$ for parton flavour b with fractional momentum x is given by the coupled set of partial-differential equations

$$Q^2 \frac{\partial}{\partial Q^2} f_{\mathrm{b}}(x, Q^2) = \frac{\alpha_s(Q^2)}{2\pi} \sum_{\mathrm{c}} \int_x^1 \frac{\mathrm{d}z}{z} P_{\mathrm{a}\to\mathrm{bc}}(z) f_{\mathrm{a}}(\frac{x}{z}, Q^2) \tag{3.15}$$

known as the DGLAP equations [103–106], where the kernels $P_{\mathrm{a}\to\mathrm{bc}}$ correspond to the splitting functions (3.14), with the collinear divergences regularized using overall flavour and momentum conservation.

The choice of the evolution variable Q^2 in (3.15) is not unique. In the `PYTHIA` parton-shower algorithm this scale is associated with the squared mass m_{a}^2 of the mother parton in the transition a $\to$ bc. This choice ensures a mass ordering in the evolution, and is in the program further extended to take into account colour-coherence effects between the partons, which reduce soft-gluon emission. Other choices for the evolution scale, like in the `HERWIG` [107] parton-shower implementation, naturally take into account coherence effects by using angular ordering.

Using the branching evolution in (3.15), the final-state radiation in an event can be traced from an upper scale Q^2_{max} down to a lower cut-off scale Q^2_{min} at which the perturbative regime needs to be abandoned. This way the final-state radiation of the partons produced in the hard interaction can be modelled. The shower evolution can be traced in the same way for the incoming particles. The partons taken from the parton density functions can undergo initial-state radiation before they enter the hard interaction. In practise initial showers are simulated in reverse direction, from the scale of the interaction up to the initial scale at which the parton is extracted from the parton density functions.

In simulations two parameters in the parton-shower model are of particular importance: the scale Q^2_{max}, at which the parton shower is stopped for the initial-state or started for the final-state simulation, and Λ_{QCD}, controlling the absolute scale of the strong coupling α_s. The cut-off Q^2_{min} at which the shower is stopped does not influence the shower itself, but rather the subsequent non-perturbative hadronization. Typically the scale Q^2_{max} is chosen to match the scale of the hard process [92]. For the final state

$Q^2_{\text{max}} = 4Q^2_{\text{hard}}$ is preferred, which corresponds roughly to the maximally allowed virtuality of the parton shower in massless $2 \to 2$ processes. For initial-state showers the matching scale is typically set to $Q^2_{\text{max}} = Q^2_{\text{hard}}$, unless a heavy resonance is produced in the s channel.

In Table 3.2 the parameters are shown that govern the described initial- and final-state parton-shower simulation in the PYTHIA implementation, optimized for the CTEQ6L parton distribution functions [101]. For the evaluation of systematic uncertainties related to the parton shower, first all three Λ_{QCD} parameters need to be simultaneously changed within the confidence interval. Next, also the matching-scale parameters need to be altered within reasonable limits, to estimate the uncertainty from the somewhat arbitrary choice of these lower and upper scales for respectively initial- and final-state parton showers. The given intervals reflect the current understanding [101]. The experimental effects of these parameters on the parton shower and the possible future improvement of the confidence intervals is discussed in more detail in Section 6.2.3.

Parton-shower parameter	PYTHIA parameter	Value	Confidence interval
Λ_{QCD} general	PARJ(81)	0.25 GeV	$0.15 - 0.35$ GeV
Λ_{QCD} for ISR	PARP(61)	0.25 GeV	$0.15 - 0.35$ GeV
Λ_{QCD} for FSR	PARP(72)	0.25 GeV	$0.15 - 0.35$ GeV
$Q^2_{\text{max}}/Q^2_{\text{hard}}$ for ISR	PARP(67)	1	$0.25 - 4$
$Q^2_{\text{max}}/Q^2_{\text{hard}}$ for FSR	PARP(71)	4	$1 - 16$

Table 3.2: PYTHIA parton-shower parameters.

3.3.2 Heavy quarks from gluon splitting

The presence of heavy quarks in parton showers is not only due to their direct production in the hard event. As discussed in Section 3.3.1, gluons can split into a quark–anti-quark pair. If a splitting into heavy quarks happens with a sufficiently-small opening angle, the heavy-quark pair is fully contained in a single experimentally-observed jet, although the jet could nevertheless originate from a gluon or light quark. The distinction between heavy-flavour content from directly-produced heavy quarks or from gluon splitting inside jets plays an important role in the interpretation of efficiencies of experimental heavy-flavour identification algorithms. It gives rise to the notion of the physical definition and the algorithmic definition of a heavy-flavour jet, depending respectively on whether the jet originates from a directly-produced heavy quark, or whether the jet contains a heavy-flavour quark independent of the production mechanism.

A gluon that splits into a pair of heavy quarks $\text{Q}\bar{\text{Q}}$ must have a virtuality $Q^2 > 4m^2_{\text{Q}}$, and therefore perturbative methods should be applicable for a sufficiently large quark mass m_{Q}, as is the case for b but also for c quarks. The calculation of the mean number $R_{\text{Q}\bar{\text{Q}}}$ of heavy quark pairs $\text{Q}\bar{\text{Q}}$ per jet consists of two parts. First the number of gluons $n_{\text{g}}(E, Q^2)$ with virtuality Q^2 is determined inside a gluon-initiated jet produced with

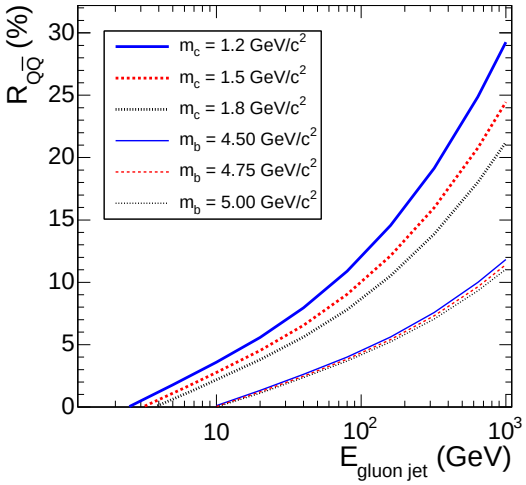


Figure 3.8: Average abundance $R_{Q\bar{Q}}$ of charm and bottom quark pairs in gluon jets.

energy E in a hard process [108–110]. Then the number of heavy-quark pairs, calculated in analogy to (3.15), becomes

$$R_{Q\bar{Q}}(E) = \int_{4m^2}^{E^2} \frac{dQ^2}{Q^2} \frac{\alpha_s(Q^2)}{2\pi} \int_{z_-}^{z_+} \frac{dz}{2} \left[z^2 + (1-z)^2 + \frac{2m_Q^2}{Q^2} \right] n_g\left(E, Q^2\right), \qquad (3.16)$$

where the splitting function $\hat{P}_{g\to q\bar{q}}$ from (3.14) is generalized to massive quarks. Due to the kinematics of the decay $g \to Q\bar{Q}$, the integration over the longitudinal momentum fraction z of the heavy quark is limited by $z_\pm = (1 \pm \beta)/2$, with $\beta = \sqrt{1 - 4m_Q^2/Q^2}$. Performing this integration over z, one obtains

$$R_{Q\bar{Q}}(E) = \frac{1}{6\pi} \int_{4m^2}^{E^2} \frac{dQ^2}{Q^2} \alpha_s(Q^2) \left[1 + \frac{2m_Q^2}{Q^2} \right] \sqrt{1 - \frac{4m_Q^2}{Q^2}}\ n_g\left(E, Q^2\right). \qquad (3.17)$$

The predicted average number of charm and bottom quark pairs per gluon-initiated jet is shown in Figure 3.8. About 4% of primary gluons with an energy of 100 GeV give rise to a $b\bar{b}$ quark pair, while about 10% are expected to create a $c\bar{c}$ pair in the parton shower.

3.3.3 Jet fragmentation

Due to the increase of the strong coupling constant at small energies, the perturbative evolution of the parton shower eventually has to be followed by a non-perturbative description of fragmentation or hadronization of the partons into jets of colour-neutral hadrons. This hadronization process is not yet understood from first principles, but several probabilistic phenomenological models describing the process are available.

One of the widely used models for fragmentation in Monte-Carlo simulations is the Lund string model [111], implemented in the `PYTHIA` program. This model relates to the intuitive idea of linear confinement of quarks, where the growing colour

field between quarks that move away from each other eventually gives rise to a new quark pair, hence confining the quarks in colour-neutral hadrons. In the Lund model coloured partons are mathematically connected by a one-dimensional string with constant tension, representing the colour field as a cylindrically-symmetric flux tube with a linearly-rising potential. The string constant κ, corresponding to the amount of energy per unit length, is estimated to be $\kappa \approx 1\,\mathrm{GeV/fm}$ from experiment and in agreement with the value obtained from calculations.

The string-fragmentation process takes the transverse momenta $p_{\mathrm{T,q}}$ of the quarks in the created pairs with mass m_{q} from a Gaussian probability distribution

$$\exp\left(-\frac{\pi m_{\mathrm{q}}^2}{\kappa}\right)\exp\left(-\frac{\pi p_{\mathrm{T,q}}^2}{\kappa}\right), \tag{3.18}$$

implying a suppression of heavy quark production $\mathrm{u:d:s:c} \approx 1:1:0.3:10^{-11}$. Hence charm- and bottom-quark generation are neglected in the non-perturbative hadronization process. Longitudinal momenta of the created hadrons are generated according to the fragmentation function $f(z)$. This function describes the probability of the fractional momentum z to be taken by the hadron that is split off from the string, leaving fractional momentum $1-z$ for the remainder of the string. Imposing the requirement on $f(z)$ to be independent of the direction of fragmentation on the string, the Lund symmetric fragmentation function

$$f(z) \propto \frac{1}{z}(1-z)^a \exp\left(-\frac{b(m_{\mathrm{h}}^2 + p_{\mathrm{T,h}}^2)}{z}\right) \tag{3.19}$$

is obtained for u, d and s quarks, where a (the so-called Lund a) and b (the so-called Lund b) are parameters that need to be fit to experimental observations. The dependence on the mass m_{h} of the created hadron indicates a harder spectrum for heavier hadrons.

For fragmentation of strings with massive charm or bottom quarks at the string endpoint, the function (3.19) is to be corrected by a factor $z^{-bm_{\mathrm{q}}^2}$. This correction results in a better description for spectra of B mesons in jets. Experimental data, however, indicate the need for a harder fragmentation function for heavy flavours. The best known description is derived from the energy difference due to fragmentation of a rapidly-moving heavy quark into a heavy-flavoured meson and a light quark, given by the Peterson function [112]

$$f(z) \propto \frac{1}{z\left(1-\frac{1}{z}-\frac{\epsilon_{\mathrm{q}}}{1-z}\right)^2}, \tag{3.20}$$

where $\epsilon_{\mathrm{q}} \equiv \epsilon_{\mathrm{c}}$ and $\epsilon_{\mathrm{q}} \equiv \epsilon_{\mathrm{b}}$ are free parameters that are expected to scale like $\epsilon_{\mathrm{q}} \sim 1/m_{\mathrm{q}}^2$.

In Figure 3.9 the normalized Peterson function (3.20) is shown for b quarks, along with the normalized Lund fragmentation function (3.19) for the creation of a collinear hadron with mass $m_{\mathrm{h}} = m_{\mathrm{B}} = 5.28\,\mathrm{GeV}/c^2$, assuming both a massless quark or a b quark at the string endpoint. For the function parameters the default `PYTHIA` settings are used. Lund a takes the value 0.3 by default, Lund b is fixed to $0.58\,\mathrm{GeV}^{-2}c^2$, and the c- and b-quark Peterson functions are defined with $\epsilon_{\mathrm{c}} = 0.05$ and $\epsilon_{\mathrm{b}} = 0.005$.

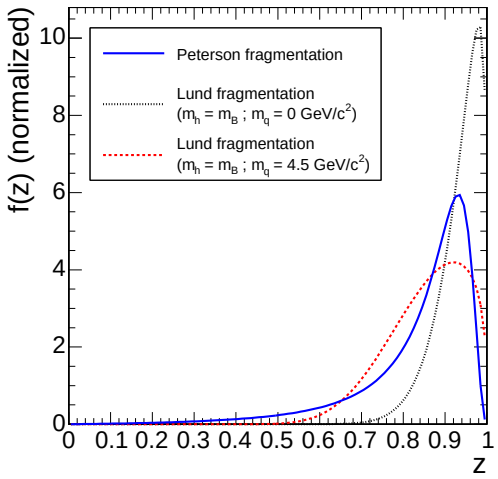


Figure 3.9: Normalized light and massive quark fragmentation functions.

In the study of systematics related to jet fragmentation, the uncertainties on the fragmentation parameters from a fit or tune to experimental data are used. Assuming universality of jet fragmentation between different colliders, the tune from DELPHI [101, 113] is used for the systematics studies in Section 6.2. In this tune the assumptions $\Lambda_{\mathrm{QCD}} = 297\,\mathrm{MeV}$ and $Q_{\mathrm{min}} = 1.34\,\mathrm{GeV}$ lead to the fragmentation parameters and uncertainties shown in Table 3.3.

Fragmentation parameter	PYTHIA parameter	Value	Uncertainty σ
$p_{\mathrm{h}x}$ and $p_{\mathrm{h}y}$ widths	PARJ(21)	$0.401\,\mathrm{GeV}/c$	$0.008\ \mathrm{GeV}/c$
Lund a	PARJ(41)	0.409	0.022
Lund b	PARJ(42)	$0.850\,\mathrm{GeV}^{-2}c^2$	0 (corr. to σ_a)
Peterson ϵ_{c}	-PARJ(54)	0.0372	0.0017
Peterson ϵ_{b}	-PARJ(55)	0.00284	0.00013

Table 3.3: PYTHIA fragmentation parameters from the DELPHI tune [101, 113].

Apart from the described Lund model, other fragmentation models exist. An often used alternative is the cluster model, implemented in HERWIG. After the perturbative shower and non-perturbative splitting of remaining gluons, the partons are clustered into colour singlet clusters. These clusters are then decayed isotropically to form the final-state mesons and baryons.

3.3.4 Jets with heavy-flavour content

After the fragmentation of a c or a b quark, jets are formed containing heavy mesons (D, $\mathrm{D_s}$, B, $\mathrm{B_s}$, etc.) and baryons (Λ_{c}, Λ_{b}, etc.). Due to the hard fragmentation function $f(z)$ for heavy quarks, these heavy hadrons carry a large fraction of the jet energy, and are thus strongly boosted in the jet direction for sufficiently-large jet energies. In the PYTHIA fragmentation implementation [111] the relative abundance of these heavy

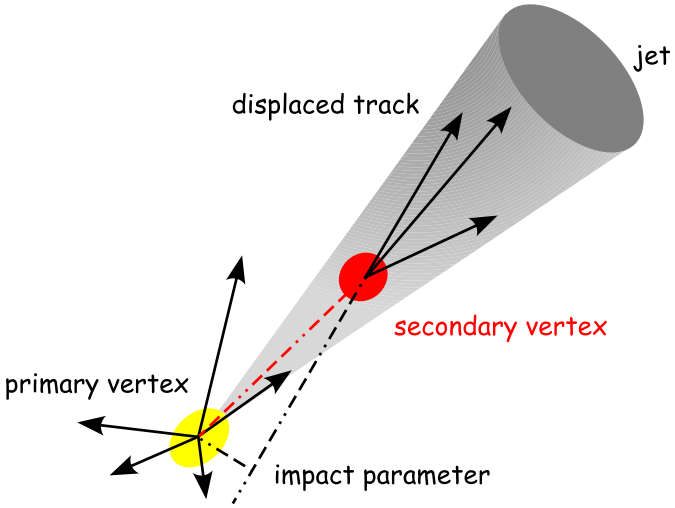


Figure 3.10: Schematic representation of the characteristics of a heavy-flavour jet.

hadrons follows from physics-motivated rules concerning for example the hadron spin. These rules are complemented in specific cases with additional parametrizations that enhance the agreement with experimental observations.

To a first approximation the decays of heavy-flavour hadrons can be considered in the so-called spectator model, where the quarks accompanying the heavy quark Q in the hadron are assumed not to take part in the weak $\mathrm{Q} \to \mathrm{W}^{-*}\mathrm{X}$ decay. As a consequence the lifetime of heavy-flavour hadrons is expected to depend only on the flavour type. For all b hadrons a lifetime of about 1.5 ps is indeed measured, corresponding to $c\tau \approx 450\,\mu\mathrm{m}$. For c hadrons this universality is less pronounced, because diagrams causing deviations from the spectator model contribute more for a lower heavy-quark mass. This long lifetime for heavy-flavour hadrons reflects experimentally in displaced decay vertices and tracks not compatible with the primary vertex. These properties will be the main ingredients for the b-identification algorithms described in Section 5.2.1 and 5.2.2, and are visualized in Figure 3.10.

Jets with heavy flavour exhibit several other properties that distinguish them from other jets. In particular in b jets a high track multiplicity of on average about five charged tracks is expected from b-hadron decays, due to the direct $\mathrm{b} \to \mathrm{W}^{-*}\mathrm{X}$ decay or $\mathrm{b} \to \mathrm{W}^{-*}\mathrm{c}, \mathrm{c} \to \mathrm{W}^{+*}\mathrm{X}$ cascades. The weak decay can also give rise to non-isolated leptons in the jet from a leptonic $\mathrm{W}^* \to \ell\nu_\ell$ decay. Considering the b-hadron decays in the spectator model as quark decays, three b-quark decay channels can create these leptons: the direct decay $\mathrm{b} \to \mathrm{W}^{-*}\mathrm{X}, \mathrm{W}^{-*} \to \ell^-\bar{\nu}_\ell$ with branching fraction $\sim 10.7\%$, the cascade decay $\mathrm{b} \to \mathrm{W}^{-*}\mathrm{c}, \mathrm{c} \to \ell^+\nu_\ell\mathrm{X}$ with branching fraction $\sim 8.0\%$ and the so-called wrong-sign cascade $\mathrm{b} \to \mathrm{W}^{-*}\mathrm{X}, \mathrm{W}^{-*} \to \mathrm{q}'\bar{\mathrm{c}}, \bar{\mathrm{c}} \to \ell^-\bar{\nu}_\ell\mathrm{X}$ with branching fraction $\sim 1.6\%$ [8, 114]. In total an inclusive $\mathrm{b} \to \ell\mathrm{X}$ branching fraction of 19% is obtained for each lepton family. The leptons that are produced via these mechanisms have properties that allow them to be distinguished from other sources of leptons inside jets, like their momentum and relative momentum with respect to the jet axis. A b-jet identification algorithm that exploits $\mathrm{b} \to \ell$ decays is described in Section 5.2.3.

3.4 Examples of Parton-Level Phenomenology

In this section a few examples of physics processes are presented at the parton level: inclusive top-quark pair production, top quarks produced with one or two additional partons, and charged Higgs-boson production. These processes are chosen because they are of particular interest for the analyses presented in Chapter 6 and 7. The discussion aims to build the bridge between the theoretical framework presented in this chapter, and the experimental aspects of the channels relevant to the analyses.

3.4.1 Inclusive top-quark pair production

Theoretical aspects of the cross-section calculation for heavy-quark pair production at leading order (LO) and next-to-leading order (NLO) are discussed in Sections 3.2.1 and 3.2.2. Experimentally, top-quark detection has only been made possible since the advent of the Tevatron proton–anti-proton collider. With a centre-of-mass energy of $\sqrt{s} = 1.8\,\text{TeV}$ during the Run-I phase, a top-quark pair cross section of 4.8 pb is predicted at the next-to-leading order. During the still ongoing Run-II phase, the energy increased to 1.96 TeV, resulting in an expected $\mathrm{t\bar{t}}$ cross section of 6.6 pb. An overview of the current cross-section measurement results in the different decay channels, found to be in agreement with the predictions, can be found in [115]. These cross sections correspond to a fraction of about 10^{-10} of the total inelastic cross section. This should be contrasted to the LHC where the centre-of-mass energy of the colliding protons will reach 14 TeV. At the NLO the $\mathrm{t\bar{t}}$ cross section rises to 830 pb, compared to 560 pb at the LO [116]. With the total inelastic cross section expected to stay about constant, the ratio of both increases to about 10^{-8}, significantly facilitating the suppression of QCD backgrounds. It is interesting to note the anti-correlation between the $\mathrm{t\bar{t}}$ cross section and the top-quark mass, allowing for an additional indirect measurement of this mass [116].

For the creation of an on-shell top-quark pair to occur in proton–(anti-)proton collisions, the partonic centre-of-mass energy $\sqrt{\hat{s}}$ needs to exceed $2m_\mathrm{t} \approx 350\,\text{GeV}/c^2$. At this threshold it is found that $\hat{s} = x_1 x_2 s = (2m_\mathrm{t})^2$, and the factorization scale, chosen as the momentum transfer in the hard scattering, amounts to m_t. At the Tevatron collider $x_1 x_2 \approx 0.03$, resulting in $\sim 90\%$ of the $\mathrm{t\bar{t}}$ pairs created through quark annihilation. At the LHC on the other hand $x_1 x_2 \approx 6 \times 10^{-4}$. In this range the gluon density of the proton dominates and the situation is reversed: $\sim 90\%$ of the $\mathrm{t\bar{t}}$ production is gluon-initiated.

In the Standard Model the top quark is predicted to decay weakly for $\sim 99.8\%$ of the cases through $\mathrm{t} \to \mathrm{Wb}$ from the measurement of the unitary CKM quark-mixing matrix [30]. Neglecting other top-quark decays and the mass of the b quark, the total width for $m_\mathrm{t} = 175\,\text{GeV}$ is given by

$$\Gamma\left(\mathrm{t} \to \mathrm{Wb}\right) = \frac{G_F m_\mathrm{t}^3}{8\pi\sqrt{2}} \left|V_\mathrm{tb}\right|^2 \left[1 + \mathcal{O}\left(\frac{m_\mathrm{W}^4}{m_\mathrm{t}^4}\right)\right] \approx 1.5\,\text{GeV}. \tag{3.21}$$

This results in an expected lifetime $\tau_\mathrm{t} = 4.3 \times 10^{-25}\,\text{s}$ for the top quark, and $c\tau_\mathrm{t} = 0.12\,\text{fm}$. A direct consequence of this short lifetime is that top quarks decay before

they hadronize, and top mesons or baryons are not formed. Since in addition the lifetime is also short compared to the time scale associated to a spin flip of the top quark, spin correlations between pair-produced top quarks can be probed through the W helicities.

Due to the full branching into the $t \to Wb$ decay, the final state of top quark pairs is fully governed by the decays of the $W^{\pm}$ bosons. Because $W \to qq' \approx 2/3$ and $W \to \ell\nu_\ell \approx 1/9$ ($\ell = e, \mu, \tau$), three experimentally different final states can be distinguished. When both W bosons decay hadronically, the $t\bar{t}$ pair is said to decay fully hadronically (44.4%); with one hadronic and one leptonic W-boson decay, the lepton+jets or semi-leptonic final state is obtained (44.4%); with two leptons from the W bosons in the final state, the decay is called fully leptonic (11.1%).

3.4.2 Top-quark pair production with additional partons

Top-quark pairs are often produced along with additional partons. To a first approximation the resulting additional jets can be generated by initial- and final-state branching in parton-shower models as found in the `PYTHIA` generator. This is often sufficient in studies where the final state of interest does not involve these extra partons. This approximation deteriorates when more additional partons are expected, however, especially when heavy flavour is involved.

The four Feynman diagrams involved in the leading-order $pp \to t\bar{t}$ matrix-element calculation, become 206 diagrams when one additional parton is required. Most of these diagrams, contributing a large part of the total cross section, can be directly identified with a branching of an initial- or final-state parton. This shows that the parton shower indeed approximates a full matrix-element treatment, although under the assumption that the extra branching is uncorrelated to the $t\bar{t}$ production itself. Further deviations are due to the extra diagrams taken into account in the full matrix element, that cannot be reproduced by the parton shower. An example is the gluon branching from the top-quark propagator in the t- or u-channel production.

For simulation purposes, the implementation of individual matrix element calculations with multi-particle final states is tedious work. Instead, several event generators have been developed to dynamically generate and calculate Feynman diagrams. As an example `MadGraph/MadEvent` is used to generate, with exact matrix-element calculations, top-quark pair events with one additional b or non-b parton. These events are further referred to as $t\bar{t}b/t\bar{t}j$ events. To avoid collinear and soft divergences in initial gluon splitting, as described in Section 3.2.2, a threshold needs to be applied on the kinematics of the additional parton. For the analyses presented in this thesis, b identification of jets is a crucial tool to distinguish signal from background. As this b-tagging relies on the reconstruction of tracks within the tracker acceptance (see Section 5.2), $t\bar{t}b/t\bar{t}j$ events have been generated with a cut on the pseudorapidity $|\eta| < 2.5$ of the extra parton accompanying the top quarks. An additional cut on the transverse momentum $p_T > 10\,\text{GeV}/c$ of this parton is imposed to regulate the divergences discussed in Section 3.2.2. This generation resulted in a cross section of 678 pb with negligible statistical uncertainty. In Figures 3.11 and 3.12, the transverse-momentum and pseudorapidity distributions are shown for the extra parton accompanying the top-quark pair. For the use in the analysis in Chapter 7, these events are interfaced to `PYTHIA` for

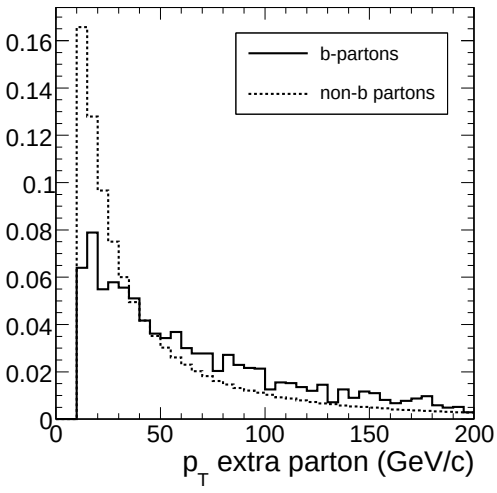


Figure 3.11: Distribution of the p_{T} of the extra parton, produced in $pp \to t\bar{t}b/t\bar{t}j$ in association to the top-quark pair.

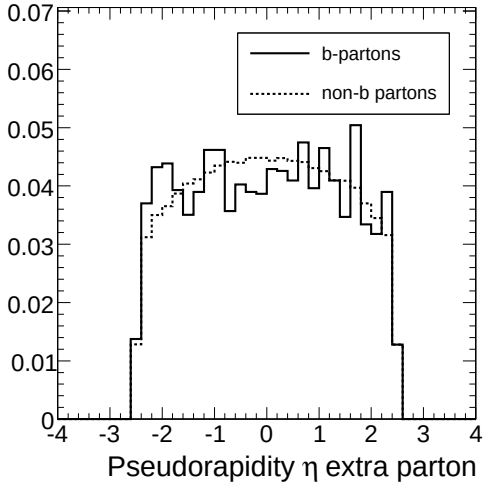


Figure 3.12: Distribution of the pseudorapidity η of the extra parton, produced in $pp \to t\bar{t}b/t\bar{t}j$ in association to the top-quark pair.

parton showering, decay and hadronization, after the simulation of the hard interaction in `MadGraph/MadEvent`.

It has been shown that the kinematics of one extra light parton are well described by the parton shower up to a transverse momentum $p_{\mathrm{T}} \sim 175\,\mathrm{GeV}/c$ [117]. For the description of two additional partons, deviations between the shower and matrix-element picture become larger, and the extra partons are poorly described by parton-shower branchings only. In physics analyses exploiting directly the kinematics of the extra partons, the hard interaction needs to be simulated with an exact matrix element generator. In the analysis presented in Chapter 7, event samples are used that have been generated with the `CompHEP` simulation program. Top-quark pair production with two additional light or heavy partons has been simulated separately at leading order for the $pp \to t\bar{t}b\bar{b}$ and the $pp \to t\bar{t}jj$ processes. On the partons produced in association with the $t\bar{t}$ pair, the generator level cuts $p_{\mathrm{T}} > 15\,\mathrm{GeV}/c$ and $|\eta| < 3.0$ have been applied, and a separation cut $\Delta R > 0.3$ between these partons in (η, ϕ) space has been imposed. These cuts regulate the collinear singularities, resulting in a cross section of 3.285 pb for the $pp \to t\bar{t}b\bar{b}$ process and 507.8 pb for $pp \to t\bar{t}jj$ production. In Figures 3.13 and 3.14, respectively the p_{T} and η distributions are shown for the extra partons, for b and non-b partons separately. For the use in the analysis in Chapter 7, as in the case of the $t\bar{t}b/t\bar{t}j$ generation, the $t\bar{t}b\bar{b}$ and $t\bar{t}jj$ events are interfaced to `PYTHIA` for parton showering, decay and hadronization after the simulation of the hard interaction in `CompHEP`.

Between the $pp \to t\bar{t}b\bar{b}$ and $pp \to t\bar{t}jj$ processes double counting of events can occur when a gluon in the latter process splits into a $b\bar{b}$ pair in the parton shower. To avoid this in a rigorous way, a next-to-leading order event generator should be used

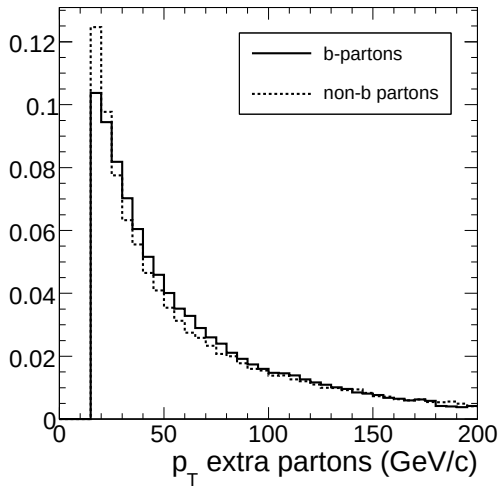


Figure 3.13: Distribution of the p_{T} of the extra partons, produced in $\mathrm{pp} \to \mathrm{t\bar{t}b\bar{b}}$ and $\mathrm{pp} \to \mathrm{t\bar{t}jj}$ in association to the top quark pair.

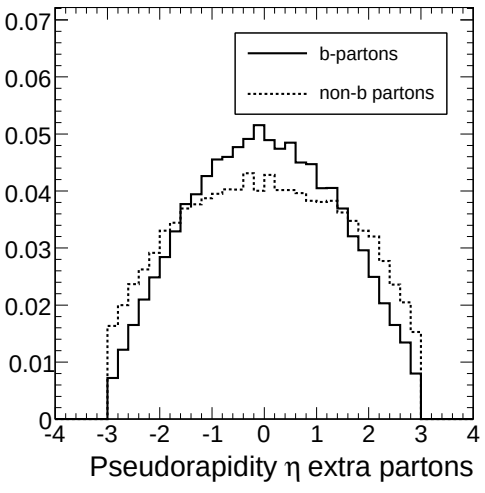


Figure 3.14: Distribution of the pseudorapidity η of the extra partons, produced in $\mathrm{pp} \to \mathrm{t\bar{t}b\bar{b}}$ and $\mathrm{pp} \to \mathrm{t\bar{t}jj}$ in association to the top quark pair.

that simultaneously simulates both processes and subtracts overlaps between them at higher orders. As this is currently not possible, this type of double counting is avoided in the above by rejecting $\mathrm{t\bar{t}jj}$ events where a $\mathrm{b\bar{b}}$ pair is found that passes the previously mentioned cuts on transverse momentum, pseudorapidity and angular separation.

Another problem of double counting, that occurs in multi-jet processes at leading order, involves the transition between the matrix element for the hard interaction and the parton shower in the case of massless partons. When a matrix-element generator is used to generate $\mathrm{t\bar{t}}$ events with n extra partons which satisfy certain thresholds, these events are subsequently processed by a parton-shower algorithm to generate extra partons from soft-gluon emissions not accounted for in the matrix elements. The number of experimentally-observable jets m in the final state of such events is in general larger than the number of hard partons n from the matrix element calculation. For certain events however, the thresholds on the partons after the parton shower, or overlaps between these partons giving rise to a single jet, can cause the event to obtain $m < n$ jets in the final state. Such events are double counted, since they should not be accounted for in a $\mathrm{t\bar{t}}$ matrix-element calculation with n extra partons. A jet-matching technique between matrix-element and parton-shower jets can be applied, such that all $\leq n$ exclusive parton final states are transformed from exact matrix-element calculations into an exclusive n-jet final state. Such a matching further enhances the compatibility of the leading-order simulations with parton shower to a full next-to-leading order description. The jet-matching technique recently deployed within the CMS collaboration uses the combination of the `ALPGEN v2.05` [95] and `PYTHIA` Monte-Carlo programs. Among other processes a large amount of $\mathrm{t\bar{t}}$ events have been generated in several non-overlapping samples. The final-state partons in the generated events have been required to satisfy $p_{\mathrm{T}} > 20\,\mathrm{GeV}$, $|\eta| < 5.0$ and $\Delta R > 0.7$. In Table 3.4 an overview is given of the various exclusive samples that have been generated, and their respective

leading-order cross sections.

In the case of the $t\bar{t} + 2\,\text{jet}$ production, a comparison has been performed applying the same generator cuts on the CompHEP $t\bar{t}jj$ samples as on the inclusive $t\bar{t}jj$ ALPGEN sample. This comparison has revealed that a correction factor of 0.64 on the CompHEP cross section is needed to correct for the double counting inclusively. This correction is applied throughout the thesis. The remaining kinematic differences between the two samples have been neglected.

Final state	Cross section (pb)
$t\bar{t} + 0\,\text{jets}$	190
$t\bar{t} + 1\,\text{jets}$	170
$t\bar{t} + 2\,\text{jets}$	100
$t\bar{t} + 3\,\text{jets}$	40
$t\bar{t} + \geq 4\,\text{jets}$	61

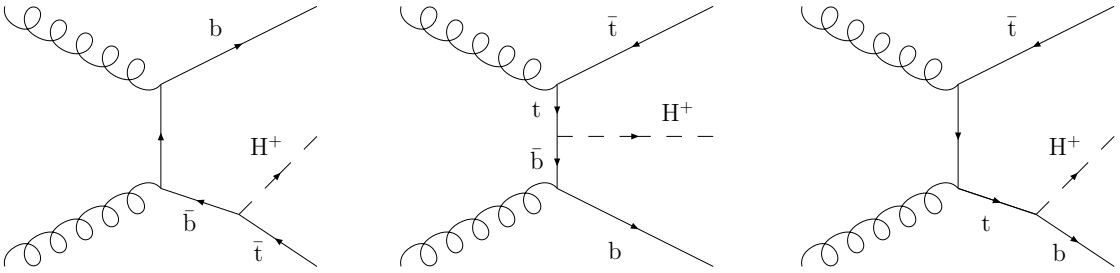
Table 3.4: Exclusive cross sections for $t\bar{t} + n\,\text{jets}$ final states, obtained after ME/PS matching between LO ALPGEN and PYTHIA.

3.4.3 Charged Higgs-boson production

As discussed in Section 1.2.4, charged Higgs bosons have not been observed yet. In this thesis charged Higgs-boson identification at the LHC is studied in the $H^{\pm} \to tb$ decay channel. In the mass range above the kinematic threshold for this decay, a charged Higgs boson can be produced either by quark annihilation or through gluon fusion. The quark-initiated s-channel creation $q\bar{q}' \to H^{\pm}$ with a cross section of about 60 fb for $m_{H^{\pm}} = 311\,\text{GeV}/c^2$ is expected to be overwhelmed by backgrounds from Standard Model processes in the subsequent $H^{\pm} \to tb$ decay. The charged Higgs boson can also be pair produced in a Drell–Yan-like interaction, but only with a very low cross section of about 4 fb for $m_{H^{\pm}} = 311\,\text{GeV}/c^2$.

The gluon-initiated production mode produces charged Higgs bosons in association with a top quark and a bottom quark. In Figure 3.15 the contributing diagrams of the gluon-initiated hard interaction are shown. Although these diagrams are not gauge invariant separately [118], the diagrams 3.15(a) and 3.15(b) dominate for a charged Higgs-boson mass well above the top-quark mass, and can be approximated by the $2 \to 2$ process $gb \to tH^{\pm}$ if the inclusive $tH^{\pm} + X$ final state is considered. For a charged Higgs-boson mass sufficiently below the top-quark mass, the outgoing top and anti-top quark in diagram 3.15(c) are on the mass shell. This diagram is dominant in this mass range, and charged Higgs-boson production can be described by on-shell top-quark pair production followed by the $t \to H^{\pm}b$ decay. In the intermediate range $m_{H^{\pm}} \approx m_t$ a correct description of the charged Higgs-boson production requires all three diagrams in Figure 3.15 to be taken into account.

For the studies presented in Chapter 7 only the gluon-initiated production modes are considered. The spectator b quark, remaining from the initial gluon splitting, is expected to be emitted along the beam line. Charged Higgs-boson searches in this channel can either be performed inclusively, looking for the $H^{\pm}$ and the associated top



(a) (b) (c)

Figure 3.15: Leading-order Feynman diagrams of gluon-initiated charged Higgs-boson production.

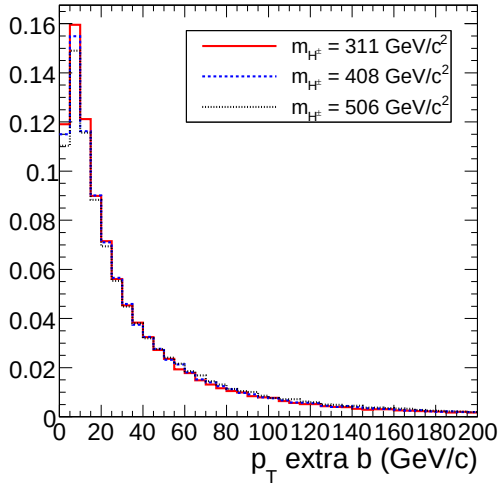


Figure 3.16: Distribution of the p_{T} of the spectator b quark in $\mathrm{pp} \to \mathrm{tbH}^{\pm}$ charged Higgs-boson production.

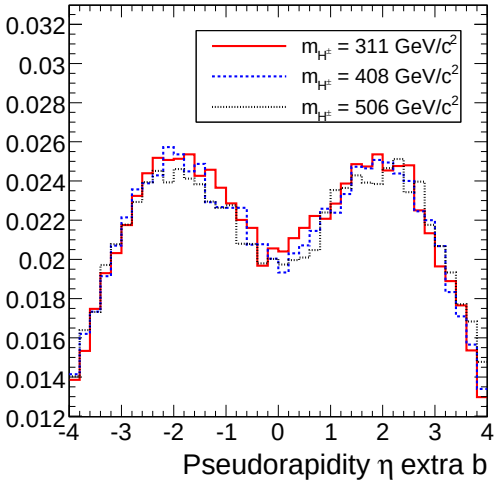


Figure 3.17: Distribution of the pseudorapidity η of the spectator b quark in $\mathrm{pp} \to \mathrm{tbH}^{\pm}$ charged Higgs-boson production.

quark, or exclusively, resolving in addition the forward spectator b. In Figures 3.16 and 3.17, respectively the p_{T} and η distributions are shown for the spectator b quark in $\mathrm{pp} \to \mathrm{tbH}^{\pm}$ charged Higgs-boson production for $m_{\mathrm{H}^{\pm}} = 311$, 408 and 506 GeV/c^2. These forward and soft spectra can be contrasted to the corresponding hard and central p_{T} and η distributions for the b quark originating from the charged Higgs-boson decay. The latter distributions are shown respectively in Figures 3.18 and 3.19 for the same masses $m_{\mathrm{H}^{\pm}} = 311$, 408 and 506 GeV/c^2.

For a description of the inclusive channel at leading order, the cross section for the charged Higgs-boson production should be considered in the $2 \to 2$

$$\mathrm{gb} \to \mathrm{tH}^{\pm} \tag{3.22}$$

channel. The initial b quark is taken as a massless parton from the corresponding parton density in the proton. As discussed in Section 3.1.2, this description sums up large logarithms that arise from long-distance interactions where the spectator b

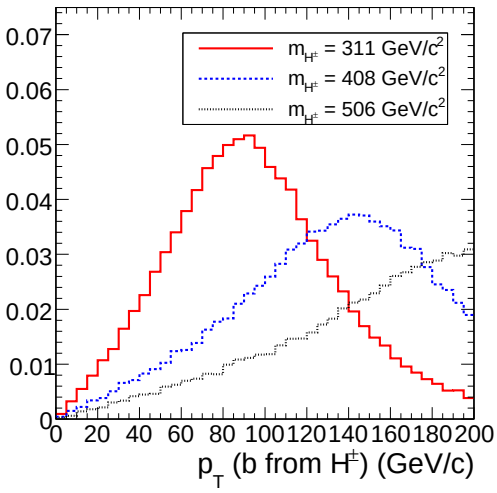


Figure 3.18: Distribution of the p_{T} of the b quark from the $\mathrm{H}^{\pm}$ decay in $\mathrm{pp} \rightarrow \mathrm{tbH}^{\pm}$ charged Higgs-boson production.

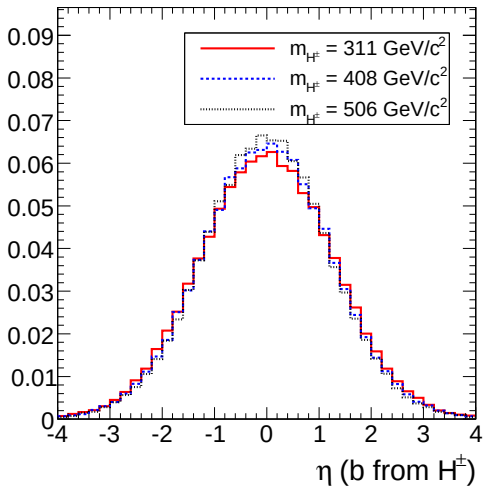


Figure 3.19: Distribution of the pseudorapidity η of the b quark from the $\mathrm{H}^{\pm}$ decay in $\mathrm{pp} \rightarrow \mathrm{tbH}^{\pm}$ charged Higgs-boson production.

quark has small transverse momentum. When on the other hand charged Higgs-boson production is studied in the exclusive mode, the $2 \rightarrow 3$

$$\mathrm{gg} \rightarrow \mathrm{tbH}^{\pm} \tag{3.23}$$

channel should be used at leading order, to appropriately describe the kinematics of the spectator b quark, which is taken to be massive in the calculations. In this case the spectator b quark is produced via gluon splitting in the matrix element.

In a next-to-leading order description both processes (3.22) and (3.23) are to be properly combined. In the `PYTHIA` event generator that is used to simulate charged Higgs-boson production in this thesis, such a treatment at NLO is not possible. Recent developments allow to include a matching between both processes [119], but this is not considered here. The theoretical calculation of the total cross section has been performed at NLO [120], starting from the process (3.22). This calculation contains the process (3.23) as one of the NLO corrections. When calculating the cross section for both processes to all orders one should obtain the same result, as they both describe the same physics. Therefore processes (3.22) and (3.23) have been used to generate the inclusive and exclusive channels respectively, but the cross sections used are in each case rescaled to the NLO theoretical calculation for $\mathrm{pp} \rightarrow \mathrm{tH}^{\pm}\mathrm{X}$.

In Section 1.2.2 and 1.2.3 the Higgs sector in a general 2HDM and in the MSSM is described respectively with six or two free parameters. Charged Higgs-boson production, however, is in both cases only sensitive to the two parameters $\tan\beta$ and $m_{\mathrm{H}^{\pm}}$ at tree level. Since the $\bar{\mathrm{t}}\mathrm{bH}^{+}$ Yukawa term in the Lagrangian can be written as

$$\frac{g_2 V_{\mathrm{tb}}}{\sqrt{2}\, m_{\mathrm{W}^{\pm}}} \mathrm{H}^{+} (m_{\mathrm{t}} \cot\beta\, \bar{\mathrm{t}}\, \mathrm{b}_L + m_{\mathrm{b}} \tan\beta\, \bar{\mathrm{t}}\, \mathrm{b}_R), \tag{3.24}$$

the cross section is enhanced at small and large values of $\tan\beta$, with a minimum at $\tan\beta = \sqrt{m_{\mathrm{t}}/m_{\mathrm{b}}} \approx 6$. The cross section decreases rapidly with rising $m_{\mathrm{H}^{\pm}}$. Typically,

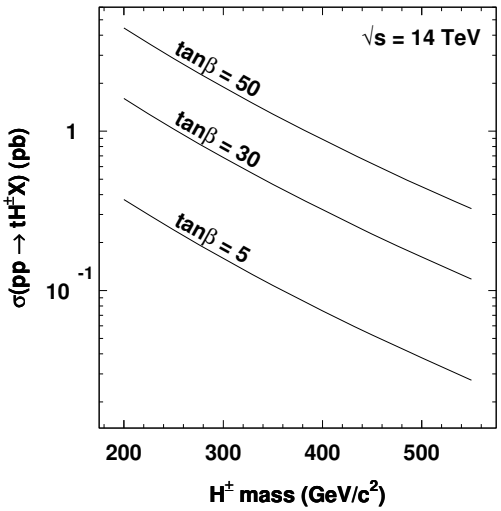


Figure 3.20: NLO cross section for pp → $tH^{\pm}X$ as a function of $m_{H^{\pm}}$.

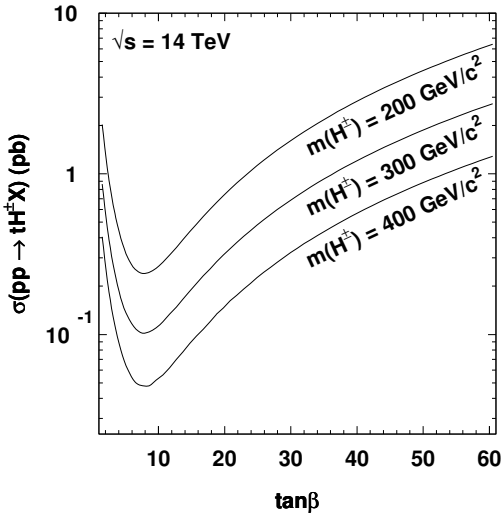


Figure 3.21: NLO cross section for pp → $tH^{\pm}X$ as a function of $\tan\beta$.

it decreases an order of magnitude as $m_{H^{\pm}}$ increases from $250\,\mathrm{GeV}/c^2$ to $500\,\mathrm{GeV}/c^2$, and when going from $\tan\beta = 30$ down to the minimum $\tan\beta \approx 6$. These cross-section dependencies on $\tan\beta$ and $m_{H^{\pm}}$ are shown in Figures 3.20 and 3.21 for the NLO pp → $tH^{\pm}X$ cross section. As a reference, a production cross section of 1 pb corresponds to about 17 events per day when running constantly under low-luminosity conditions.

Chapter 4

CMS Detector, Simulation, and Data Handling

The exploration of the rich and diverse physics opportunities at the LHC requires large-scale particle detectors of great complexity, in order to be able to reconstruct the proton collisions in as much detail as possible. The design of such detectors must find a balance between a multitude of competing requirements. The general purpose Compact Muon Solenoid detector has been designed to be simple, with a strong magnetic field, yet equipped with powerful and technologically very challenging subdetectors.

In this chapter the design and operation of the CMS detector is outlined. Details on the subdetectors' design and the corresponding low-level reconstruction in Section 4.1 are followed by a discussion on detector simulation in Section 4.2. The CMS software and computing environments are considered next. Finally, in Section 4.3, the online selection procedures in the CMS trigger system are briefly examined. Throughout the chapter, emphasis is put on the physics building blocks, computing tools and event samples needed to perform the high-level object and event reconstruction used in the following chapters.

4.1 The Compact Muon Solenoid

The design of the CMS detector is geared by the envisaged physics program, as introduced in Chapter 2. For the CMS experiment a compact design with a strong central magnet has been chosen, with the goal to retain a resolution of 10% on muon momenta measurements at $\approx 1\,\text{TeV}$. The superconducting solenoid, 13 m long and with 5.9 m inner diameter, will create a magnetic field along the LHC beam line of 4 T. With the nominal current of 19.5 kA it will store 2.7 GJ of energy in its field. The diameter of the magnet's coil has been chosen large enough such that inner tracker and calorimetry systems could be accomodated inside. On the outside an iron return yoke closing the field hosts the CMS muon spectrometer. A detailed overview of the layered layout of the CMS detector is shown in Figure 4.1.

This section discusses the design of the various CMS subdetectors and the reconstruction of low-level objects that serve as building blocks for the global event reconstruction. The focus is on some aspects relevant for the physics studies in this thesis.

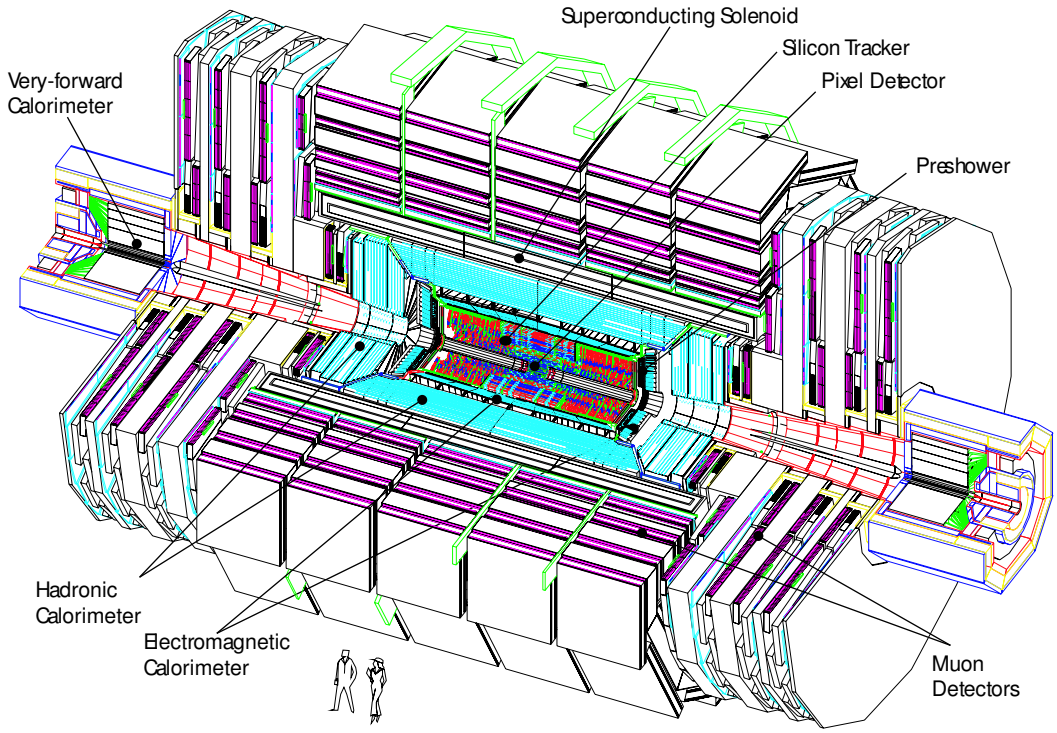


Figure 4.1: The Compact Muon Solenoid (CMS) detector layout.

4.1.1 The CMS tracker

The CMS tracking system is designed to balance many competing requirements [121, 122]. One of the important constraints comes from the dense track environment expected with the nominal LHC luminosity. Under such conditions, highly efficient track and vertex reconstruction requires a tracking device with very high granularity close to the interaction point. For the CMS experiment, a semi-conductor detection technology has been chosen for the complete tracker. With a radius of 110 cm and a length of 540 cm, in total 200 m^2 of silicon sensors will give tracking coverage up to $|\eta| < 2.5$.

In Figure 4.2 an (r, z) view is given of a fourth of the tracker, showing the constituent detector components. Close to the interaction point three barrel layers of pixel detectors and two endcap pixel disks are foreseen. They contain a total of 66 million pixels, with a size of 100 μm $\times$ 150 μm, yielding a single point resolution of 10 μm in (r, ϕ) and 20 μm in z. Further from the beamline silicon strip detectors are used in the Silicon Strip Tracker. The barrel part constitutes of the Tracker Inner Barrel with four layers of silicon strips, and the Tracker Outer Barrel with six longer layers. A few layers with "stereo" modules provide measurement in (r, ϕ) as well as in (r, z) coordinates. The pitch size, ranging between 80 and 180 μm, is chosen larger further from the beam line, where a lower detector occupancy is expected. The resulting resolution in (r, ϕ) ranges between 23 and 52 μm and in z between 230 and 530 μm. The forward region is occupied by nine disks in each Tracker Endcap and three layers of smaller Tracker Inner Disks, placed between the Inner Barrel and the endcaps. The detectors on these disks are arranged in concentric rings, and have strips pointing to the beamline, with variable

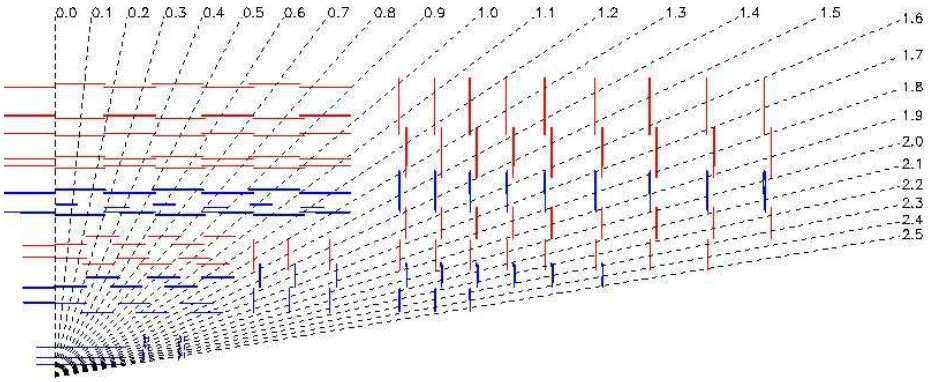


Figure 4.2: Layout of the detector components of 1/4th of the CMS tracker in (r, z) view. The dashed lines correspond to directions of the indicated pseudorapidity.

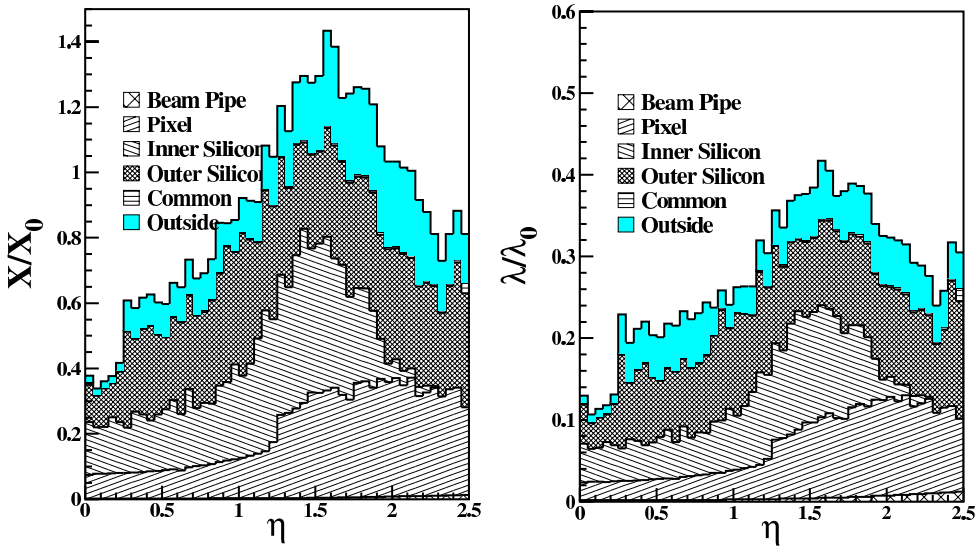


Figure 4.3: Tracker material budget as a function of pseudorapidity, in units of radiation length (left) and in units of interaction length (right).

pitch and hence variable resolution. Also in the endcaps some rings are equipped with stereo modules. The number of silicon strips in the tracker amounts to 9.6 million in total.

With the chosen design, a low detector occupancy of a few percent is obtained in the different subsystems of the Silicon Strip Tracker, down to 10^{-4} in the pixels. This low occupancy is needed for efficient track reconstruction, both to avoid blurred hit position resolution due to particles traversing a sensitive detector at the same location, and to suppress ambiguities in the association of detector hits to a track. The large tracker detection surface also reflects in a large amount of material for cabling, cooling, support and electronics. In Figure 4.3 the material budget is shown as a function of η, in units of radiation length and interaction length. Track finding hence needs to take possible interaction of particles within the tracker material into account.

4.1.2 Track and vertex reconstruction

Track reconstruction can be decomposed into hit reconstruction, seed generation, trajectory building, ambiguity resolution and the final track fit. For the pixel hits the simulated reconstruction inefficiency is expected to be below 0.5%, with less than 0.01% ghost hits [123]. Due to this efficient hit reconstruction and the low occupancy, the pixel detector is particularly useful for track seeding. Alternative seed generation methods using hits in the silicon strips, although less efficient, will be employed at LHC startup when the pixels will not yet be included into the CMS detector. To build track seeds from the hits, at least three hits, or two hits with a beam constraint, must be combined into a five-parameter helix description of a track. The parameters and covariance matrix of a track seed are estimated from the hit position and its estimated uncertainties.

Trajectory building starts from a set of track seeds formed from combinations of hits. For each seed the trajectory is extrapolated along a helix to the first compatible detector layer, taking into account multiple scattering and energy loss in the traversed material. For each compatible hit found, a trajectory candidate is created, adding one trajectory with a fake hit to account for cases where the track did not leave a hit on the considered layer. Each trajectory is then updated with the corresponding hit according to the combinatorial Kalman filter formalism [124]. This procedure is then repeated until the outermost layer of the tracker is reached. For each extra measurement included, the parameters of the trajectory are further constrained. During the trajectory-building stage, conditions are needed to avoid the number of trajectories to grow exponentially. Trajectories are discarded as soon as they no longer satisfy the imposed conditions, like on the number of fake hits and the fit's normalized χ^2. These quality conditions are tunable, depending on timing constraints.

After the trajectory-building process, ambiguities must be resolved to avoid double counting of tracks. These ambiguities can arise when more than one seed is associated to a trajectory, or when more than one valid trajectory is built from the same seed. Trajectories sharing too many hits are discarded, keeping the highest-quality trajectory. To avoid biases on the track parameters and covariance matrix due to constraints applied during the trajectory building, all valid tracks are refitted after the trajectory building. To obtain an optimal estimation of the track parameters and uncertainties at the innermost and outermost states, a new fit away from the beamline is complemented by a smoothing fit from the exterior towards the beamline.

Simulation results of track-reconstruction efficiency, fake rate and parameter resolutions are summarized in [85, 124]. For isolated muons about 99% reconstruction efficiency is obtained, except in the region $\eta \approx 0$ and for $\eta > 2$. For tracks in b-flavoured jets with $120 < p_{\mathrm{T,jet}} < 170\,\mathrm{GeV}/c$, 90% tracking efficiency is obtained centrally, down to 65% in the forward region, with fake-track probability below the per-mille level. The transverse-momentum resolution on isolated muons with p_{T} up to $100\,\mathrm{GeV}/c$ is at the percent level centrally, growing to about 7% in the most forward region. Tracks with too low transverse momentum are so strongly curved by the magnetic field that they do not reach the outer tracker layers but rather spiral towards the forward detectors. Reconstruction of these tracks is therefore very inefficient. In Figure 4.4 the average number of generated charged particles and reconstructed tracks within the tracker ac-

ceptance is shown versus their p_{T} per minimum-bias event. The cut-off due to the magnetic field is indeed revealed at about $0.8\,\mathrm{GeV}/c$. The relative ratio of both curves deviates from unity due to an interplay of several effects, like track-reconstruction inefficiencies, additional tracks from decays, interactions in the tracker material and fakes, etc.

The reconstruction of the interaction vertex, relevant for resolving several simultaneous pile-up collisions, is performed in two steps. First reconstructed tracks are grouped into vertex candidates. The vertex-finding algorithm used for this track grouping can be very different depending on the physics goal, for example for primary or secondary vertex finding. Secondly, vertices are fit in order to obtain a best estimate of the vertex' kinematic parameters.

For fast primary-vertex finding purposes, pixel tracks from pixel-hit triplets can be used to find efficiently the vertex positions along the beamline. In offline analyses, where timing is not an issue, fully reconstructed tracks are used to provide precise estimations of vertex positions and the corresponding covariance matrices derived from the list of associated tracks. The default CMS primary-vertex finder [125] starts from all reconstructed tracks in an event. First, a preselection is performed on the tracks' distance of closest approach to the beamline, which is shown in Figure 4.5 for reconstructed tracks in three classes of transverse momentum in $\mathrm{t\bar{t}}$ events with low-luminosity pile-up conditions. The long tails in the distributions are due to tracks from secondary decay vertices. The significance of the transverse impact parameter of tracks, defined as the ratio of the impact parameter and its uncertainty, is required to be smaller than 3. After this selection the tracks are grouped according to the mutual z separation at their point of closest approach. The primary-vertex candidates are then fitted with a least-squares based Kalman fitter, discarding incompatible tracks. Poor fits and vertices incompatible with the beam line are excluded. The primary vertex of the hard event is finally taken as the first in the list of vertices, ordered with decreasing $\sum p_{\mathrm{T}}^2$ of the associated tracks.

In the signals considered in the analyses in Chapter 6 and 7, the many tracks with high transverse momentum originating from the hard interaction make primary-vertex finding almost fully efficient ($\sim 99\%$ in top-quark pair events). The resolution on the primary-vertex position in $\mathrm{t\bar{t}}$ events is about $13\,\mu\mathrm{m}$ in x and y and $18\,\mu\mathrm{m}$ in the z direction.

To distinguish between tracks from the hard interaction and from pile-up collisions, a custom longitudinal track association criterion is defined:

- if no primary vertex is reconstructed, no track association is performed;
- if only one primary vertex is found, a track is considered associated to this vertex if the distance $d_z(\mathrm{Tr}, \mathrm{PV})$ in z between the track's point of closest approach to the beamline and the fitted vertex is smaller than 2 mm;
- if more than one primary vertex is found, the same absolute cut of 2 mm on the track–vertex distance $d_z(\mathrm{Tr}, \mathrm{PV})$ is required for the association of the track to the vertex. In addition, to resolve ambiguities with other primary vertices within 2 mm in z distance, the distance $d_z(\mathrm{Tr}, \mathrm{PV})$ needs to be smaller than 60% of the

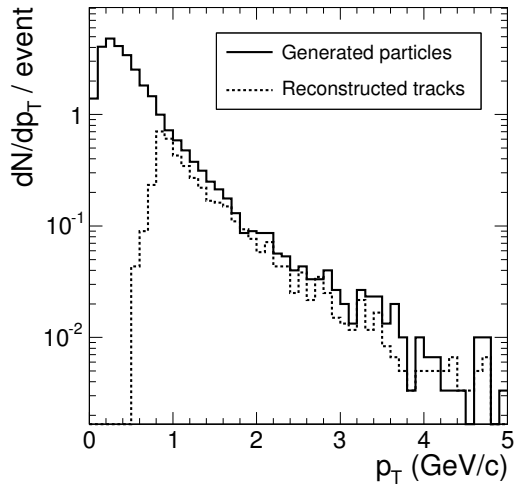


Figure 4.4: Average number of generated charged particles and reconstructed tracks versus p_{T} in minimum-bias events.

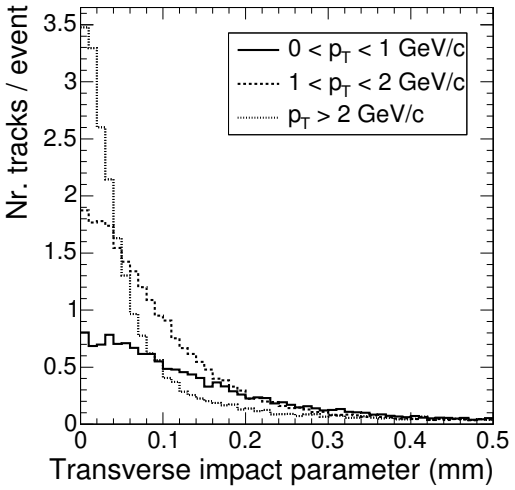


Figure 4.5: Distance of closest approach to the beamline for reconstructed tracks in top-quark pair events with low-luminosity pile-up conditions.

minimal distance $d_z(\mathrm{PV}, \mathrm{PV}')$ between the hard interaction vertex PV and other primary vertices PV$'$ at the same side of the hard vertex.

In Figure 4.6 the distance $d_z(\mathrm{Tr}, \mathrm{PV})$ is shown for top-quark pair events with low-luminosity pile-up, with exactly one reconstructed primary vertex. Figure 4.7 shows this distance relative to the distance to the closest primary vertex in case more than one primary vertex is reconstructed. On both figures the thresholds used for track association are illustrated.

The excellent accuracy of reconstructed track parameters in CMS is reflected in the possibility to find secondary vertices in jets with high efficiency and purity. Such secondary vertices are a discriminating signature for heavy-flavour jets, as discussed in Section 3.3.4. A particular difficulty in secondary-vertex finding comes from the large probability of b-c cascades in heavy-flavour jets, giving rise to a tertiary vertex. If tracks from a tertiary vertex are also used in the fit of a secondary vertex, the measured flight distance is biased to a higher value.

The default secondary-vertex reconstruction algorithm starts by finding and fitting a vertex from a given collection of input tracks as described for the primary vertex. Incompatible tracks are then fed into subsequent iterations. As most vertex finders are sensitive to both primary and secondary or tertiary vertices, a vertex filter is needed after vertex fitting. In the default implementation, the closest transverse distance from the beam line has to exceed $100\,\mu$m (see Figure 4.5), but should be smaller than 2 cm to avoid contamination from interactions with the beampipe, first pixel layer, etc. This transverse distance also needs to have a significance larger than 3. Finally, the total invariant mass of the tracks associated to the vertex must be smaller than $6.5\,\mathrm{GeV}/c^2$.

For simulated vertices with a transverse distance from the primary vertex between $100\,\mu$m and 2 cm, an efficiency of 60% for finding the secondary vertex is obtained in b jets, for a purity of 90% expressing the fraction of reconstructed secondary vertices

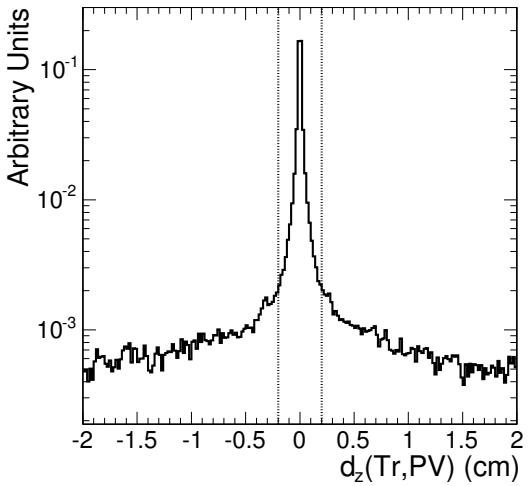


Figure 4.6: Distance $d_z(\mathrm{Tr}, \mathrm{PV})$ in z between the track's point of closest approach to the beamline and the fitted vertex in $t\bar{t}$ events with low-luminosity pile-up.

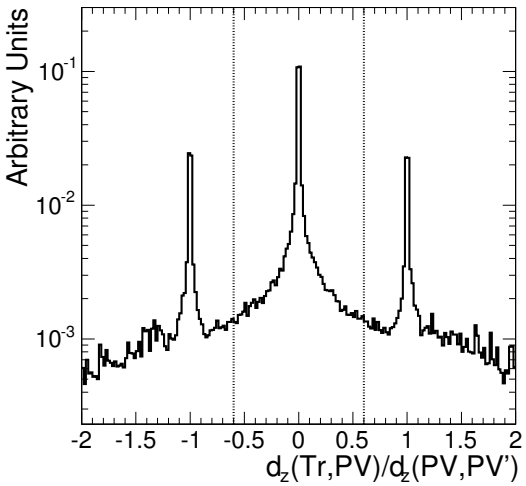


Figure 4.7: Ratio of the track–vertex distance $d_z(\mathrm{Tr}, \mathrm{PV})$ and the minimal vertex–vertex distance $d_z(\mathrm{PV}, \mathrm{PV}')$ at the same side of the vertex in $t\bar{t}$ events with low-luminosity pile-up.

well associated with the simulated one. For c jets an efficiency of 20% is reached with a purity of 65%. More details on the performance of secondary-vertex finding in the framework of flavour identification of jets can be found in [126].

The influence of the misalignment of the different substructures of the tracking system on the vertex reconstruction performance is studied in [125, 127] for various alignment scenarios defined in [85]. In the event simulations performed for this thesis a perfect alignment is assumed.

4.1.3 The CMS calorimeter system

The CMS experiment is equipped with two calorimeters to perform energy measurements of final-state particles leaving the tracking volume. The electromagnetic calorimeter (ECAL) is placed just outside the tracker, and aims to measure the energy of electrons and photons [128]. The hadronic calorimeter (HCAL) surrounds the ECAL and is designed to measure charged- and neutral-hadron energies [129]. Both systems infer the particle's energy from the energy contained in the particle shower, that is caused by the inelastic interaction of the incident particle with the calorimeter material.

Lead tungstate ($PbWO_4$) scintillating crystals have been chosen as the detection material of the ECAL, because of the short radiation length, the high scintillation speed and the radiation hardness. This choice allows for a compact calorimeter with high granularity. In the barrel region ($|\eta| < 1.479$) the crystals are tilted such that they almost point towards the nominal vertex position. A 3° displacement in both θ and ϕ reduces the longitudinal energy loss in the cracks. The crystals have a front-face cross section covering 0.0174 (22 mm) both in $\Delta\phi$ and $\Delta\eta$. In the endcaps ($1.479 < |\eta| < 3.0$)

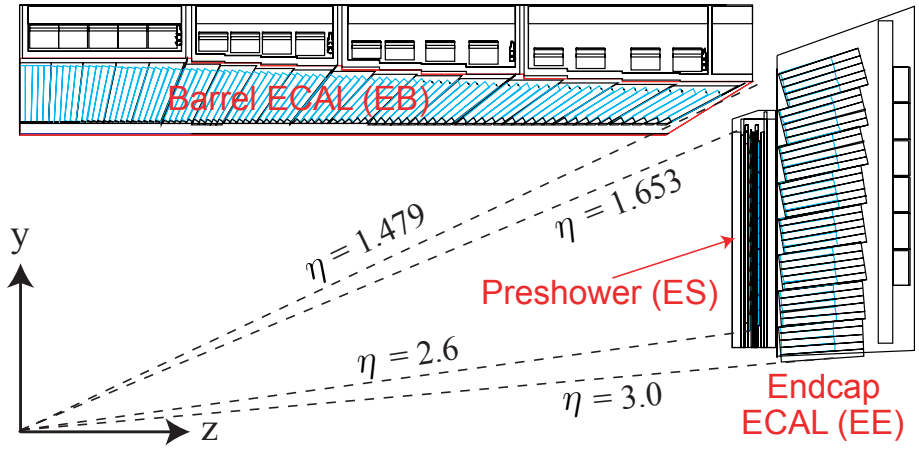


Figure 4.8: View in (r, z) of a quarter of the ECAL layout.

the crystals also off-point slightly from the nominal vertex position, and are arranged 5-by-5 in an x-y grid, rather than an η-ϕ configuration. In this forward region the crystals measure $28.6 \times 28.6\,\mathrm{mm}^2$ at the front face. Both barrel and endcap crystals have a length corresponding to about 25 radiation lengths. In front of most of the endcap range a preshower device is placed, consisting of two layers of lead absorbers and silicon strip detectors. Its primary goal is the identification of neutral pions faking a photon signal, but also improves the identification of electrons and the position determination of both electrons and photons due to its superior granularity. The layout of a quarter of the ECAL is shown in an (r, z) view in Figure 4.8. The intrinsic energy resolution measured on ECAL crystals in test beams is $\sigma_E/E = 2.8\%/\sqrt{E/\mathrm{GeV}} \oplus 0.26\%$ with a negligible noise term. An intercalibration precision of 0.6% between the barrel crystals in the CMS detector environment is envisaged with $10\,\mathrm{fb}^{-1}$ [130].

The barrel and endcap hadron calorimeters completely surround the ECAL and are as well immersed in the magnetic field from the solenoid, maximizing the amount of interaction lengths contained inside the coil. They are hermetically joined, with the barrel part extending out to $|\eta| = 1.4$ and the endcaps covering the range $1.3 < |\eta| < 3.0$. The HCAL design of the barrel and the endcaps consists of respectively 17 and 19 active layers of plastic scintillator tiles, interspersed between brass absorber plates. Brass has been chosen as main absorber because of its relatively short interaction length, its ease of use and because it is non-magnetic. For the barrel HCAL stainless steel is used for the innermost and outermost absorber plates to provide additional structural strength. The sampling of absorber material and active detector layers ensures that sufficient interaction lengths can be reached in a compact and affordable design. The segmentation of the HCAL into calorimeter towers is shown in Figure 4.9 in an (r, z) view. Up to tower 20 ($|\eta| < 1.740$), the tower size is $\Delta\eta = 0.087$, with $\Delta\phi = 5°$. Further forward the tower size in η increases, and the azimuthal segmentation becomes $\Delta\phi = 10°$. The performance of the barrel and endcap HCAL with single-particle test beams is detailed in [131]. The resolutions relevant for the analyses in this thesis are those on the measurement of jet energies and missing energy. These are discussed in Section 5.1.4.

The HCAL barrel and endcap systems are complemented with two extra subdetec-

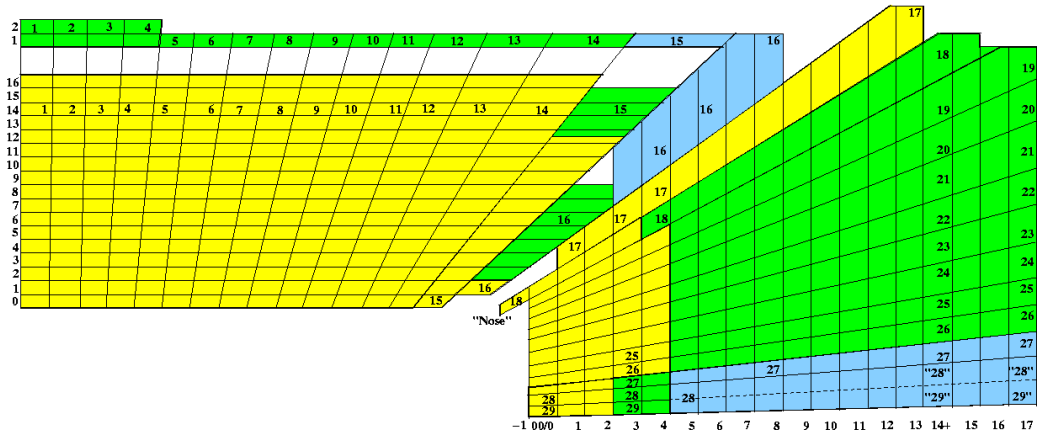


Figure 4.9: View in (r, z) of a quarter of the HCAL tower layout.

tors. In addition to the described barrel HCAL, an outer barrel calorimeter will be placed outside the magnet coil, covering $|\eta| < 1.26$. It consists of iron absorber material instrumented with scintillators, closely following the segmentation of the barrel muon system. Its purpose is to sample the energy leakage of penetrating hadron showers out of the rear of the calorimeter, hence reducing the tails in the hadron energy resolution function and missing transverse-energy measurements. Coverage for high pseudorapidities ($3.0 < |\eta| < 5.0$) is provided by the forward hadron calorimeter. The choice of steel as an absorber material in this detector leads to narrow and short showers, which is particularly useful in the dense environment of the very forward region. Quartz fibres parallel to the beam line act as Čerenkov detectors, and provide a segmentation into 13 towers of $\Delta\eta \approx 0.175$ and $\Delta\phi = 10°$, except for the towers at lowest and highest pseudorapidity, which are respectively less and more coarsely segmented.

4.1.4 Energy reconstruction in the calorimeters

Energy deposits in the ECAL due to electromagnetic electron and photon showers are spread over several crystals. Approximately 94% of the incident energy of a single electron or photon is contained in a combination of 3×3 crystals, and 97% in a 5×5 crystal array around the point of impact. The presence of material in front of the calorimeter, however, can result in bremsstrahlung and photon conversions in the tracker. The magnetic field spreads the energy of such interacting electrons and photons in superclusters extended in ϕ, and therefore simple rectangular energy clustering is not optimal for reconstruction of incident particle energies.

Two methods are currently employed for ECAL energy clustering into superclusters [132]. The Hybrid algorithm is used in the barrel, exploiting the η-ϕ arrangement of the crystals. In a fixed width in η of 3 or 5 crystals, a dynamical search for separated energy is performed in the ϕ direction. The Island algorithm, mostly used for energy clustering in the endcaps, is more flexible. This method first selects seeds by searching for crystals above an energy threshold. Around the seeds a search in ϕ and then η clusters the crystals with decreasing energy. These clusters can then in turn be combined into superclusters, grouping nearby clusters around the most energetic cluster in a narrow η and wide ϕ window.

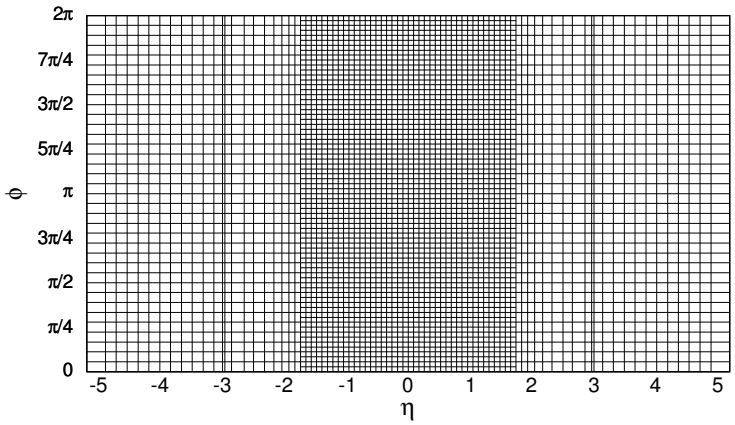


Figure 4.10: Projection in (η, ϕ) of the combined calorimeter-tower layout, including the forward hadron calorimeter ($|\eta| > 3$).

Energy measurements of electromagnetic showers are subject to several effects that change the collected cluster energies, and that can be corrected for using simulations and test-beam data. Variations are expected as a function of the impact position of the incident particle. In addition, showers close to the cracks induced by the ECAL support structures are expected to exhibit large rear leakage as the effective calorimeter depth is considerably reduced. A third large effect to be corrected is the already described energy spread due to showering from interaction with the tracker material.

For the description of jets, causing energy deposits within both hadronic and electromagnetic components, measurements from the ECAL and HCAL need to be combined. Since the ECAL segmentation in crystals is much finer than the HCAL tower sizes, hadronic showers are described in calorimeter towers (ECAL plus HCAL), that are formed by addition of signals in (η, ϕ) bins corresponding to individual HCAL cells. In total there are 4176 such towers, including the forward hadron calorimeter. In Figure 4.10 the pattern of these towers is shown, unfolded into (η, ϕ).

The reconstruction performance of physics objects like jets, built from calorimeter towers, depends strongly on the way the various sources of calorimeter noise are treated in the reconstruction of these lower-level energy towers. Regular calorimeter noise can be suppressed by applying an energy cut on all towers. In addition, soft deposits will be made by particles from pile-up collisions and the underlying event that are energetic enough to reach the calorimeter. One minimum-bias collision generates on average about 850 GeV of energy within $|\eta| < 3$ and about 480 GeV within the tracker acceptance $|\eta| < 2.4$. Although these deposits arise from physical sources, they are unwanted and are expected to be efficiently suppressed by a cut on the transverse energy of the towers. Several schemes of noise suppression are available for calorimeter-tower reconstruction in CMS, either with only a fixed cut $E_{\mathrm{T}} > 0.5\,\mathrm{GeV}$, either combined with a fixed energy cut $E > 0.8\,\mathrm{GeV}$, or more complex schemes taking into account the variable tower sizes. In Figure 4.11 the generated energy per minimum-bias collision is shown differentially as a function of η, for three calorimeter threshold schemes applied on the particles' energies. Figure 4.12 shows the corresponding average reconstructed calorimeter energy per minimum-bias collision for the same threshold schemes, applied

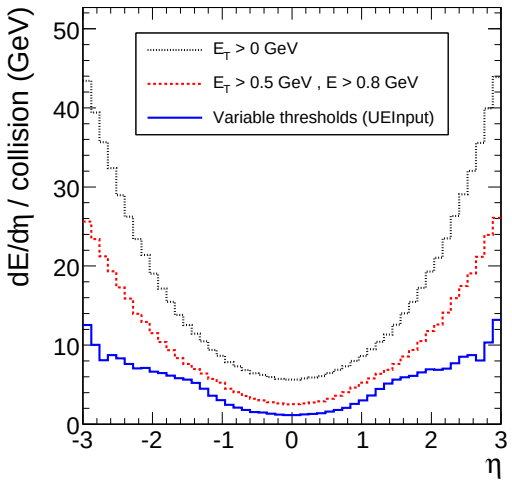


Figure 4.11: Average generated energy flux as a function of η from minimum-bias collisions for three threshold schemes.

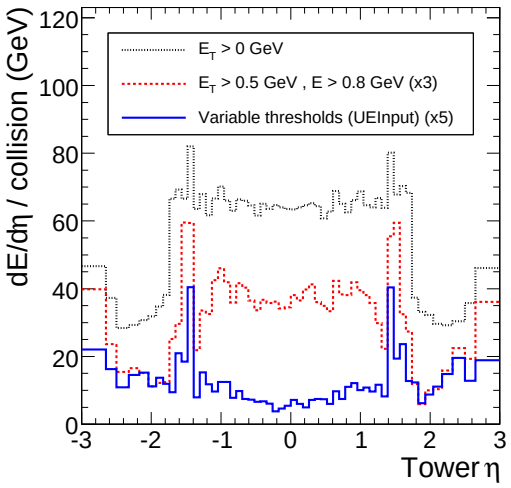


Figure 4.12: Average reconstructed energy flux as a function of η from minimum-bias collisions for three threshold schemes.

on the towers' energies. The first scheme, $E_{\mathrm{T}} > 0\,\mathrm{GeV}$, suppresses nothing in addition to the online zero-suppression thresholds. With $E_{\mathrm{T}} > 0.5\,\mathrm{GeV}$ and $E > 0.8\,\mathrm{GeV}$ both noise and minimum-bias suppression are aimed for. The underlying-event input scheme consists of tower-specific energy thresholds, that increase in the forward direction, and has been constructed in the framework of the CMS $\mathrm{t\bar{t}H}$ analyses [133]. The sharp rise in energy per tower at $|\eta| \sim 1.4$ is attributed to the increase of calorimeter cell sizes in the HCAL endcaps, while the drop at $|\eta| \sim 1.7$ is due to the doubling of the tower sizes in ϕ. Even though small deposits from minimum-bias events are efficiently suppressed, still sizable amounts of energy from minimum-bias pile-up collisions will be superimposed on the hard collisions. Effects on the high-level event reconstruction due to deviations from the assumed nominal underlying event and pile-up contributions will further be accounted for as systematic uncertainties.

4.1.5 The CMS muon spectrometer

The CMS muon system provides full geometric coverage for muon identification up to $|\eta| = 2.4$. The detectors are embedded in the iron construction of the magnet's return yoke, such that muon momentum and charge measurements can also exploit the strong magnetic return field. This is particularly important for muons with transverse momentum in the TeV range, for which the complementary tracker measurements degrade and reach resolutions comparable to the momentum measurements from the muon system only. In analogy to the previously described subdetectors, the CMS muon spectrometer is composed of a barrel part ($|\eta| < 1.2$) and a forward region ($0.9 < |\eta| < 2.4$). The barrel consists of five wheels, in which drift-tube (DT) detectors and resistive-plate chambers (RPC) are placed in concentric muon stations around the beam line, with radius between about 4 m and 7 m. The detector modules are arranged with respect to each other, such that a muon traverses at least three of the

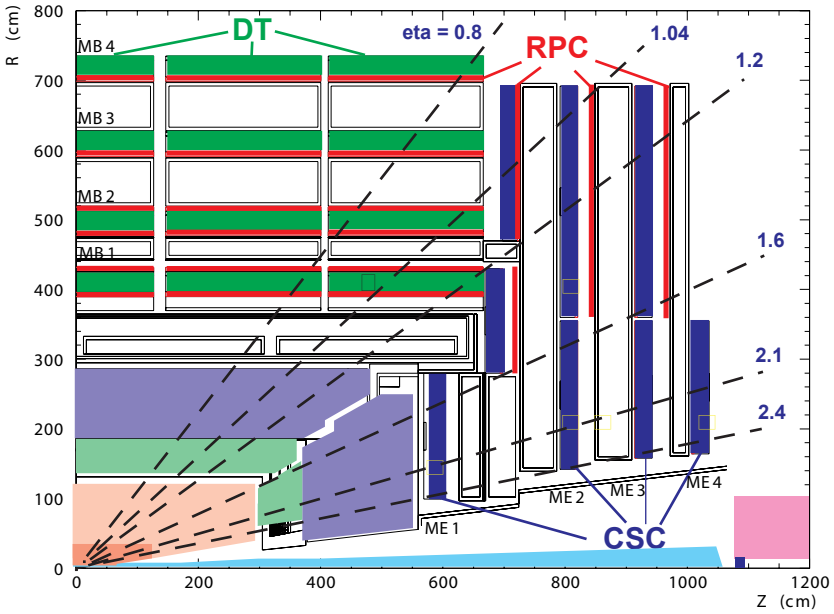


Figure 4.13: View in (r, z) of the layout of a quarter of the CMS muon system at the start-up of the experiment.

four detection layers. In the forward region cathode-strip chambers (CSC) and RPC's are mounted perpendicular to the beamline in overlapping rings on the endcaps. In Figure 4.13 an (r, z) view is given of the different parts of a quarter of the CMS muon system.

The CMS muon system uses three types of gaseous detectors. The choice of detector technologies has been driven by the need for fast triggers, excellent resolution, coverage of a very large total surface, operation in dense radiation environments, etc.

Drift tubes In the barrel range the muon rate and background rate from neutrons is expected to be low. Also the residual magnetic field will be low in the gaps of the return yoke. This allows for the use of drift tubes. The CMS DT's are long aluminium cells of a few centimetres wide, filled with gas, and with an anode wire in the centre that collects ionization charges when a charged particle traverses the tube. In a DT chamber many of these drift tube cells are arranged in three (two in the outermost muon station) superlayers, consisting each of four layers of drift tubes that are shifted by half the width of a cell with respect to each other. Two of these superlayers have anode wires parallel to the beam line, providing a measurement of the r and ϕ coordinates; the third superlayer is placed perpendicular between the others, and provides the z-coordinate measurement. Each station is designed to measure muon positions with better than 100 μm precision and the direction in ϕ with about 1 mrad resolution.

Cathode-strip chambers The CMS cathode-strip chambers each consist of six gas layers, each gas gap having radial cathode strips and a plane of anode wires running almost perpendicular to the strips. The gas ionization and subsequent electron avalanche caused by a charged particle traversing a layer, produces a fast

charge deposit on the anode wire and a slower image charge on a group of cathode strips. Hence each of the six layers in a CSC provides a position measurement of all r, ϕ and z coordinates. The best spatial resolution is provided by the centre of the charge distribution on the strips, and is typically about 200 μm; the resolution in ϕ is of the order of 1 mrad.

Resistive-plate chambers The RPC detectors in CMS consist of double-gap Bakelite chambers, with the 2 mm spacings filled with gas. The position resolution from the RPC's is coarser than for the DT's and CSC's, but the collection of charges on the strips is very fast, with 3 ns time resolution. Therefore these chambers serve mainly in the trigger, where they provide information complementary to the DT's and CSC's. The RPC's can in particular identify unambiguously the correct bunch crossing. In addition they complement the other muon tracking detectors in regions with reduced geometrical coverage, especially in the barrel–endcap overlap region. In the barrel an RPC is placed in front of every DT, and in the first two layers also behind. In the forward region three layers of RPC's are present behind the CSC's, but only up to $\eta = 1.6$. The RPC's at larger rapidities have been staged for the low-luminosity period of the LHC.

4.1.6 Standalone muon reconstruction

Reconstruction of tracks in the muon system can roughly be divided into three steps: determination of the hit position, segment reconstruction, and track finding. The position and segment finding are specific to the considered subdetector, only at the track finding step the information from different detector modules is combined.

The reconstruction of hit positions within a DT cell requires the knowledge of the drift time. Therefore a detailed synchronization mechanism is needed to account for the time-of-flight of the muon from the interaction point to the DT cell, the time for propagation of the signal on the anode wire and the time offset due to the electronics and the cables. With the drift time the drift distance can be inferred from models, depending on the local magnetic field and the particle's incident angle. The hit positions serve as input to the segment reconstruction within one DT chamber. Sets of aligned hits are searched in the r-ϕ and r-z projections independently, considering both left–right hypotheses for each cell. Hit combinations are linearly fitted to form segments, solving conflicts and suppressing ghosts. The hit reconstruction is then updated with information from the segments and the segments themselves are refitted. Finally all combinations of segments from both orthogonal projections are retained.

In the CSC's the hit position is determined from the strips by fitting the charges collected on typically three to five strips to an expected charge-distribution parametrization. This position is then combined with compatible anode hits within two bunch crossings of each other to form a two-dimensional reconstructed hit. The hence obtained hits in all six layers of a CSC are then searched for aligned hits using a linear fit. After quality requirements on the fit the reconstructed local track segments are obtained.

In the RPC's the much less precise hit positions are inferred by clustering adjacent fired strips, and calculating the "centre of gravity" of the clusters. Next, these hit

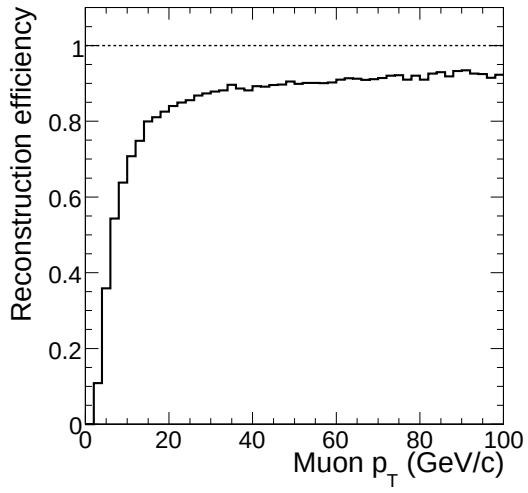


Figure 4.14: Standalone muon reconstruction efficiency in $t\bar{t}$ events as a function of generated muon p_T.

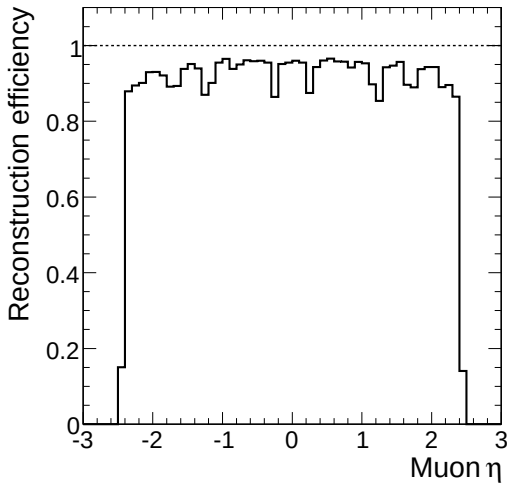


Figure 4.15: Standalone muon reconstruction efficiency in $t\bar{t}$ events as a function of generated muon η.

positions need to be combined with the track segments reconstructed in the DT's and CSC's. The state vectors from the segments on the innermost detection layers are used to seed the muon trajectories, that are built using the combinatorial Kalman filter, introduced in Section 4.1.2. In the barrel the parameters of the reconstructed track segments are used in the Kalman-filter procedure, while in the endcaps the individual three-dimensional CSC hits constituting the segments are directly used because of the inhomogeneous magnetic field. Propagation of the trajectories in the muon system takes into account energy loss in the material, multiple scattering and the non-uniform magnetic field. Trajectory parameters and uncertainties are updated at each inclusion of a new measurement. When the outermost layer is reached, the procedure is repeated in the reverse direction to obtain the best estimation of the track parameters at the innermost station. In the so-called standalone muon reconstruction these trajectories are further extrapolated to the nominal interaction point, where a beam constraint is applied. In Figures 4.14 and 4.15 the reconstruction efficiencies are shown as a function of the generated muon p_T and η, respectively, for muons in top-quark pair events, reconstructed with the standalone muon-reconstruction algorithm. A significant part of the inefficiency is checked to be due to large tails in the muon angular resolution which hamper the identification with the generated truth. The remaining loss is explained by acceptance effects, track quality cuts, the beam constraint, etc.

In general, trajectories in the muon chambers are combined with tracks in the silicon tracker, greatly improving the momentum resolution. More details on this cross-detector global muon reconstruction can be found in Section 5.3.2.

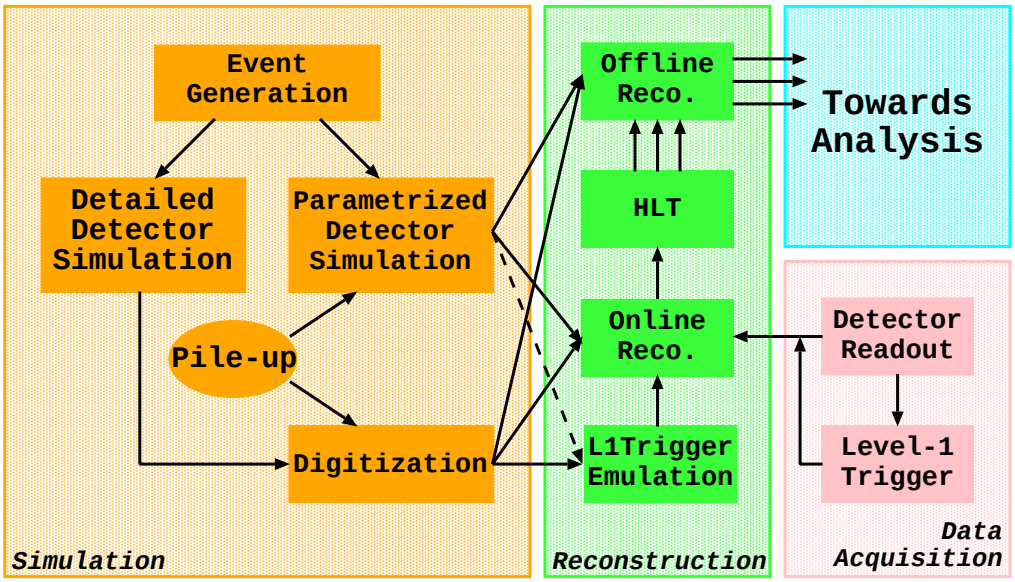


Figure 4.16: Overview of the event data flow for the CMS experiment. Event data are propagated through the various steps following the arrows. The dashed arrow indicates a step which is foreseen but does not exist yet.

4.2 The CMS Software and Computing Environment

The complexities of present-day collider experiments, like CMS at the LHC, translate directly in the demands on the software that is used to process and analyse the produced electronic signals. To profit from the diverse superior detection qualities of the experiment, advanced reconstruction, calibration and detector-alignment techniques need to be conceived and implemented prior to the start-up of the experiment. For the development of such techniques, detailed detector simulations are required for validation of observed test-beam data as well as for reliable estimations of the performance of the detector. These simulations serve also as input for the preparation of the multitude of physics analyses. Once data are being taken, the software will be used to calibrate and align the various detector components. The hence obtained understanding of the detector can then be used to validate the simulation software and ultimately the physics.

In Figure 4.16 the event data flow through the CMS software is visualized. In this section the detector-simulation software is discussed first. Next, the requirements on the software and the resulting implementation within the CMS software framework are detailed. Finally, the deployment of this software framework in the overall CMS computing model is dealt with, emphasizing the use of the CMS software in the context of this thesis. The principles of the crucial online-selection system are treated separately in Section 4.3.

4.2.1 Detailed detector simulation

The detailed CMS detector simulation propagates generated particles from the interaction point through the magnetic field and the CMS detector geometry, allowing for in-flight decays. It describes energy loss in the detector material and simulates in detail electromagnetic and hadronic interactions, yielding simulated hits. In addition it accounts for timing and position issues, which are relevant at the later stage of the detector read-out simulation. For these simulations, a perfect detector geometry is assumed. Expected deviations from this are currently taken into account as systematic uncertainties. Prior to 2004 the detailed detector simulation was performed with the `GEANT3`-based program `CMSIM` [134]. From then on the transition has been made towards the `OSCAR` [135] package, based on the `GEANT4` [136] toolkit and a more flexible geometry description.

The simulation of the interaction of the particles with the traversed detector material is a very time-consuming process. On present-day computers the simulation of a single proton–proton collision for CMS takes several minutes. Due to this large simulation time, the addition on top of the hard interactions of the luminosity-dependent number of pile-up collisions, is performed by mixing these events only after the detector simulation. Moreover, in the simulation it is necessary to account for out-of-time pile-up, coming from bunch crossings before and after the triggered event. The number of crossings considered before and after the nominal one is subdetector dependent, as it is related to the specific response time.

During the mixing, pile-up interactions are randomly chosen from a large sample of minimum-bias collisions, simulated with the same procedure as the hard interaction. Care is taken such that the same sequence of pile-up collisions is never used twice. In addition the sample of minimum-bias collisions is filtered such that none of these interactions pass any trigger thresholds (see Section 4.3). Although this procedure excludes the rare possibility to have two simultaneous collisions firing the trigger, it ensures no low-statistics bias is introduced due to the recycling of the filtered events.

The final step in the detector simulation involves the emulation of the electronic read-out systems of the detector, also referred to as digitization. This process starts from the simulated hit positions and energy losses in the sensitive detectors, and aims to produce sampled signals that mimic the real data-acquisition conditions of CMS.

- In the tracker, energy losses inside the silicon detectors are distributed along a path between the entry and exit points of a track. The resulting charges are then drifted towards the detector surface, taking into account the Lorentz drift and diffusion. Charges on pixels and strips are then integrated, and Gaussian noise is added on all channels that exceed a given threshold, taking into account coupling between channels. The signal is then converted to digital counts using the gain of the detector and the time with respect to the signal bunch crossing.

- In both calorimeters, the simulated energies deposited in the scintillators are converted to photo-electrons, adding noise and folding in the contributions from out-of-time pile-up collisions. In the ECAL, the variation of the light collection with respect to the crystal depth is taken into account when generating the signal pulse as a function of time for each simulated hit.

- Digitization in the muon DT system translates the smeared drift time, inferred from the impact position, muon direction and magnetic field, into the corresponding electronics signal, subtracting an expected average time offset. The CSC signals from the strips and wires are digitized independently, taking into account the ion drift time, background hits from other beam crossings, noise, etc. Due to the small time resolution, the digitization of the RPC signals is simply described by the digital information on the signal strips, adding intrinsic noise.

4.2.2 Parametrized detector simulation

Apart from the detailed detector simulation, a fast, parametrized simulation program, called `FAMOS` [137], has been developed for CMS. It aims for a simulation speed of about one event per second, while maintaining an agreement within 1% with the detailed simulation. The program is conceived to deliver reconstructed objects, like tracks, calorimetry towers and muons, of the same type as the full reconstruction, such that analysis software can be run transparently with both detailed and parametrized simulations.

Also for the fast simulation the generated final state is propagated in the magnetic field and unstable particles are decayed. In the tracker, bremsstrahlung, photon conversion, multiple scattering and energy loss by ionization are simulated. A simplified version of the tracker geometry is used, with uniform pure-silicon cylinders and disks with tuned thickness. To save additional time, no pattern recognition is performed to search for tracks, but hits are rather fit to tracks along the simulated trajectories. Electromagnetic showers in the ECAL are simulated with a dedicated shower parametrization assuming a homogeneous calorimeter. As a second step, the energy distribution is sliced into a small array of crystals. Also in the HCAL the response is simulated using a dedicated parametrization for hadron showers, tuned on the full detector simulation with charged pions. For muons a simple parametrization of the detailed simulation efficiencies and resolutions is used, dependent on the muons' momenta and pseudorapidities.

4.2.3 The CMS software framework

The CMS software framework used for the studies in this thesis consists of a synergy of many CMS-specific and non-specific tools and programs. This includes the physics packages, like event generators and detector-simulation programs, but also many services and tools, like visualization, database services, validation and analysis tools, code management and compilation programs, etc. The main components to perform physics studies have been decided to be written in object-oriented `C++` code on the Linux platform, which has become the standard in modern particle-physics experiments. An important choice made is the integration of all software into a single framework. Abstraction has been used to maintain flexibility, which is needed to deploy the software in versatile environments, like test beams, large-scale Monte-Carlo event production or end-user analysis. In Figure 4.17 a sketch is shown of the main components in the CMS software framework, their dependencies, and the most important external tools.

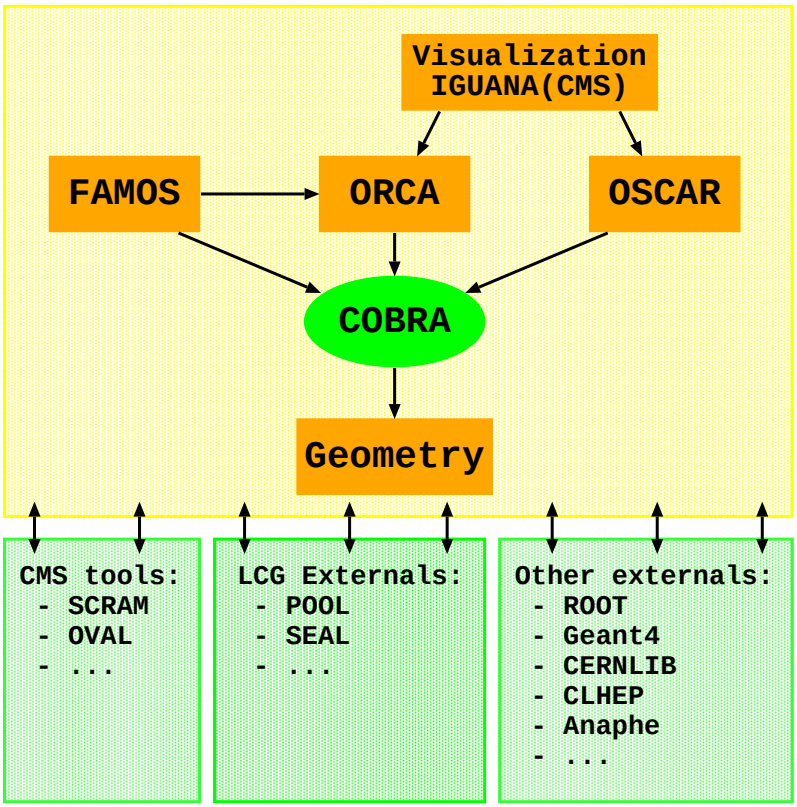


Figure 4.17: Overview of the CMS software framework. The arrows in the top diagram show the inter-dependencies of the packages. In the bottom part the most important examples of external tools are given.

The core of the framework is embedded in the COBRA package, which hosts the basic services to be able to read out the detector or simulated events, process data from different sources and at different levels of reconstruction, and write results back to storage elements. COBRA also implements the powerful on-demand reconstruction, caching and persistency facilities, providing transparency to developers and users. These mechanisms allow to ignore whether intermediate reconstruction steps have already been performed previously and stored either in memory or in persistent objects.

From the end-user's perspective the ORCA package plays the role of the central analysis component. It groups all algorithms for reconstruction in CMS, from the digitization of the detector hits simulated by OSCAR and the reconstruction of low-level building blocks, up to the reconstruction of high-level compound physics objects, of which some will be described in detail in Chapter 5. Next to ORCA, used for analysis of the detailed simulations and actual data, the fast simulation program FAMOS is conceived such that it produces the same type of objects as encountered in ORCA. Although typically only part of the objects' information is simulated, the great advantage enters through the high-level reconstruction and analysis algorithms, that can be run transparently on the results from parametrized or detailed simulation.

In 2005 a redesign has been proposed for the central components of the CMS software into a new Event Data Model. At the time of writing, the transition to this new

framework has only just started, and therefore all results in this thesis rely on the software described above. More information on the new software framework can be found in [85].

4.2.4 Distributed CMS computing

Simulation and data analysis are of paramount importance for the CMS experiment. The organization of the general computing facilities is a project on its own, because of the experiment's software complexity, because of the magnitude of the calculations to be performed and data to be stored, but also because of the international context of the experiment, with many users and developers worldwide. The overall requirements and organization of computing resources and tools is referred to as the CMS computing model [138].

Distributed computing is a key ingredient of the computing model, and will be implemented as a system called the computing grid. Grid-enabled computing technology aims to provide a fully transparent interface to the user, from which grid jobs can be submitted to any computing resource matching specified conditions, without prior knowledge on location, availability, etc. End users do not need local accounts but are authenticated based on credentials within virtual organizations. For this abstraction to work, a wide variety of services and tools have been developed in recent years, to provide security, authentication, job management, monitoring, accounting, and much more. This has been done within the context of more general grid-research programs [139], but also LHC and CMS specific needs are being addressed [140]. This mostly involves problems due to the massive amount of data to be stored and transferred. Indeed, an expected $1 - 1.5\,\mathrm{MB}$ of raw data per event will stream out of the CMS online event-selection system at $\mathcal{O}(100\,\mathrm{Hz})$. Events simulated with the described simulation programs, on the other hand, take up about $0.5\,\mathrm{MB}$ per event after detailed detector simulation, and an additional $0.7\,\mathrm{MB}$ after digitization.

The CMS computing model is conceived as a layered structure. A Tier-0 computing centre at CERN will be directly connected to the experiment for initial processing and archiving of the huge stream of raw data. The largest amount of the event reconstruction will be performed at a few remote Tier-1 sites, that at the same time provide services like archiving, calibration and data reduction. A more numerous amount of Tier-2 centres provides capacity for data analysis, calibration tasks and Monte-Carlo simulation to the Tier-1 centre they are connected to. Further refinements consist of small local resources and interfaces.

At the IIHE[1] CMS computing started in 2002, when a first cluster of 10 processors was conceived and installed with the CMS software framework. Later, this cluster has been more than doubled in size. This early presence of a significant amount of computing power allowed, on the one hand, the detailed Monte-Carlo simulation of hundreds of thousands of events, and on the other hand it offered the possibility of rapid in-house code and analysis development and testing on previously simulated

[1]The *Inter-university Institute for High Energies* (IIHE) was created in 1972, and groups the experimental particle physics teams of the *Vrije Universiteit Brussel* (VUB) the *Université Libre de Bruxelles* (ULB), the *Universiteit Antwerpen* (UA) and the *Université de Mons-Hainaut* (UMH).

samples. In Table 4.1 an overview is given of the samples that have been produced and used for this thesis. They have all been simulated with `OSCAR_2_4_6`, digitized with `ORCA_7_6_1` and reconstructed with `ORCA_8_7_1`. Details on the generation of the various samples are given in Section 3.4. Low-luminosity pile-up conditions are discussed in Section 2.3.2.

Dataset	Pile-up (low lumi.)	Cross section (pb)	Nr. of events (approx.)	Integr. lumi. (fb^{-1})
Generation with `PYTHIA`, LO cross section				
$\text{t}\bar{\text{t}} \to \text{bbqq}'\mu\nu$	yes	81.3	200 000	2.5
$\text{t}\bar{\text{t}} \to \text{bbqq}'\mu\nu$	no	81.3	200 000	2.5
$\text{t}\bar{\text{t}} \to \text{bbqq}'\text{e}\nu$	yes	81.3	200 000	2.5
$\text{t}\bar{\text{t}} \to \text{bbqq}'\text{e}\nu$	no	81.3	200 000	2.5
$\text{t}\bar{\text{t}} \to \text{bb}\ell_1\nu_1\ell_2\nu_2$ ($\ell = \text{e}, \mu$)	yes	27.7	20 000	0.72
$\text{t}\bar{\text{t}} \to \text{bb}\ell_1\nu_1\ell_2\nu_2$ ($\ell = \text{e}, \mu$)	no	27.7	20 000	0.72
Generation with `PYTHIA`, scaled to NLO cross section				
$\text{gg} \to \text{tbH}^{\pm} \to \text{ttbb}$ ($m_{\text{H}^{\pm}} = 263\,\text{GeV}/c^2$)	yes	0.850	20 000	24
$\text{gg} \to \text{tbH}^{\pm} \to \text{ttbb}$ ($m_{\text{H}^{\pm}} = 311\,\text{GeV}/c^2$)	yes	0.570	20 000	35
$\text{gg} \to \text{tbH}^{\pm} \to \text{ttbb}$ ($m_{\text{H}^{\pm}} = 359\,\text{GeV}/c^2$)	yes	0.377	20 000	53
$\text{gg} \to \text{tbH}^{\pm} \to \text{ttbb}$ ($m_{\text{H}^{\pm}} = 408\,\text{GeV}/c^2$)	yes	0.251	20 000	80
$\text{gg} \to \text{tbH}^{\pm} \to \text{ttbb}$ ($m_{\text{H}^{\pm}} = 457\,\text{GeV}/c^2$)	yes	0.169	20 000	118
$\text{gg} \to \text{tbH}^{\pm} \to \text{ttbb}$ ($m_{\text{H}^{\pm}} = 506\,\text{GeV}/c^2$)	yes	0.116	20 000	172

Table 4.1: Listing of privately-simulated samples, with detailed `OSCAR_2_4_6` detector simulation.

With the advent in 2005 of grid computing in the framework of CMS data analysis, also many centrally-simulated samples have become available for remote analysis. In Table 4.2 the samples are listed that have been accessed remotely within the scope of this thesis. The simulation program and version are indicated in the table. Digitization and analysis have been performed as in the case of the locally-produced samples. More details on the generation of some samples can be found in Section 3.4.

At the same time grid infrastructure started to be deployed at the IIHE. The often almost exclusive availability of about 100 processors allowed for the use of the earlier privately-simulated event samples in an even more efficient way. For this to work, private changes have been applied to the CMS job submission tool `CRAB` [141]. Some of these adaptions have later been incorporated into this maturing program, along with many improvements from extensive testing. Some more private changes allowed in addition to submit `FAMOS` jobs on the IIHE grid resources, well before this possibility was incorporated in `CRAB` itself.

With these tools in place, about 6.5 million events have been processed with `FAMOS_1_3_2` at the IIHE for the studies presented in Chapters 6 and 7. In Table 4.3 an overview is given of the concerned samples. All of these have been simulated with low-luminosity pile-up included. More details on the generation of the samples are given

Dataset	Pile-up (low lumi.)	Cross section (pb)	Nr. of events (approx.)	Integr. lumi. (fb^{-1})
Generation with `PYTHIA`, LO cross section; simulation with `OSCAR_2_4_6`				
$\mathrm{t\bar{t}}$ inclusive	yes	560.0	3 431 000	6.1
W+jets $(10 < \hat{p_\mathrm{T}} < 100\,\mathrm{GeV}/c)^2$	yes	27 550.0	514 000	0.019
W+jets $(75 < \hat{p_\mathrm{T}} < 500\,\mathrm{GeV}/c)^2$	yes	1 216.0	282 000	0.23
W+jets $(350 < \hat{p_\mathrm{T}} < 2\,200\,\mathrm{GeV}/c)^2$	yes	4.9	71 000	15
$\mathrm{t\bar{t}} \to \mathrm{bb}\ell_1\nu_1\ell_2\nu_2$ $(\ell = \mathrm{e}, \mu, \tau)$	yes	57.4	958 000	17
WW inclusive[3]	yes	190.0	483 000	2.5
ZW inclusive	yes	26.8	277 000	10
Z+jets $(0 < \hat{p_\mathrm{T}} < 40\,\mathrm{GeV}/c)^2$	yes	31 720.0	199 000	0.0063
Z+jets $(42.5 < \hat{p_\mathrm{T}} < 300\,\mathrm{GeV}/c)^2$	yes	575.7	284 000	0.49
Z+jets $(200 < \hat{p_\mathrm{T}} < 1\,400\,\mathrm{GeV}/c)^2$	yes	7.0	45 000	6.4
Generation with `CompHEP`, LO cross section; simulation with `CMSIM133`				
$\mathrm{t\bar{t}b\bar{b}}$	yes	2.4	390 000	163
$\mathrm{t\bar{t}jj}$	yes	235.8	1 304 000	5.5

Table 4.2: Listing of remotely-accessed simulation samples.

in Section 3.4. The remarks for the $\mathrm{t\bar{t}}$ samples point to the particular property in the generation that has been altered for the systematic-uncertainty studies in Section 6.2.

4.3 The CMS Online-Selection System

The very high LHC bunch-crossing rate of 40 MHz, and the limited storage capacity corresponding to some 100 to 200 events per second, requires a severe yet flexible online-selection or trigger system. In CMS a choice has been made for a programmable Level-1 trigger system implemented in dedicated custom electronics, followed by a fully software-based High-Level Trigger. The CMS trigger system is a very complex and challenging part of the experiment, due to the many requirements: efficiency, inclusiveness, robustness, speed, rate, adaptability, etc. In this section the online selection will be described only briefly, with focus on the lepton triggers used for the analyses in this thesis.

4.3.1 The Level-1 trigger

The Level-1 trigger (L1) consists of the on-detector electronics and the trigger electronics housed in the underground service cavern. The latency to transmit the signal to and from the trigger logic and to reach the L1 decision is fixed at 3.2 μs, requiring a

[2]The range $\hat{p_\mathrm{T}}$ limits the allowed transverse momentum obtained in the rest frame of the hard interaction.

[3]The WW inclusive sample is generated together with another non-contributing process, causing an artificially large cross section.

Dataset	Cross section (pb)	Nr. of events (approx.)	Integr. lumi. (fb^{-1})	Remark
Generation with PYTHIA, LO cross section				
$\mathrm{t\bar{t}} \to \mathrm{bbqq'}\mu\nu$	81.3	158 000	1.94	reference
$\mathrm{t\bar{t}} \to \mathrm{bbqq'}\mu\nu$	81.3	162 000	1.99	CTEQ6 p.d.f.'s
$\mathrm{t\bar{t}} \to \mathrm{bbqq'}\mu\nu$	81.3	158 000	1.94	ISR/FSR max.
$\mathrm{t\bar{t}} \to \mathrm{bbqq'}\mu\nu$	81.3	162 000	1.99	ISR/FSR min.
$\mathrm{t\bar{t}} \to \mathrm{bbqq'}\mu\nu$	81.3	153 000	1.88	UE max.
$\mathrm{t\bar{t}} \to \mathrm{bbqq'}\mu\nu$	81.3	157 000	1.93	UE min.
$\mathrm{t\bar{t}} \to \mathrm{bbqq'}\mu\nu$	81.3	156 000	1.92	$m_{\mathrm{t}} = 170\,\mathrm{GeV}$
$\mathrm{t\bar{t}} \to \mathrm{bbqq'}\mu\nu$	81.3	166 000	2.04	$m_{\mathrm{t}} = 180\,\mathrm{GeV}$
$\mathrm{t\bar{t}} \to \mathrm{bbqq'e}\nu$	81.3	159 000	1.96	reference
$\mathrm{t\bar{t}} \to \mathrm{bbqq'e}\nu$	81.3	169 000	2.08	ISR/FSR max.
$\mathrm{t\bar{t}} \to \mathrm{bbqq'e}\nu$	81.3	154 000	1.89	ISR/FSR min.
$\mathrm{t\bar{t}} \to \mathrm{bb}\ell_1\nu_1\ell_2\nu_2$ $(\ell = \mathrm{e}, \mu, \tau)$	57.4	75 000	1.31	reference
$\mathrm{t\bar{t}} \to \mathrm{bb}\ell_1\nu_1\ell_2\nu_2$ $(\ell = \mathrm{e}, \mu, \tau)$	57.4	83 000	1.45	CTEQ6 p.d.f.'s
$\mathrm{t\bar{t}} \to \mathrm{bb}\ell_1\nu_1\ell_2\nu_2$ $(\ell = \mathrm{e}, \mu, \tau)$	57.4	78 000	1.36	ISR/FSR max.
$\mathrm{t\bar{t}} \to \mathrm{bb}\ell_1\nu_1\ell_2\nu_2$ $(\ell = \mathrm{e}, \mu, \tau)$	57.4	73 000	1.27	ISR/FSR min.
$\mathrm{t\bar{t}} \to \mathrm{bb}\ell_1\nu_1\ell_2\nu_2$ $(\ell = \mathrm{e}, \mu, \tau)$	57.4	79 000	1.38	UE max.
$\mathrm{t\bar{t}} \to \mathrm{bb}\ell_1\nu_1\ell_2\nu_2$ $(\ell = \mathrm{e}, \mu, \tau)$	57.4	78 000	1.36	UE min.
$\mathrm{t\bar{t}} \to \mathrm{bb}\ell_1\nu_1\ell_2\nu_2$ $(\ell = \mathrm{e}, \mu, \tau)$	57.4	71 000	1.24	b fragm. max.
$\mathrm{t\bar{t}} \to \mathrm{bb}\ell_1\nu_1\ell_2\nu_2$ $(\ell = \mathrm{e}, \mu, \tau)$	57.4	71 000	1.24	b fragm. min.
$\mathrm{t\bar{t}} \to \mathrm{bb}\ell_1\nu_1\ell_2\nu_2$ $(\ell = \mathrm{e}, \mu, \tau)$	57.4	80 000	1.39	udsg fragm.
Generation with PYTHIA, scaled to NLO cross section				
$\mathrm{gb} \to \mathrm{tH}^{\pm} \to \mathrm{ttb}$	0.850	80 000	94	$m_{\mathrm{H}^{\pm}} = 263\,\mathrm{GeV}$
$\mathrm{gb} \to \mathrm{tH}^{\pm} \to \mathrm{ttb}$	0.570	70 000	123	$m_{\mathrm{H}^{\pm}} = 311\,\mathrm{GeV}$
$\mathrm{gb} \to \mathrm{tH}^{\pm} \to \mathrm{ttb}$	0.377	78 000	207	$m_{\mathrm{H}^{\pm}} = 359\,\mathrm{GeV}$
$\mathrm{gb} \to \mathrm{tH}^{\pm} \to \mathrm{ttb}$	0.251	75 000	299	$m_{\mathrm{H}^{\pm}} = 408\,\mathrm{GeV}$
$\mathrm{gb} \to \mathrm{tH}^{\pm} \to \mathrm{ttb}$	0.169	80 000	473	$m_{\mathrm{H}^{\pm}} = 457\,\mathrm{GeV}$
$\mathrm{gb} \to \mathrm{tH}^{\pm} \to \mathrm{ttb}$	0.116	85 000	733	$m_{\mathrm{H}^{\pm}} = 506\,\mathrm{GeV}$
Generation with MadGraph/MadEvent, LO cross section				
$\mathrm{t\bar{t}b}/\mathrm{t\bar{t}j}$	678	3 521 000	5.19	

Table 4.3: Listing of privately-simulated samples with parametrized FAMOS_1_3_2 detector simulation.

buffer of 128 beam crossings in the read-out electronics. Of the total latency, the time allocated to L1-trigger calculations is less than 1 μs. After the L1 trigger the event rate should be reduced to about 100 kHz, which will be limited to 50 kHz at detector start-up.

The Level-1 triggers involve the calorimetry and muon systems. Data with coarse granularity and resolution are used to form so-called trigger-primitive objects, while the high-resolution data awaits the trigger decision in buffers. For the muons the three subsystems, DT, CSC and RPC, each have their own trigger logic which delivers the four best muon candidates for combination in the global muon trigger. In the HCAL barrel and endcaps the trigger-primitive geometry corresponds to the tower layout shown in Figure 4.9. In the ECAL trigger primitives are built combining 5×5 crystal arrays in the barrel, and in the endcaps a layout is used following approximately the HCAL trigger towers. Level-1 trigger decisions are based on the presence of primitives such as calorimeter clusters, muon or jet candidates, all with E_{T} or p_{T} above a certain threshold. Also global scalar or vectorial sums over several primitives of E_{T} are used.

In `ORCA` an emulation of the Level-1 trigger has been implemented. With this simulation, detailed studies on trigger rates and efficiencies have been performed [86, 123, 142]. From these studies trigger thresholds have been derived for different streams, that keep the event rate sufficiently under control. For the triggers that are used in this thesis, the Level-1 primitive thresholds are shown in Table 4.4 for low-luminosity conditions, along with the most recent expectations for the corresponding rates [86]. To characterize the turn-on curve associated with the given trigger-primitive threshold, a point on this curve is chosen as the so-called working point, for which the efficiency and corresponding threshold are also given in Table 4.4. In `FAMOS` an implementation of the Level-1 trigger is not present yet.

Trigger stream	Chosen working-point efficiency	Working-point p_{T} threshold (GeV/c)	p_{T} threshold L1 prim. (GeV/c)	Rate (kHz)
inclusive muon	90%	14	14	2.5
di-muon	90%	3/3	3/3	4.0
inclusive electron	95%	29	23	3.9
di-electron	95%	17/17	12/12	1.0

Table 4.4: Low-luminosity Level-1 trigger thresholds and expected rates for the streams relevant in this thesis. A working point is chosen to characterize the trigger turn-on curve.

Currently much effort is put into new trigger studies for the CMS experiment. The L1 trigger allows much more flexibility than the current layout [86], with up to 128 available trigger streams. Moreover, the operation at start-up still needs to be defined, as well as the L1 trigger streams needed for calibration and monitoring.

4.3.2 The High-Level Trigger

The CMS High-Level Trigger (HLT) processes all events that are accepted by the Level-1 trigger in a single processor farm with several thousand nodes. It is intended

to bring the event rate from the Level-1 output down to $\mathcal{O}(100\,\mathrm{Hz})$, which corresponds to the available bandwidth towards mass storage for an expected event size of $\sim 1\,\mathrm{MB}$. Between 10% and 30% of the bandwidth will need to be reserved for dedicated calibration and monitoring triggers, running at low-energy thresholds with appropriate scaling factors. Although the HLT is a single entity, the selection of events is optimized by rejecting events as early as possible within the trigger algorithms. A minimal amount of detector information needed to reject background events is reconstructed, using regional reconstruction and virtual trigger levels.

The HLT electron reconstruction first searches for a supercluster in the calorimeter according to the algorithms described in Section 4.1.4. Then matching pixel hits are sought, by extrapolating from the cluster position through the magnetic field to the beamline, without constraints on the longitudinal position of the hits, and taking into account both possible charges. Finally the reconstruction of the electron track is performed, seeded by the associated hits, but with large freedom on the track parameters. Electrons not satisfying some quality constraints are removed. At each stage of the reconstruction events without remaining electron candidates are rejected, gaining overall almost 50% of computation time in the case of jet backgrounds.

The muon selection for the HLT proceeds in two steps. First, muons are reconstructed in the muon chambers as described in Section 4.1.6, starting from the Level-1 trigger primitives. This reconstruction confirms the Level-1 trigger decision and refines the p_T measurement. The calorimeters are used for isolation at this step by cutting on the energy sum in a cone around the muon direction. In the second step, the muon trajectories are extended into the tracker, which further refines the p_T measurement. Another isolation criterion is then applied, which vetoes additional pixel tracks in a region around the muon trajectory. This tracker isolation suppresses muons from b, c, π and K decays.

For the studies in this thesis the single and double electron and muon triggers have been employed, as implemented into `ORCA` and `FAMOS`. In Table 4.5 the thresholds and most recent rate expectations are given for these triggers under low-luminosity conditions. The efficiency and threshold are also given for the chosen working points, which characterize the turn-on curves at the given trigger thresholds.

Trigger stream	Chosen working-point efficiency	Working-point p_T threshold (GeV/c)	p_T threshold HLT (GeV/c)	Rate (Hz)
inclusive muon	90%	19.0	19.0	25.8
di-muon	90%	7.0/7.0	7.0/7.0	4.8
inclusive electron	95%	29.0	26.0	23.5
di-electron	95%	17.0/17.0	14.5/14.5	1.0

Table 4.5: Low-luminosity HLT trigger thresholds and expected rates for the streams relevant in this thesis. A working point is chosen to characterize the trigger turn-on curve.

In the case of $\mathrm{t\bar{t}}$ events the single-muon trigger is found to be 62.0% efficient on semi-leptonic $\mathrm{t\bar{t}} \to \mathrm{b\bar{b}qq'}\mu\nu$ events, while the single-electron trigger thresholds cause a 48.4% efficiency for semi-leptonic $\mathrm{t\bar{t}} \to \mathrm{b\bar{b}qq'}\mathrm{e}\nu$ events. These efficiencies correspond

to respective rates of 0.051 Hz and 0.040 Hz using the leading-order $t\bar{t}$ cross section for the low luminosity $\mathcal{L} = 2 \times 10^{33}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$.

Apart from electrons and muons, many other physics objects will be used in the HLT, such as photons, jets, τ candidates, missing energy and b jets. Some trigger streams combine several of these objects or apply topological cuts, to allow for lower thresholds. The most recent trigger table consists of 37 streams [86], but this scheme is under active development and rates are being re-evaluated in the framework of the new CMS software. Special attention is being paid to calibration streams, and cross triggers or topological triggers with particular physics motivations, like the $e + \mu$ trigger.

Chapter 5

Event Reconstruction

In Chapter 4 the basic reconstruction of sub-detector specific objects has been discussed. To perform physics studies with these building blocks, they need to be combined into higher-level physics objects, that allow to identify the final states of analysed channels. For most of these physics objects, information from various sub-detectors is combined. The reconstruction needs to balance the trade-off between efficiency and fake rate, and identification of the high-level objects often involves analysis-specific quality requirements on the reconstruction. In this chapter the techniques are described for reconstruction of the physics objects that are used in the analyses in Chapters 6 and 7. Details on the reconstruction of other objects, like photons and τ leptons, can be found in [85]. Throughout the chapter emphasis is put on the performance of the discussed techniques, in terms of efficiencies and resolutions. Results are presented such that they can serve as input to an integrated view of the event reconstruction.

In Section 5.1 the complex problem is tackled of the association of calorimeter activity with jets of hadrons originating from quarks and gluons. Many detailed properties of reconstructed jets and the corresponding tracks are used in Section 5.2 to build algorithms for b-jet identification. Section 5.3 deals with the offline reconstruction of electrons and muons. A channel-specific technique is described that allows to identify with high purity the leptons from W-boson decays in $t\bar{t}$ events. In Section 5.4, the reconstruction of the transverse missing energy is described. In Section 5.5, finally, a technique is discussed to fit the kinematics of an event's final-state objects to a hypothesized event topology, which enhances measurements by exploitation of correlated information in a global event reconstruction.

5.1 Jet Reconstruction

The association of measured energy clusters with originating scattered partons is a complex problem, because of the multitude of physics and detector effects that need to be accounted for: gluon radiation, underlying event, pile-up and noise contributions, uncertainties in jet-fragmentation models, out-of-cone showering and loss of low-momentum charged particles due to the magnetic field, energy loss due to dead material, cracks, interactions with tracker material, rear leakage, etc.

In this section various experimental aspects of jet reconstruction in CMS are treated

in detail. First a few algorithms for jet clustering are presented, along with a discussion on clustering efficiency and fake rate. Next, several methods are presented to determine jet energy-scale corrections that compensate for the effects mentioned above. The CMS performance of jet energy reconstruction is characterized next with jet resolution functions. Finally, a method is presented for association of jets with the signal vertex.

5.1.1 Jet clustering

When a quark or gluon fragments into a jet, the produced hadrons are collimated in the direction of the initial parton, due to its high boost. Therefore, a straightforward method to reconstruct jets is to cluster energy deposits in a cone, of which the direction is associated to the parton's initial direction. Because of Lorentz invariance, as explained in Section 2.3.1, these cones are to be described in (η, ϕ) space rather than Euclidean (θ, ϕ) space, with pseudorapidity η given by (2.6). The metric is defined as

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}. \tag{5.1}$$

For CMS two cone-based methods, the iterative cone and midpoint cone algorithms, and one cluster-based method, the inclusive k_{T} algorithm, have been implemented so far [133]. These algorithms can cluster several types of input objects, for example generator-level particles at the vertex, excluding muons or neutrinos if desired, leading to so-called particle-level jets. These can be used in comparisons with phenomenological studies and to study in detail physics consequences of a jet definition. In realistic analyses calorimetry deposits can be used, taking into account effects from material interactions, magnetic field, pile-up, etc.

Iterative cone The iterative cone algorithm starts by ordering the input objects according to decreasing E_{T}. The first object in the list is taken as jet seed, provided it exceeds a specified seed threshold. By clustering objects in a given cone around the seed, a so-called proto-jet is constructed. The direction of this proto-jet is used as new seed, and this procedure is iterated until the energy and the direction of the proto-jet stabilizes. At that point the proto-jet is added to the list of jets, and the objects in the proto-jet are deleted from the list of input objects. The whole procedure is then repeated until no more input objects are found exceeding the seed threshold. The jet cone size and the seed threshold are the algorithm's parameters.

Midpoint cone The midpoint cone algorithm has been designed to improve the treatment of overlapping jets. It uses the same procedure as the iterative cone algorithm to find proto-jets, but in contrast no input objects are removed from the list of seeds. Once all proto-jets are constructed, the momenta of pairs of overlapping proto-jets are combined into so-called midpoints, which serve as additional seeds for new proto-jets. On all the obtained proto-jets, finally, a splitting and merging procedure is applied, starting with the highest E_{T} proto-jet. This proto-jet becomes a jet if no input objects are shared with other proto-jets. Otherwise, a comparison is made between the shared transverse energy with the highest-E_{T} neighbour and the total E_{T} of this proto-jet. If this fraction is larger than a given

value, typically 50%, the proto-jets are merged, otherwise the common objects are assigned depending on which of the two proto-jets is closest. This procedure is repeated, always starting from the highest E_{T} proto-jet, until no proto-jets are left. This algorithm has as parameters the jet cone size, the seed threshold, the threshold on the shared-energy fraction for jet merging, and also a maximum on the number of proto-jets used to generate midpoints.

Inclusive k_{T} The k_{T} algorithm implemented for CMS calculates for each input object i and each pair (i, j) the quantities

$$d_i = E_{\mathrm{T},i}^2 R^2 \qquad \text{and} \qquad d_{ij} = \min\left\{E_{\mathrm{T},i}^2, E_{\mathrm{T},j}^2\right\} \Delta R_{ij}^2, \tag{5.2}$$

where R is a dimensionless parameter, and ΔR is the metric defined in (5.1). An iterative procedure searches the smallest value of all d_i and d_{ij} over all objects and pairs. If a d_{ij}-type value is smallest, the objects i and j are removed, and merged into a new input object. If a d_i value is smallest, the input object i is added to the list of found jets, and deleted from the input list. When all input objects are transformed into jets, a new iteration merges all jets i and j with distance $\Delta R_{ij} < R$. It follows that $\Delta R_{ij} < R$ for all i and j, and hence R can be interpreted as a similar parameter as the jet cone size for the cone-based algorithms.

Two different options are available to combine clustered input objects into jets. Either the jet constituents are added as four-vectors, which results in massive jets. This is called the energy or E scheme. In contrast, the so-called E_{T} scheme produces massless jets by equating the jet transverse momentum to the ΣE_{T} of the constituents. For the two cone-based algorithms, the choice of this recombination scheme is only relevant in the final jet determination. During the iterative steps of jet finding the E_{T} scheme to add input objects is always used.

The described jet clustering algorithms have several parameters that leave freedom to tune the jet clustering to the needs of the analysis it is used for. Signatures with very collimated high-energy jets, for example, will profit from a reduced cone size when using cone algorithms. A detailed study has been performed with the three presented jet clustering algorithms based on particle-level jets [143]. A combined reconstruction quality parameter is defined, that provides a measure for the event-by-event reconstruction quality of the kinematics of the primary partons. This quality measure is determined as a function of the algorithm's parameters, for top-quark related signatures involving two, four, six and eight primary partons from the hard interaction in the final state. In Figures 5.1 and 5.2 the quality measure is shown, respectively as a function of the jet cone size for the iterative cone algorithm, and as a function of the R parameter for the k_{T} jet algorithm. A higher percentage of the measure reflects a better kinematic reconstruction. Currently similar results are being obtained on samples with full detector simulation.

In this thesis, involving final states with 2 up to 6 jets, the iterative-cone algorithm is used with the default $\Delta R = 0.5$ setting for the cone size, which is close to the optimal value. For the studies presented in Chapters 6 and 7 jet reconstruction is limited to the tracker acceptance of $|\eta| < 2.5$.

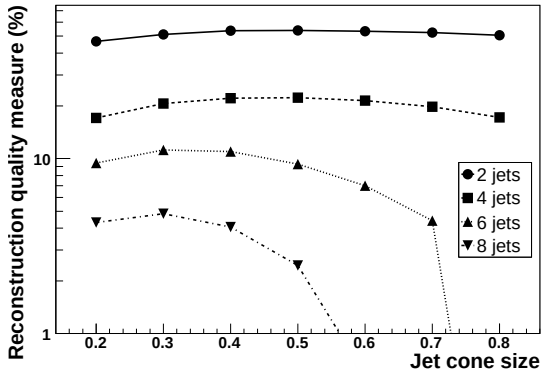


Figure 5.1: Jet-reconstruction performance for the iterative cone algorithm as a function of the jet cone size.

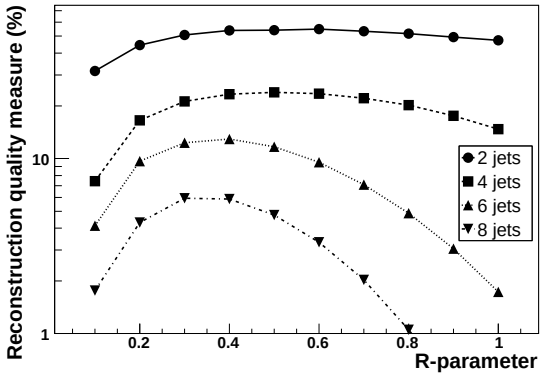


Figure 5.2: Jet-reconstruction performance for the k_{T} jet clustering as a function of the algorithm's R parameter.

5.1.2 Jet-reconstruction efficiency and fake rate

It has been shown in Section 4.1.4 that pile-up interactions deposit sizable amounts of energy in the calorimeter system. Also electronic noise gives additional energy contributions, as well as the channel-dependent underlying event. Several schemes have been presented that can be used to suppress these contributions to combined electromagnetic and hadronic calorimetry towers. Such a suppression results in a lower jet fake rate, but at the same time in a lower jet reconstruction efficiency for low energies. The choice of a tower threshold scheme needs to find a balance between both.

To compare different suppression schemes, a common working point is chosen for the jet efficiency, defined as the percentage of all particle-level jets that match with a reconstructed jet within a cone $\Delta R = 0.3$. This efficiency is forced to be 50% for a 20 GeV particle-level jet with $|\eta| < 2.5$, by adjusting the minimal threshold on the reconstructed transverse jet energies $E_{\mathrm{T,min}}^{\mathrm{rec}}$. In Figure 5.3 the jet reconstruction efficiency is shown at the common working point, for several tower-threshold schemes and as a function of pseudorapidity η. These results are obtained on fully-simulated di-jet samples in a wide range of $\hat{p_{\mathrm{T}}}$ bins ($0 < \hat{p_{\mathrm{T}}} < 4000\,\mathrm{GeV}/c$), with low-luminosity pile-up. The reconstruction efficiency in the forward regions of the detector is much higher, because jets satisfying the $E_{\mathrm{T,min}}^{\mathrm{rec}}$ threshold must necessarily have higher energies in that region, and are hence more efficiently reconstructed. In the central region the enhancement near $\eta \approx 0$ stems from the electronic noise. Tower-threshold schemes that eliminate this behaviour are more effective in suppressing this noise contribution. As a function of jet E_{T} all threshold schemes show very similar behaviour.

A corresponding behaviour is found in the jet fake rate, which is defined by the number of reconstructed jets above $E_{\mathrm{T,min}}^{\mathrm{rec}}$ threshold, that do not match any particle-level jet with $E_{\mathrm{T}} > 10\,\mathrm{GeV}$ from the hard interaction. In Figure 5.4 the jet fake rate at the working point is shown for the same sample and same tower thresholds, as a function of η. The peaks at $\eta \approx 0$ indeed correspond to insufficient noise suppression. Also at $|\eta| \approx 3$ large peaks can be observed. These have been understood to be due to the coarser calorimeter-tower geometry in this region, which results in more energy collected in a single cell, and hence the corresponding tower energies more easily pass

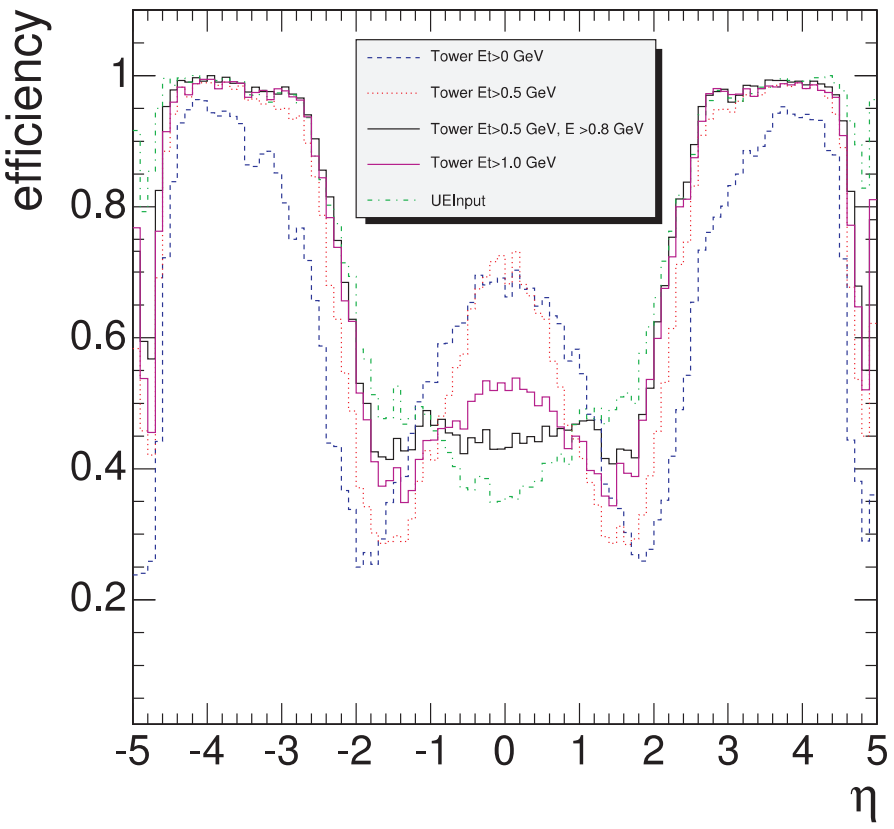


Figure 5.3: Jet reconstruction efficiency in QCD events in a wide range of $\hat{p_{\mathrm{T}}}$ bins, for several tower-threshold schemes defined in Section 4.1.4 and as a function of pseudo-rapidity η.

the noise and seed thresholds in the jet clustering. Recently an attempt has been made to split tower energies in this region into a finer virtual geometry, indeed suppressing the peaks in the jet fake rate. This splitting of towers has not been used for jet reconstruction in the studies for this thesis, however, because it mostly affects jets in the forward region, in contrast to the jets of interest in the tracker acceptance used in this thesis. In addition, the doubling of the cell size in ϕ at $\eta = 1.740$ is not accounted for in the split-tower geometry.

5.1.3 Jet energy-scale corrections

Many simultaneous effects complicate the translation of a measured jet energy for a given jet definition into a corresponding originating parton energy. A first class of effects relates to the physics of jets in hadron collisions. Final-state radiation blurs the connection between the initial parton and the corresponding jet. Hence, on average, the jet definition will necessarily exclude some of the showered energy from the clustering. At the same time the underlying event and the pile-up collisions give additional external contributions to jet energy measurements. A second group of effects stems

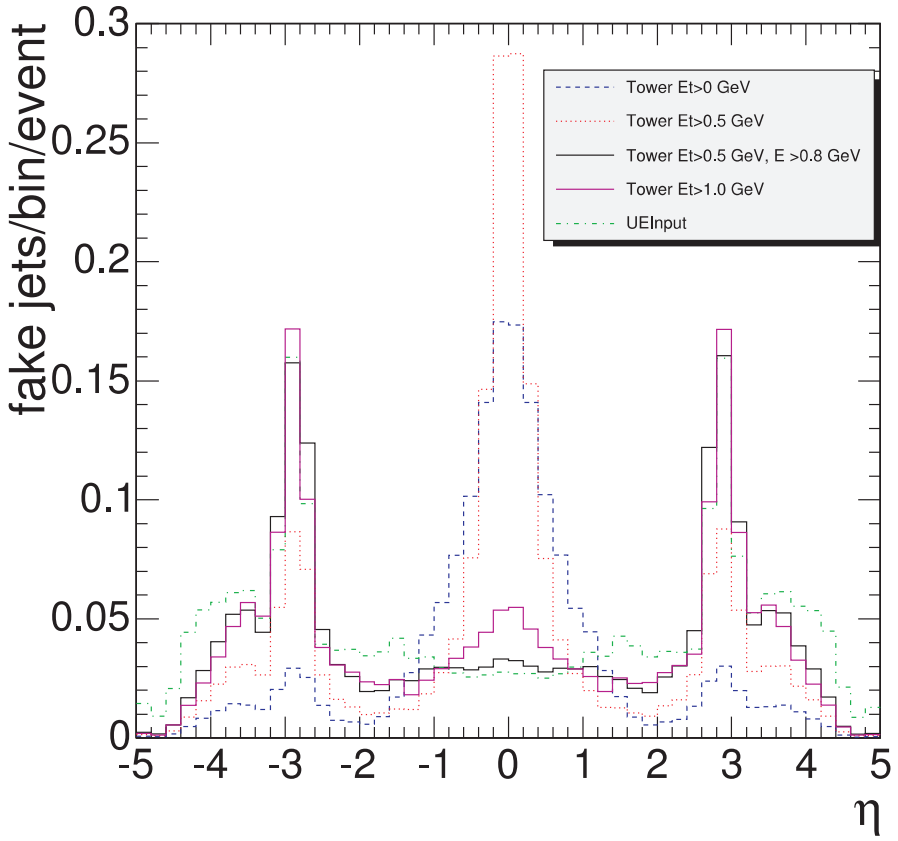


Figure 5.4: Jet fake rate in QCD events in a wide range of $\hat{p_{\mathrm{T}}}$ bins, for several tower-threshold schemes defined in Section 4.1.4 and as a function of pseudorapidity η.

from instrumental complications. Low-momentum charged particles are swept out of jet cones due to the magnetic field. Electronic noise, on the other hand, will generate additional contributions to jet energy measurements. Also unequal calorimeter response to electromagnetic and hadronic showers, and energy loss due to dead material, cracks, interactions with material in front of the calorimeters and rear leakage render jet energy reconstruction a highly non-trivial task.

Depending on the purpose of the jet energy-scale correction scheme, either the reconstructed jet energy is calibrated to match the corresponding particle-level jet, or the calibration is performed directly to the originating parton. In the former case some channel dependency and effects due to the choice of the jet algorithm are factorized out; in the latter case final-state gluon radiation needs to be carefully accounted for. Currently several methods for jet energy-scale corrections are available in CMS. One Monte-Carlo based method has been implemented, correcting the reconstructed jet energy $E_{\mathrm{T}}^{\mathrm{rec}}$ back to the particle-level jet energy. Three other methods have been explored, that will use directly data to calibrate the jet energy scale.

Monte-Carlo corrections This Monte-Carlo based method [133] assumes that the calorimeter simulation has been accurately tuned to the single-particle response measurements from test beams. Jets are reconstructed in simulations, on the one

hand at the particle level, excluding muons and neutrinos, and on the other hand with calorimeter towers, using the same settings for the considered jet clustering algorithm. A matching between the jets from the two types of input is made, requiring $\Delta R < 0.2$. For each jet pair the ratio $E_{\mathrm{T}}^{\mathrm{rec}}/E_{\mathrm{T}}^{\mathrm{MC}}$ is calculated and added to histograms in several bins of the jet $E_{\mathrm{T}}^{\mathrm{MC}}$ and η^{MC}. The mean of these distributions renders correction factors k, which are fitted as a function of $E_{\mathrm{T}}^{\mathrm{MC}}$. This results in the correction function $k(E_{\mathrm{T}}^{\mathrm{MC}}, \eta^{\mathrm{MC}})$ in several bins of η^{MC}. In Figure 5.5 the correction curves are shown as a function of η^{MC} for several bins of $E_{\mathrm{T}}^{\mathrm{MC}}$ in the case of the iterative cone algorithm with cone size $\Delta R = 0.5$. These results are obtained from fully-simulated di-jet samples on a wide range of $\hat{p_{\mathrm{T}}}$ bins, with addition of low-luminosity pile-up collisions.

Di-jet balancing Transverse-momentum balancing in QCD di-jet events is a powerful technique to measure a relative jet response only from data [144]. The di-jet balance b is defined as

$$b = 2\,\frac{p_{\mathrm{T}}^{\mathrm{probe}} - p_{\mathrm{T}}^{\mathrm{ref}}}{p_{\mathrm{T}}^{\mathrm{probe}} + p_{\mathrm{T}}^{\mathrm{ref}}} \tag{5.3}$$

for a probed jet with respect to a reference jet, which is chosen inside the barrel, $|\eta| < 1$, for convenience. The relative response is then given by the ratio $2\,\langle b\rangle\,/(2 - \langle b\rangle)$, where $\langle b\rangle$ is the mean of the di-jet balance distribution. The measurement of this response as a function of η can be used to derive relative calibration functions. In Figure 5.6 the relative jet response is shown for raw jets and jets with Monte-Carlo corrections applied, for $120 < (p_{\mathrm{T}}^{\mathrm{probe}} + p_{\mathrm{T}}^{\mathrm{ref}})/2 < 250\,\mathrm{GeV}/c$. With a dedicated prescaled calibration trigger stream with 2.5 Hz HLT rate, it takes only one hour of recorded data to obtain Figure 5.6. This relative calibration technique is therefore particularly suited for monitoring and online calibration tasks.

Photon-jet balancing At leading order, the prompt photon produced in the $\mathrm{qg} \to \mathrm{q}\gamma$ and $\mathrm{q\bar{q}} \to \mathrm{g}\gamma$ processes has a transverse momentum that is exactly balanced by the accompanying jet. The excellent resolution on photon energy measurements with the electromagnetic calorimeter serves as the basis of this jet calibration method [145]. The relation $p_{\mathrm{T}}^{\mathrm{jet}}/p_{\mathrm{T}}^{\gamma} \approx p_{\mathrm{T}}^{\mathrm{jet}}/p_{\mathrm{T}}^{\mathrm{parton}}$, however, is in general only approximately satisfied, due to initial-state radiative corrections. It has been found better to determine calibration corrections directly in bins of p_{T}^{γ}, in which case the p_{T} balance in the γ-parton system is broken. The peak position in the asymmetric $p_{\mathrm{T}}^{\mathrm{parton}}$ histogram at fixed p_{T}^{γ}, however, is found to correctly return the calibration constants $k(p_{\mathrm{T}}^{\mathrm{parton}}, \eta) = p_{\mathrm{T}}^{\mathrm{jet}}/p_{\mathrm{T}}^{\gamma}$. Systematic uncertainties for this method arise from QCD backgrounds with fake photons, and from differences in quark- and gluon-jet response. This method is expected to provide the first absolute jet energy-scale estimations due to the high statistics, especially at low transverse jet momentum.

Calibration with the W-boson mass constraint In carefully selected samples of jets from hadronic W-boson decays, the well-measured W-boson mass can be used to extract the absolute jet energy scale for jets originating from quarks [146]. Top-quark pair events with one leptonically- and one hadronically-decaying W boson

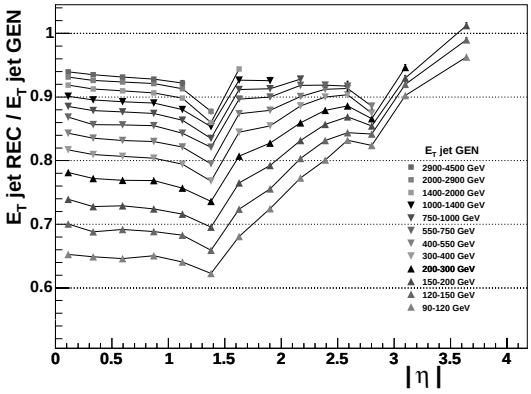


Figure 5.5: Monte-Carlo jet energy correction curves as a function of η^{MC} for several bins of $E_{\mathrm{T}}^{\mathrm{MC}}$, determined on QCD samples with the iterative cone algorithm with cone size $\Delta R = 0.5$.

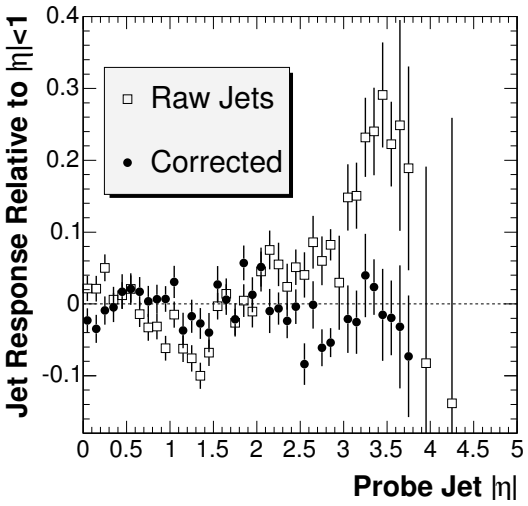


Figure 5.6: Relative jet response as a function of η^{probe} of the probed jet, for raw jets and jets with Monte-Carlo corrections applied, in a QCD sample with $120 < (p_{\mathrm{T}}^{\mathrm{probe}} + p_{\mathrm{T}}^{\mathrm{ref}})/2 < 250\,\mathrm{GeV}/c$.

provide an ideal environment to extract such a sample, as backgrounds can be efficiently suppressed with the presence of an isolated lepton and two b quarks, and the b-tagging makes the identification possible of the jets from the W-boson decay. In Figure 5.7 the W-boson mass spectrum is shown for events with an isolated muon and four non-overlapping jets within tracker acceptance, of which two are tagged as b jet, and two are anti-tagged and are used to reconstruct the W-boson mass. Backgrounds from other sources than top-quark production are negligible. After selection of the sample of W-boson candidates, the measured four-momenta of the jets from the W-boson decays are inclusively scaled by factors $1 + \Delta C$ in a wide range of ΔC. For each value of ΔC the W-boson mass spectrum is fitted with a Gaussian function. The mean of this function with corresponding statistical uncertainty is shown as a function of ΔC in Figure 5.8 for about $5\,\mathrm{fb}^{-1}$ of integrated luminosity. The relation between the reconstructed W-boson mass and the applied inclusive scale factor proves to be linear over a wide range, with very small statistical uncertainty. Imposing the W-boson mass as a constraint, finally, the inclusive calibration factor is retrieved with only 0.6% uncertainty for $1\,\mathrm{fb}^{-1}$ of data, and with only small bias.

Several systematic effects have been considered, influencing the absolute jet energy-scale determination with the W-boson mass constraint. The only significant effect found is due to the presence of the low-luminosity pile-up, yielding a 3.1% shift between being included or left out. The pile-up dependence is no surprise, of course, and can even be used to make a pile-up dependent calibration. The exploration of this method still needs to be complemented with robustness studies, for example concerning the problem of jet overlap. Also a differentiation with respect to E_{T} and η is still to be developed, which is complicated by the

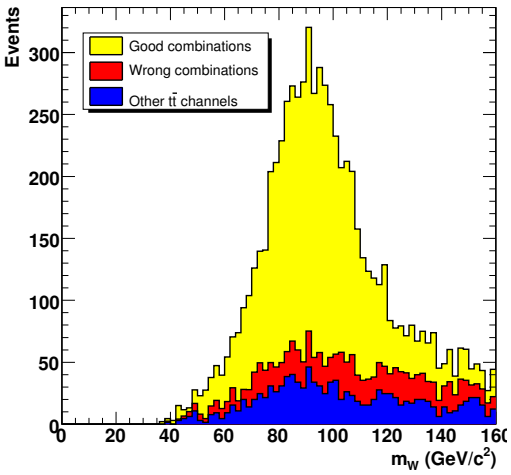


Figure 5.7: The W-boson mass spectrum for all events with an isolated muon and four jets, of which two are tagged as b, and two are anti-tagged and are used to reconstruct the W-boson mass.

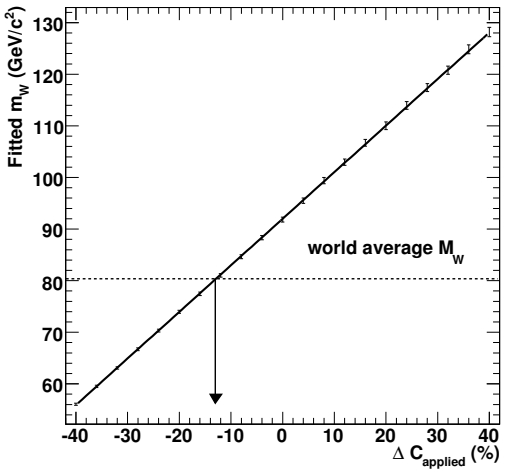


Figure 5.8: Fitted W-boson mass as a function of the applied inclusive jet energy shift ΔC on the constituting jets.

presence of two jets per event with different kinematics, but connected by only a single constraint. Finally, this method is only applicable on light-quark jets, but is easily extended to b quarks, by application of the top-quark mass constraint on the two corrected light quarks and one of the b quarks.

Once CMS is gathering data, the jet energy-scale corrections will be determined and cross checked with several methods. Depending on the available amount of data, on the jet type and on the energy and pseudorapidity range considered, different methods will deliver the best estimate of the needed jet energy-scale corrections [147]. It is expected that with functional b-tagging algorithms ($\mathcal{L} \gtrsim 1\,\mathrm{fb}^{-1}$), the jet energy-scale uncertainty will quickly be reduced to a few percent with the method exploiting the W-boson mass, except in the kinematic region with small, $E_\mathrm{T} < 30\,\mathrm{GeV}$, or at very large transverse energies. For small transverse energy the initial calibration from the $\gamma+\mathrm{jet}$ method is expected to be improved by the similar $\mathrm{Z}+\mathrm{jet}$ balance method, which needs more statistics but is much cleaner. Additional improvements can be awaited from energy-flow techniques, which improve energy measurements by combining calorimeter measurements with tracker information, on a jet-by-jet basis rather than with average correction factors.

5.1.4 Jet resolutions

The same matching $\Delta R < 0.2$ between calorimeter and particle-level jets as for the determination of the Monte-Carlo jet corrections, is also used to investigate the jet resolutions. The distributions of the ratio $E_\mathrm{T}^\mathrm{rec}/E_\mathrm{T}^\mathrm{MC}$ between matched jet pairs, in several bins of E_T^MC and η^MC, are fitted with a Gaussian distribution on the peak. Instead of the offset from 1 of the mean of the Gaussian fit, which is used to determine

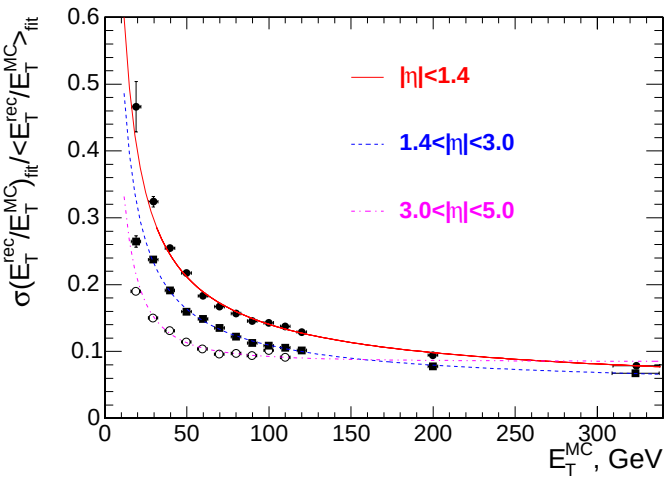


Figure 5.9: Jet resolution functions for the iterative cone algorithm with cone size $\Delta R = 0.5$ in the barrel, the endcap and the forward region.

the Monte-Carlo corrections, the width of the fit function is determined as a function of $E_{\mathrm{T}}^{\mathrm{MC}}$. The obtained resolution graph is then fitted with the functional form

$$\frac{\sigma\left(E_{\mathrm{T}}^{\mathrm{rec}}/E_{\mathrm{T}}^{\mathrm{MC}}\right)}{\langle E_{\mathrm{T}}^{\mathrm{rec}}/E_{\mathrm{T}}^{\mathrm{MC}}\rangle} = \frac{a}{E_{\mathrm{T}}^{\mathrm{MC}}} \oplus \frac{b}{\sqrt{E_{\mathrm{T}}^{\mathrm{MC}}}} \oplus c \tag{5.4}$$

to obtain the jet resolution function, where the first term is due to fixed energy fluctuations in the jet, like electronic noise, pile-up and underlying event energy, the second term originates from the stochastic response of the calorimeter measurements, and the last term is the constant term from residual non-uniformities in the detector response. The terms are added quadratically to obtain the combined resolution. In Figure 5.9 the jet resolution functions are shown for the iterative cone algorithm with cone size $\Delta R = 0.5$ in the barrel, the endcap and the forward region, in the case of QCD di-jet samples from several bins in a wide range of $\hat{p_{\mathrm{T}}}$, with the thresholds $E_{\mathrm{T}} > 0.5\,\mathrm{GeV}$ and $E > 0.8\,\mathrm{GeV}$ applied on the input towers. In the barrel region $a = 5.6\,\mathrm{GeV}$, $b = 1.25\,\mathrm{GeV}^{1/2}$ and $c = 0.033$ are obtained from the fit with the resolution function (5.4).

In many physics studies resolution parametrizations are employed in statistical estimators. For the studies presented in Chapters 6 and 7, jet resolutions are used in covariance matrices of jet objects in a kinematic fit, which will be introduced in Section 5.5. In these studies with crowded multi-jet events, the jet resolutions will be different from the resolutions obtained from di-jet events, due to channel-specific properties like the underlying event. In addition, to use the resolution functions in practise, they have to be constructed as a function of the observable reconstructed jet energy, and they need to be obtained by comparison of the calorimeter jet with the corresponding generated parton, instead of the matched particle-level jet. In Figure 5.10 several jet resolution functions are shown, that are obtained in fully simulated $\mathrm{t\bar{t}}$ events, using the underlying-event input thresholds introduced in Section 4.1.4 and with low-luminosity pile-up collisions included. The functions have been determined for the whole $|\eta| < 2.4$ range inclusively, using the iterative cone algorithm with cone size $\Delta R = 0.5$.

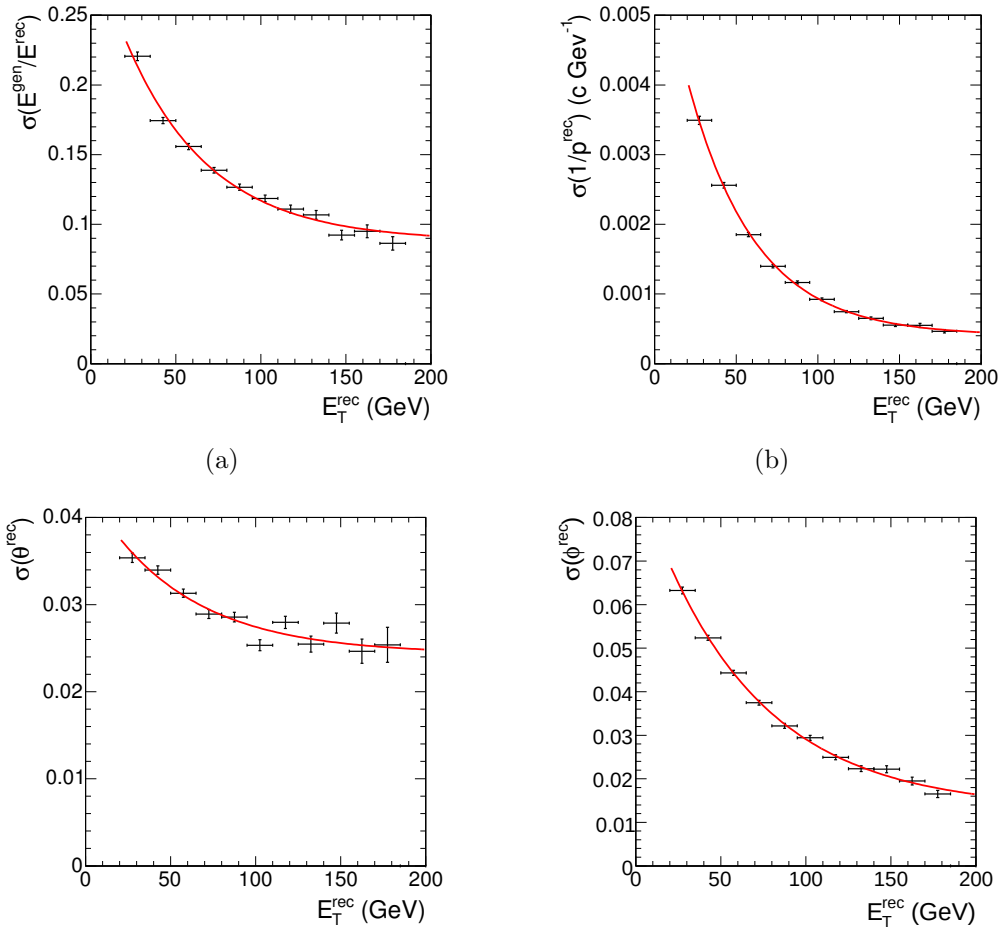


Figure 5.10: Absolute jet resolutions of $E^{\mathrm{gen}}/E^{\mathrm{rec}}$, $1/p^{\mathrm{rec}}$, θ^{rec} and ϕ^{rec}, obtained for iterative-cone jets with $\Delta R = 0.5$, reconstructed in $\mathrm{t\bar{t}}$ events with low-luminosity pile-up in the $|\eta| < 2.4$ range, applying the underlying-event input thresholds.

Uncertainties on the jet resolution play an important role in many analyses. At detector start-up the jet energy resolution is estimated to be accurate up to 20% based on test-beam data and simulation studies [147]. Once data arrive, jet resolutions can be constrained from the spread of the di-jet balance b, which has been introduced in Section 5.1.3. This method is limited by systematics, especially on the use of a veto on a third jet, and is expected to reach a relative accuracy of the jet resolution up to 10% [144]. Energy-flow methods, using tracker information to correct calorimeter measurements on a jet-by-jet basis, can give further improvements.

5.1.5 Association of jets with the signal vertex

The presence of pile-up collisions generates a sizable amount of additional soft contributions in the detector. Since these collisions typically involve soft-jet production from a primary vertex well separated from the hard interaction vertex, track-based

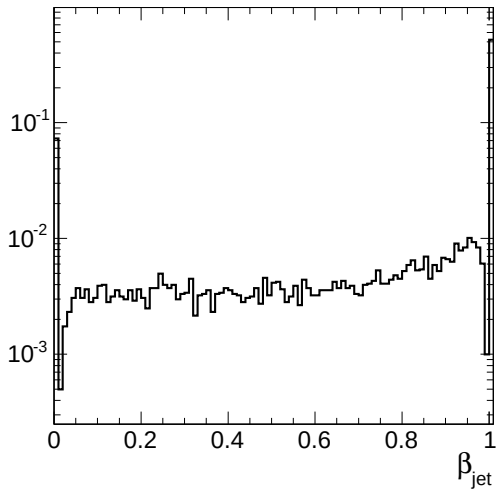


Figure 5.11: Normalized distribution of the β_{jet} parameter, defined in equation (5.5), for all reconstructed jets in $\mathrm{t\bar{t}}$ events including low-luminosity pile-up collisions.

criteria have been developed to identify these jets originating from pile-up vertices. For CMS, currently two simple methods have been explored, based on the calculation of the variables

$$\alpha_{\text{jet}} = \frac{\sum p_{\text{T}}(\text{PV tracks in jet})}{E_{\text{T}}(\text{jet})} \qquad \text{and} \qquad \beta_{\text{jet}} = \frac{\sum p_{\text{T}}(\text{PV tracks in jet})}{\sum p_{\text{T}}(\text{all tracks in jet})}. \tag{5.5}$$

In the original exploratory study [148], the association of the tracks to primary-vertex jets is done with respect to a hard isolated muon in the events. More generally the association of the tracks can be performed with respect to the reconstructed primary vertex associated to the hard interaction, as discussed in Section 4.1.2. For the jet reconstruction in this thesis, the β_{jet} discriminator is used, which ideally equals 1 for jets from the hard interaction, and vanishes for pile-up jets. This parameter is shown in Figure 5.11 for all jets in a sample of $\mathrm{t\bar{t}}$ events, which are fully simulated including low-luminosity pile-up collisions.

Jets not compatible with the primary vertex can be suppressed by placing a threshold on β_{jet}. Throughout the studies in this thesis jets have been selected with $\beta_{\text{jet}} > 0.04$. In Figures 5.12 and 5.13, the effect of the application of the described primary vertex constraint is shown for semi-leptonically decaying $\mathrm{t\bar{t}}$ events, respectively on the number of selected jets in the tracker acceptance, and on the p_{T} spectrum of these jets[1]. Jets originating from pile-up collisions are numerous, but indeed have small transverse momentum.

5.2 b-Jet Identification Tools

Heavy flavour jets, originating from c or b quarks, exhibit characteristics that allow to discriminate between these jets and others. In Section 3.3.4 several of these properties

[1]These events have been calibrated with the γ + jet corrections, which showed poor performance extrapolated to small p_{T}. The calibration method is however of no importance here.

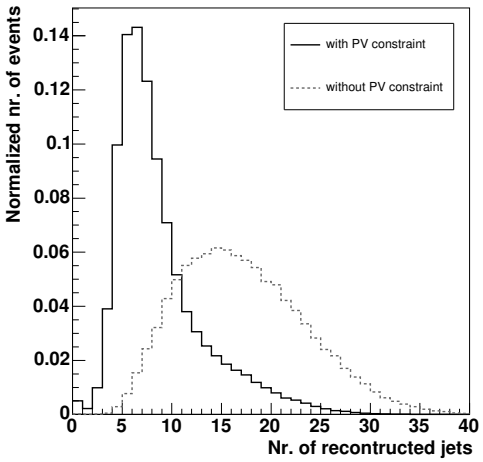


Figure 5.12: The effect of the jet–vertex association constraint $\beta_{\text{jet}} > 0.04$ on the number of reconstructed jets for semi-leptonically decaying $\mathrm{t\bar{t}}$ events including low-luminosity pile-up collisions.

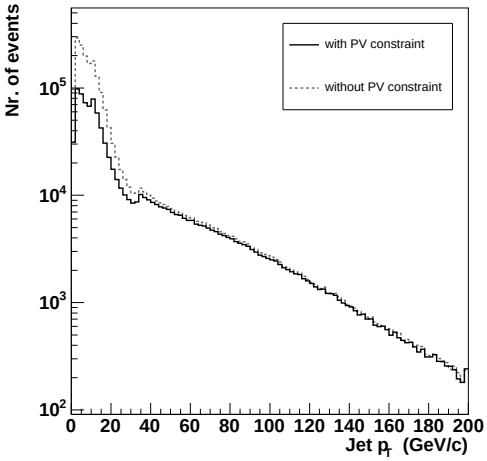


Figure 5.13: The effect of the jet–vertex association constraint $\beta_{\text{jet}} > 0.04$ on the jet p_{T} spectrum for semi-leptonically decaying $\mathrm{t\bar{t}}$ events including low-luminosity pile-up collisions.

are described. The production of heavy long-lived hadrons in the fragmentation of c and b quarks, causes the presence of displaced secondary decay vertices and therefore of tracks not compatible with the primary vertex. In addition, a significant fraction of heavy-flavour jets contains a muon or electron from weak decays in the fragmentation. In this section several algorithms and their performance are presented that tag jets as originating from b quarks, based on diverse experimental characteristics and techniques.

When studying the performance of b-identification algorithms, the results can be presented either in the algorithmic or in the physics definition of the jet flavour. As discussed in Section 3.3.2, the distinction arises from the possible presence of heavy-flavour particles inside gluon jets. For general performance studies of b-tagging algorithms, it is natural to use the algorithmic definition, in which gluon jets containing $\mathrm{b\bar{b}}$ or $\mathrm{c\bar{c}}$ pairs are catalogued as the corresponding heavy-flavour jet. For results on specific channels, the physics definition can be used, which makes the visualization possible of the effect of the heavy-flavour content in gluon jets compared to light-quark jets.

To attribute a flavour to a reconstructed jet, a matching needs to be made between the jet and the originating parton, by analysing the generated parton content in a cone around the jet direction. For the presented performance studies, a reconstructed jet is associated to a generated parton from the hard interaction or the parton shower, if they are within a distance $\Delta R < 0.4$. Unassociated jets, produced for example in pile-up collisions, are discarded.

5.2.1 Track impact-parameter tagging

The presence of a b hadron with significant lifetime in jets originating from a b quark can be exploited in different ways to identify b jets. Tracks coming from a b-hadron

decay originate from a displaced vertex, and have therefore a large impact parameter with respect to the primary vertex of the collision.

The impact parameter of a track, being the smallest distance between the track and the considered vertex, can be computed either two-dimensionally in the transverse plane, or by using full three-dimensional information. In the former case the calculation of the impact parameter can be done analytically, since the trajectory is circular in the transverse plane, while in the latter case an approximation is made. First the point of closest approach of the track to the jet direction is determined, after which the track is extrapolated linearly from that point. Then the three-dimensional impact parameter is calculated as the point of closest approach between the linearized track and the primary vertex.

As discussed in Section 4.1.2, the accuracy of the transverse position reconstruction of primary vertices is of the order of the transverse beam size of $15\,\mu$m, while the longitudinal accuracy is typically of the order of $20\,\mu$m. The accuracy on the transverse and longitudinal impact parameters are both expected to be about $100\,\mu$m for central tracks of $1\,\text{GeV}/c$. Hence, the addition of longitudinal information can add significantly to the impact-parameter discrimination power. In addition, the impact parameter acquires a positive or negative sign, depending on whether the track originates from a decay downstream or upstream from the primary vertex, respectively. This sign is calculated as the sign of the scalar product of the impact-parameter direction with the jet direction.

To take into account the experimental resolutions, the significance of the track impact parameter is commonly used, which is defined as the ratio between the impact parameter and its uncertainty. In Figure 5.14 the three-dimensional signed impact-parameter significance distribution is shown for tracks from udsg, c and b jets in the algorithmic flavour definition in jets from $\mathrm{t\bar{t}}$ events. Significance values of tracks from udsg jets are almost symmetrically distributed, consistent with originating from the primary vertex within the experimental resolution. Tracks from b and c jets, on the other hand, show a clearly asymmetric distribution, reflecting the long lifetimes of beauty and charm hadrons. The negative tails of the distributions are due to resolution effects and fake tracks, and are in principle flavour independent for jet samples with similar kinematic properties. The difference between the flavours for negative impact-parameter significance is understood to be arising from the jet angular resolution, which can generate a wrong sign assignment of the impact-parameter significance for tracks where the impact parameter is almost perpendicular to the jet direction.

Two impact-parameter based methods, the track-counting method and a probabilistic method, have been implemented for CMS. In addition, an adapted version of the track-counting method has been developed for use in the online selection.

Track-counting b-tagging Track counting is a simple and therefore fast and robust method for b-jet identification [149]. For each selected track in a jet, the three-dimensional impact-parameter significance is computed. In addition, the distance of closest approach of a track to the jet axis is required to be smaller than 0.7 mm. The jet is considered tagged as b jet if the number of tracks with impact-parameter significance exceeding a given threshold is greater or equal than a given value n. Hence, ordering the tracks with decreasing signed impact-

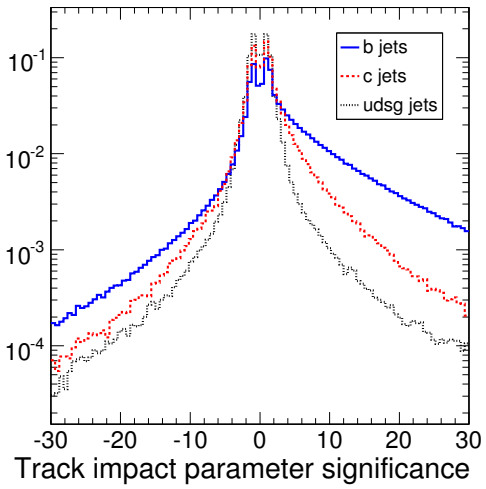


Figure 5.14: Three-dimensional signed impact-parameter significance distribution for tracks from udsg, c and b jets in $t\bar{t}$ events.

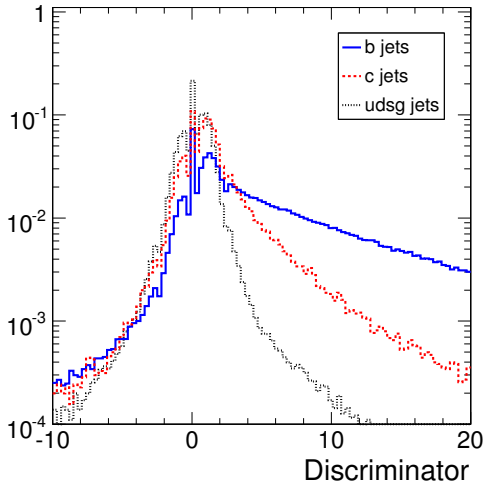


Figure 5.15: Three-dimensional signed impact-parameter significance of the n'th track with $n = 2$, for tracks from udsg, c and b jets in $t\bar{t}$ events.

parameter significance, the significance of the n'th track serves as discriminator for this algorithm. In Figure 5.15 the distribution of this discriminator is shown in the case of $n = 2$ for jets from $t\bar{t}$ events. Jets for which less than $n = 2$ tracks are reconstructed are given the discriminator value 0, causing the central peak in the figure.

Probability b-tagging This method computes a jet probability, combining signed track probabilities for all the tracks associated to a jet [149]. These track probabilities express the compatibility of the track with the primary vertex. They are calculated with respect to the negative part of the impact-parameter significance distribution. This negative tail consists mainly of tracks from the primary vertex or from fake tracks. Hence, it is possible to calibrate the unsigned significance distribution R for primary-vertex tracks with the negative tail of the signed significance distribution obtained with tracks from data. The signed probability for a track to come from the primary vertex is defined as the integral of the distribution R, multiplied by the sign of the significance S,

$$P_{\mathrm{tr}}(S) = \mathrm{sgn}(S) \int_{|S|}^{+\infty} \mathrm{d}x R(x). \tag{5.6}$$

The resulting probability distribution, shown in Figure 5.16 for jets from $t\bar{t}$ events, is supposed to be flat for tracks coming from the primary vertex, and is positive and concentrated near 0 for tracks with large impact-parameter significance. The deviation from flat probability stems from the use of a distribution R obtained from a Monte-Carlo simulation of QCD jets rather than from $t\bar{t}$ data.

The combined jet probability P_{jet} is calculated as the probability that a set of N tracks, presumed to come from the primary vertex, produces the observed or any

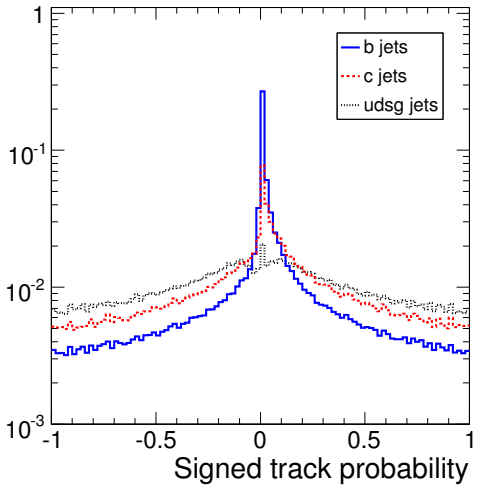


Figure 5.16: Distribution of the signed track probability $P_{\mathrm{tr}}(S)$, defined in (5.6), for tracks from udsg, c and b jets in $\mathrm{t\bar{t}}$ events.

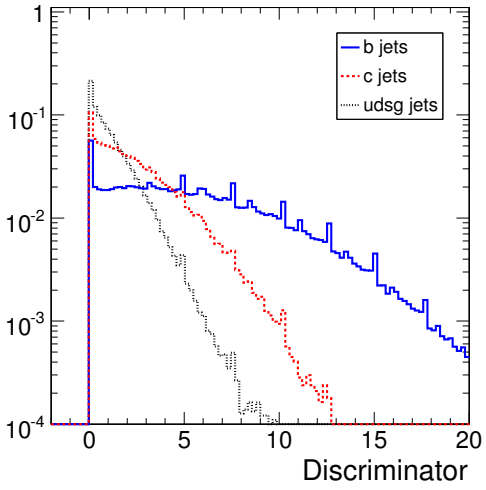


Figure 5.17: Distribution of the track-probability algorithm's discriminator, defined in (5.7), for tracks from udsg, c and b jets in $\mathrm{t\bar{t}}$ events.

less likely value of track probability. It is given by

$$P_{\mathrm{jet}} = \sum_{j=0}^{N-1} \frac{e^{-\mathcal{L}_{\mathrm{jet}}} (\mathcal{L}_{\mathrm{jet}})^j}{j!} \qquad \text{with} \qquad \mathcal{L}_{\mathrm{jet}} = -\sum_{i=1}^{N} \ln \hat{P}_{\mathrm{tr}}(S_i), \tag{5.7}$$

where $\hat{P}_{\mathrm{tr}}(S) = P_{\mathrm{tr}}(S)/2$ for tracks with positive impact parameter S, and $\hat{P}_{\mathrm{tr}} = 1 + P_{\mathrm{tr}}/2$ for negative S. Figure 5.17 shows the distribution of the algorithm's discriminator $-\log P_{\mathrm{jet}}$ for b, c and udsg jets in $\mathrm{t\bar{t}}$ events. The peak structure in the distribution is understood to be due to a minimal requirement on the track probability (5.6), which is needed because the distribution R consists of a finite number of entries. With tracks from data this distribution can be refined such that the minimal-probability requirement is taken even lower. The entries in the peaks would then be distributed at higher values of the discriminator.

b-tagging in the HLT Algorithms in the context of the High-Level Trigger need to comply with restrictive time constraints, and should be robust and only weakly dependent on calibration. Therefore, for online b identification, the simplest methods are preferred. Two algorithms have been set up, that are meant to function at different stages of the HLT [150]. The first algorithm acts early in the HLT, coping with rates of di-jet events up to 1 kHz. The emphasis is on speed, and therefore track seeds from the pixel layers are used rather than the full track reconstruction. A drawback of the use of pixel tracks is the poor impact-parameter resolution, expected to be $\approx 150\,\mu$m in the transverse plane and $120\,\mu$m longitudinally for $1\,\mathrm{GeV}/c$ tracks. In addition, the primary-vertex finding needs to be performed differently. A simple longitudinal algorithm with the pixel seeds is complemented with the beam spot finding, yielding a good efficiency and accuracy. This algorithm uses then the track-counting method

with $n = 2$, tuned to reach a rejection factor up to five. Events passing this first algorithm are fed into a second algorithm, that uses full track reconstruction regionally around the jets identified as b jets in the previous step. The track-counting algorithm is applied with $n = 3$, reaching another rejection of a factor four.

5.2.2 Secondary-vertex tagging

The excellent performance of track reconstruction in CMS makes it possible to efficiently reconstruct secondary vertices from long-lived hadron decays in heavy-flavour jets. Information on reconstructed secondary vertices can then be used to perform b-jet identification. For CMS a combined algorithm has been developed, exploiting different topological and kinematic vertex variables, together with track impact-parameter significances, to distinguish b-quark jets from non-b jets [151].

As an input to secondary-vertex reconstruction, tracks are required to have a transverse impact parameter smaller than 2 mm, with respect to the primary vertex they are associated to. This threshold suppresses tracks from longer-distance effects, like V^0 decays, photon conversions and nuclear interactions in the beampipe and first pixel layers. Reconstructed secondary vertices from these tracks are filtered with a quality criterion. The transverse distance L_t between a secondary and the primary vertex needs to be larger than 100 μm, but must not exceed 2.5 cm, and its significance L_t/σ_{L_t} should exceed 3. Moreover, the invariant mass of charged particles associated to a secondary vertex must not exceed 6.5 GeV/c^2 and should not be compatible with the K^0 mass within a window of 50 MeV.

Based on the result of the secondary-vertex reconstruction, three jet categories are defined. The first category contains jets that have at least one reconstructed secondary vertex that passed all selection criteria. For this category all accepted vertices in a jet are used in the computations further on. The second category groups jets where no secondary vertex is found, but in which at least two tracks are present, with a signed transverse impact-parameter significance larger than 2. These tracks, not fully compatible with the primary vertex, are treated as originating from a so-called pseudovertex. The third category contains all other jets.

Several variables are identified that allow to discriminate between heavy- and light-flavour jets:

- the invariant mass of the tracks associated to the secondary vertex;
- the multiplicity of tracks associated to the secondary vertex;
- the significance L_t/σ_{L_t} of the transverse distance between the primary and the secondary vertex;
- the ratio of the energy of the charged particles associated to the secondary vertex and the energy of all charged particles within the jet;
- the rapidities $y = \frac{1}{2}\ln\left(\frac{E+p_\parallel}{E-p_\parallel}\right)$ of all tracks associated to the secondary vertex with respect to the jet direction.

- the tracks' transverse impact-parameter significances;
- the transverse impact-parameter significance of the track which moves the invariant mass, computed with all tracks with larger or equal significance, in the region of charmed hadrons; this variable is useful to improve charm rejection.

If a secondary vertex is reconstructed, all variables are combined, while in the pseudovertex case all but the transverse distance are used. For the jet category without reconstructed vertex only the tracks' impact-parameter significances are taken into account.

The construction of a combined discriminator is done starting from the probability density functions $f_\alpha^{\mathrm{b,c,udsg}}(x_i)$ of the discriminating variable x_i in jet category α, for the different jet flavours b, c and udsg. The functions $f_\alpha^{\mathrm{b,c,udsg}}(x_i)$ are obtained from simulated QCD jet samples, and depend on the transverse jet energy and pseudorapidity, but also on the choice of jet definition for some of the variables. For the determination of the functions $f_\alpha^{\mathrm{b,c,udsg}}(x_i)$ the iterative-cone jet algorithm has been used with cone size $\Delta R = 0.5$, which is also the choice in the studies throughout this thesis.

The probability distributions of all variables are multiplied,

$$P^{\mathrm{b,c,udsg}} = f^{\mathrm{b,c,udsg}}(\alpha) \times \prod_i f_\alpha^{\mathrm{b,c,udsg}}(x_i), \tag{5.8}$$

taking into account the sample-specific weight $f^{\mathrm{b,c,udsg}}(\alpha)$ for each jet category α. The final discriminator is defined as[2]

$$\zeta_{\mathrm{b}} = \log\left[f_{\mathrm{c}}\frac{P^{\mathrm{b}}}{P^{\mathrm{c}}} + f_{\mathrm{udsg}}\frac{P^{\mathrm{b}}}{P^{\mathrm{udsg}}}\right], \tag{5.9}$$

where f_{c} and f_{udsg} are the expected priors for the c and udsg content in non-b jets, related by $f_{\mathrm{c}} + f_{\mathrm{udsg}} = 1$. By default the free parameter for the c content is chosen as $f_{\mathrm{c}} = 0.25$, which corresponds to the expected fraction of c jets in hadronic decays of W bosons [30], as found in the context of top-quark pair events.

The probability distribution of the discriminator ζ_{b} is shown in Figure 5.18 for b, c and udsg jets in semi-leptonic $\mathrm{t\bar{t}}$ events, according to the algorithmic flavour definition. A b-jet probability $P_{\mathrm{b}}(\zeta_{\mathrm{b}})$ is derived from these distributions. It is defined as the ratio $S/(S+B)$, where S is the number of b jets in a sample at a given ζ_{b}, and B the corresponding number of non-b jets. The fitted distribution of P_{b} as a function of ζ_{b} is shown in Figure 5.19.

5.2.3 Soft-lepton tagging

The branching fraction for the direct and cascade decays of b hadrons into electrons and muons is large, about 19% for each lepton family, as discussed in Section 3.3.4. This can be exploited by identifying heavy-flavour jets based on the reconstruction of a non-isolated electron or muon in the jet [114]. Electron and muon candidates are

[2] In the most recent implementation of this algorithm, the definition has been changed to a likelihood ratio, to confine the value of the discriminator to the $[0, 1]$ interval.

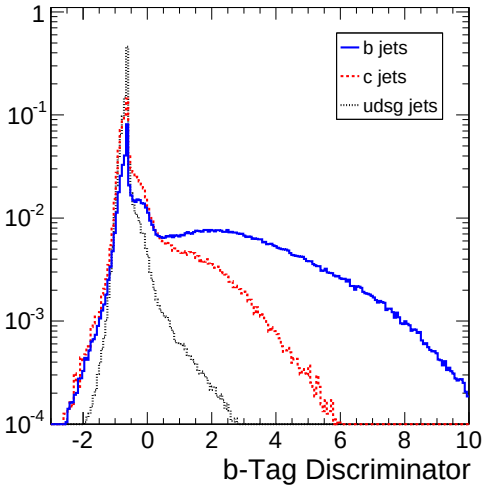


Figure 5.18: Distribution of the combined b-tag discriminator ζ_b, defined in (5.9), for udsg, c and b jets in semi-leptonic $t\bar{t}$ events.

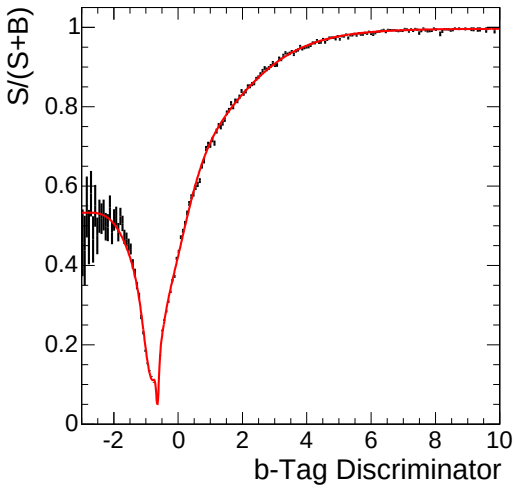


Figure 5.19: Distribution of the b-jet probability $P_b(\zeta_b)$ in semi-leptonic $t\bar{t}$ events, as a function of the combined b-tag discriminator ζ_b.

first selected as clusters in the electromagnetic calorimeter and tracks in the muon detectors, respectively, and are then associated to reconstructed tracks in the central tracker to ensure an accurate determination of the lepton's momentum and direction. Reconstruction impurities arise from light-meson decays to muons ($\pi^{\pm} \to \mu^{\pm}\nu_{\mu}$, $K^{\pm} \to \mu^{\pm}\nu_{\mu}X$), by photon conversions to electron pairs, from fakes, etc. In this section the specific electron- and muon-reconstruction methods for the soft-lepton algorithm are briefly described. A more elaborate description of the standard offline electron and muon reconstruction in CMS can be found in Section 5.3.

A non-standard approach is used for electron identification, geared towards reconstruction of soft electrons inside jets. Energy clusters in the ECAL are matched with tracks, extrapolated to the calorimeter surface, which satisfy $p_T > 2\,\mathrm{GeV}/c$. Because of the differences in shower development between electromagnetic and hadronic showers, the cluster-shape information can be used as a discriminator. Also the repartition of the deposited energy between the ECAL and the HCAL can be used to distinguish electrons from hadrons. In addition, the individual crystals within a cluster are correlated with one another, which justifies the use of the covariances of the cluster energy distribution. Finally, also the ratios of the clusters' energy and tracks' momenta suppress backgrounds. Nine of these discriminating variables in total are used as input to a neural network. A threshold on the output of the neural network, trained on non-isolated electrons in $t\bar{t}$ signal samples and on QCD background, is used to select showers associated to signal electrons.

Muon identification for soft-lepton b-tagging is performed with the standard muon-reconstruction algorithm, described in Section 5.3.2. To match a reconstructed muon track to tracks contained in a jet, an additional requirement is imposed on both tracks to share 70% of the reconstruced hits. The main limitation of this approach is the low efficiency achieved for muons with low p_T, because of the strong bending in the

magnetic field. Muon reconstruction inside jets is further complicated by the matching of the track in the muon chambers to the central tracks contained in dense jets. An alternative approach [85] reconstructs muons from the inside outwards, taking into account additional calorimeter information and unassociated muon detector hits. This recently-developed method greatly improves muon reconstruction in jets, and will be used in future implementations.

Leptons from heavy-hadron decays in jets possess kinematic properties that allow them to be distinguished from background leptons. Four discriminating variables are identified:

- the lepton transverse momentum $p_{\mathrm{T}}^{\mathrm{rel}}$, relative to the lepton-excluded jet axis, which is calculated as the p_{T}-weighted average of the other charged tracks inside the jet;
- the distance ΔR between the lepton and the lepton-excluded jet axis;
- the ratio of the lepton track momentum and the jet energy;
- the significance of the three-dimensional impact parameter of the lepton with respect to the event's primary vertex.

In the case of muons also the jet energy and pseudorapidity are taken into account.

The variables are combined with a neural-network technique into a single observable, which can be used to balance the efficiency and purity of the algorithm. The neural network is trained for electrons and muons separately with b jets from $\mathrm{t\bar{t}}$ events as signal and non-b QCD jets as background. In Figures 5.20 and 5.21 the distributions of the resulting discriminators are shown for jets in $\mathrm{t\bar{t}}$ events, respectively in the case of electrons and muons. Improvements on separation power can be envisaged by taking into account the various possible origins of the lepton. In the current implementation the abundance of leptons from the various direct or cascade decays is fixed by the b-, c- and udsg-jets content in the neural-network training samples.

5.2.4 Performance of b-identification algorithms

Each of the presented b-identification algorithms returns a single discriminating variable for each reconstructed jet, optimized for the iterative-cone jet reconstruction with cone size $\Delta R = 0.5$, as used throughout the thesis. This discriminator is used to test the hypothesis that a jet originates from a b quark. Two types of errors can be made in this test. The hypothesis can be rejected while it is true, or it can be accepted in case it is wrong. These errors are reflected respectively in the identification efficiency of true b jets and the mis-identification efficiency of non-b jets. By imposing a cut on the discriminating observable, the working point of the considered algorithm is chosen, fixing at once the b-tagging and mis-identification efficiency. The algorithm's performance can be scanned by changing the cut on the observable, and by showing the dependence of the non-b-jet efficiencies as a function of the b-jet efficiency. In Figures 5.22, 5.23 and 5.24 the b-jet identification performance is shown, respectively

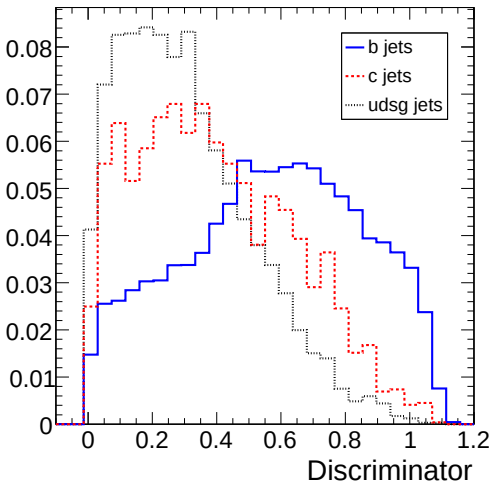


Figure 5.20: Neural-network discriminator for the soft-electron b-tagging method in $t\bar{t}$ events.

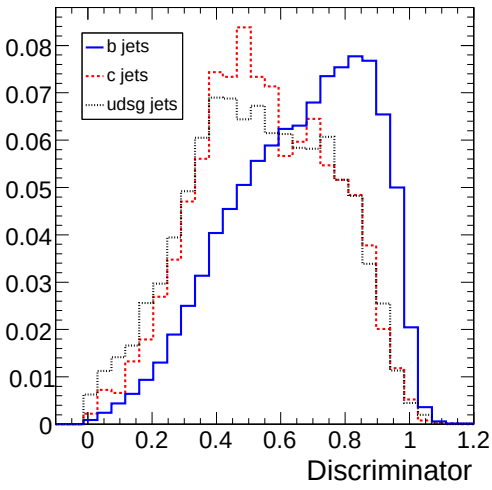


Figure 5.21: Neural-network discriminator for the soft-muon b-tagging method in $t\bar{t}$ events.

for the track-counting, the track-probability and the combined secondary-vertex algorithms, for reconstructed udsg, c and b jets in $t\bar{t}$ events, applying the algorithmic flavour definition.

All three of the considered algorithms show similar behaviour[3]. At a typical working point of 50% b-jet tagging efficiency, a c-jet rejection factor of about 10 is obtained, together with a uds and gluon rejection of ~ 100. The b-tag efficiencies of the uds-jets and the gluon jets show very similar behaviour, which is expected from the choice of the algorithmic flavour definition. The remaining difference is attributed to the kinematical differences in $t\bar{t}$ events between the two jet types, since the discrepancies are shown to dissapear for uds and gluon jets from kinematically comparable QCD jet samples [151]. Finally, the stagnation of the rejection power at very low b-tagging efficiency for the track-probability algorithm is due to statistical limitations on the simulated input to the algorithm.

The dependence of the b-tagging performance on the transverse momentum and pseudorapidity of the jet's initial parton are illustrated, respectively in Figures 5.25 and 5.26 for the b-jet identification efficiencies, and in Figures 5.27 and 5.28 for the mis-identification rates. These figures show the efficiencies for a fixed average b-jet tagging efficiency of 50%. The results are obtained with the combined secondary-vertex b-tagging algorithm, on jets from $t\bar{t}$ events, in the algorithmic flavour definition. The performance degrades at larger pseudorapidities, mainly because of the increase in traversed material and the worse resolutions on tracking. The decrease in performance for large transverse momenta is caused by an enhanced track multiplicity from fragmentation, increasingly boosted hadron decays and the more difficult pattern recognition associated with dense jets. The steep fall in performance at relatively low transverse momenta is mainly due to increased multiple scattering, which results in less separation

[3]The most recent implementation of the combined secondary-vertex algorithm performs significantly better than the impact-parameter based algorithms [151, 152].

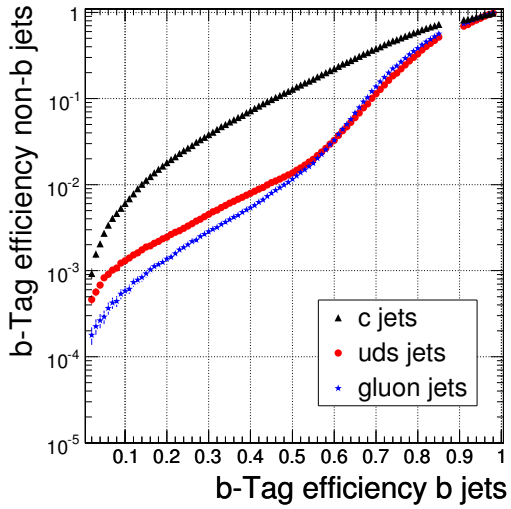


Figure 5.22: Performance of the track-counting b-tag algorithm with $n = 2$, for gluon, uds, c and b jets in $t\bar{t}$ events, in the algorithmic flavour definition.

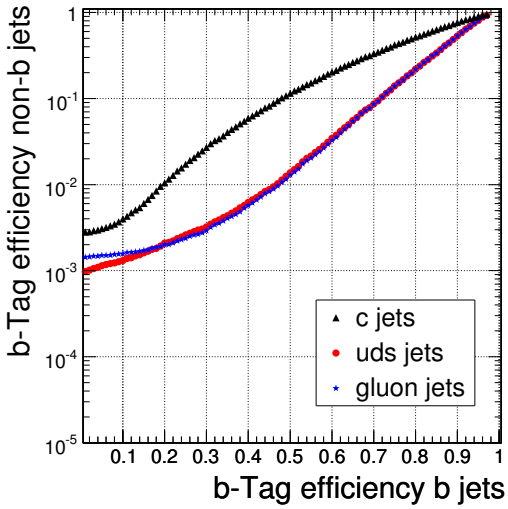


Figure 5.23: Performance of the track-probability b-tag algorithm, for gluon, uds, c and b jets in $t\bar{t}$ events, in the algorithmic flavour definition.

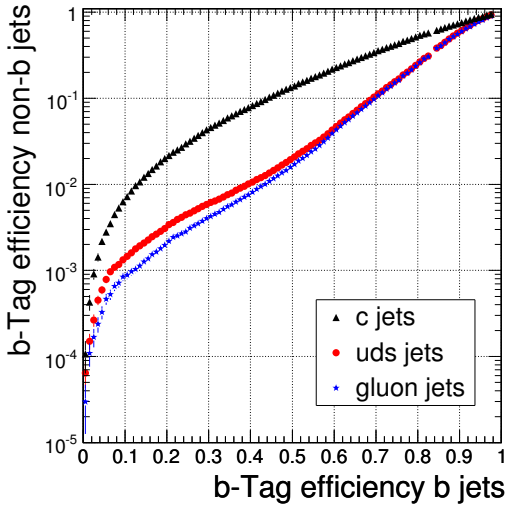


Figure 5.24: Performance of the combined secondary-vertex b-tag algorithm, for gluon, uds, c and b jets in $t\bar{t}$ events, in the algorithmic flavour definition.

power between primary and secondary vertices. Although for c rejection, the p_T and η dependence is still moderate, the udsg mis-identification dependence is large. An order of magnitude difference in performance is found between the most central and the most forward jets, and almost as much between jets with optimal $p_T \approx 70\,\text{GeV}/c$ and the softest and hardest part of the jet-p_T spectrum. Additionally, the b-jet identification rate decreases for jets in the same range of low and high transverse momentum and in the forward region. The overall b-tagging performance of the algorithm is hence optimal for central jets with $p_T \approx 100\,\text{GeV}/c$.

The performance of the soft-lepton b-tag algorithms are investigated in the same

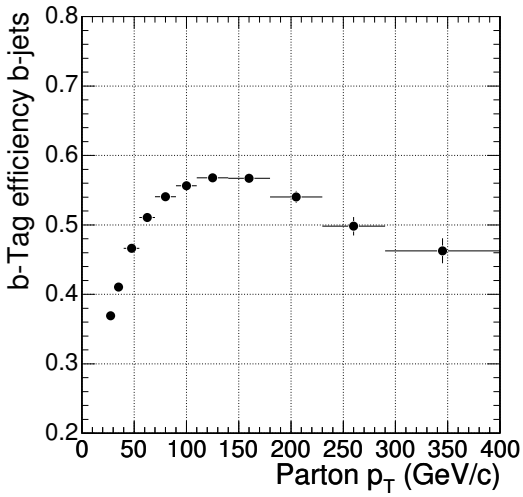


Figure 5.25: b-Jet identification efficiency as a function of the p_{T} of the originating b parton in $\mathrm{t\bar{t}}$ events, for the combined secondary-vertex algorithm in the algorithmic flavour definition, at a fixed average b-jet tag efficiency of 50%.

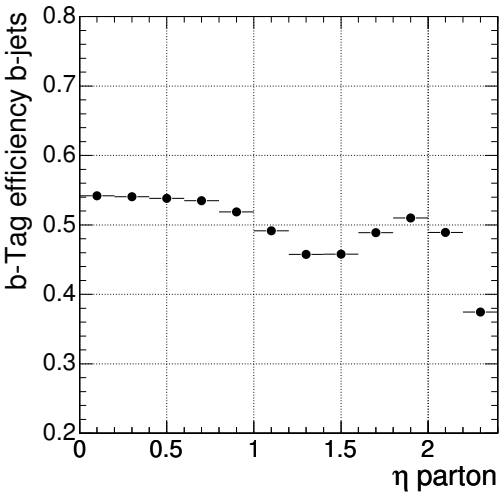


Figure 5.26: b-Jet identification efficiency as a function of η of the originating b parton in $\mathrm{t\bar{t}}$ events for the combined secondary-vertex algorithm in the algorithmic flavour definition, at a fixed average b-jet tag efficiency of 50%.

way, by scanning over the algorithms' discriminators. The distributions of the non-b-jet mis-identification rate versus the b-jet tag efficiency are shown in Figures 5.29 and 5.30, respectively for the electron and muon algorithms. The efficiency to identify b-jets is given relative to the complete sample of b jets, and is limited by the total leptonic branching fractions of b hadrons, discussed in Section 3.3.4, and by the reconstruction efficiency of the soft non-isolated electrons and muons. The different behaviour of the performance with respect to the jet types is due to the different nature of the backgrounds to electron and muon identification in jets.

5.2.5 Correlations between b-identification algorithms

Although the use of the soft-lepton b-identification algorithms is limited by the total leptonic branching fraction of b hadrons, these algorithms are nevertheless interesting because of the expected low correlation with the lifetime-based algorithms. In Table 5.1, the linear correlation coefficients are shown between the discriminators of the five algorithms considered in this section. Only those jets are considered that originate from b-quarks. In the case of the soft-lepton algorithms, only those jets are used in which a soft lepton of the indicated flavour is identified.

As expected from the similar physics exploited, the track-counting, track-probability and combined secondary-vertex algorithms are strongly correlated. Due to observed anti-correlations for the negative tails of the track-counting and combined algorithms, the correlations are in practise even higher. The soft-electron and soft-muon algorithms, for which a rather uncorrelated behaviour is expected, are nevertheless found to exhibit important correlations with the three lifetime-based algorithms. These correlations

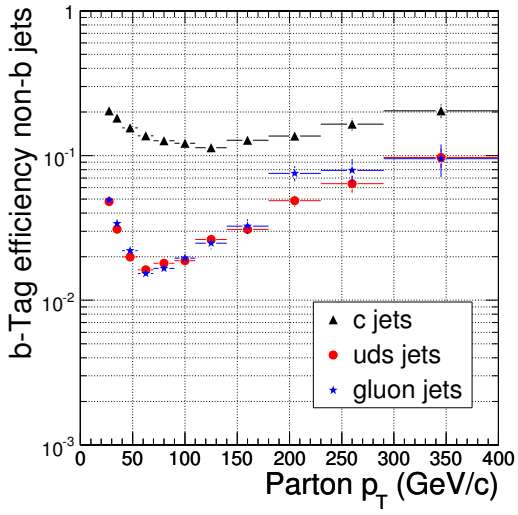


Figure 5.27: Mis-identification efficiency as a function of the p_T of the originating parton in $t\bar{t}$ events, for the combined secondary-vertex algorithm in the algorithmic flavour definition, at a fixed average b-jet tag efficiency of 50%.

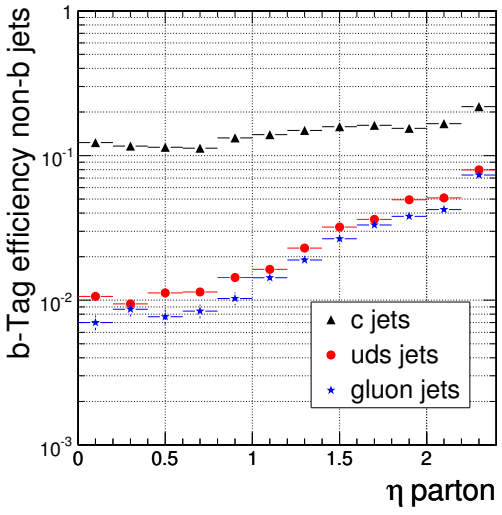


Figure 5.28: Mis-identification efficiency as a function of η of the originating parton in $t\bar{t}$ events for the combined secondary-vertex algorithm in the algorithmic flavour definition, at a fixed average b-jet tag efficiency of 50%.

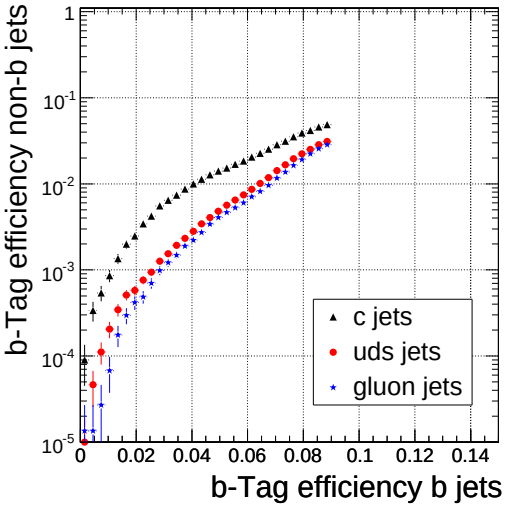


Figure 5.29: Performance of the soft-electron b-tag algorithm, for gluon, uds, c and b jets in $t\bar{t}$ events, in the algorithmic flavour definition.

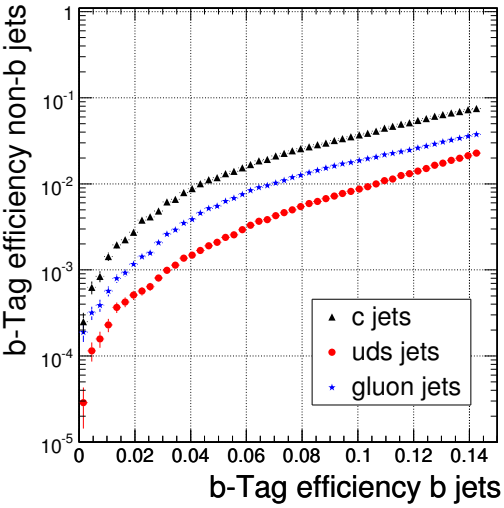


Figure 5.30: Performance of the soft-muon b-tag algorithm, for gluon, uds, c and b jets in $t\bar{t}$ events, in the algorithmic flavour definition.

increase when when the b-tagging thresholds are tightened. Such a behaviour suggests a too strong weight of the electron's and muon's impact-parameter significance in the neural network, and could be reduced by the exclusion of this observable from the algorithm for purposes where a minimal correlation is desirable. The correlation between the soft-lepton algorithms is not significant due to the very low number of jets

	Track counting	Track probability	Combined sec. vtx.	Soft electron	Soft muon
Track counting	1	0.83	0.57	0.16	0.14
Track probability	-	1	0.65	0.22	0.28
Combined sec. vtx.	-	-	1	0.33	0.41
Soft electron	-	-	-	1	(0.41)
Soft muon	-	-	-	-	1

Table 5.1: Linear correlations coefficients between the discriminators of the various considered b-identification algorithms, for jets originating from b quarks.

with both a reconstructed electron and muon.

The possible low correlations between the soft-lepton algorithms and the ones exploiting the long lifetime of b hadrons open the possibility for improvement of the b-tagging performance by combining the results from the two types of algorithms. This is in particular interesting for c-jet rejection, since the soft-lepton algorithms obtain high purities with respect to c jets. Another application can be found in the calibration of one type of algorithm on b-jet samples that are selected with another type of algorithm. This last possibility is also briefly discussed in Section 6.4.

5.3 Electron and Muon Identification

The building blocks to perform reconstruction of electrons and muons have been introduced in Chapter 4. Electron reconstruction is based on the association of a supercluster with a track, while muon reconstruction matches tracks from the standalone muon reconstruction with compatible tracks in the silicon tracker. In this section the CMS offline electron- and muon-reconstruction algorithms are described. The performance is illustrated with leptons, being electrons or muons, from top-quark pair events, with low-luminosity pile-up included. After the discussion on lepton reconstruction, a powerful likelihood ratio method is presented for the selection and identification of isolated electrons and muons from W-boson decays in $\mathrm{t\bar{t}}$ events. More details on this method, like for example the effect of omitting pile-up collisions in the simulation, can be found in [153].

5.3.1 Offline electron reconstruction

Electron reconstruction in CMS is complicated by the interactions in the tracker material in front of the electromagnetic calorimeter. Electrons traversing the tracker layers radiate bremsstrahlung photons, such that the electron energy reaches the ECAL azimuthally spread because of the solenoidal magnetic field. The amount of radiated energy can be very large. Electrons with $p_{\mathrm{T}} = 10\,\mathrm{GeV}/c$ uniformly distributed in η and ϕ, radiate about 50% of their energy in 50% of the cases, and up to 95% in 10% of the cases. The bremsstrahlung photons can in turn convert into $\mathrm{e^+e^-}$ pairs, which may lead to electromagnetic shower patterns and significant energy loss in the tracker.

Offline reconstruction of electrons in CMS starts from the ECAL superclusters, as detailed in Section 4.1.4. These superclusters are used to drive the pixel-seed finding for the primary electron tracks in a similar way as in the HLT algorithm, which is briefly discussed in Section 4.3.2. The extrapolation of the cluster position through the magnetic field towards the beamline, takes advantage of the fact that the impact point of the electron and associated bremsstrahlung photons, obtained from a cluster-energy weighted average, coincides with the impact point of a non-radiating electron with the same momentum, provided all photons are collected. For both charge hypotheses, the first and second matched pixel hits are sought consecutively tightening the $\Delta\phi$ and Δz search windows. These windows are chosen wider than in the HLT algorithm, however, to increase efficiency, especially for low transverse-momentum electrons.

Trajectory building, as described in Section 4.1.2, is used in an adapted way for electron tracking. Emphasis is put on the early stages of the electron-track evolution, which contains the most significant information on the electron's initial momentum and direction. Recent developments with the non-linear Gaussian-Sum Filter reconstruction technique [154] allow to efficiently collect track hits up to the end of the tracking volume, and result in a significantly increased electron-tracking performance at low transverse momentum.

Reconstructed superclusters and associated electron tracks are in the next step combined into electron candidates. Referring to the supercluster with "SC" and to the track at either the production vertex or the calorimeter surface as "TR", a candidate is required to satisfy $E^{\mathrm{SC}}/|\vec{p}|^{\mathrm{TR}} < 3$, $|\eta^{\mathrm{SC}} - \eta^{\mathrm{TR}}| < 0.1$, $|\phi^{\mathrm{SC}} - \phi^{\mathrm{TR}}| < 0.1$, and $H^{\mathrm{SC}}/E^{\mathrm{SC}} < 0.2$, where H^{SC} denotes the energy deposited in the HCAL just behind the electromagnetic seed cluster. For each obtained candidate the information of the track's momentum and the cluster's energy is combined to improve the estimate of the electron momentum at the interaction vertex. The resolution on the energy measured by the ECAL improves with increasing electron energy, while the energy from the track measurement shows a worse resolution for rising energies. Below electron energies of about 20 GeV, improvement can be reached by combining tracker and ECAL information. At the higher electron energies used in the analyses in this thesis, however, the ECAL provides by far the best energy measurement, and the electron's four-vector is determined from the track's direction and the supercluster's energy.

The list of electron candidates reconstructed by the above recipe still needs to be cleaned for duplicates, which arise when multiple reconstructed tracks get associated with the same supercluster. Cleaning of these duplicates is performed by erasing all candidates for a given supercluster, except for the one with $E/|\vec{p}|$ closest to 1. From the remaining list of electron candidates the reconstruction efficiency of isolated electrons from W decays in a semi-leptonic $\mathrm{t\bar{t}}$ sample is determined in the range $|\eta| < 2.4$. This efficiency is shown in Figures 5.31 and 5.32, as a function of the generated electron p_{T} and η respectively.

5.3.2 Offline muon reconstruction

The CMS global muon reconstruction consists in extending the muon trajectories from the standalone muon reconstruction, described in Section 4.1.6, to include hits in the central silicon tracker. Starting from a standalone muon, without an applied beam

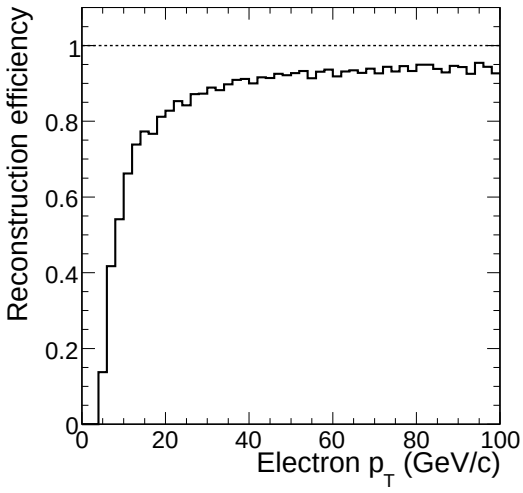


Figure 5.31: Reconstruction efficiency of the isolated electron from semi-leptonic $t\bar{t}$ events as a function of the generated electron p_T.

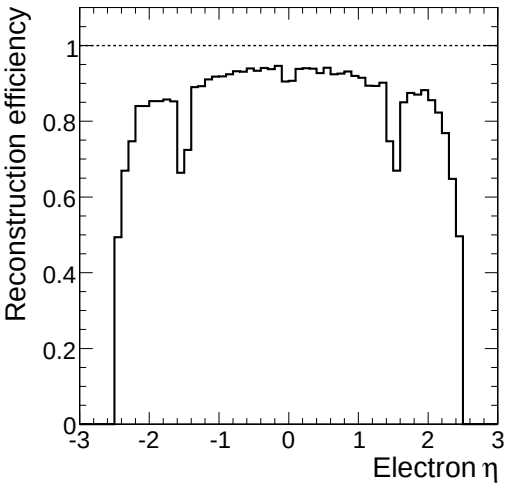


Figure 5.32: Reconstruction efficiency of the isolated electron from semi-leptonic $t\bar{t}$ events as a function of the generated electron η.

constraint, the trajectory is extrapolated from the innermost muon station to the outer tracker surface, taking into account energy loss in the material and the effect of multiple scattering. Next, track reconstruction is performed in a region of interest, defined by the muon trajectory's parameters and uncertainties and the assumption that the muon originates from the interaction point. Tracker trajectory building with the Kalman-filter technique, discussed in Section 4.1.2, is seeded with compatible hit pairs found in the region of interest. In the final step all reconstructed tracks are fitted once again, using both the hits from the standalone muon track and from the silicon tracker. Ambiguities due to tracks sharing several detector hits are resolved by a cleaning step, which selects the final muon candidates on the basis of a quality requirement on the χ^2 of the track fit.

The efficiency of the offline muon reconstruction, determined on isolated muons from W decays in semi-leptonic $t\bar{t}$ events, is shown in Figures 5.33 and 5.34, as a function of the generated muon's transverse momentum and pseudorapidity respectively. These efficiencies can be directly compared to the efficiency of muon reconstruction with the muon system alone in Figures 4.14 and 4.15. The observed inefficiencies are explained by the same effects as in Section 4.1.6. In the global muon-reconstruction algorithm, seeded by the standalone muon candidates, no beam constraint is applied, however. As a result, the reconstruction efficiency is enhanced, since standalone muons that fail the beam constraint are in some cases recovered by the tracker matching in the global reconstruction. In addition, the improved angular resolutions of global muons artificially increase the reconstruction efficiency compared to the standalone muon results.

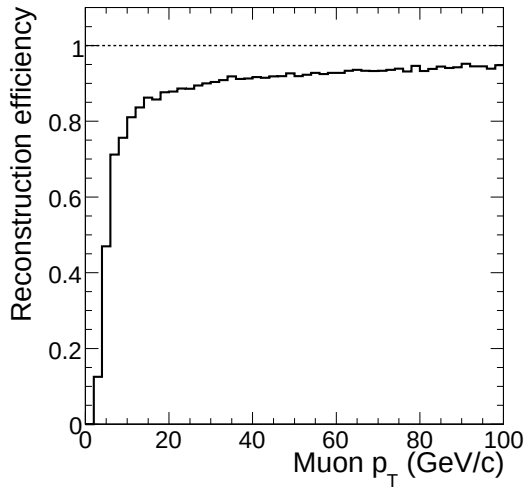


Figure 5.33: Reconstruction efficiency of the isolated muon from semi-leptonic $t\bar{t}$ events as a function of the generated muon p_T.

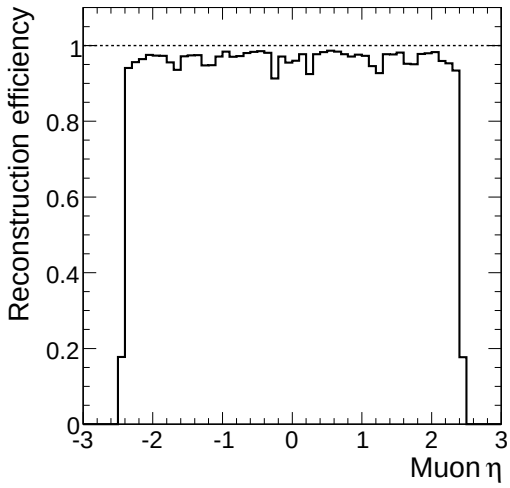


Figure 5.34: Reconstruction efficiency of the isolated muon from semi-leptonic $t\bar{t}$ events as a function of the generated muon η.

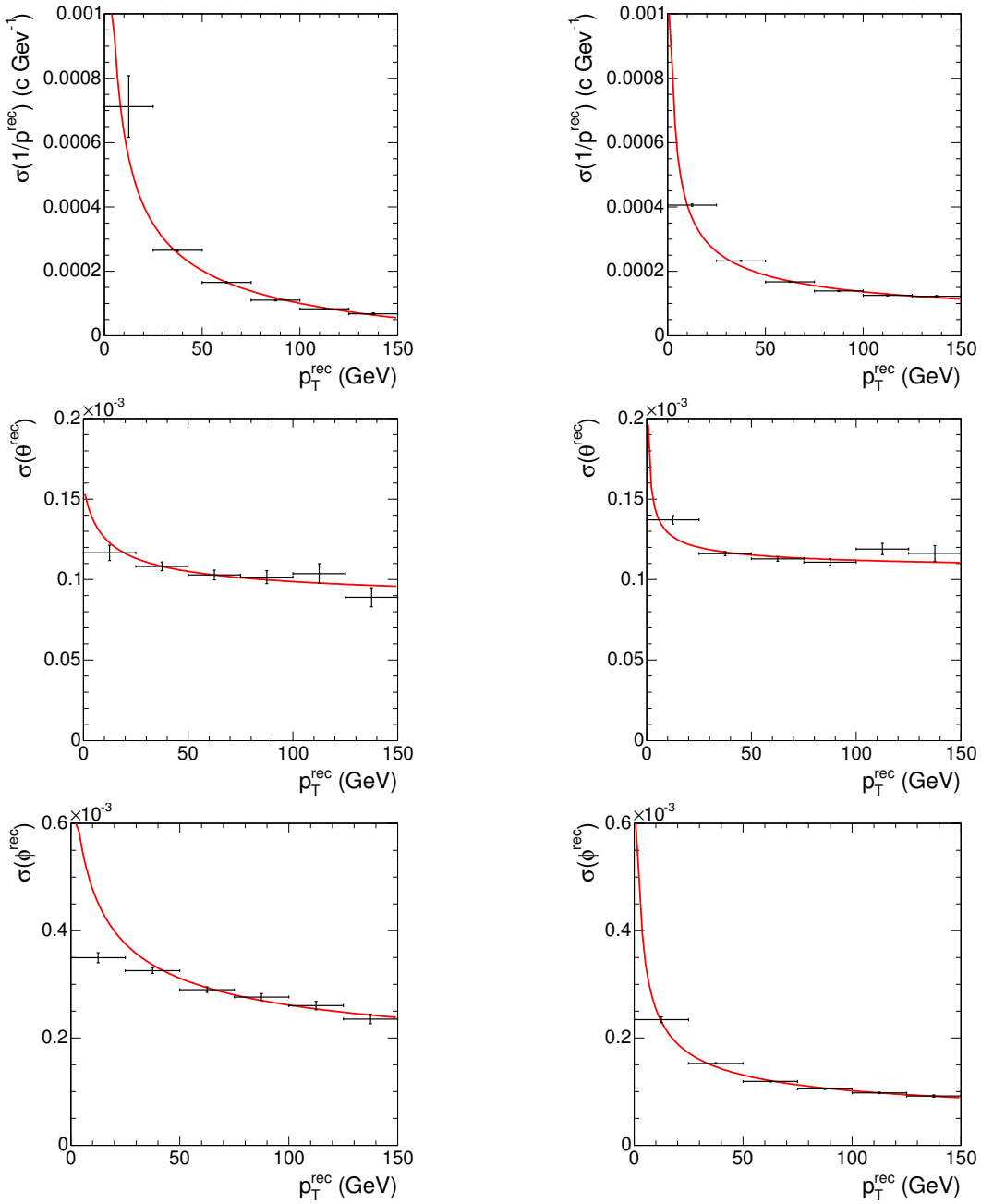
5.3.3 Lepton resolutions

In CMS electron studies, four classes of electrons are distinguished, referred to as golden, big brem, narrow and showering, depending on the signature that results from interactions with the tracker material. The abundance of each category depends strongly on the amount of material to be traversed, and hence on the pseudorapidity. The performance of electron reconstruction is typically presented separately for these classes of electrons [154].

For the purpose of using sample-specific leptons in a kinematic fit, which will be introduced in Section 5.5, the resolutions on several observables are estimated directly on the sample of interest. For both electrons and muons a simple angular criterion $\Delta R < 0.2$ is used to match the generated and reconstructed leptons in (η, ϕ) space. For the various variables, the residuals are put in histograms for various bins of reconstructed lepton p_T. The width of a Gaussian fit on the residuals peak returns the resolution of the variable in the considered bin of p_T. In Figures 5.35(a) and 5.35(b) the resolutions on $1/p$, θ and ϕ are shown, for electrons and muons respectively, as a function of the reconstructed lepton p_T.

5.3.4 Lepton-identification observables in $t\bar{t}$ events

Isolated leptons can be produced from top quarks through leptonic $t \to bW \to b\ell\nu_\ell$ decays. In Sections 5.3.1 and 5.3.2 the offline electron- and muon-reconstruction algorithms have been described and have been shown to efficiently reconstruct sufficiently hard leptons. The number of lepton candidates returned by these algorithms, however, is in general not unity in semi-leptonic $t\bar{t}$ events, which leaves an ambiguity in the identification of the lepton of interest. In Figures 5.36 and 5.37 the number of electron candidates returned by the offline reconstruction algorithm is shown for semi-leptonic



(a) Electrons (b) Muons

Figure 5.35: Absolute resolutions on $1/p$, θ and ϕ for electrons and muons as a function of the reconstructed lepton p_{T}.

$\mathrm{t\bar{t}}$ events with one isolated electron or muon respectively within $|\eta| < 2.4$ in the generated final state. In Figures 5.38 and 5.39 the number of globally reconstructed muons is shown for semi-leptonic $\mathrm{t\bar{t}}$ events with one isolated muon or electron respectively within $|\eta| < 2.4$ in the generated final state. Some of these reconstructed leptons are genuine non-isolated leptons in jets, while others are fake lepton candidates, especially in the case of electrons. To identify the isolated electrons and muons that originate

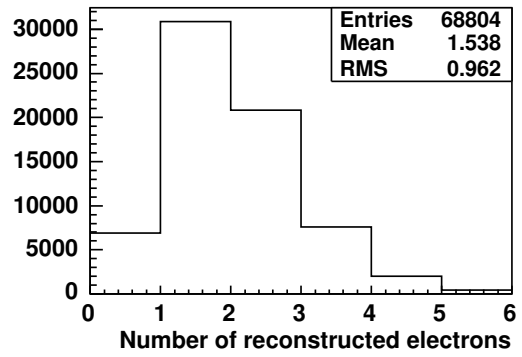


Figure 5.36: Number of offline electron candidates in semi-leptonic $t\bar{t}$ events with one isolated electron within $|\eta| < 2.4$ in the generated final state.

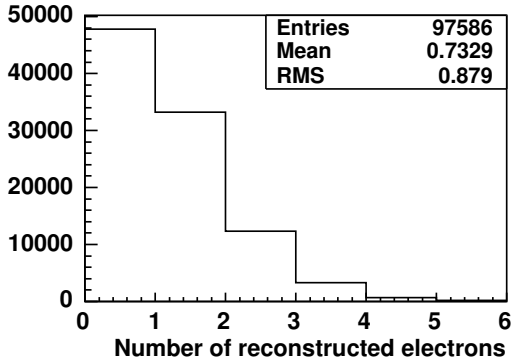


Figure 5.37: Number of offline electron candidates in semi-leptonic $t\bar{t}$ events with one isolated muon within $|\eta| < 2.4$ in the generated final state.

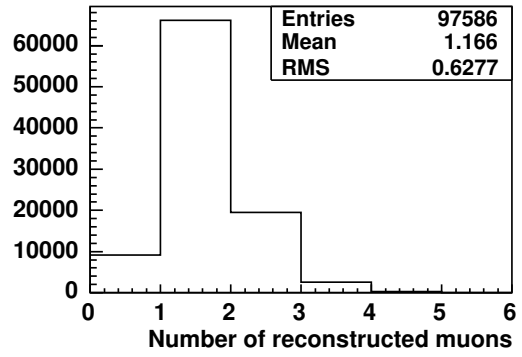


Figure 5.38: Number of globally reconstructed muons in semi-leptonic $t\bar{t}$ events with one isolated muon within $|\eta| < 2.4$ in the generated final state.

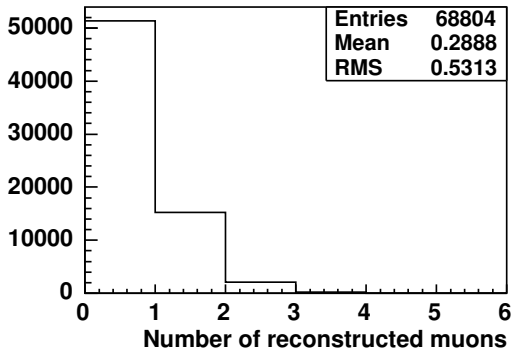


Figure 5.39: Number of globally reconstructed muons in semi-leptonic $t\bar{t}$ events with one isolated electron within $|\eta| < 2.4$ in the generated final state.

from the W-boson decays, properties of the leptons must be identified which differentiate between the correct and the wrong candidates among all reconstructed leptons in the final state of the event. For each reconstructed lepton in the final state, the following list of discriminating observables is determined:

Isolation energy E_{iso} The charged lepton track is propagated in the magnetic field to the surface of the calorimeter. In Figures 5.41 and 5.40 the averaged energy deposits are shown around the calorimeter impact point of respectively electrons and muons from W decays. The larger spread in ϕ is caused by the magnetic field. To reveal the difference between isolated and non-isolated leptons, all energy deposits in the calorimeter towers are summed within a cone of opening angle $\Delta R = 0.3$ around the lepton direction at the calorimeter surface. This energy is called the isolation energy E_{iso}. To enhance the effect of neutral particles

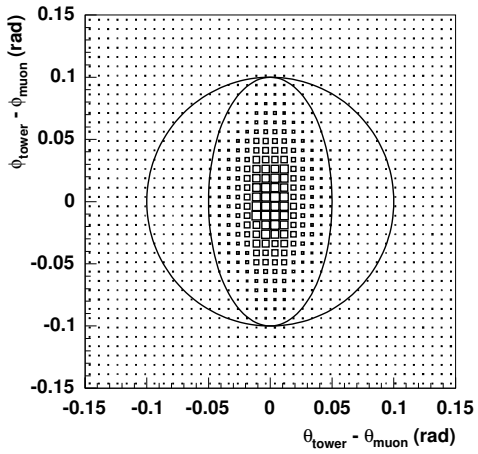


Figure 5.40: Average calorimeter energy deposits around muons from W decays in semi-leptonic $t\bar{t}$ events.

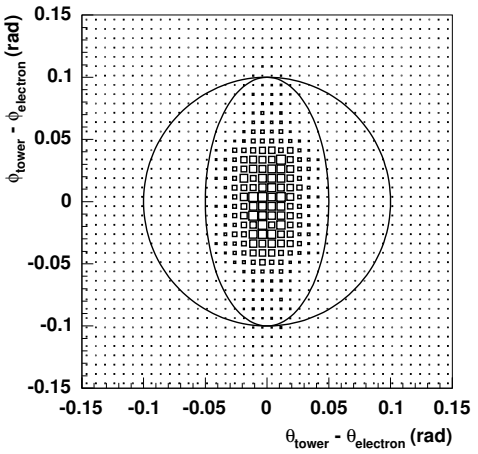


Figure 5.41: Average calorimeter energy deposits around electrons from W decays in semi-leptonic $t\bar{t}$ events.

in jets this sum is restricted to energy deposits on one side of the lepton in the azimuthal direction, depending on the lepton charge Q_{lepton}. Only when the calorimeter tower has an azimuthal angle in agreement with the criterion $(\phi_{\text{lepton}} - \phi_{\text{tower}})/Q_{lepton} < 0$, its energy is taken into account in the calculation of E_{iso}. The energy of the tower matching best the lepton according to

$$\sqrt{\left(\frac{E_{\text{lepton}} - E_{\text{tower}}}{\sigma_{E_{\text{lepton}}}}\right)^2 + \left(\frac{\Delta R(\text{lepton, tower})}{\sigma_{\Delta R}}\right)^2}, \tag{5.10}$$

is subtracted if it is contained in the one-sided cone area. The uncertainties $\sigma_{E_{\text{lepton}}}$ and $\sigma_{\Delta R}$, respectively on the energy difference and the distance in (η, ϕ), are determined from the covariance matrix of the reconstructed lepton track, propagated in the magnetic field to the calorimeter surface.

Isolation angle α_{iso} The iterative cone jet-clustering algorithm with a cone size of $\Delta R_{\text{jets}} = 0.5$ is used to cluster the calorimeter towers with an E_{T} larger than 1 GeV into jets, neglecting all towers around the lepton within an optimized cone of opening angle $\Delta R_{\text{lepton}} = 0.10$ for the muons and $\Delta R_{\text{lepton}} = 0.14$ for the electrons. The isolation angle α_{iso} is defined as the minimum of the angles between the reconstructed lepton candidate and all reconstructed jets. As a rather isotropic distribution of the jets and the leptons from W bosons is expected in $t\bar{t}$ events, the distance between the lepton and the jets is calculated in the (θ, ϕ) metric.

Logarithm of the lepton's transverse momentum $\log(p_{\text{T}})$ The logarithm is taken in order to obtain a more Gaussian probability density function.

Primary-vertex significance S_z The primary-vertex significance S_z is the ratio of the distance along the z direction between the event's reconstructed primary ver-

tex and the lepton's point of closest approach to this vertex, and the uncertainty on this distance.

Logarithm of the isolation transverse momentum $\log(p_{\mathrm{T}}^{\mathrm{iso}})$ At the event's primary vertex a cone with size $\Delta R = 0.3$ is constructed around the lepton. Tracks are considered associated to this primary vertex following the criteria described in Section 4.1.2. The isolation transverse momentum $p_{\mathrm{T}}^{\mathrm{iso}}$ is defined as the scalar sum of the transverse momentum p_{T} of all tracks associated to the vertex. To enhance the sensitivity of the observable, the p_{T} of the track, matching best the lepton track according to

$$\sqrt{\left(\frac{E_{\mathrm{lepton}} - E_{\mathrm{track}}}{\sigma_{E_{\mathrm{lepton}}}}\right)^2 + \left(\frac{\Delta R(\mathrm{lepton, track})}{\sigma_{\Delta R}}\right)^2}, \tag{5.11}$$

is subtracted. The uncertainties $\sigma_{E_{\mathrm{lepton}}}$ and $\sigma_{\Delta R}$ are determined from the covariance matrix of the reconstructed lepton track at the vertex. When no track is present in the cone area at the vertex, the value of $\log(p_{\mathrm{T}}^{\mathrm{iso}})$ is set equal to -1.

Electron quality discriminator Q_{e} The collection of reconstructed electron candidates does not have any additional quality requirements applied yet. Electron identification requirements, possibly analysis dependent, need to be envisaged to suppress electrons faked by for example soft jets or pion conversions. Two methods have been explored for CMS so far [155, 156]. For electron identification in this thesis, the second, likelihood-based approach is followed. It consists of the combination of several variables that allow to distinguish between signal electrons from $\mathrm{H} \rightarrow \mathrm{ZZ}$ decays, and fake electrons from ECAL trigger-primitive enriched QCD samples. The variables used, $E^{\mathrm{SC}}/|\vec{p}|^{\mathrm{TR}}$, $|\eta^{\mathrm{SC}} - \eta^{\mathrm{TR}}|$, $H^{\mathrm{SC}}/E^{\mathrm{SC}}$ and two shower-shape variables, are similar to the variables that are loosely constrained to form electron candidates, and also to the ones used in Section 5.2.3 for electron identification. In contexts where electron identification becomes crucial, for example where correct charge identification is needed at the per mille level [157], these reconstruction-quality variables, especially $E^{\mathrm{SC}}/|\vec{p}|^{\mathrm{TR}}$, play an essential role in the correct matching of tracks with superclusters. The combined likelihood discriminator Q_{e} of this method receives values in the interval $[0, 1]$.

The probability density function, for each observable i defined above, is denoted as $P_i^S(x)$ for W-like leptons, referred to as good leptons, and $P_i^B(x)$ for non-W-like leptons, which are called bad leptons, according to the matching criterion $\Delta R < 0.2$ between reconstructed and generated leptons. The corresponding distributions for the first five variables are shown in Figures 5.42 and 5.43, for electrons and muons respectively. The distribution of the reconstruction-quality variable Q_{e} is shown in Figure 5.44.

For each observable i the likelihood ratio is defined as

$$\mathcal{L}_i(x_i) = \frac{P_i^S(x_i)}{P_i^S(x_i) + P_i^B(x_i)}, \tag{5.12}$$

which takes values in the interval $[0, 1]$. The likelihood ratios, also shown in Figures 5.42, 5.43 and 5.44, are determined by calculating $S/(S+B)$ bin-by-bin, where S

and B denote respectively the number of W-like and non-W-like electrons or muons in a given bin. Rather than fitting both probability density functions $P_i^S(x_i)$ and $P_i^B(x_i)$, only the likelihood-ratio function is parametrized with single or double sigmoid functions. The goodness of those parametrizations is demonstrated by comparing the histogram of the bin-by-bin likelihood ratio distributions with the superimposed functions. The values of the χ^2 divided by the number of degrees of freedom for the fits indicate that this sigmoid model is sufficient.

In Figures 5.46 and 5.47 the correlation is shown between the isolation observables E_{iso} and $\log(p_{\text{T}}^{\text{iso}})$, for the correct W-like electrons and muons respectively. No large correlation is observed between both variables, hence the information of both can be used simultaneously to enhance the separation power between W-like and non-W-like leptons. The estimated linear-correlation coefficients between the observables for W-like leptons are given in Table 5.2.

	E_{iso}	$\log(p_{\text{T}}^{\text{iso}})$	α_{iso}	$\log(p_{\text{T}})$	S_z	Q_{e}
Electrons						
E_{iso}	1	0.15	-0.21	0.13	-0.05	-0.13
$\log(p_{\text{T}}^{\text{iso}})$	-	1	-0.32	-0.02	0.01	-0.13
α_{iso}	-	-	1	0.06	0.07	0.09
$\log(p_{\text{T}})$	-	-	-	1	0.02	0.14
S_z	-	-	-	-	1	-0.08
Q_{e}	-	-	-	-	-	1
Muons						
E_{iso}	1	0.41	-0.35	-0.04	-0.02	-
$\log(p_{\text{T}}^{\text{iso}})$	-	1	-0.35	-0.06	-0.02	-
α_{iso}	-	-	1	0.12	-0.06	-
$\log(p_{\text{T}})$	-	-	-	1	-0.01	-
S_z	-	-	-	-	1	-

Table 5.2: Correlations between the sensitive observables, estimated from the simulated W-like leptons. For the electrons also the reconstruction quality variable Q_{e} is taken into account.

5.3.5 Lepton-identification efficiencies in $\mathrm{t\bar{t}}$ events

The individual observables, E_{iso}, θ_{iso}, $\log(p_{\text{T}})$, S_z and $\log(p_{\text{T}}^{\text{iso}})$, with additionally Q_{e} for electrons, are combined into one observable by multiplying their likelihood ratios $\mathcal{L}_i(x_i)$ according to

$$\mathcal{L} = \prod_{i=1}^{5 \text{ or } 6} \left[\frac{\mathcal{L}_i(x_i)}{\max\{\mathcal{L}_i(x_i)\}} \right]^{1/\sum_{j=1}^{5 \text{ or } 6} |c_{ij}|}, \tag{5.13}$$

which also takes values in the interval $[0, 1]$. The correlations c_{ij} between observables i and j are taken into account in an ad-hoc way by the power $1/\sum_{j=1}^{5 \text{ or } 6} |c_{ij}|$, which reduces the importance of each observable according to the mutual correlations. The

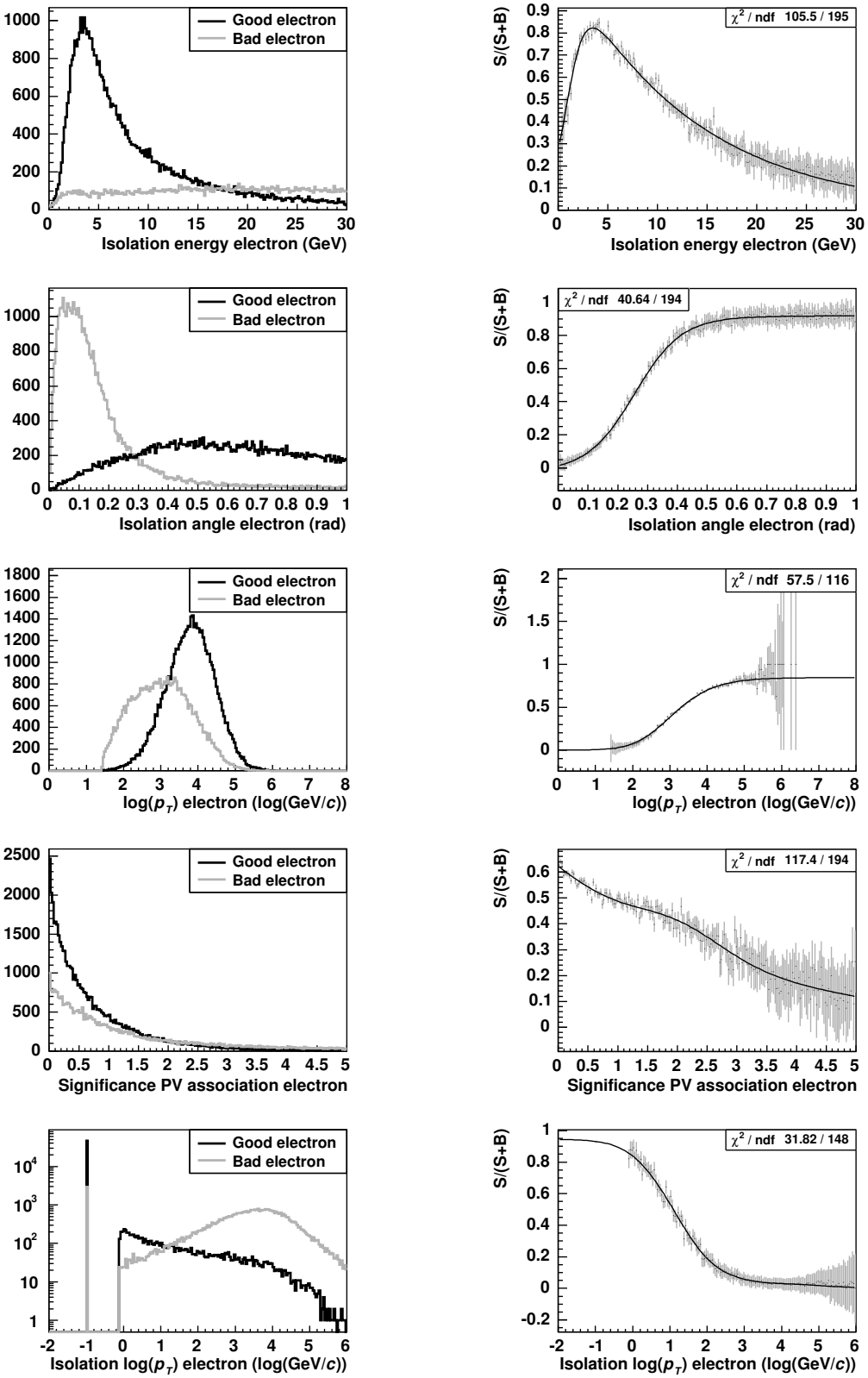


Figure 5.42: Distributions of the observables E_{iso}, α_{iso}, $\log(p_{\mathrm{T}})$, S_z and $\log(p_{\mathrm{T}}^{\mathrm{iso}})$, along with their respective fitted likelihood-ratio distributions $S/(S+B)$, where S and B denote bin-by-bin the number of W-like (good) and non-W-like (bad) electrons.

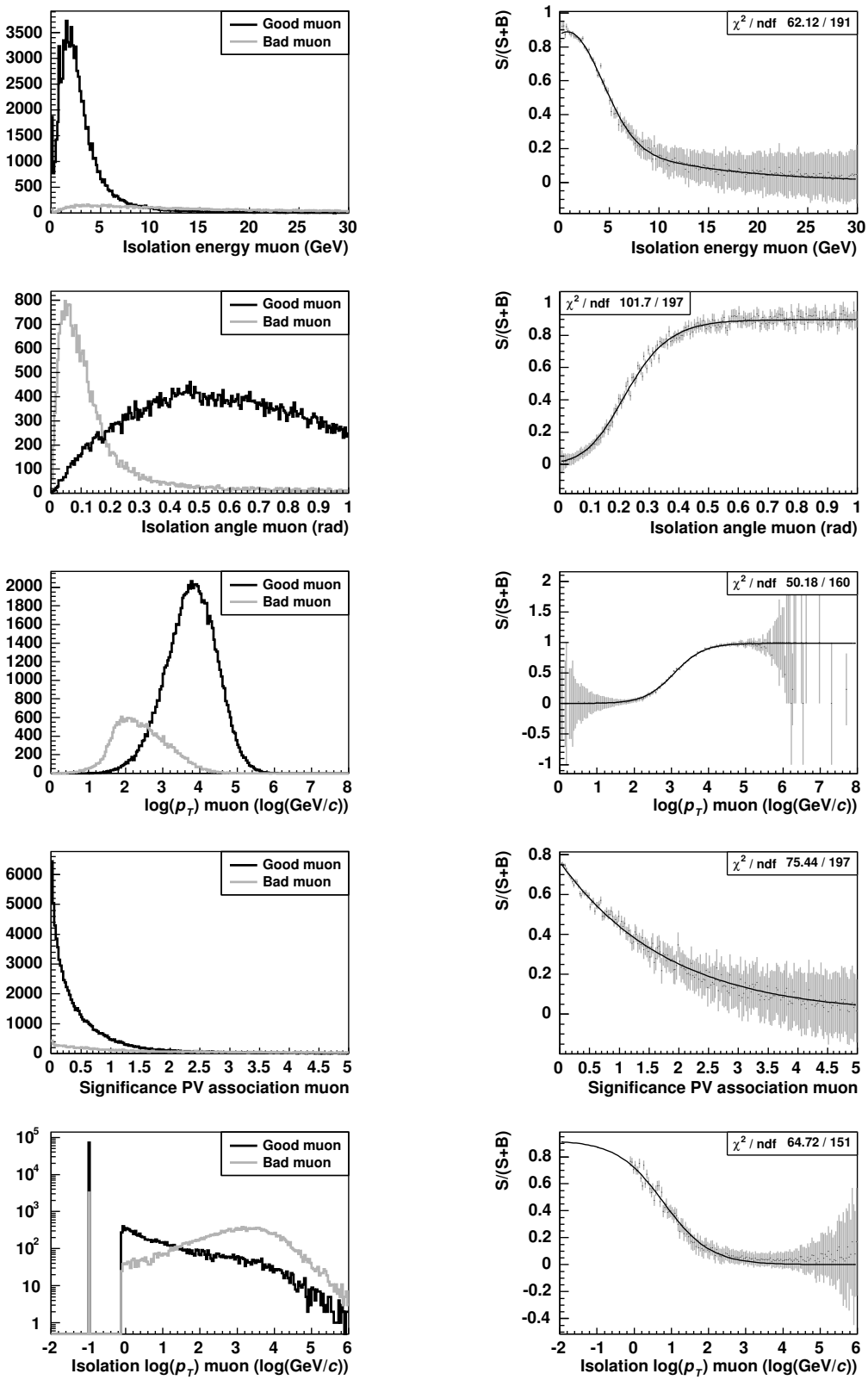


Figure 5.43: Distributions of the observables E_{iso}, α_{iso}, $\log(p_{\text{T}})$, S_z and $\log(p_{\text{T}}^{\text{iso}})$, along with their respective fitted likelihood-ratio distributions $S/(S+B)$, where S and B denote bin-by-bin the number of W-like (good) and non-W-like (bad) muons.

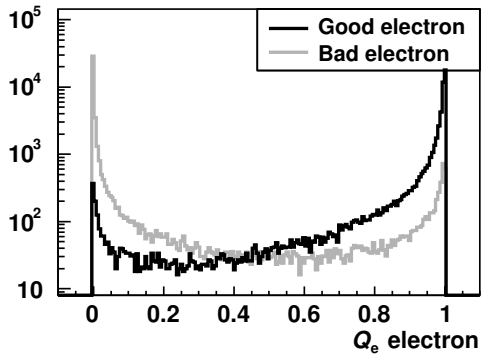


Figure 5.44: Distribution of the electron-reconstruction quality variable Q_e in a single-electron $t\bar{t}$ sample.

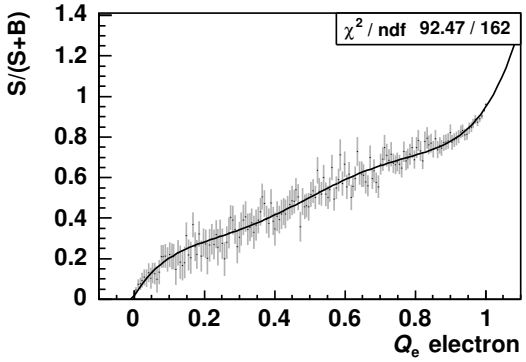


Figure 5.45: Likelihood-ratio distribution of the electron-reconstruction quality variable Q_e in a single-electron $t\bar{t}$ sample.

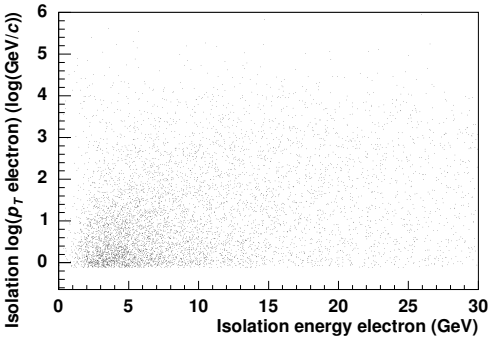


Figure 5.46: The correlation between the observables E_{iso} and $\log(p_{\mathrm{T}}^{\mathrm{iso}})$ for W-like electrons.

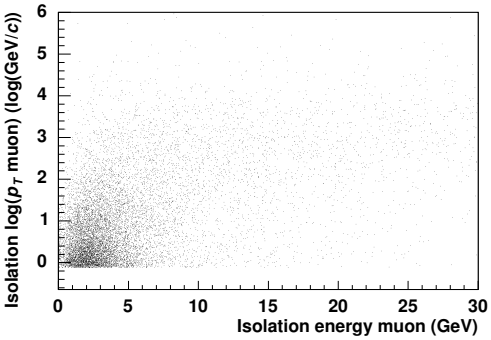


Figure 5.47: The correlation between the observables E_{iso} and $\log(p_{\mathrm{T}}^{\mathrm{iso}})$ for W-like muons.

combined observable $\mathcal{L}$ is shown for electrons and muons in Figures 5.48 and 5.49 respectively.

On the basis of this combined likelihood-ratio observable, the expected correct W-like lepton is selected as the one with the highest value of $\mathcal{L}$ in the event. The efficiencies for selecting in this way the correct lepton in events with at least one candidate are summarized in Table 5.3 for semi-leptonic $t\bar{t}$ events including low luminosity pile-up collisions, before and after applying the single-lepton trigger criteria.

The combined likelihood ratio value $\mathcal{L}$ of the selected lepton can also be used as an event selection variable. Placing a threshold on the maximum value of $\mathcal{L}$ for all leptons in the event will reduce the $t\bar{t}$ event selection efficiency, but increase the efficiency for selecting the correct W-like lepton. A scan of the efficiencies in this two-dimensional plane is shown in Figures 5.50(a) and 5.50(b) for electrons and muons, respectively. The results are compared to other methods which aim to select the correct W-like lepton in the final state of the event.

Selecting the W-like lepton in an event as the one corresponding to the only lepton in

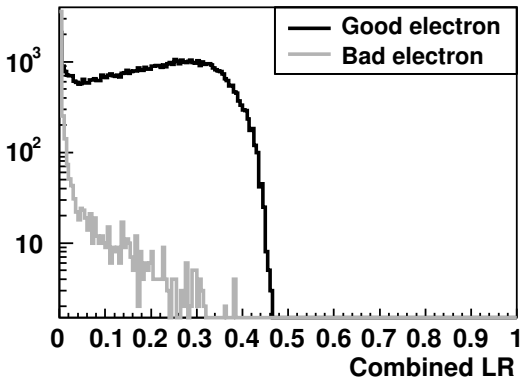


Figure 5.48: Distribution of the combined likelihood-ratio variable for electrons in semi-leptonic $t\bar{t}$ events, either from the W decay, or originating from jet fragmentation or fakes.

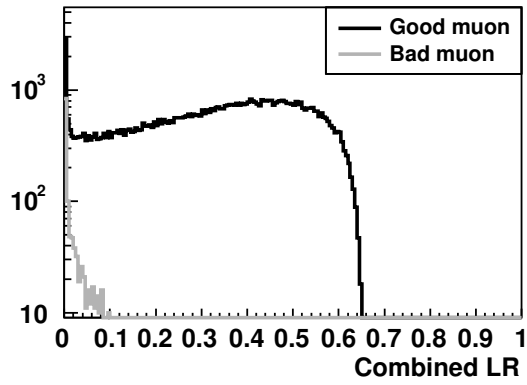


Figure 5.49: Distribution of the combined likelihood-ratio variable for muons in semi-leptonic $t\bar{t}$ events, either from the W decay, or originating from jet fragmentation or fakes.

	without HLT		with HLT	
	e	μ	e	μ
Correct electron selected	93.3%	-	99.0%	-
Correct muon selected	-	97.9%	-	98.9%

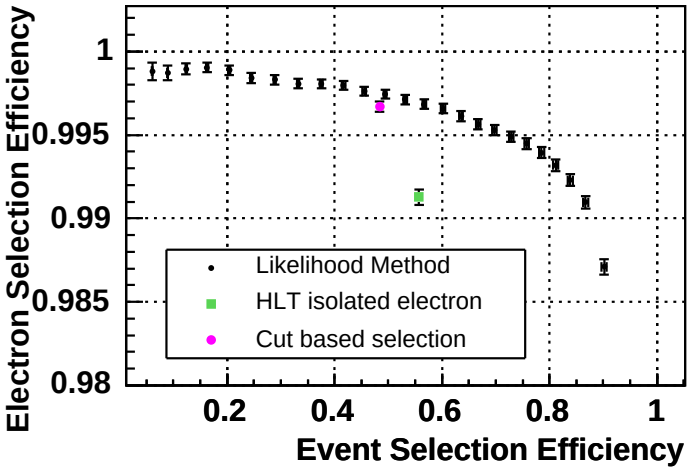
Table 5.3: Efficiencies for the selection based on the combined likelihood-ratio observable, for $t\bar{t}$ events with at least one reconstructed and identified electron or muon, before and after applying the HLT criteria.

agreement with the HLT single-lepton criteria, a point in the two-dimensional plane of the $t\bar{t}$ event selection efficiency and the W-like lepton selection efficiency is obtained, as shown in Figures 5.50(a) and 5.50(b). A simple sequential cut-based analysis is compared on the same figures. Hard thresholds are chosen on the observables that enter the likelihood ratio, rather than using the inter-dependent cuts that follow from cutting on the combined likelihood ratio directly. In Table 5.4 the various cuts are defined. In most of the cases there is only one lepton remaining after applying these sequential cuts. If several leptons remain, the one with the highest transverse momentum is chosen.

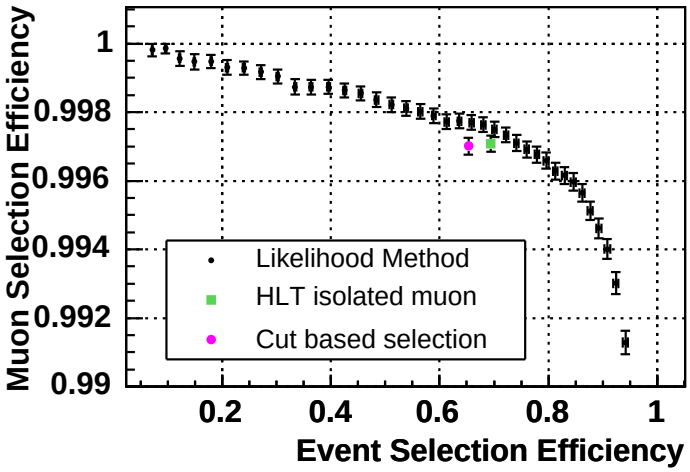
	E_{iso} (GeV)	α_{iso} (rad)	$\log(p_{\text{T}})$ ($\log(\text{GeV}/c)$)	S_z	$\log(p_{\text{T}}^{\text{iso}})$ ($\log(\text{GeV}/c)$)	Q_{e}
Electron	< 13	> 0.25	> 3	< 2	< 1.3	> 0.36
Muon	< 6	> 0.2	> 2.8	< 1.2	< 1.0	-

Table 5.4: Definition of the independent cut values applied as alternative lepton selection.

For the muon events the resulting efficiencies of the HLT isolation criteria and the



(a) Electrons



(b) Muons

Figure 5.50: Comparison of the relation between the $t\bar{t}$ event selection efficiency and the W-like lepton selection efficiency for different methods. The event selection efficiencies are relative to the total amount of $t\bar{t}$ events with at least one reconstructed lepton, before applying the single-lepton HLT criteria.

sequential cut-based approach are comparable to those obtained with the likelihood-ratio method. This is not surprising, since the room for improvement is very small. For the electron events a significant improvement is obtained using the likelihood-ratio method. For the same electron selection efficiency as obtained by the HLT criteria, the event selection efficiency is increased from about 55% to about 90% when using the likelihood-ratio method. Hence, by applying the likelihood-ratio selection rather than the HLT criteria, 30 to 40% of the $t\bar{t}$ electron events, which would otherwise not have been collected by CMS, can be recovered provided sufficient computation time in the

HLT and control over the background rates.

As a reference, the efficiencies are also given for two simple selection criteria which give by default only one lepton. The first criterion selects naively events with exactly one reconstructed lepton, while the second method chooses the lepton with the highest transverse momentum out of all reconstructed leptons. The performance of these naive selection criteria, however, is significantly worse compared to the other methods. The selection of the always existing highest p_{T} lepton is 82% and 96.4% efficient, for the electron and muon case respectively. The requirement of exactly one reconstructed lepton, on the other hand, is respectively 92% and 98.7% efficient, both at an event selection efficiency of only 35%.

5.4 Transverse missing-energy reconstruction

The CMS detector has been designed to be as hermetic as possible from calorimetric point of view. The main motivation for this is to cover as much of the detectable final state in an event as possible, such that also apparent missing energy can be reconstructed. The compositeness of the protons and the large amount of energy sent out in the direction of the beam pipe, make it impossible to directly measure longitudinal missing energy. The transverse energy balance, however, can be measured with an accuracy that is sufficient to help identify physics signatures involving non-interacting particles, like neutrinos.

Accurate transverse missing-energy $E_{\mathrm{T}}^{\mathrm{miss}}$ measurements in CMS profit from the high granularity and good hermeticity of the calorimeter system. Many effects degrade the $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction accuracy, on the other hand, like the presence of pile-up collisions, the difference in calorimeter response between photons and pions and the bending of tracks in the magnetic field. The dominating limiting factor on the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution of CMS, however, is expected to be the calorimeter energy resolution.

The simplest way to reconstruct the transverse missing energy in an event, is to compute the vectorial sum of all the energy deposits in the calorimeters. Doing so for minimum-bias collisions without pile-up, the resolution can be measured directly from the E_x^{miss} distribution, because no intrinsic missing energy is expected in these events. This results in a total $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution of 6.1 GeV, including a 4.8 GeV stochastic contribution from calorimeter resolution and 3.8 GeV due to electronic noise. To examine the behaviour of the missing energy for harder collisions with larger energy deposits, several samples of QCD jets are analysed, simulated in different bins of $\hat{p_{\mathrm{T}}}$. In Figures 5.51 and 5.52, respectively the average reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$ and the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution are shown as a function of the scalar sum ΣE_{T} of the calorimeter energy deposits in the event. These distributions are parametrized by

$$\left\langle E_{\mathrm{T}}^{\mathrm{miss}} \right\rangle = 5.4\,\mathrm{GeV} \oplus 1.23\,\mathrm{GeV}^{1/2}\sqrt{\Sigma E_{\mathrm{T}}} \oplus 0.019\,\Sigma E_{\mathrm{T}} \tag{5.14}$$

for the average reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$, and by

$$\sigma_{E_{\mathrm{T}}^{\mathrm{miss}}} = 3.8\,\mathrm{GeV} \oplus 0.97\,\mathrm{GeV}^{1/2}\sqrt{\Sigma E_{\mathrm{T}}} \oplus 0.012\,\Sigma E_{\mathrm{T}} \tag{5.15}$$

for the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution. The resolution on the azimuthal angle ϕ is found to improve

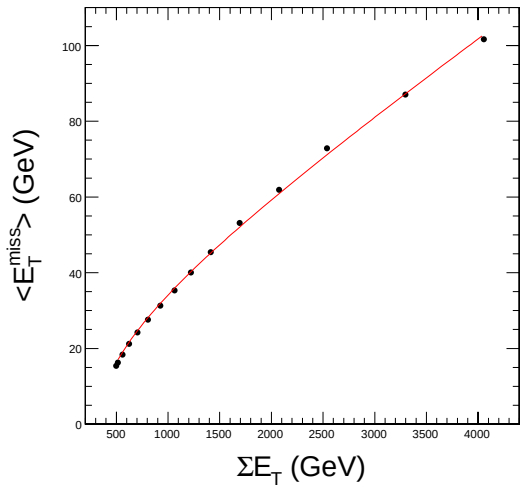


Figure 5.51: Average reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$ in QCD jet samples as a function of ΣE_{T}.

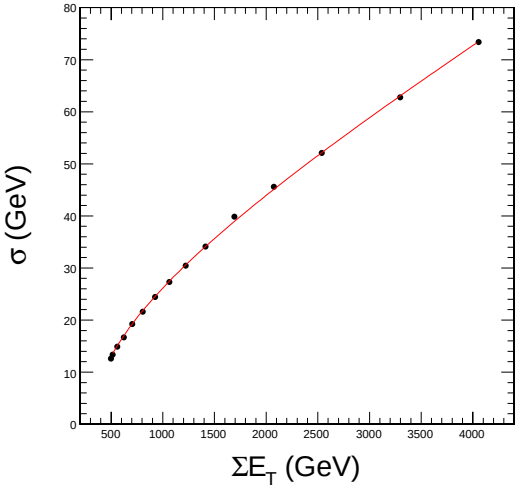


Figure 5.52: Resolution on $E_{\mathrm{T}}^{\mathrm{miss}}$ in QCD jet samples as a function of ΣE_{T}.

with increasing $E_{\mathrm{T}}^{\mathrm{miss}}$. As an input to the studies in this thesis, the ϕ resolution is determined inclusively on semi-leptonic $\mathrm{t\bar{t}}$ events, and is found to be 0.54 rad.

Several techniques can be employed to improve the calorimeter-tower computation of the missing energy. A straightforward correction is the addition of reconstructed muons, which deposit very little energy in the calorimeters. For QCD events, a correction of the jet energies with the techniques introduced in Section 5.1.3, does not significantly improve the transverse missing-energy resolution, because the reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$ is largely an artifact of the detector response. In events with genuine missing energy, on the other hand, the $E_{\mathrm{T}}^{\mathrm{miss}}$ is underestimated, and jet energy corrections can be applied to calibrate at the same time the missing energy [158]. It has been shown that such corrections can improve the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution by 15% at high $E_{\mathrm{T}}^{\mathrm{miss}}$ in $\mathrm{t\bar{t}}$ events. Other possible improvements include the use of only jet and lepton energies in hard interactions, or the application of energy-flow techniques, which are starting to be explored within the CMS collaboration.

For the purpose of missing-energy reconstruction in the studies presented in this thesis, the simple reconstruction from calorimeter towers is used, correcting for muons. Instead of pursuing an $E_{\mathrm{T}}^{\mathrm{miss}}$ estimation as accurate as possible, it is rather obtained by applying kinematic constraints on events as a whole, where only one neutrino in the final state is assumed. This is done with a kinematic-fitting technique as described in Section 5.5. This technique also delivers an event-by-event estimation of the missing energy along the beam direction.

5.5 Constraining Events with a Kinematic Fit

5.5.1 Principles of kinematic fitting

When an analysis is performed for a chosen signal, certain reconstructed objects are expected to be measured in the final state, while other unmeasured quantities can

be associated to undetected objects. For a considered event hypothesis the measured as well as the unmeasured quantities are related by certain kinematic constraints. Due to the uncertainties on the measured quantities, these constraints are not exactly fulfilled. With a kinematic-fitting technique all measured and unmeasured objects are kinematically altered such that they exactly fulfil the external constraints deduced from the event hypothesis. The method employed in the studies for this thesis follows the concept of linearized iterative least-square fitting, together with the application of Lagrange multipliers. It results in a χ^2 test of the event hypothesis providing a fit probability, but it also improves estimators of the underlying kinematics for a given event. The technique has been used already extensively in an experimental context. It allowed for significant improvements of the W-boson mass measurement at LEP [159] and the determination of the CKM matrix elements V_{cb} and V_{ub} in BaBar [160–162]. It has also been applied for the verification of hypothetical kinematic topologies, for example in the search for the Standard Model Higgs boson at LEP [22].

The n measured parameters $\vec{y}$ and the p unmeasured parameters $\vec{a}$ are required to fulfil the m constraint equations $\vec{f}$,

$$\begin{aligned} f_1(a_1, \ldots, a_p, y_1, \ldots, y_n) &= 0 \\ &\vdots \\ f_m(a_1, \ldots, a_p, y_1, \ldots, y_n) &= 0. \end{aligned} \tag{5.16}$$

By definition, these requirements will be fulfilled for the true parameters $\bar{y}$ and $\bar{a}$. In general, however, the measured values $\vec{y}$ do not solve the constraints. Corrections $\Delta\vec{y}$ have to be calculated, such that $\vec{y} + \Delta\vec{y}$ solves (5.16). For these corrections, the weighted sum

$$S(\vec{y}) = \Delta\vec{y}^{\mathrm{T}}\mathcal{V}^{-1}\Delta\vec{y} \tag{5.17}$$

should be minimal, with $\mathcal{V}$ being the covariance matrix of the measured parameters. Finding the minimum of $S(\vec{y})$ under the constraints (5.16) can be expressed equivalently as finding the minimum of

$$L(\vec{y}, \vec{a}, \vec{\lambda}) = S(\vec{y}) + 2\sum_{k=1}^{m} \lambda_k f_k(\vec{y}, \vec{a}), \tag{5.18}$$

where $\vec{\lambda}$ are Lagrange multipliers. With linear constraints the solution can be calculated analytically, otherwise it can be found by linearizing the constraint functions $\vec{f}(\vec{y}, \vec{a})$ and solving iteratively, until all constraints are fulfilled within the required accuracy. The details of this linearization procedure, as well as examples of the application of the described kinematic-fitting technique in a wide range of physics use-cases, can be found in [163, 164].

5.5.2 Applications of kinematic fitting

A specific illustration of the application of kinematic constraints concerns the use of a kinematic fit to build a top-quark mass estimator in semi-leptonic $\mathrm{t\bar{t}}$ events [163]. In such events the final state consists of four jets, an isolated lepton and missing energy.

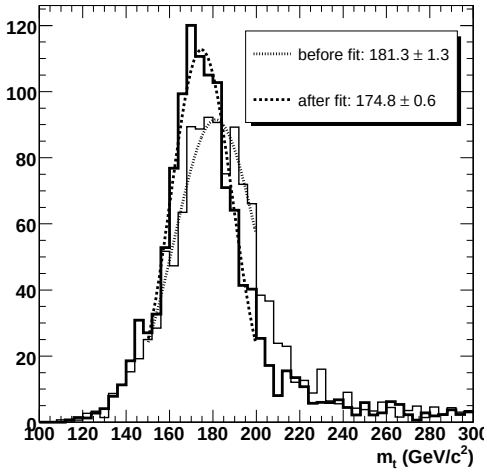


Figure 5.53: The effect of the kinematic fit on the top-quark mass distribution of hadronically-decaying top-quark candidates.

The event reconstruction and selection first identifies the four jets and the hard lepton. For each possible jet association, a kinematic fit is used to impose both W-boson mass constraints on the event. Jet associations are retained, if they result in a kinematic-fit probability larger than 0.2, an absolute value of the difference between the fitted and reconstructed W-boson masses smaller than $35\,\mathrm{GeV}/c$, a fitted hadronic top-quark mass smaller than $350\,\mathrm{GeV}/c^2$ and a fitted leptonically-decaying top-quark mass larger than $125\,\mathrm{GeV}/c^2$. The best jet association is finally chosen using the b-tagging information of all jets, the angle between the lepton and the jet considered to originate from the leptonically-decaying top quark, etc.

The top-quark mass is estimated from the mass distribution of the hadronically decaying top-quark candidates, shown in Figure 5.53. The presented reconstruction and selection resulted in a purity larger than 70% of correct jet associations around the top-quark mass. A Gaussian fit of the top-quark mass spectrum obtained from the reconstructed objects and from the kinematics of these objects after the kinematic fit, resulted in respectively $m_{\mathrm{t,rec}} = 181.3 \pm 1.3\,\mathrm{GeV}/c^2$ and $m_{\mathrm{t,fit}} = 174.8 \pm 0.6\,\mathrm{GeV}/c^2$. In order to obtain the same precision on the top-quark mass without applying a kinematic fit, one would need five times more data compared to when a kinematic fit is applied.

The currently most advanced top-quark mass measurement technique proposed for CMS [165] takes the kinematic fit one step further. On an event-by-event basis an ideogram is constructed, that expresses the resolution function of the estimated top-quark mass for the considered event. This ideogram is obtained by using a range of top-quark masses as input hypothesis to the kinematic fit that constrains both W-boson and top-quark masses. In this way the probability of the event kinematics to match a given top-quark mass hypothesis is scanned event by event, yielding very good uncertainty estimations, which is reflected in the excellent statistical properties of the top-quark mass estimator based on this technique.

The described kinematic-fit technique is applied throughout Chapters 6 and 7 to

constrain the studied event topologies. In these analyses, for final states containing a semi-leptonically decaying top-quark pair, both W-boson and both top-quark mass constraints are imposed on the invariant mass relations of the constituting final-state objects. The mass constraints are applied assuming a zero-width resonance, which is a good approximation, since the event-by-event resolutions on the reconstructed W-boson and top-quark masses are much larger than the intrinsic width of the W and top resonances. The fit is not used in fully-leptonic decays, because in that case the presence of two neutrinos renders the event kinematics underconstrained.

In the fit the jets are declared as objects with a free mass, while the muon and the neutrino have an imposed fixed mass. The objects' kinematics are parametrized by their inverse momentum $1/p$, their polar angle θ, their azimuthal angle ϕ and in the case of the jets also their energy parameter $d = E_{\text{fitted}}/E_{\text{measured}}$. Apart from the objects' Lorentz vectors, also the covariance matrices for the objects in this parametrization are needed. As a simplification, only the diagonal elements of the covariance matrices are determined, corresponding to the resolution functions presented in Section 5.1.4 and 5.3.3. The neutrino object is also treated as a measured object. Its transverse momentum is identified with the transverse missing-momentum vector. The resolutions on θ and $1/p$ of the neutrino object are taken very large, to give full freedom in the fit to the unknown z component. The resolution of the angle ϕ of the neutrino object is taken as discussed in Section 5.4.

The use of all four mass relations in the fit is possible because these studies are not biased by W- or top-mass deviations, explicitly shown in Section 6.2. Further constraints on the transverse components of the energy deposited in the detector are found to be useless. Due to the complexities of proton collisions at the LHC, like the presence of pile-up collisions, the accuracy on the determination of the underlying event contribution is insufficient.

Chapter 6

Measurement of b-Jet Identification Efficiencies

The CMS b-tagging algorithms and their performance on Monte-Carlo simulated events are presented in Section 5.2. The efficiencies and purities of these algorithms, however, need to be calibrated once the detector starts taking data. At the TEVATRON accelerator, this is done by measuring b-tagging efficiencies on back-to-back b-jet pairs from quark–anti-quark annihilation. A relative uncertainty on the b-tagging efficiency of about 8% is obtained in the CDF experiment [166]. In the proton–proton collision environment of the LHC such an identification of b-jet pairs is hampered by the overwhelming multi-jet QCD production. Due to the high energy and luminosity of the LHC, however, a completely new method to calibrate b-tagging algorithms can be exploited. Already during the first year of low-luminosity operation, the LHC is expected to produce more than 8 million $t\bar{t}$ pairs. With the branching fraction $t \to Wb$ being $\sim 100\%$, these $t\bar{t}$ events can be used to isolate jet samples with a highly enriched b-jet content. The calibration of b-jet identification algorithms can then be performed on these jet samples [167].

In this chapter the measurement of b-jet identification efficiencies on jet samples from $t\bar{t}$ events is explored. In Section 6.1 the reconstruction and selection of b-enriched jet samples from $t\bar{t}$ events is described. The enrichment of the jet samples is intended to suppress the systematical uncertainties on the b-jet purity. This suppression is shown explicitly in Section 6.2, where the systematic uncertainties are estimated in detail. With the knowledge of the systematic uncertainties, the measurement of the b-tagging efficiency is performed at a point of minimal total uncertainty, following the procedure described in Section 6.3. The hence obtained expected uncertainties serve as direct input for physics analyses that rely on b-tagging, like for example $H \to b\bar{b}$ searches. In Section 6.4 an outlook is given on other possibilities to measure the performance of b-tagging algorithms on data.

6.1 b-Enriched Jet Samples from Top-Quark Events

In this section, the selection of b-jet enriched jet samples from top-quark pair events is described. Two $t\bar{t}$ final states are considered from which jet samples are extracted:

- the semi-leptonic $t\bar{t}$ decay with final state $b\bar{b}qq'\ell\nu$, with the lepton ℓ being either a muon or an electron. The large W+jets background is suppressed by requiring one b-tagged jet, and the combinatorial ambiguities are handled using the kinematic constraints in these events;
- the fully-leptonic $b\bar{b}\ell_1\nu_1\ell_2\nu_2$ final state, with the leptons ℓ_1 and ℓ_2 being a muon or an electron, which is combinatorically cleaner compared to the semi-leptonic channel but suffers from lower statistics and an important contamination from WW and $Z + \text{jets}$ backgrounds.

After the reconstruction of the physics objects in the final state with the tools studied in Chapters 4 and 5, the event selection criteria for semi-leptonic and fully-leptonic $t\bar{t}$ events are described. With the events passing the initial selection, jet samples with an enriched b content are constructed by means of a combined likelihood ratio. This likelihood ratio allows to tune with a single observable the b-jet purity of the sample as a function of the event selection efficiency.

6.1.1 Event reconstruction

In Chapters 4 and 5 all tools are described that are necessary to reconstruct the above-mentioned final states of semi-leptonic and fully-leptonic $t\bar{t}$ events. In this section, a brief overview is given of the application of these reconstruction tools in the analysis presented in this chapter.

The single-muon or single-electron Level-1 trigger thresholds and the corresponding High-Level Trigger requirements are applied prior to the main event reconstruction, as discussed in Sections 4.3.1 and 4.3.2. Offline isolated electrons and muons are reconstructed as described in Sections 5.3.1 and 5.3.2 respectively. For electrons a cut $Q_e > 0.15$ is imposed on the reconstruction quality variable. To identify the leptons that result from $W^{\pm}$ decays in the $t\bar{t}$ events, the discriminator $\mathcal{L}_{\text{muon}}$ or $\mathcal{L}_{\text{electron}}$, defined in (5.13), is calculated for each muon or electron candidate, depending on the final state considered. The one or two muons or electrons associated to the $W^{\pm}$ decays are then selected to be the candidates with the largest $\mathcal{L}_{\text{muon}}$ or $\mathcal{L}_{\text{electron}}$.

Jets are clustered using the iterative cone algorithm, described in Section 5.1.1, with cone size $\Delta R = 0.5$ and a cut on the E_T of the jet seeds of 2 GeV. Calorimeter towers are used as input to the clustering algorithm, applying the underlying-event input thresholds as defined in Section 4.1.4. The jet clustering is performed in a region excluding the leptons considered to originate from $W^{\pm}$ decays. To accomplish this, all calorimeter towers in an isolation cone around the muon or electron candidates, shown on average in Figures 5.40 and 5.41, are deleted from the input to the jet-clustering algorithm prior to clustering. Jet energy-scale corrections are applied to the reconstructed jets using the CMS Monte-Carlo jet-energy corrections, introduced in Section 5.1.3. Finally, jets emerging from pile-up collisions are vetoed using the association criterion from Section 5.1.5.

The last part of the event reconstruction deals with the transverse missing energy. As explained in Section 5.4, this quantity is calculated by making the vectorial sum of the transverse components of all the energy deposits in the calorimeters, correcting for the transverse energy carried by the muons.

6.1.2 Event selection for semi-leptonic $t\bar{t}$ events

The experimental signature of the final state for the semi-leptonic decay $t\bar{t} \to b\bar{b}qq'\ell\nu$ consists of an isolated lepton, at least four jets and transverse missing momentum. In the following two distinct cases are considered in which the lepton is either a muon or an electron. The former final state is called the semi-muon case, the latter is referred to as the semi-electron final state. Detailed information on the simulation of the analysed event samples is given in Section 4.2.4.

An initial event selection is applied to select semi-leptonic $t\bar{t}$ events and suppress backgrounds. One isolated muon or electron is demanded in the range $|\eta| < 2.5$ and with lepton identification likelihood ratio $\mathcal{L}_{\text{lepton}} > 0.01$, along with minimally four jets with calibrated $E_{\text{T}} > 25\,\text{GeV}$ in $|\eta| < 2.4$. The four jets with the highest E_{T} are selected to reconstruct the kinematics of the event. No cuts are applied on the transverse missing energy. Once the physics objects are reconstructed, the event is considered as a whole. The four selected jets can be assigned to the four final-state partons in twelve different ways, taking into account the interchangeability of the partons from the hadronically decaying $W^{\pm}$. The lepton and the neutrino do not introduce additional combinatorial ambiguities. In the next step the reconstructed objects and their resolutions are used as input to a kinematic fit, constrained by the W-boson and top-quark masses, as described in Section 5.5. The kinematic-fitting procedure is applied to the event for each of the twelve particle combinations. As a result, a χ^2 probability is returned if the fit converged, along with the fitted kinematic variables. This output is used in Section 6.1.3 to choose the best jet association.

When constructing a b-enriched jet sample from $t\bar{t}$ events, contributions from other processes not containing b quarks should be suppressed. In the case of the semi-leptonic decay mode, a few backgrounds might contribute significantly to the impurity of the constructed jet sample. These backgrounds are the other $t\bar{t}$ decays, mostly involving τ leptons, and W + jets production with a leptonic W decay. Other processes are considered negligible, as is also shown in the context of the top-quark mass and $t\bar{t}$ cross-section analyses in the semi-leptonic channel [165, 168]. In Tables 6.1 and 6.2 the cross sections for the considered processes are shown in the second column, along with the efficiencies of the initial reconstruction and selection in the third column.

After the initial selection, the W + jets background is still relatively large. To further suppress it, an extra requirement is imposed on each combinatorial solution to the event reconstruction. The jet considered to come from a top quark together with the hadronically-decaying W, needs to be tagged as a b jet with the secondary-vertex algorithm discussed in Section 5.2.2. A rather loose tag is imposed, asking the discriminating variable $\zeta_{\text{b}} > 1.0$. Events not having any jet combination satisfying this requirement are discarded. The effect of this cut on the selected cross sections of the signal and background processes is shown in the fourth column of Tables 6.1 and 6.2. Considering all $t\bar{t}$ decays as signal, as they all have two b jets in their final state, this cut improves the signal-to-noise ratio from 15 to 59 in the case of the muon analysis and from 9 to 49 in the case of semi-leptonic events with an isolated electron.

For the generation of the W + jets backgrounds, the samples produced in different bins of $\hat{p_{\text{T}}}$ overlap partly. Therefore, the $75 < \hat{p_{\text{T}}} < 500\,\text{GeV}/c$ sample is analysed, along with a lower and upper overlapping bin. This way an estimate is made of the

Process (Simulation details in Section 4.2.4)	Total LO cross section (pb)	Initial reconstruction and selection (fb)	After b-tag requirement (fb)
$gg/q\bar{q} \to t\bar{t} \to b\bar{b}qq'\mu\nu$	81.5	20 345	15 188
$gg/q\bar{q} \to t\bar{t} \to$ other	478.5	6 782	4 886
W + jets ($75 < \hat{p_T} < 500\,\mathrm{GeV}/c$)	1216.0	1 818	343

Table 6.1: Leading-order cross sections and reconstruction and selection efficiencies for signal from semi-muon $t\bar{t}$ events and background events.

Process (Simulation details in Section 4.2.4)	Total LO cross section (pb)	Initial reconstruction and selection (fb)	After b-tag requirement (fb)
$gg/q\bar{q} \to t\bar{t} \to b\bar{b}qq'e\nu$	81.5	15 854	11 873
$gg/q\bar{q} \to t\bar{t} \to$ other	478.5	4 762	3 398
W + jets ($75 < \hat{p_T} < 500\,\mathrm{GeV}/c$)	1216.0	2 239	309

Table 6.2: Leading-order cross sections and reconstruction and selection efficiencies for signal from semi-electron $t\bar{t}$ events and background events.

importance of the neglected part of the process. In Table 6.3 the cross sections after event selection are given in the third column for the muon final state. The remaining cross section after the full analysis, which is described in Section 6.3, is shown in the fourth column. No W + jets events remain after the full analysis, even when the selection thresholds are relaxed. Therefore, the quoted upper limits for the cross-section are expected to become more stringent when a larger sample of simulated events is available. It is concluded that the analysis of only the $75 < \hat{p_T} < 500\,\mathrm{GeV}/c$ bin of the W + jets background is sufficient to show that the background can indeed be suppressed.

Process (Simulation details in Section 4.2.4)	Total LO cross section (pb)	Cross section after event selection (fb)	Cross section after analysis (fb)
$t\bar{t}$ signal	560.0	20 074	759
W + jets ($10 < \hat{p_T} < 100\,\mathrm{GeV}/c$)	27 550.0	295	< 59[1]
W + jets ($75 < \hat{p_T} < 500\,\mathrm{GeV}/c$)	1 216.0	343	< 6[1]
W + jets ($350 < \hat{p_T} < 2200\,\mathrm{GeV}/c$)	4.9	21	< 0.5[1]

Table 6.3: Remaining leading-order W + jets cross section after event selection and full analysis in different bins of $\hat{p_T}$.

[1] The quoted upper limits are the 68% confidence levels, calculated assuming an uncertainty of one event when no events remain after the selection.

6.1.3 b-Enriched jet sample from semi-leptonic $t\bar{t}$ events

To extract a b-enriched jet sample from the selected semi-leptonic $t\bar{t}$ events, only the b coming from the leptonic top-quark decay $t \to bW \to b\ell\nu$ can be used, because one jet is already tagged as b in the system of the hadronically-decaying top quark. To enhance the fraction of b jets in the jet sample of interest, the correct jet association should be selected out of the twelve solutions, hence suppressing the combinatorial background.

Several observables are identified that discriminate between correct jet associations and combinatorial background. These observables can be grouped into a few categories.

- Observables associated to the output of the kinematic fit, namely the χ^2 probability of the fit (Figure 6.1) and the mass difference $m(t_{hadr,fit}) - m(t_{hadr,rec})$ of the hadronically-decaying top quark after and before the fit (Figure 6.2). The distribution of the χ^2 probability is peaked towards 1 for good associations, indicating an overestimation of the resolutions of the objects entering the fit. The mass difference for good associations receives a negative bias due to the residual miscalibration of the Monte-Carlo corrected jet-energy scale. The mass-difference distribution for the bad associations receives its peculiar shape from inversion of the top-mass distribution for wrong jet pairings.

- The transverse momenta $p_T(t_{hadr})$ (Figure 6.3) and $p_T(t_{lept})$ (Figure 6.4) of both the reconstructed top quarks, and $p_T(b_{lept})$ of the b jet from the leptonically decaying top (Figure 6.5).

- The differences in angle $\Delta\phi\,(b_{hadr}, b_{lept})$ (Figure 6.6), $\Delta\phi\,(b_{hadr}, W_{hadr})$ (Figure 6.7), $\Delta\theta\,(b_{hadr}, W_{hadr})$ (Figure 6.8), $\Delta\phi\,(b_{lept}, W_{lept})$ (Figure 6.9), $\Delta\phi\,(q, q')$ (Figure 6.10) and $\Delta\theta\,(q, q')$ (Figure 6.11), where q and q′ in the last variables originate from the hadronically-decaying W_{hadr}. The angle $\Delta\phi\,(b_{hadr}, b_{lept})$ shows the preference for the two correct b jets to be back to back, while the other considered differences $\Delta\phi$ in azimuthal angle take typical shapes for decay products from heavy particles. The polar angle differences $\Delta\theta$ are used rather than $\Delta\eta$, since the jets from top-quark decays are expected to be rather isotropically distributed.

- The ratio $E_{T,jet5}/E_{T,jet4}$ between the fifth and the fourth entry in the E_T-ordered list of jets, or zero if only four jets pass the jet selection criteria (Figure 6.12). This event observable is useful to suppress events where a hard gluon is radiated from the final-state partons, blurring the event kinematics and complicating the jet association. Events without good jet associations are indeed found to have more often a fifth jet with significant transverse energy.

To construct the distributions of these discriminating observables only events are considered with a correct association between the relevant reconstructed and generated physics objects according to a criterion on the separation in (η, ϕ) space. In these events the number of correct and wrong associations is normalized to unity, such that each event with a correct association contributes a unity weight in both the distributions.

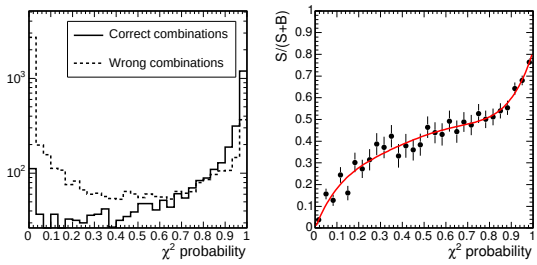


Figure 6.1: χ^2 probability

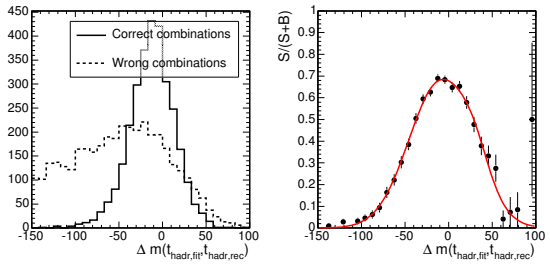


Figure 6.2: $m(\mathrm{t_{hadr,fit}}) - m(\mathrm{t_{hadr,rec}})$

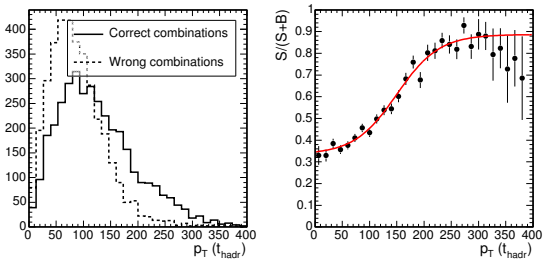


Figure 6.3: $p_{\mathrm{T}}(\mathrm{t_{hadr}})$

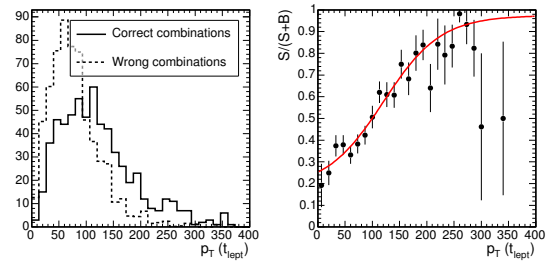


Figure 6.4: $p_{\mathrm{T}}(\mathrm{t_{lept}})$

Some of the identified discriminating variables, like the number of constituents in a jet, are not used to suppress combinatorial background, as they would bias the sample. For the retained observables the distributions $S/(S+B)$ are determined bin-by-bin, with S the number of correct and B the number of wrong associations in the corresponding bin. These distributions are then parametrized with a likelihood ratio function $\mathcal{L}_i(x_i)$ for each observable x_i. The distributions of the variables used for the semi-muon final state and the corresponding likelihood-ratio functions can be found in Figures 6.1 through 6.12. In the semi-electron case these variables show quasi identical behaviour.

With the likelihood-ratio functions $\mathcal{L}_i(x_i)$, a combined likelihood ratio $\mathcal{L}_{\mathrm{semilep}}(\vec{x})$ is calculated for each combinatorial solution in the event reconstruction,

$$\mathcal{L}_{\mathrm{semilep}} = \prod_i \left[\frac{\mathcal{L}_i(x_i)}{\max\{\mathcal{L}_i(x_i)\}} \right]^{1/\sum_j |c_{ij}|}, \tag{6.1}$$

taking into account the correlations c_{ij} between all variables i and j in analogy to (5.13). The jet association considered to give the best reconstruction of the event is chosen as the association with the highest value of $\mathcal{L}_{\mathrm{semilep}}$. The distribution of this maximal combined likelihood ratio is shown, respectively in the case where the chosen jet association yields a correct or wrong association of the jet from the leptonically-decaying top, in Figures 6.13 and 6.14 for the semi-muon final state, and in Figures 6.15 and 6.16 for the semi-electron case. This jet from the leptonically-decaying top is considered well associated if it is flagged as a jet originating from any b parton, using the matching method introduced in Section 5.2, in the algorithm flavour definition. The test whether the considered jet coincides with a b quark before its hadronization is performed with a cone size of $\Delta R = 0.4$.

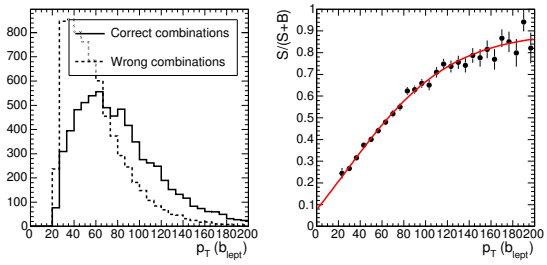


Figure 6.5: $p_{\mathrm{T}}(\mathrm{b_{lept}})$

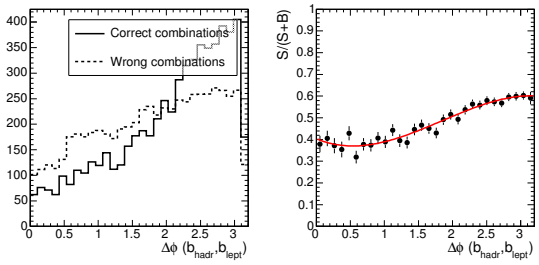


Figure 6.6: $\Delta\phi\,(\mathrm{b_{hadr}}, \mathrm{b_{lept}})$

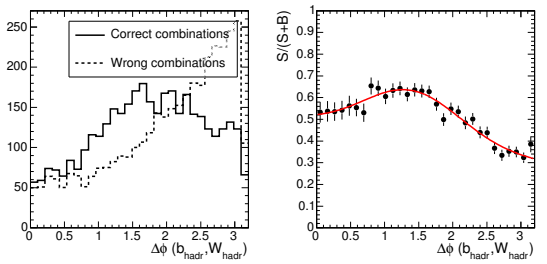


Figure 6.7: $\Delta\phi\,(\mathrm{b_{hadr}}, \mathrm{W_{hadr}})$

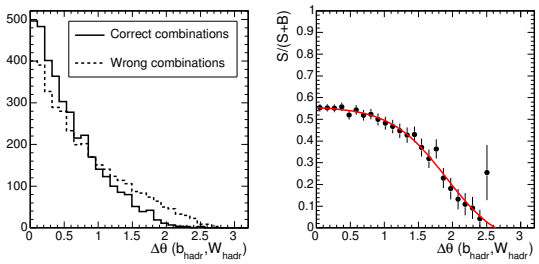


Figure 6.8: $\Delta\theta\,(\mathrm{b_{hadr}}, \mathrm{W_{hadr}})$

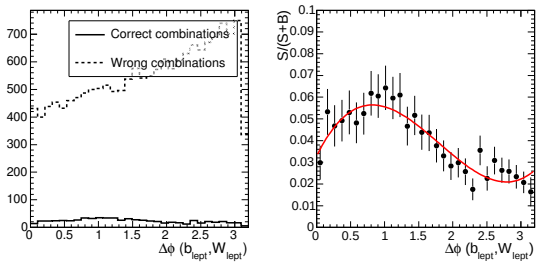


Figure 6.9: $\Delta\phi\,(\mathrm{b_{lept}}, \mathrm{W_{lept}})$

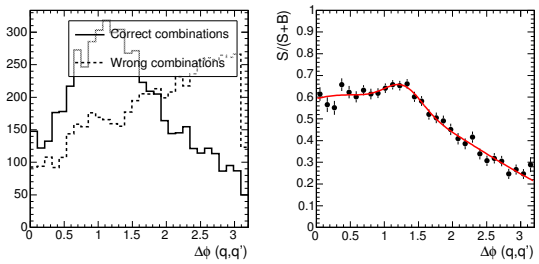


Figure 6.10: $\Delta\phi\,(\mathrm{q}, \mathrm{q}')$

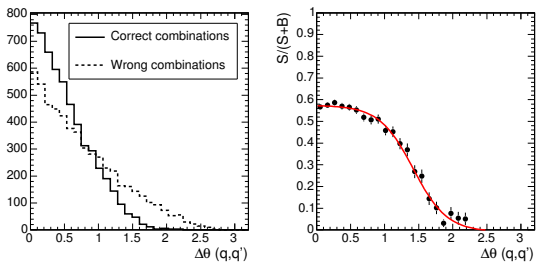


Figure 6.11: $\Delta\theta\,(\mathrm{q}, \mathrm{q}')$

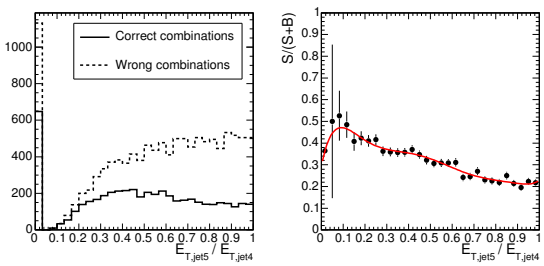


Figure 6.12: $E_{\mathrm{T,jet5}}/E_{\mathrm{T,jet4}}$

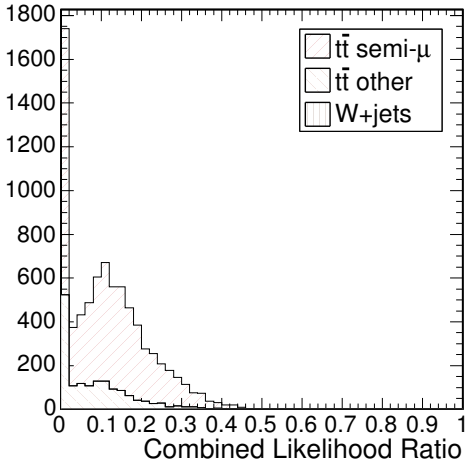


Figure 6.13: Combined likelihood-ratio distributions for the semi-muon final state for correct jet associations.

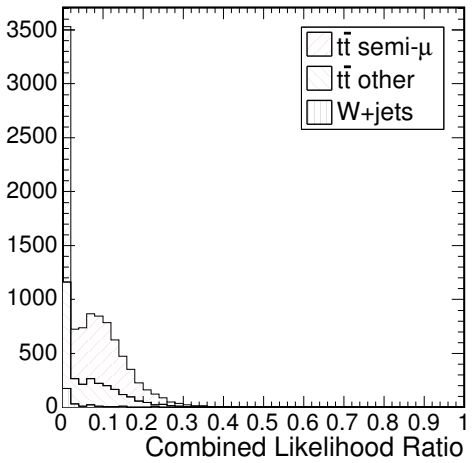


Figure 6.14: Combined likelihood-ratio distributions for the semi-muon final state for wrong jet associations.

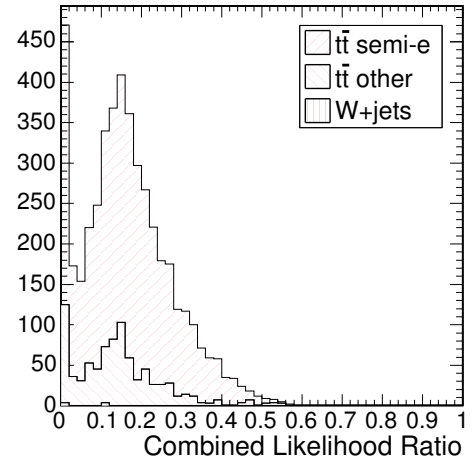


Figure 6.15: Combined likelihood-ratio distributions for the semi-electron final state for correct jet associations.

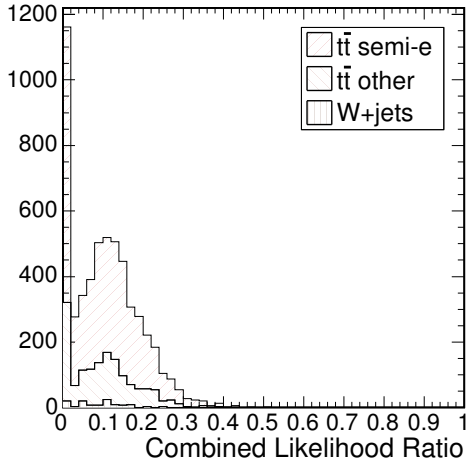


Figure 6.16: Combined likelihood-ratio distributions for the semi-electron final state for wrong jet associations.

The presence of the peak near zero in the combined likelihood-ratio distributions is due to events where the kinematic fit converges, but with a χ^2 probability very close to zero, even for the best association. The overall difference between the electron and muon sample is explained as the result of the different thresholds of the single-electron and single-muon triggers, being 26 GeV and 19 GeV respectively, as discussed in Section 4.3.2. The higher p_{T} cut on the HLT electron candidates selects events with harder kinematics compared to the single-muon trigger, explaining also the relative reduction of the peak near zero.

By imposing a selection threshold on $\mathcal{L}_{\mathrm{semilep}}$ it is possible to increase the b content of the jet sample. In Figure 6.17 the b-jet purity of the sample is shown as a function of the cut on the combined likelihood ratio for the semi-muon sample. In Figure 6.18

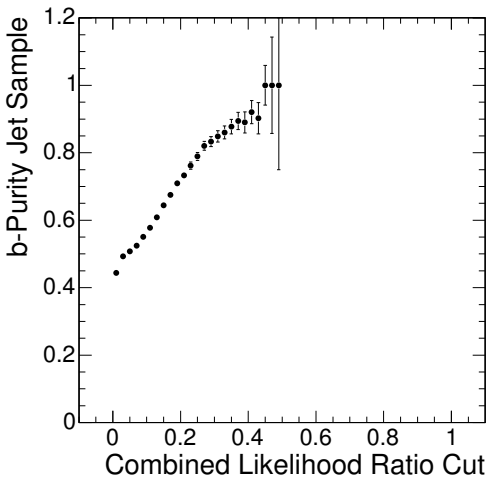


Figure 6.17: b-Jet purity of the semi-muon jet sample as a function of the cut on $\mathcal{L}_{\text{semilep}}$.

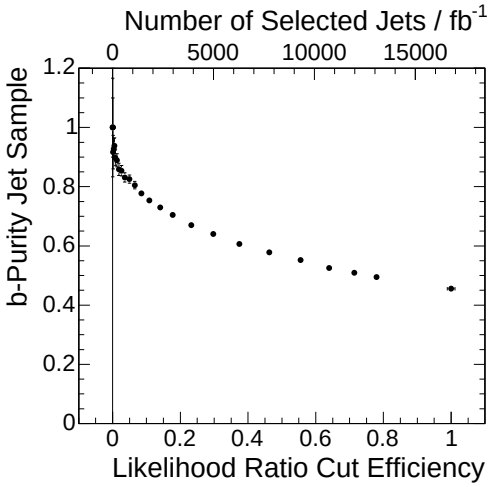


Figure 6.18: b-Jet purity of the semi-muon jet sample as a function of the efficiency of the cut on $\mathcal{L}_{\text{semilep}}$ and the total number of jets remaining in the sample for $1\,\text{fb}^{-1}$.

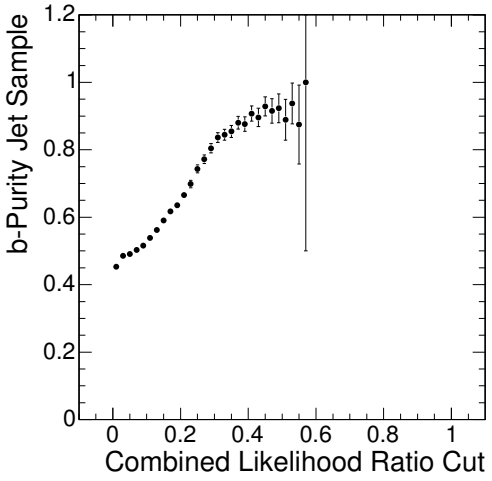


Figure 6.19: b-Jet purity of the semi-electron jet sample as a function of the cut on $\mathcal{L}_{\text{semilep}}$.

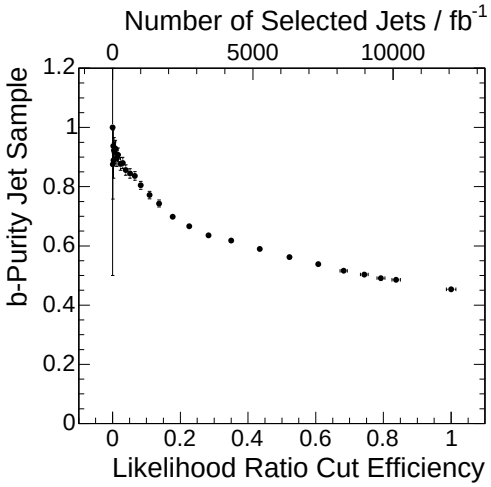


Figure 6.20: b-Jet purity of the semi-electron jet sample as a function of the efficiency of the cut on $\mathcal{L}_{\text{semilep}}$ and the total number of jets remaining in the sample for $1\,\text{fb}^{-1}$.

this purity is shown again, but this time as a function of the efficiency of the cut on $\mathcal{L}_{\text{semilep}}$ and of the total number of jets remaining in the sample for $1\,\text{fb}^{-1}$ of integrated luminosity. Figures 6.19 and 6.20 show the corresponding results for the semi-electron sample.

A cut on the combined likelihood ratio enriches the samples in b jets, even up to almost 100% purity. The description of the purity at values larger than about 90% is statistically limited, both by the number of Monte-Carlo simulated events used to

estimate the purity, and by the precision on the determination of the likelihood-ratio functions. For a b-purity of typically 80%, about 1 000 jets are expected both in the semi-muon and the semi-electron case for an integrated luminosity of $1\,\mathrm{fb}^{-1}$. For lower purities, the jet sample extracted from semi-electron events performs a little worse than the one extracted from semi-muon events, because the initial harder selection on the lepton, and hence the complete event, does not purify the sample as much as the combined likelihood-ratio cut in the muon case.

6.1.4 Event selection for fully-leptonic $\mathrm{t\bar{t}}$ events

In the case of fully-leptonic $\mathrm{t\bar{t}} \to \mathrm{b\bar{b}}\ell_1\nu_1\ell_2\nu_2$ decays, two isolated leptons, being electrons or muons, two jets and transverse missing momentum will define the experimental signature in the detector. For the initial event selection the events are first passed through the Level-1 trigger, and are then asked to fulfil the requirements of at least the single-electron or single-muon trigger of the High-Level Trigger. The two leptons with highest discriminator values $\mathcal{L}_{\mathrm{lepton}}$ are chosen as the two final-state leptons. They are demanded to satisfy the same cut $\mathcal{L}_{\mathrm{lepton}} > 0.01$ as in the semi-leptonic case. From the reconstructed and calibrated jets, the two jets with the highest E_{T} are selected, satisfying $E_{\mathrm{T}} > 25\,\mathrm{GeV}$. The transverse missing momentum, finally, is not constrained.

With the presence of the two isolated leptons in the final state, it is possible to suppress contaminations from other processes without requiring a b-tagged jet in the event selection. Other cuts are however necessary to suppress the important Z + jets and di-boson backgrounds. Due to the important off-shell tail in the invariant-mass distribution of the two leptons from $\mathrm{Z} \to \ell^+\ell^-$ decays, the exclusion of a di-lepton mass window around the Z-boson mass peak is insufficient. Therefore such a mass cut is not applied. Instead, only events are accepted where one electron and one muon are selected as the two leptons with the highest likelihood ratio, passing the initial lepton likelihood-ratio threshold[2]. In addition, the charges of the leptons are also required to be opposite, which further suppresses fake leptons and also the same-charge component of the WW background. In Table 6.4 an overview is given of the efficiencies of the initial reconstruction and of the described cuts on the cross sections of the signal and the relevant backgrounds. Application of the cuts increases the signal-to-noise ratio from < 0.1 to ~ 7. Other background sources are considered negligible [169]. Detailed information on the simulation of the analysed event samples is given in Section 4.2.4.

For the Z+jets background, the same overlaps are present between the various bins in $\hat{p_{\mathrm{T}}}$, as in the W+jets case. In this case the sample $42.5 < \hat{p_{\mathrm{T}}} < 300\,\mathrm{GeV}/c$ is found to be dominant. The lower bin that is analysed to check the importance of the neglected part of the cross section does not overlap, but the missing part is expected to be small. In Table 6.5 the cross sections after the event selection of the fully-leptonic sample are shown in the third column. The remaining cross section after the full analysis, described in Section 6.3, is shown in the fourth column. The events remaining after analysis in the two upper $\hat{p_{\mathrm{T}}}$ bins have a large probability to come from the overlap region due to the steeply falling Z + jets differential cross section, and can in any case be neglected compared to the signal. No Z + jets events remain in the lower $\hat{p_{\mathrm{T}}}$ bin

[2]The possible use of the discarded same-flavour di-lepton final state is discussed in Section 6.4.

Process (Simulation details in Section 4.2.4)	Total LO cross section (pb)	Initial reconstruction and selection (fb)	After $e + \mu$ and charge cuts (fb)
$gg/q\bar{q} \to t\bar{t} \to b\bar{b}\ell_1\nu_1\ell_2\nu_2$ ($\ell = e, \mu$)	25.6	10 795	5 197
$gg/q\bar{q} \to t\bar{t} \to$ other leptonic	31.8	2 899	1 221
WW inclusive[3]	190.0	1 649	406
ZW inclusive	26.8	292	17
Z + jets ($42.5 < \hat{p_T} < 300\,\mathrm{GeV}/c$)	575.7	36 692	534

Table 6.4: Leading order cross sections and reconstruction efficiencies for signal from fully-leptonic $t\bar{t}$ events and background events.

after the full analysis, even when the selection thresholds are relaxed. Therefore, as in the case of the W + jets background, the quoted upper limits for the cross-section are expected to become more stringent when a larger sample of simulated events is available. It is concluded that the consideration of only the $42.5 < \hat{p_T} < 300\,\mathrm{GeV}/c$ sample is sufficient for this analysis.

Process (Simulation details in Section 4.2.4)	Total LO cross section (pb)	Cross section after event selection (fb)	Cross section after analysis (fb)
$t\bar{t}$ fully-leptonic signal	57.4	6 418	1 264
Z + jets ($0 < \hat{p_T} < 40\,\mathrm{GeV}/c$)	31 720.0	552	< 276[4]
Z + jets ($42.5 < \hat{p_T} < 300\,\mathrm{GeV}/c$)	575.7	534	6
Z + jets ($200 < \hat{p_T} < 1400\,\mathrm{GeV}/c$)	7.0	50	6

Table 6.5: Remaining leading-order Z + jets cross section after event selection and full analysis in different bins of $\hat{p_T}$.

6.1.5 b-Enriched jet sample from fully-leptonic $t\bar{t}$ events

In the reconstruction of fully-leptonic $t\bar{t}$ events, wrong jet associations are due to the presence of extra jets. Since no jet is tagged as b in the selection, both the b-jet candidates can be used to construct a b-enriched jet sample. Contrary to the semi-leptonic case where the combinatorial background is the most important source of impurities, in the fully-leptonic case mainly the process backgrounds play this role.

Due to the presence of two neutrinos in the final state, the kinematics of the event are underconstrained. Nevertheless, it is still possible to use the measured kinematic properties to distinguish between correct and wrong jet associations in signal events and select an enriched b-jet sample. The same kinematic properties can also be used to

[3]The WW inclusive background has been generated together with another non–contributing process, causing an artificially-large cross section.

[4]The quoted upper limit is the 68% confidence level, calculated assuming an uncertainty of one event when no events remain after the selection.

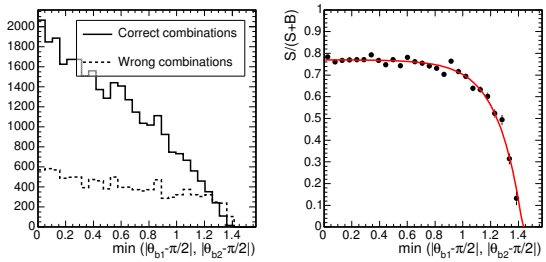


Figure 6.21: $\min\left(\left|\theta_{\mathrm{b1}}-\frac{\pi}{2}\right|,\left|\theta_{\mathrm{b2}}-\frac{\pi}{2}\right|\right)$

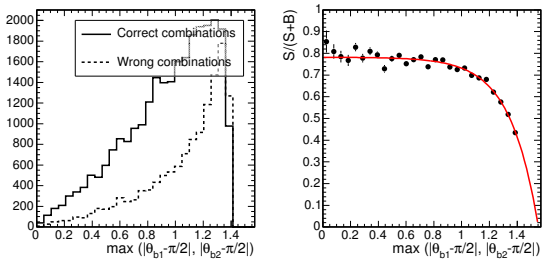


Figure 6.22: $\max\left(\left|\theta_{\mathrm{b1}}-\frac{\pi}{2}\right|,\left|\theta_{\mathrm{b2}}-\frac{\pi}{2}\right|\right)$

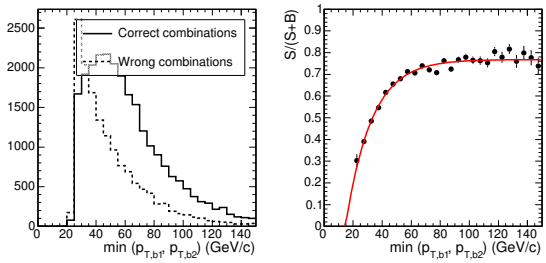


Figure 6.23: $\min\left(p_{\mathrm{T,b1}}, p_{\mathrm{T,b2}}\right)$

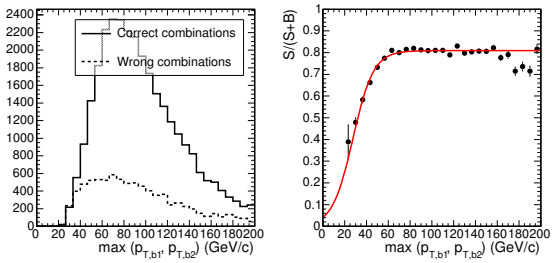


Figure 6.24: $\max\left(p_{\mathrm{T,b1}}, p_{\mathrm{T,b2}}\right)$

discriminate between signal and background events. In the following, all background events are treated as wrongly associated events, applying proper weights according to the relative cross sections. The following observables are identified to contain discrimination power for distinguishing between correct and wrong associations.

- The minimal and maximal deviation of the polar angle from the central direction, $\min\left(\left|\theta_{\mathrm{b1}}-\frac{\pi}{2}\right|,\left|\theta_{\mathrm{b2}}-\frac{\pi}{2}\right|\right)$ (Figure 6.21) and $\max\left(\left|\theta_{\mathrm{b1}}-\frac{\pi}{2}\right|,\left|\theta_{\mathrm{b2}}-\frac{\pi}{2}\right|\right)$ (Figure 6.22), for the two b-jet candidates. These variables reflect the reconstruction difficulties for forward jets and hence the easier mis-association.

- The minimal and maximal p_{T} of the b jets, $p_{\mathrm{T,b1}}$ and $p_{\mathrm{T,b2}}$ (Figures 6.23 and 6.24), and the leptons, $p_{\mathrm{T},\ell 1}$ and $p_{\mathrm{T},\ell 2}$ (Figures 6.25 and 6.26). Higher transverse momenta for the b jets ease the correct jet identification, since extra jets in the event from the underlying event and pile-up collisions have a soft spectrum. The p_{T} of the leptons is used to suppress events with incorrectly-identified soft leptons inside jets.

- The angles $|\Delta\phi\,(\mathrm{b1},\mathrm{b2})|$ (Figure 6.27) and $|\Delta\theta\,(\mathrm{b1},\mathrm{b2})|$ (Figure 6.28) between the b-jet candidates. The same preference for a back-to-back configuration of correctly-associated b jets is observed as in the semi-leptonic case.

- The minimal and maximal absolute angle differences $|\Delta\theta\,(\mathrm{b},\ell)|$ (Figures 6.29 and 6.30) and $|\Delta\phi\,(\mathrm{b},\ell)|$ (Figures 6.31 and 6.32) between the two b jets and the two leptons. These variables reveal the tendency of the b jet and the lepton from a leptonic top decay to be closer to each other than mis-associated combinations. The relative increase of mis-associated entries for small angles is due to mis-identified leptons inside b jets.

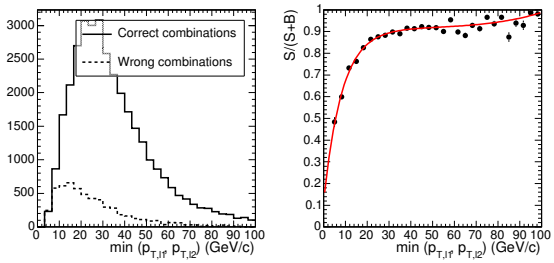


Figure 6.25: $\min(p_{\mathrm{T},\ell 1}, p_{\mathrm{T},\ell 2})$

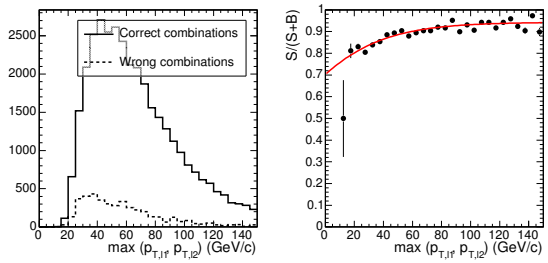


Figure 6.26: $\max(p_{\mathrm{T},\ell 1}, p_{\mathrm{T},\ell 2})$

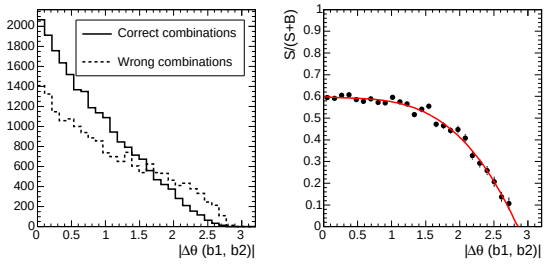


Figure 6.27: $|\Delta\theta(\mathrm{b1}, \mathrm{b2})|$

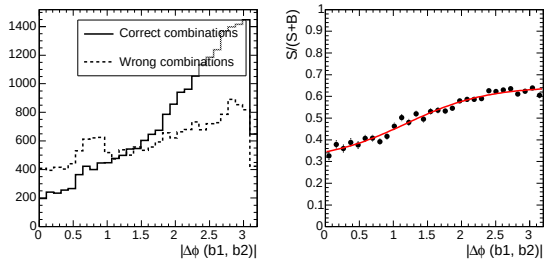


Figure 6.28: $|\Delta\phi(\mathrm{b1}, \mathrm{b2})|$

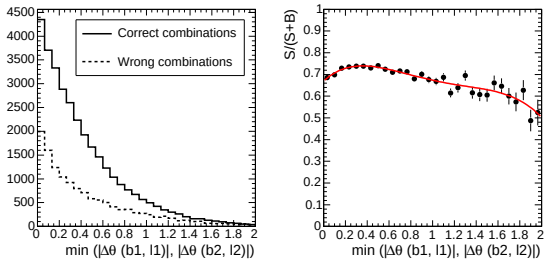


Figure 6.29: $\min(|\Delta\theta(\mathrm{b1}, \ell 1)|, |\Delta\theta(\mathrm{b2}, \ell 2)|)$

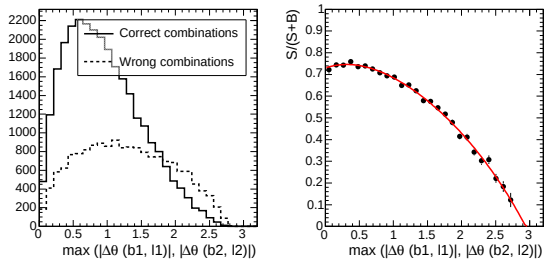


Figure 6.30: $\max(|\Delta\theta(\mathrm{b1}, \ell 1)|, |\Delta\theta(\mathrm{b2}, \ell 2)|)$

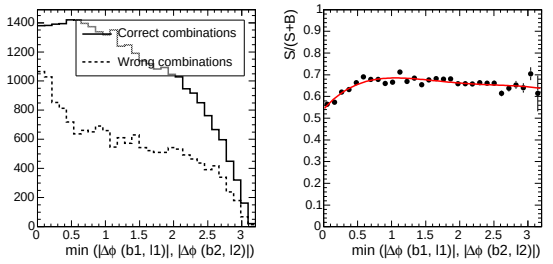


Figure 6.31: $\min(|\Delta\phi(\mathrm{b1}, \ell 1)|, |\Delta\phi(\mathrm{b2}, \ell 2)|)$

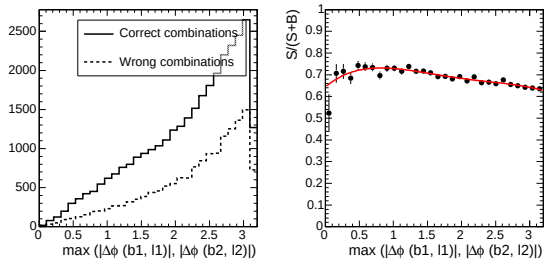


Figure 6.32: $\max(|\Delta\phi(\mathrm{b1}, \ell 1)|, |\Delta\phi(\mathrm{b2}, \ell 2)|)$

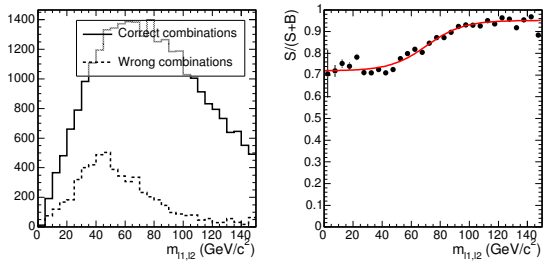


Figure 6.33: $m_{\ell 1,\ell 2}$

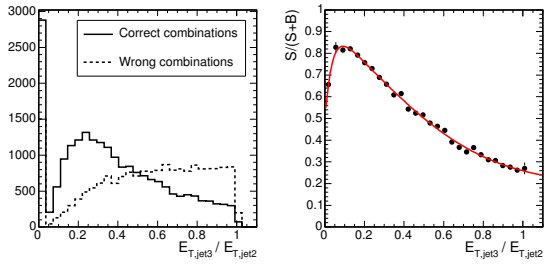


Figure 6.34: $E_{\mathrm{T,jet\ 3}}/E_{\mathrm{T,jet\ 2}}$

- The invariant mass $m_{\ell 1,\ell 2}$ of the leptons (Figure 6.33). This variable allows to further suppress mis-identified leptons and background events.
- The ratio $E_{\mathrm{T,jet3}}/E_{\mathrm{T,jet2}}$ between the third and the second entry in the E_{T}-ordered list of jets, or zero in the case only two jets pass the selection criteria (Figure 6.34). As for the semi-leptonic case, this variable makes it possible to suppress hard final-state gluon radiation.

All signal and background events are used to construct the distributions of these discriminating observables, properly weighted with the relative cross sections. Association of objects is performed between the relevant reconstructed and generated physics objects according to a criterion on the separation in (η, ϕ) space. The only combinatorial ambiguity in the events stems from the association of the b quarks and the leptons to the originating top quarks. For the observables sensitive to this ambiguity both possible associations are accounted for separately, each with a weight of 1/2.

For the observables in the above-mentioned list the $S/(S+B)$ values are determined bin-by-bin, with S representing the number of correct and B the number of wrong associations in the corresponding bin. From the distribution of these values the likelihood-ratio functions $\mathcal{L}_i(x_i)$ are determined for each observable x_i. The distributions of the considered variables and the corresponding likelihood-ratio functions are shown in Figures 6.21 through 6.34.

Using the likelihood-ratio functions $\mathcal{L}_i(x_i)$, a combined likelihood ratio $\mathcal{L}_{\mathrm{fulllep}}(\vec{x})$, similarly as defined in (6.1), is calculated for the two possible associations of the b quarks and the leptons to the originating top quarks. The association with the highest value of both is retained. The distributions of this highest combined likelihood ratio for signal and backgrounds are shown in Figures 6.35 and 6.36, respectively for correct and wrong associations. Correctly associated events have both jets flagged as originating from a b parton, using the same association criteria as in the semi-leptonic case. Most of the wrong associations originate from signal $\mathrm{t\bar{t}}$ events, where at least one jet is misidentified and does not originate from a b quark. The process backgrounds are smaller but more problematic, as both the b-jet candidates in these events contribute to the sample's impurity.

The fraction of b jets in the sample as a function of the selection cut on the combined likelihood ratio $\mathcal{L}_{\mathrm{fulllep}}$ is shown in Figure 6.37. The same purity, this time as a function of the efficiency of the cut and of the total number of jets remaining in the sample, is shown in Figure 6.38, for an integrated luminosity of $1\,\mathrm{fb}^{-1}$.

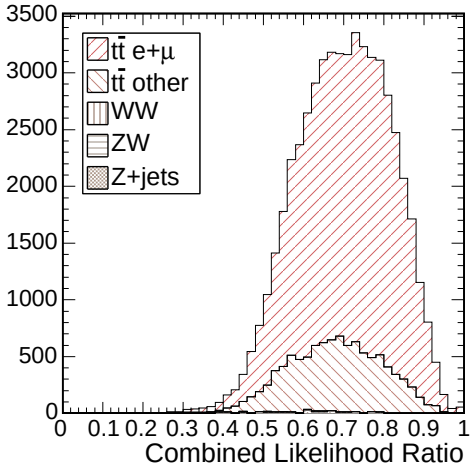


Figure 6.35: Combined likelihood-ratio distributions for the fully-leptonic final state with correct jet association.

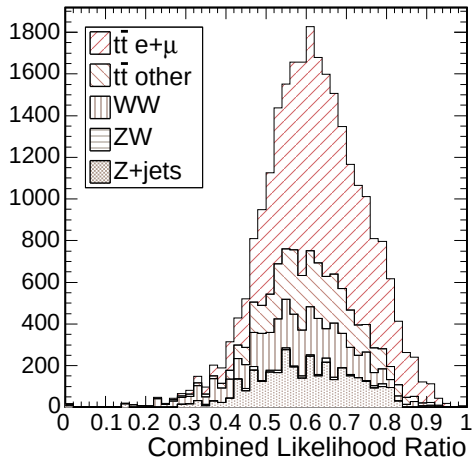


Figure 6.36: Combined likelihood-ratio distributions for the fully-leptonic final state with wrong jet association.

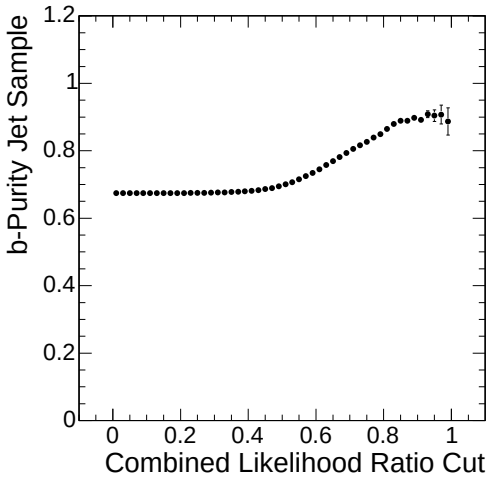


Figure 6.37: b-Jet purity of the fully-leptonic jet sample as a function of the cut on $\mathcal{L}_{\text{fulllep}}$.

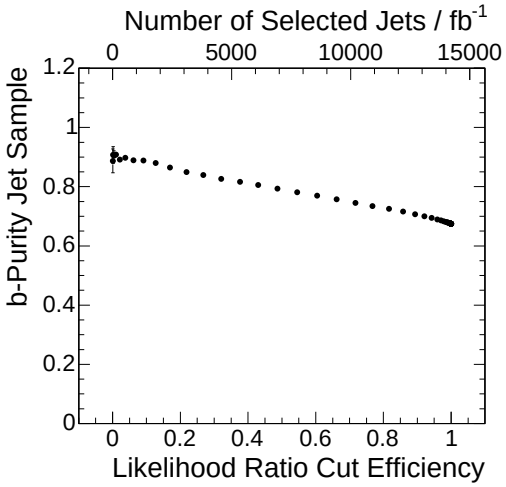


Figure 6.38: b-Jet purity of the fully-leptonic jet sample as a function of the efficiency of the cut on $\mathcal{L}_{\text{fulllep}}$ and the total number of jets remaining in the sample for $1\,\text{fb}^{-1}$.

For a typical b-jet purity of 80%, about 6 000 jets are expected to remain in the fully-leptonic sample for an integrated luminosity of $1\,\text{fb}^{-1}$. This b-jet purity reaches maximally $\sim 90\%$ for a hard cut on the combined likelihood ratio cut. As in the semi-leptonic case, the b-purity estimation at such a high cut value is statistically limited in this analysis, both by the number of Monte-Carlo simulated events used to estimate the purity, and by the precision on the determination of the likelihood-ratio functions. In addition, for a high likelihood-ratio threshold, wrong reconstructions of $\text{t}\bar{\text{t}}$ events with one mis-identified jet constitute the largest part of the background. The combined likelihood ratio, however, is designed to also suppress the process backgrounds, which

have very different characteristics. Therefore it can be assumed that a more detailed treatment of the high likelihood-ratio events further increases the b-jet purity.

6.2 Systematic Uncertainties on the b-Jet Purity

The relation between the b purity of the jet samples and the threshold on the corresponding combined likelihood ratios $\mathcal{L}_{\text{semilep}}$ and $\mathcal{L}_{\text{fulllep}}$ is obtained from Monte-Carlo simulations, and has therefore systematic uncertainties resulting from the uncertainties on the underlying models or from the assumptions on the detector operation. In particular all the effects causing the purity to be smaller than 100% will contribute a systematic uncertainty to this estimated purity. In this section the possible theoretical and experimental contributions to the systematic uncertainty on the samples' b-jet purities are examined in detail. The results are evaluated as a function of the likelihood-ratio observables for each considered jet sample. For the evaluation of these systematic uncertainties many Monte-Carlo event samples have been generated with either parametrized or detailed detector simulation. Details on the simulation of these samples are given in Section 4.2.4.

6.2.1 Phenomenological systematic uncertainties

Several systematic uncertainties result from making use in the simulations of phenomenological models and previously-obtained experimental data.

ISR/FSR Initial- and final-state gluon radiation (ISR/FSR) gives rise to extra jets in the event and causes the kinematics of the reconstructed jets to be different than expected. The effect due to the uncertainties in the model have been evaluated as discussed in Section 3.3.1, changing simultaneously the scales Λ_{QCD} and Q^2_{max} between their respective uncertainties. Various samples have been simulated at the limits of the confidence interval $[0.15, 0.35]$ GeV for Λ_{QCD}, $[0.25, 4]Q^2_{\text{hard}}$ for Q^2_{max} in initial-state radiation and $[1, 16]Q^2_{\text{hard}}$ for Q^2_{max} in final-state radiation. The fundamental QCD scale Λ_{QCD} and the scales Q^2_{max} and Q^2_{hard}, used for matching the initial- and final-state parton showers to the matrix element, are introduced respectively in Sections 1.1.3 and 3.3.1. Half of the difference in b-jet purity of the jet samples is taken as the resulting uncertainty. The results for this systematic effect from ISR/FSR are summarized and discussed in detail in Section 6.2.3.

Signal and background cross sections The theoretical or experimental uncertainties on the signal and background cross sections imply an uncertainty on the b-jet purities. In the semi-muon and semi-electron case, the background from $\mathrm{W}+\text{jets}$ is very small and hence its cross-section uncertainty has negligible impact on the b-jet purity. Therefore, a scaling of the signal cross section does not contribute either, for the same reason why the analysis is insensitive to uncertainties on the luminosity, as discussed in detail in Section 6.2.2. In the fully-leptonic case, however, the presence of the background cannot be neglected. Due to this, the b-jet purity of the sample is sensitive to the uncertainty on the ratio between the

signal and background cross section. In the initial phase of the LHC, the uncertainties on the measured $t\bar{t}$ or background cross sections will be large [86, 169]. Also some of the uncertainties from the currently best theoretical signal and background cross-section calculations are still sizable [170–173]. Therefore a conservative 20% uncertainty on the total background cross section is applied to account for the yet unknown uncertainty on the signal-to-background ratio. The resulting uncertainty on the b-jet purity of the fully-leptonic jet sample is discussed in Section 6.2.3.

Contamination from other $t\bar{t}$ final states In all of the semi-muon, semi-electron and fully-leptonic event samples, a significant amount of $t\bar{t}$ events remain after the event selection, with a decay chain different from the considered signal. This is mostly due to the presence of a τ lepton in the final state. Although these final states also contain the b jets from the top-quark decay, the efficiency for correct jet association is lower and hence the b-jet purity in these subsamples is lower than in the real signal. An uncertainty on the ratio of the number of events from the correct and from the other $t\bar{t}$ final states will therefore result in a systematic uncertainty on the b-jet purity. Such an uncertainty comes for example from uncertainties on the W-boson branching fractions and from uncertainties on the event selection efficiencies obtained from Monte-Carlo simulations. It has been checked that a change of 20% on the ratio of the number of events with correct and other $t\bar{t}$ final states results in a negligible systematic uncertainty of less than 1%, compatible with zero, on the absolute b purity for all the considered samples.

Pile-up collisions The multiple inelastic collisions happening during the same bunch crossing give rise to extra jets. These jets originate from other primary vertices in the event, and this fact is used to suppress their contribution as described in Section 5.1.5. The remaining systematic uncertainty due to the presence of pile-up collisions has been estimated with detailed detector simulation. The absolute difference in purity between the samples without pile-up collisions and with low-luminosity pile-up conditions is found to be smaller than 1% and compatible with zero. This effect is thus negligible, since the pile-up collisions will be monitored to a much better accuracy once data arrive, as discussed in Sections 2.2.2 and 2.3.2.

Underlying event The underlying event, composed of the protons' remnants and the effects of multiple interactions, also contributes extra jets in the event. The systematic influence due to the uncertainties on the model have been evaluated as described in Section 3.1.3, varying the assumptions on the partonic matter distribution in the proton and the parameters that regulate the $2 \to 2$ partonic cross section at very low momentum transfer. The resulting absolute uncertainty on the b-jet purity is found to be compatible with zero and smaller than 0.5%, and hence negligible compared to other effects.

Parton density functions The systematic uncertainty originating from the proton's parton density functions are evaluated following the prescription discussed in Section 3.1.2 in the case of the fully-leptonic sample. Starting from the central value of the CTEQ6M fit, the sample purity is evaluated for each of the sets of

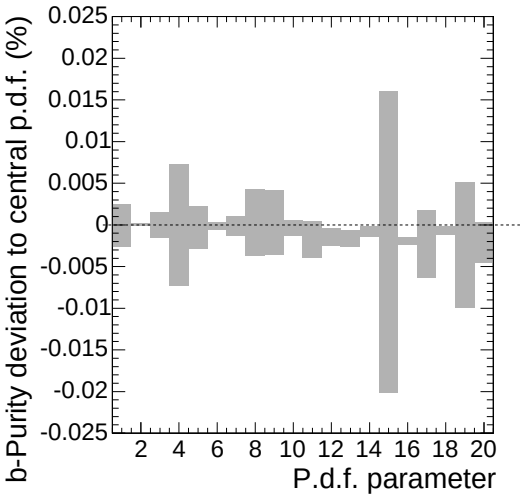


Figure 6.39: Absolute deviations of the b-jet purity of the jet sample extracted from fully-leptonic $t\bar{t}$ events due to uncertainties on the 20 parameters in the CTEQ6M fit of the parton density functions (p.d.f.) of the proton.

parton distributions associated with the uncertainties on the 20 parameters of the fit. This is performed by calculating a weight on an event-by-event basis, starting from the initial partons' fractional momenta x_1 and x_2, their flavour (gluon or type of quark) and the event's $Q^2 \approx m_t^2$. In this way 40 deviations in sample purity are obtained, corresponding to the positive and negative variation of the uncertainty on the 20 parameters. The overall normalization due to these parton density-function uncertainties is found to be of the order of a few percent, but this does not influence the sample purity as is discussed in detail in Section 6.2.2 for the uncertainties from the luminosity. The actual effect on the sample purity due to the uncertainties from parton density functions, which is the observable of interest, is shown in Figure 6.39 for the fully-leptonic sample, without a minimal requirement on $\mathcal{L}_{\text{fulllep}}$. Each of the deviations is found to be smaller than 0.1%, and hence the systematic effect from the combination of all can be neglected. The effect is also expected to be negligible for the samples extracted from the semi-leptonic $t\bar{t}$ samples, since the selected events probe the parton distributions in the same kinematical range.

Top-quark mass The dependence of the kinematic fit on the top-quark mass introduces also a possible systematic uncertainty due to the top-quark mass uncertainty. To evaluate this, the effect of a deviation of 5 GeV from the nominal $m_t = 175\,\text{GeV}/c^2$ has been investigated in the semi-muon sample. Just as for the JES, which changes the jet spectra in a similar way, the absolute effect due to the top-quark mass uncertainty is smaller than 1% and compatible with zero, and can hence be neglected. This result can be directly generalized to the semi-electron sample, but also to the fully-leptonic sample which is expected to be even less affected, since no kinematic fit with explicit top-quark mass dependence is applied.

Light- and b-quark fragmentation Although the cuts used to construct the jet samples are chosen to keep the samples as unbiased as possible, correlations might exist that make the sample purity dependent on the parton fragmentation models. It has however been checked in the fully-leptonic sample that these effects are negligible, following the guidelines described in Section 3.3.3. The same result is expected to hold for the jet samples from semi-leptonic decays.

6.2.2 Detector-related systematic uncertainties

Several systematic uncertainties arise from detector-related sources:

Luminosity The uncertainties originating from the luminosity measurement will only result in an uncertainty on the overall number of events, and not on the likelihood ratio distributions. Therefore, due to the uncertainty on the luminosity, no contribution is expected to the systematic uncertainty on the samples' b purity as a function of the cut on the likelihood ratio. This has explicitly been verified by simulating an increase of 10% in the overall normalization. If on the other hand the purity would be considered as a function of the remaining number of events after a cut on the likelihood ratio, then the analysis would become sensitive to the luminosity uncertainty, which has been checked to be indeed the case.

b**-Tag efficiency** During the event selection of the semi-leptonic samples, the b jet from the hadronically-decaying top is loosely tagged, as described in Section 6.1.2. The uncertainty on the tagging efficiency is also considered as a potential systematic uncertainty on the b content of the selected jet sample. The evaluation of this uncertainty has been performed in an inclusive way, by changing the cut in the event selection on the b-identification probability from Section 5.2.2, such that an absolute upper and lower deviation of 10% on the b-tagging efficiency is obtained for the considered sample. The resulting uncertainty is shown in Section 6.2.3 for the semi-leptonic samples.

Jet energy scale Because of the dependence of the kinematic fit on the W-boson and top-quark masses, the uncertainty on the jet energy scale (JES) contributes as well. Although a better knowledge of the JES is expected, as shown in Section 5.1.3, an inclusive 5% miscalibration has been applied on the jets from the semi-muon sample after the jet energy-scale corrections, in order to simulate the effect of the JES uncertainty. Also this effect is negligible as it is smaller than 1% in absolute scale and compatible with zero. The same is expected for the similar semi-electron sample and the fully-leptonic sample which has lower jet multiplicity.

Misalignment Detector misalignment in CMS can have large impacts on expectations of particular measurements, since in general all simulations are performed with a perfectly-aligned detector. A misalignment of the central tracker modules indeed has a sizable impact on the efficiencies of b-tagging algorithms [149, 151]. For the measurement of b-identification efficiencies presented in this chapter, however, almost no tracker-related information is used. Therefore the effect of

a possible tracker misalignment on the b-tagging efficiency will be directly measured on the data, with an accuracy that is to first order independent of it. Hence, misalignment is not considered to be a systematic uncertainty.

6.2.3 Systematic-uncertainty results

For each of the considered systematic uncertainties, the purities of the samples are evaluated at the minimal and maximal boundaries of the considered effect, as a function of the likelihood-ratio cut. The resulting systematic uncertainty on the purity is taken as half of the difference in purity for the two extremes. In Figure 6.40 an example is shown of the behaviour of the b-jet purity as a function of the cut on $\mathcal{L}_{\text{semilep}}$ in the semi-muon sample, for the boundaries that are considered to evaluate the systematic uncertainty associated to initial- and final-state gluon radiation, as described in Section 6.2.1. It is observed that the enrichment of the b-jet sample, obtained by cutting harder on $\mathcal{L}_{\text{semilep}}$, allows indeed to control and suppress systematic uncertainties.

To obtain the total systematic uncertainty, each of the considered non-negligible contributions is calculated as a function of the threshold on $\mathcal{L}_{\text{semilep}}$ or $\mathcal{L}_{\text{fulllep}}$ and summed quadratically. In Figures 6.41, 6.42 and 6.43, for the semi-muon, semi-electron and fully-leptonic samples respectively, the resulting total systematic uncertainty is shown as a function of the cut on the corresponding $\mathcal{L}_{\text{semilep}}$ or $\mathcal{L}_{\text{fulllep}}$. Parametrizations of the significant individual contributions are also shown on these figures. On the distribution of the total systematic uncertainty, a fit is performed to obtain a parametrization of the total systematic uncertainty as a function of the likelihood-ratio cut. Because of the quadratic sum, statistically insignificant negative differences in purity are added positively, making the total fit conservative in regions where an insufficient amount of simulated events is available. In Table 6.6, an overview is given of the systematic uncertainties at a typical fixed b-jet purity of 80%.

Jet sample (80% b-jet purity)	ISR/FSR systematic	Other systematics	Total systematics
semi-muon	3.7%	$< 0.5\%$	3.7%
semi-electron	3.8%	$< 0.5\%$	3.8%
fully leptonic	5.0%	0.7%	5.0%

Table 6.6: Overview of the systematic uncertainties on the b-jet purity at a fixed purity of 80%.

For all three samples, the systematic uncertainty is dominated by the initial- and final-state gluon radiation, governed by the Q^2_{max} and Λ_{QCD} scales in the `PYTHIA` simulations. To unravel the individual importance of the initial- and the final-state radiation, and of the hard and soft transition of the parton shower, the relevant `PYTHIA` parameters, discussed in Section 3.3.1, are checked separately for their systematic effect.

The variation of Λ_{QCD}, affecting the soft-gluon emission at the end of the parton shower, turns out to have no observable effect on the jet samples' purities. This is expected, since more or less soft-jet activity should only marginally hamper the identification of the hard jets of interest, as also observed for the uncertainties on the

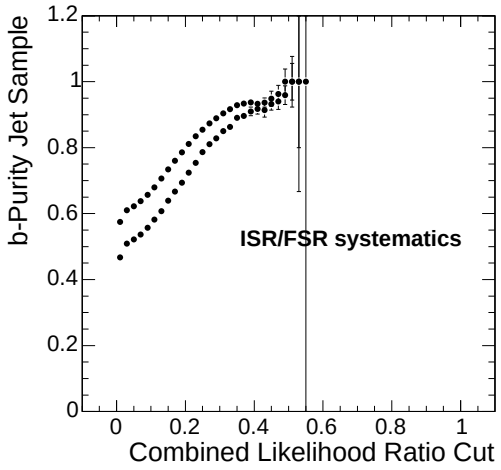


Figure 6.40: b-Jet purity as a function of the cut on $\mathcal{L}_{\text{semilep}}$ for the semi-muon jet sample, in the case of the two extremes considered to evaluate the systematic uncertainty from ISR/FSR.

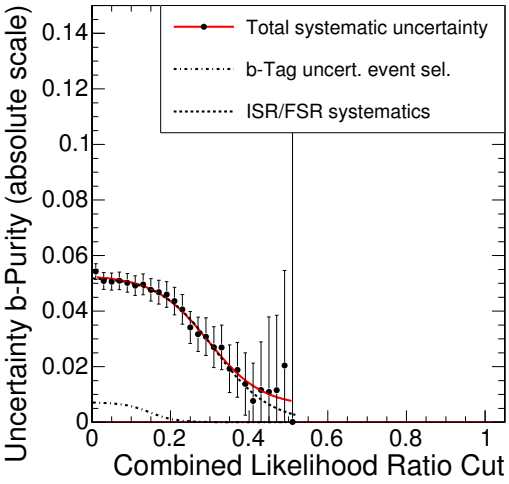


Figure 6.41: Systematic uncertainties on the b-jet purity of the semi-muon jet sample as a function of the cut on $\mathcal{L}_{\text{semilep}}$.

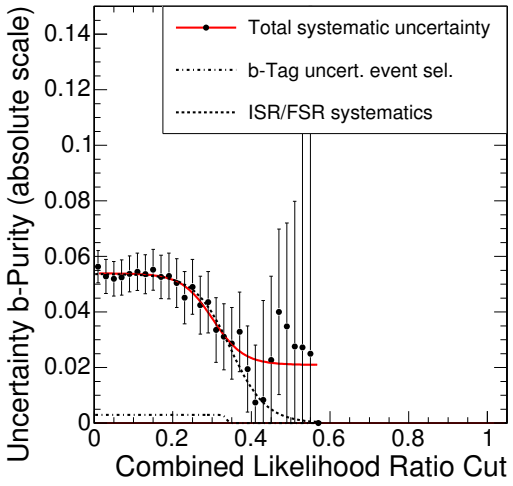


Figure 6.42: Systematic uncertainties on the b-jet purity of the semi-electron jet sample as a function of the cut on $\mathcal{L}_{\text{semilep}}$.

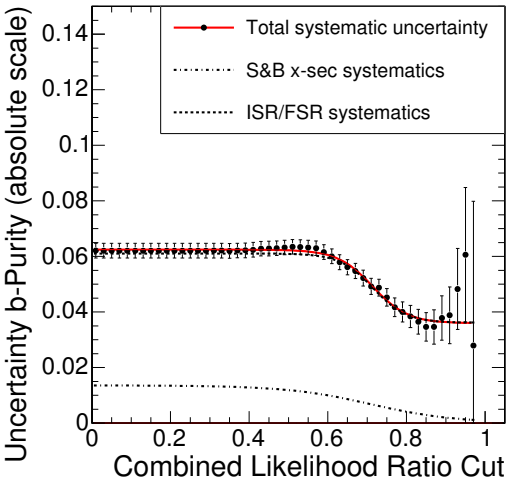


Figure 6.43: Systematic uncertainties on the b-jet purity of the fully-leptonic jet sample as a function of the cut on $\mathcal{L}_{\text{fulllep}}$.

underlying event and caused by pile-up collisions. The variation of the cut-off scale Q^2_{max}, defining the scale at which the initial- and final-state parton shower is matched to the matrix element, on the other hand, influences mainly the hard jet content in the event. In this analysis, the purity is found to be almost unaffected by that variation of the final-state matching scale, and the observed effect is completely due to uncertainties on the initial gluon radiation. The deviation of $Q^2_{\text{max}}/Q^2_{\text{hard}}$ between 0.25 and 4, which is supposed to reflect the uncertainty from the matching of the matrix element to the initial-state radiation, causes the average number of jets with $p_{\text{T}} > 25\,\text{GeV}$ to change

from 3.71 to 4.09 in the semi-leptonic sample without likelihood-ratio cuts applied. Asking for jets with $p_{\mathrm{T}} > 50\,\mathrm{GeV}$, the average number changes between 2.31 and 2.59, showing the effect increases relatively for harder jets.

Although the obtained result reflects the best current understanding [101], it is expected that these uncertainties will be better understood in the future. From the experimental side, data from the Tevatron and later from the LHC itself will constrain the currently used confidence interval [174], for instance with the measurement of the spectra of additional jets in W, Z and $\mathrm{t\bar{t}}$ production. On the phenomenological side, a better description in the simulation of the additional hard-jet production can be envisaged, by using techniques of matrix-element–parton-shower matching, or even full next-to-leading order Monte-Carlo simulation programs. The dominant uncertainty in this analysis from the initial-state radiation may possibly turn out to be controlled below the percent level, motivating future exploration of some of the neglected systematic effects. Further detailed studies are needed, however, to make more quantitative statements on these issues.

6.3 Measurement of b-Identification Efficiencies

6.3.1 Principle of the measurement

On the three jet samples obtained from data, an identical technique is used for the actual measurement of the b-identification efficiencies. In the following, the fraction of b jets in the sample is indicated with x_{b}, the fraction of c jets with x_{c}, and the fraction of jets from light quarks and gluons with x_{ℓ}. When a b-identification algorithm at a certain working point is applied on the jet sample, a fraction x_{tag} of the events will be tagged. This fraction can be written as

$$x_{\mathrm{tag}} = \epsilon_{\mathrm{b}} x_{\mathrm{b}} + \epsilon_{\mathrm{c}} x_{\mathrm{c}} + \epsilon_{\ell} x_{\ell}, \tag{6.2}$$

where ϵ_{b}, ϵ_{c} and ϵ_{ℓ} are the corresponding tagging efficiencies of the applied algorithm on the considered jet sample. Hence, the efficiency to tag b jets is

$$\epsilon_{\mathrm{b}} = \frac{1}{x_{\mathrm{b}}} \left[x_{\mathrm{tag}} - \epsilon_{\mathrm{c}} x_{\mathrm{c}} - \epsilon_{\ell} x_{\ell} \right]. \tag{6.3}$$

In this relation only x_{tag} is measured on the data. The fractions x_{b}, x_{c} and x_{ℓ} are related by

$$x_{\mathrm{b}} + x_{\mathrm{c}} + x_{\ell} = 1, \tag{6.4}$$

and have to be determined from Monte-Carlo simulations. Finally, the probabilities to tag other jet types as b jets should be obtained from another measurement. Methods for such a measurement and an estimation of the resulting expected accuracies, however, are only currently being explored in CMS. Therefore, in this analysis, these efficiencies ϵ_{c} and ϵ_{ℓ} are taken to be the efficiencies obtained from Monte-Carlo simulations. Defining

$$x_{\mathrm{o}} = x_{\mathrm{c}} + x_{\ell}, \tag{6.5}$$

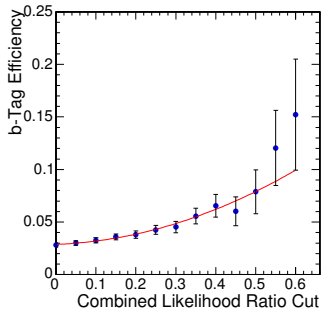


Figure 6.44: Non-b-jet tag efficiency ϵ_{o} in the semi-muon jet sample as a function of the cut on $\mathcal{L}_{\mathrm{semilep}}$.

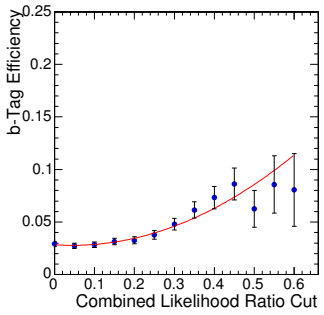


Figure 6.45: Non-b-jet tag efficiency ϵ_{o} in the semi-electron jet sample as a function of the cut on $\mathcal{L}_{\mathrm{semilep}}$.

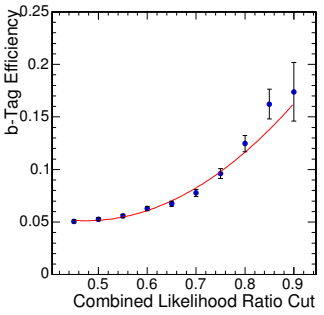


Figure 6.46: Non-b-jet tag efficiency ϵ_{o} in the fully-leptonic jet sample as a function of the cut on $\mathcal{L}_{\mathrm{fulllep}}$.

the probability ϵ_{o} to tag any non-b jet in the sample as a b jet, fulfils

$$\epsilon_{\mathrm{o}} x_{\mathrm{o}} = \epsilon_{\mathrm{c}} x_{\mathrm{c}} + \epsilon_{\ell} x_{\ell}. \tag{6.6}$$

Combining (6.4), (6.5) and (6.6), equation (6.3) can be simplified as

$$\epsilon_{\mathrm{b}} = \frac{1}{x_{\mathrm{b}}} \left[x_{\mathrm{tag}} - \epsilon_{\mathrm{o}} (1 - x_{\mathrm{b}}) \right]. \tag{6.7}$$

6.3.2 Measurement on the considered samples

To measure in practise the efficiency ϵ_{b} for an algorithm, first a choice of parameters is made for the b-tagging, defining the working point of the considered algorithm. As an example, using the combined secondary-vertex b-tag algorithm described in Section 5.2.2, the measurement is performed for a certain fixed cut $\zeta_{\mathrm{b,cut}} = 1.5$ on the b-discriminator of a jet. It should however be noted that the method is completely independent of the chosen algorithm and its parameters, and can in principle be applied to all b-tagging algorithms, offline as well as online, for any parameter setting.

For certain values of the cut on the combined likelihood ratio $\mathcal{L}$, the sample-specific mistag efficiency ϵ_{o} is determined, as shown in Figures 6.44, 6.45 and 6.46 for the semi-muon, semi-electron and the fully-leptonic samples, respectively. The uncertainties on the figures are due to the limited Monte-Carlo simulated samples, and are not accounted for, since they generate only a second-order uncertainty that will decrease with increasing simulation statistics. By means of a polynomial fit, the function $\epsilon_{\mathrm{o}}(\mathcal{L}_{\mathrm{cut}})$ is obtained for each sample at a working point with $\zeta_{\mathrm{b,cut}} = 1.5$ for the secondary-vertex based algorithm. This can be repeated for any chosen working point, fixed by $\zeta_{\mathrm{b,cut}}$. The strong dependence of the mis-identification rate on the cut on the likelihood ratio is explained by the fact that this cut selects events with overall harder kinematic properties. In addition, for the semi-leptonic cases, the cut more likely rejects soft extra jets in the event than c jets. For jets chosen in a limited η and E_{T} region, this dependence is indeed checked to become weaker.

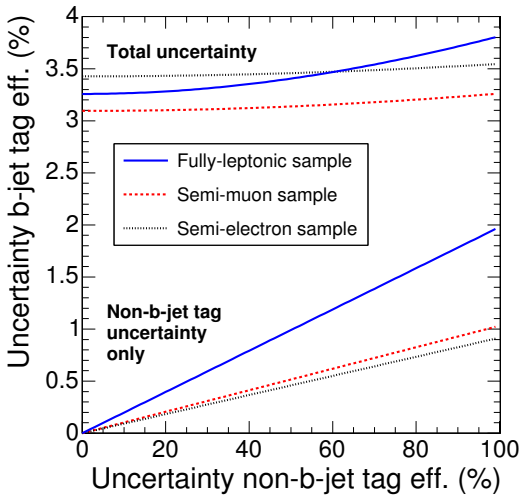


Figure 6.47: Uncertainty on the b-jet identification efficiency as a function of the uncertainty on the mis-identification rate ϵ_{o} for the optimal inclusive measurement with $1\,\mathrm{fb}^{-1}$ of integrated luminosity.

In the next step, an estimation $\hat{\epsilon_{\mathrm{b}}}(\zeta_{\mathrm{b,cut}}, \mathcal{L}_{\mathrm{cut}})$ for the b-tagging efficiency is performed using (6.7), for several cuts $\mathcal{L}_{\mathrm{cut}}$ on the likelihood ratio and for the chosen value of $\zeta_{\mathrm{b,cut}}$. To calculate the uncertainty on these estimations, the total systematic uncertainty on x_{b} from Section 6.2 is used. The fraction of tagged jets x_{tag} determines the statistical uncertainty. On the function $\epsilon_{\mathrm{o}}(\zeta_{\mathrm{b,cut}}, \mathcal{L}_{\mathrm{cut}})$, finally, a rather large uncertainty of 20% is assumed in absence of a reliable expected experimental accuracy. Due to the high purities and the low mis-identification probability, this extra systematic uncertainty only contributes as a small second-order correction. To illustrate this, the impact of the ϵ_{o} uncertainty on $\hat{\epsilon_{\mathrm{b}}}$ is visualized in Figure 6.47, for the optimal measurement defined further in this section, with $1\,\mathrm{fb}^{-1}$ of integrated luminosity. It is seen that the linearly growing contribution from the mis-identification rate is marginal, even for large uncertainties on ϵ_{o}. This uncertainty on ϵ_{o} can only become significant when the measurement of the b-jet identification efficiency reaches a sub-percent accuracy and the sample purity is still well below 100%.

In Figures 6.48, 6.49 and 6.50 the statistical, systematic and total uncertainty are shown for the different samples, as a function of the cut on $\mathcal{L}$. The total uncertainty shows a minimum at some optimal likelihood-ratio cut $\mathcal{L}_{\mathrm{opt}}$. At this cut $\mathcal{L}_{\mathrm{opt}}$, for a certain sample, at the working point defined by $\zeta_{\mathrm{b,cut}}$, the optimal estimation of the b-identification efficiency $\hat{\epsilon_{\mathrm{b}}}(\zeta_{\mathrm{b,cut}}|\mathcal{L}_{\mathrm{opt}}(\zeta_{\mathrm{b,cut}}))$ is finally obtained inclusively. For an integrated luminosity of $1\,\mathrm{fb}^{-1}$ the secondary-vertex based algorithm at the working point $\zeta_{\mathrm{b,cut}} = 1.5$ resulted for the optimal b-tagging efficiency determination in $\hat{\epsilon_{\mathrm{b}}} = 58.0\% \pm 3.1\%$ for the semi-muon sample, $\hat{\epsilon_{\mathrm{b}}} = 58.7\% \pm 3.4\%$ in the the semi-electron sample, and $\hat{\epsilon_{\mathrm{b}}} = 59.2\% \pm 3.3\%$ in the fully-leptonic sample. Although these are sample-specific estimations, the inclusive results are comparable.

The dependency of the efficiency and its uncertainty on the working point of the b-tagging algorithm, defined by $\zeta_{\mathrm{b,cut}}$, is investigated. The complete analysis on all three b-enriched jet samples is repeated for several choices of $\zeta_{\mathrm{b,cut}}$, including a re-evaluation

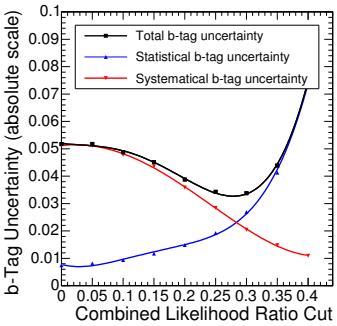


Figure 6.48: Statistical, systematic and total uncertainty in the semi-muon jet sample as a function of the cut on $\mathcal{L}_{\text{semilep}}$, for $1\,\text{fb}^{-1}$.

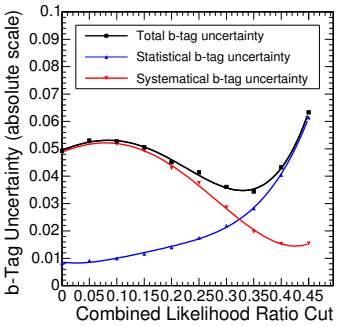


Figure 6.49: Statistical, systematic and total uncertainty in the semi-electron jet sample as a function of the cut on $\mathcal{L}_{\text{semilep}}$, for $1\,\text{fb}^{-1}$.

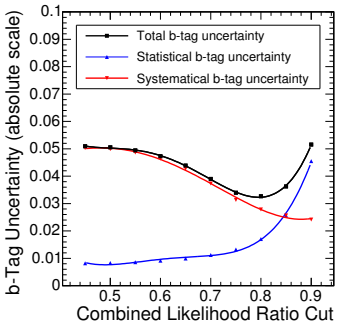


Figure 6.50: Statistical, systematic and total uncertainty in the fully-leptonic jet sample as a function of the cut on $\mathcal{L}_{\text{fulllep}}$, for $1\,\text{fb}^{-1}$.

of the Monte-Carlo mis-identification efficiency and the optimization of the total uncertainty. In Figure 6.51 the b-tagging efficiency is shown as a function of $\zeta_{\text{b,cut}}$ for the different inclusive samples evaluated at their respective points of minimal uncertainty. Only the statistical uncertainties are indicated, illustrating that the inclusive evaluation on the three independent samples agrees. In Figure 6.52 the total absolute and relative uncertainties on these efficiency estimates are shown as a function of $\zeta_{\text{b,cut}}$.

As expected, the estimated b-tagging efficiency is reduced when the value of the threshold $\zeta_{\text{b,cut}}$ is increased. At the same time, the total relative uncertainty on the efficiency estimator does increase, however, when applying tighter b-tagging criteria. The analyses developed in the context of the CMS experiment [86] need to include diverse sets of b-tagging criteria. Figure 6.52 indicates that depending on the chosen working point different uncertainties are expected on the estimation of the b-tagging efficiency measured with this method on data.

6.3.3 Parametrization of the b-identification efficiency

Up to now, the analysis of the measurement of b-tagging algorithm efficiencies is performed on a jet sample extracted from top-quark pair events with specific selection cuts applied. To make the measured values of $\hat{\epsilon_{\text{b}}}$ and the corresponding uncertainties applicable in general, a parametrization of $\hat{\epsilon_{\text{b}}}$ is needed with respect to the most important process-dependent variables it is sensitive to. In this analysis a parametrization as a function of the jet's transverse energy E_{T} and its pseudorapidity η is made. As is shown in Section 5.2.4, these variables have a large influence on the performance of b-tagging algorithms, and depend strongly on the kinematics of the channels under study. Other variables affecting the b-tagging efficiencies, like the number of jet constituents, are expected to be less channel-dependent, or to be correlated to the jet E_{T} and η. Provided enough data is available, also these dependencies can be measured, or corrections can be foreseen from Monte-Carlo simulation studies.

To accomplish the parametrization with respect to transverse energy and pseudora-

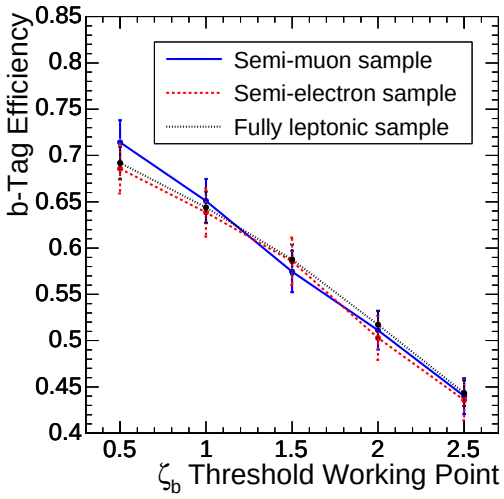


Figure 6.51: Dependence of the estimated inclusive b-tagging efficiency on the algorithm's working-point threshold $\zeta_{b,cut}$, for $1\,fb^{-1}$.

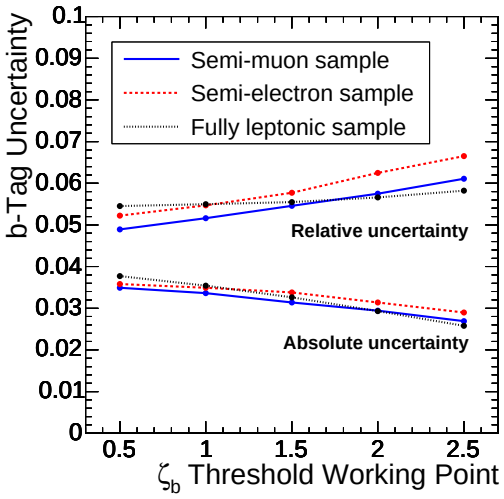


Figure 6.52: Dependence of the absolute and relative uncertainty of the estimated inclusive b-tagging efficiency on the algorithm's working-point threshold $\zeta_{b,cut}$, for $1\,fb^{-1}$.

pidity, the range in jet E_T is divided in four bins. Given the limited number of selected events for $1\,fb^{-1}$ of integrated luminosity, only two bins in η are considered, separating jets in the barrel and the endcaps of the detector. Depending on the available statistics, this binning can be refined. For each bin the analysis chain to estimate the b-tagging efficiency is repeated, using the likelihood-ratio functions built from the complete set of events. Only the jets in the considered bin are retained in the jet sample used for the efficiency estimation. The systematic uncertainties are estimated separately for each bin in E_T and η, to allow for the optimization of the total uncertainty on the b-tagging efficiency determination separately as well. In Figures 6.53, 6.55, 6.57 and 6.54, 6.56, 6.58 the dependence of ϵ_b on the transverse energy of the jet is shown for the considered samples in the barrel ($|\eta| < 1.5$) and endcap ($|\eta| > 1.5$) region respectively, for $1\,fb^{-1}$ of integrated luminosity, applying the secondary-vertex algorithm at the working point defined by $\zeta_{b,cut} = 1.5$. The statistical uncertainty is shown, together with the quadratic sum of the statistical and systematic uncertainty.

The results presented in the figures show that the b-tagging efficiency can be measured on a selected sample of b-enriched jets from data. Although the principle of this measurement is tested on simulated events, these events are in the various steps considered as if they are real data. Some expected tendencies discussed in Chapter 5 are indeed retrieved. The b-jet tagging efficiency as a function of the jet transverse energy is found to be decreasing for soft and hard jets. The efficiency ranges between 30% and 60%, with a maximum efficiency at $E_T \approx 100\,GeV$ in the barrel, coinciding with the expectations discussed in Section 5.2.4. Furthermore, the comparison of the results in the barrel and endcaps confirms the decreased performance in the forward region. The results for the three jet samples are found to be compatible with one another within their respective uncertainties. The presence of possible remaining smaller biases in the

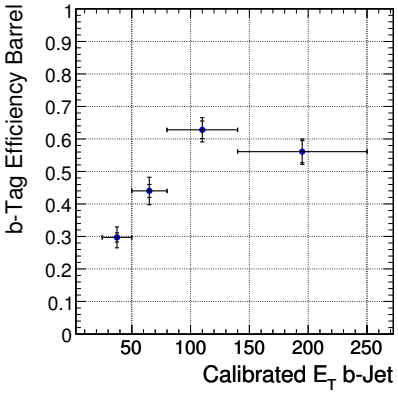


Figure 6.53: Estimated ϵ_b in the barrel ($|\eta| < 1.5$) for the semi-muon jet sample, as a function of the E_T of the jets, for $1\,\text{fb}^{-1}$.

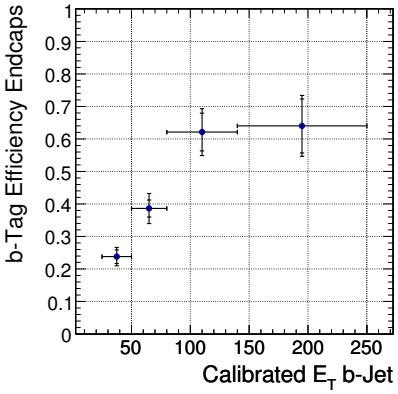


Figure 6.54: Estimated ϵ_b in the endcaps ($|\eta| > 1.5$) for the semi-muon jet sample, as a function of the E_T of the jets, for $1\,\text{fb}^{-1}$.

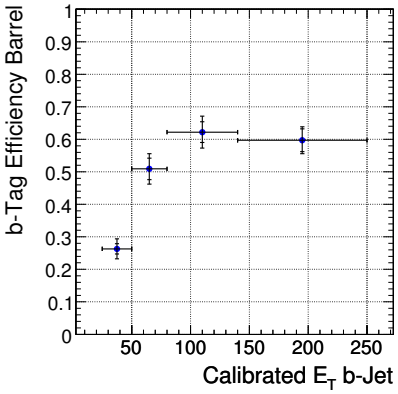


Figure 6.55: Estimated ϵ_b in the barrel ($|\eta| < 1.5$) for the semi-electron jet sample, as a function of the E_T of the jets, for $1\,\text{fb}^{-1}$.

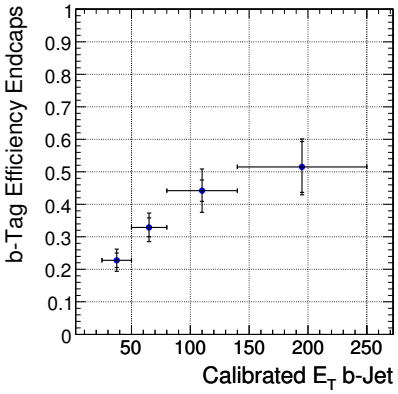


Figure 6.56: Estimated ϵ_b in the endcaps ($|\eta| > 1.5$) for the semi-electron jet sample, as a function of the E_T of the jets, for $1\,\text{fb}^{-1}$.

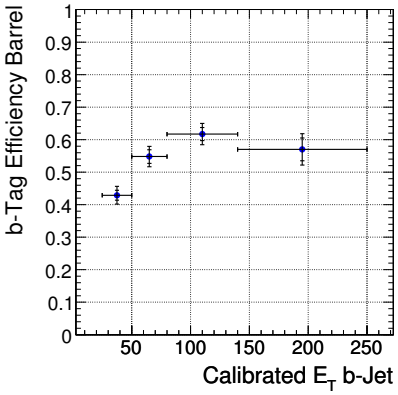


Figure 6.57: Estimated ϵ_b in the barrel ($|\eta| < 1.5$) for the fully-leptonic jet sample, as a function of the E_T of the jets, for $1\,\text{fb}^{-1}$.

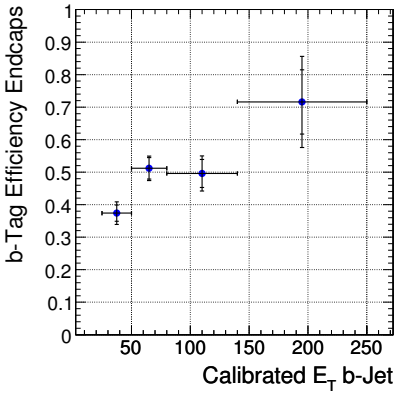


Figure 6.58: Estimated ϵ_b in the endcaps ($|\eta| > 1.5$) for the fully-leptonic jet sample, as a function of the E_T of the jets, for $1\,\text{fb}^{-1}$.

samples cannot be excluded, however, and requires more simulation statistics to be studied in detail.

6.3.4 Expected uncertainties on b-tagging efficiencies

The presented evaluation of the b-tagging efficiencies on Monte-Carlo simulated events directly provides an estimate for the uncertainties that can be obtained with the actual measurement on data. The expected absolute and relative uncertainties, optimized bin-by-bin, are summarized as a function of jet transverse energy for $\zeta_{\mathrm{b,cut}} = 1.5$ in Figures 6.59, 6.61 and 6.63 for the different samples in the barrel and Figures 6.60, 6.62 and 6.64 respectively in the endcaps.

The performance of the semi-muon and semi-electron samples is comparable, with the absolute uncertainty on the b-tag efficiency for $1\,\mathrm{fb}^{-1}$ of integrated luminosity between 3% and 5% in the barrel and ranging from 3% to 10% for increasing jet E_{T} in the endcaps. Also for $1\,\mathrm{fb}^{-1}$, the fully-leptonic sample reaches the same precision for low-E_{T} jets, and performs slightly worse for hard jets with up to 15% uncertainty for hard forward jets.

Up to here all estimations are performed for an integrated luminosity of $1\,\mathrm{fb}^{-1}$. The analysis is also extrapolated to $10\,\mathrm{fb}^{-1}$ of integrated luminosity, keeping the same expectations for the systematic uncertainties. The effect of this scaling of the integrated luminosity can be seen on the same figures. The lowest bin in E_{T} does not show much improvement, reflecting the limitation from the systematic uncertainty. As argued in Section 6.2.3, this limitation is due to the conservative fit of the total systematic uncertainty as a function of the optimization likelihood ratio.

To extrapolate even further would hence require an additional large simulation effort. The optimal working point would move further to higher values of the likelihood-ratio cut, completely into a region that is currently limited by the statistical precision of the simulations. In addition, small contributions from other effects apart from gluon radiation should be investigated further in this region, as they might dominate at some point, or for some bins in E_{T} and η of the jets.

6.3.5 Combination of the results

Supposing that the E_{T} and η parametrizations cover all sample dependency of the considered b-jet efficiency measurements, then the results for the b-identification efficiencies and uncertainties from the semi-muon, semi-electron and fully-leptonic $\mathrm{t\bar{t}}$ samples can be combined, since these samples do not overlap. Systematic uncertainties are considered fully correlated, since in all samples the dominant contribution arises from the initial- and final-state gluon radiation. No additional combined minimization of the uncertainties is performed. In Figures 6.65 and 6.66 the combined result for the efficiencies is shown for barrel jets ($|\eta| < 1.5$) and jets in the endcaps ($|\eta| > 1.5$) respectively. In Figures 6.67 and 6.68 the corresponding absolute and relative uncertainties on the b-tagging efficiency are shown for $1\,\mathrm{fb}^{-1}$ and $10\,\mathrm{fb}^{-1}$ of integrated luminosity.

The absolute uncertainty of the combined measurement on the b-tagging efficiency with $1\,\mathrm{fb}^{-1}$, is about 3% over the considered E_{T} range of jets in the barrel, going down to 2% for $10\,\mathrm{fb}^{-1}$. Because of the decreasing b-tag efficiency for soft jets, the relative

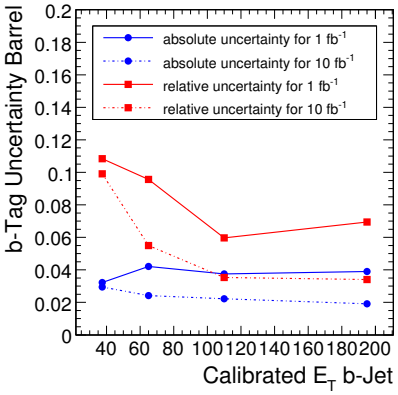


Figure 6.59: Expected ϵ_{b} uncertainty in the barrel ($|\eta| < 1.5$) for the semi-muon jet sample, as a function of the E_{T} of the jets.

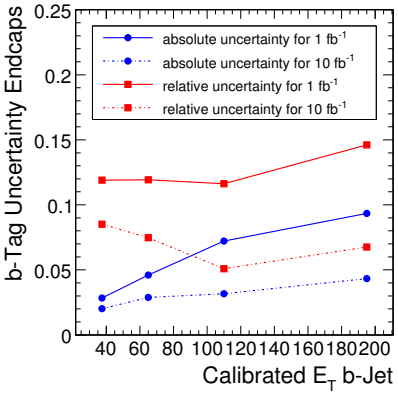


Figure 6.60: Expected ϵ_{b} uncertainty in the endcaps ($|\eta| > 1.5$) for the semi-muon jet sample, as a function of the E_{T} of the jets.

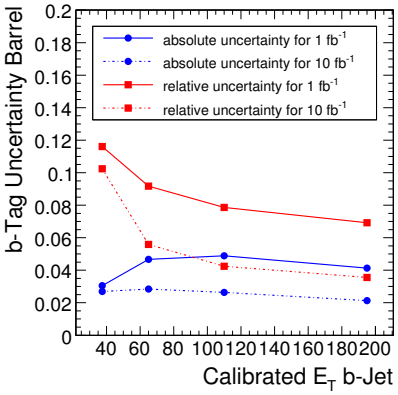


Figure 6.61: Expected ϵ_{b} uncertainty in the barrel ($|\eta| < 1.5$) for the semi-electron jet sample, as a function of the E_{T} of the jets.

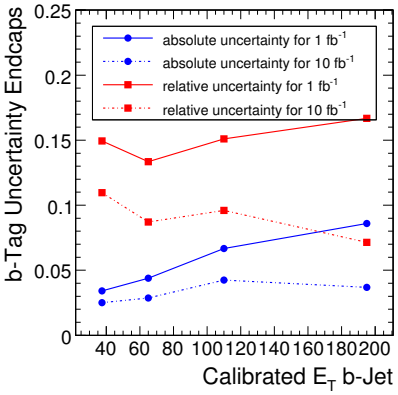


Figure 6.62: Expected ϵ_{b} uncertainty in the endcaps ($|\eta| > 1.5$) for the semi-electron jet sample, as a function of the E_{T} of the jets.

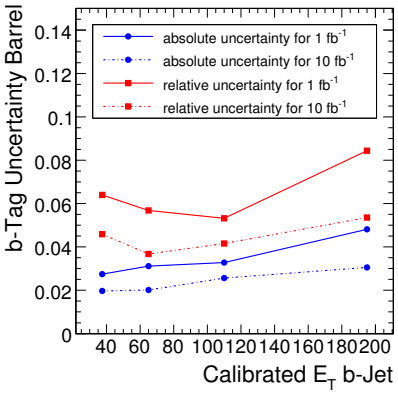


Figure 6.63: Expected ϵ_{b} uncertainty in the barrel ($|\eta| < 1.5$) for the fully-leptonic jet sample, as a function of the E_{T} of the jets.

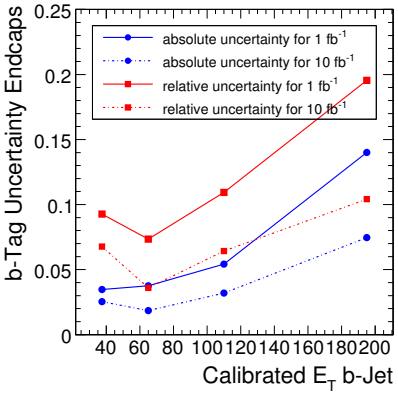


Figure 6.64: Expected ϵ_{b} uncertainty in the endcaps ($|\eta| > 1.5$) for the fully-leptonic jet sample, as a function of the E_{T} of the jets.

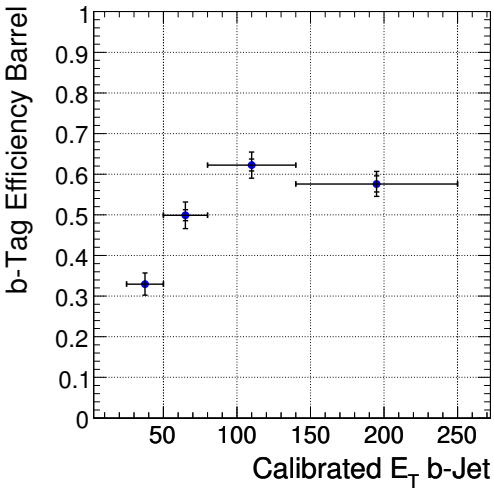


Figure 6.65: Combined estimation of ϵ_b in the barrel ($|\eta| < 1.5$), as a function of the E_T of the jets, for $1\,fb^{-1}$.

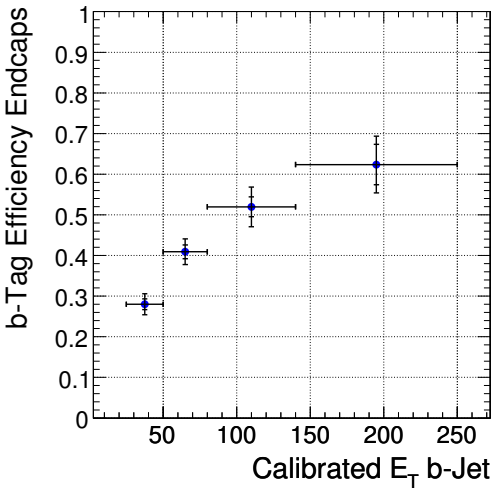


Figure 6.66: Combined estimation of ϵ_b in the endcaps ($|\eta| > 1.5$), as a function of the E_T of the jets, for $1\,fb^{-1}$.

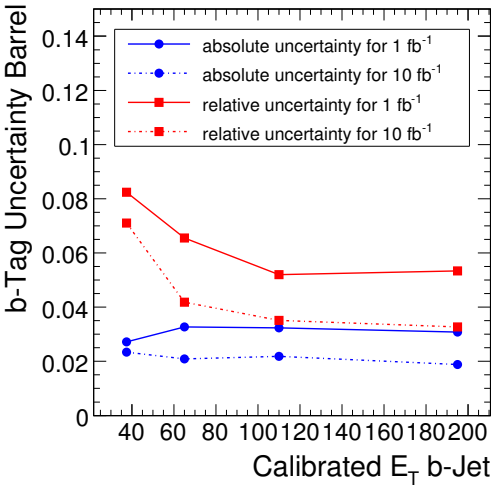


Figure 6.67: Combined expected ϵ_b uncertainty in the barrel ($|\eta| < 1.5$), as a function of the E_T of the jets.

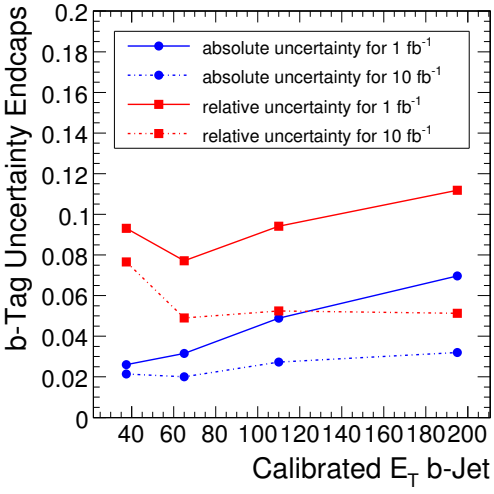


Figure 6.68: Combined expected ϵ_b uncertainty in the endcaps ($|\eta| > 1.5$), as a function of the E_T of the jets.

uncertainty at low jet E_T is at the level of 8% for $1\,fb^{-1}$ and 7% for $10\,fb^{-1}$, decreasing to respectively 5% and 4% for increasing E_T. In the endcaps the determination of the efficiencies becomes less accurate for harder jets, resulting in an expected absolute uncertainty ranging from 3% to 7% with increasing jet E_T for $1\,fb^{-1}$, and from 2% to 3% for $10\,fb^{-1}$. Relatively the uncertainty is comprised between 8% and 11% for $1\,fb^{-1}$ and between 5% and 8% for $10\,fb^{-1}$.

Analyses relying on b-tagging can use these results to estimate systematic influences due to the uncertainties on the b-tagging efficiencies. In the typical example of b-jets from light Higgs-boson $H \to b\bar{b}$ decays, with mostly central and hard jets, about 7% uncertainty on the b-tagging efficiency is expected for $1\,fb^{-1}$ of data, down to 4% with

$10\,\mathrm{fb}^{-1}$. In cross-section measurements where b jets need to be identified to select the signal from the background, as is for instance the case for the $\mathrm{t\bar{t}}$ process [168], uncertainties on the b-tagging efficiencies can be dominant.

Although the obtained results reflect the most precise efficiency estimations achievable, the expected uncertainties remain conservative. Larger samples of simulated events would improve the parametrization of the systematic uncertainties as a function of the multivariate variables. This would result in less conservative estimations for the scale of the systematic uncertainties in b-enriched jet samples at large purities. In addition, with the LHC data itself an improved understanding of initial-state radiation in proton collisions is to be achieved. Currently, this knowledge must be extrapolated from experiments at lower energies. The adoption of more accurate and dedicated event-generation tools will further reduce the uncertainties in the applied event simulations.

6.4 Alternative measurements

Although in principle the method described in this chapter is able to measure the efficiencies of any b-identification algorithm, also other methods can be used to perform similar measurements on data, which is useful for complementarity and as a cross check. Along the same line of the proposed jet samples, another sample can be extracted from fully-leptonic top decays, this time for the $e + e$ and $\mu + \mu$ final states. Application of a b-tag on one jet and possibly the exclusion of lepton pairs with invariant mass in a Z-boson mass window, should be sufficient to suppress the $\mathrm{Z} + \mathrm{jets}$ background.

Also other processes can be a source of b-enriched or theoretically well-known jet samples: $\mathrm{Z} + \mathrm{jets}$, $\mathrm{Zb\bar{b}}$, etc. Alternative approaches, exploiting the almost uncorrelated nature of soft-lepton b-tagging algorithms, discussed in Section 5.2.4, can be used to select a jet sample for calibration of lifetime-based algorithms, or vice versa. Finally, one can look at ratio's between single and double b-tag rates in samples with two b jets, possibly on a b-enriched jet sample.

Not only the identification efficiencies of b-tagging algorithms needs to be calibrated, also the knowledge from the measurement of mistag rates is crucial in many physics analyses. The presented method can also contribute in this domain. Indeed, in semi-leptonic $\mathrm{t\bar{t}}$ decays, the content of c quarks in the resulting jets from hadronic W decays is well known. Using the values of b-tag efficiencies from another measurement, or by means of a simultaneous determination, the mis-identification rate of b-tagging algorithms on c jets can be potentially measured. The measurement of tagging rates of udsg jets, on the other hand, needs different methods, but is considered less problematic.

Chapter 7

Charged Higgs-Boson Identification in the $H^{\pm} \to tb$ Channel

In the minimal extension of the Standard Model Higgs sector to a general two-Higgs-doublet model, introduced in Section 1.2.2, five physical Higgs bosons are expected after electroweak symmetry breaking, of which two are charged. Because of their charge, these bosons exhibit very different decay properties compared to the neutral states. Only the dominant Standard Model decay modes can be exploited at the LHC, namely the $H^{\pm} \to \tau\nu$ and the $H^{\pm} \to tb$ channels. The $H^{\pm} \to \tau\nu$ decay and non Standard Model decay modes are further discussed in Chapter 8. In this chapter the discovery potential for a heavy charged Higgs boson is studied in the $H^{\pm} \to tb$ decay channel. Two strategies are considered, based on the tagging of either three or four b jets. In both cases the main background from $t\bar{t}$ + jets, with either additional b jets or jets mistagged as b, is found to be very large, kinematically similar to the signal, and very hard to suppress.

The two search strategies are introduced in Section 7.1, based on triple and fourfold b-jet identification. In Sections 7.2 and 7.3 these search strategies are explored in detail. The description of the event reconstruction and the initial event selection is followed by the choice of the best jet association and the reconstruction of the Higgs-boson mass spectrum in the case of triple b tagging. Next, an optimization of the signal-to-background ratio is introduced, that allows to maximize the signal's significance. Finally the influence of systematic uncertainties is assessed.

7.1 Charged Higgs-Boson Search Strategies in the $H^{\pm} \to tb$ Channel

The branching fractions of the charged Higgs boson depend strongly on its mass, as is shown in Section 1.2.2. In Figure 1.7 the $m_{H^{\pm}}$ dependence of the branching fractions is shown for two typical values of $\tan\beta$, excluding decays to supersymmetric particles. For $m_{H^{\pm}} < m_t + m_b$, the $H^{\pm} \to \tau\nu$ decay dominates. For larger masses, the channel $H^{\pm} \to tb$ opens up.

In the analyses presented in this chapter, charged Higgs-boson production is studied in the gluon-fusion mode $gg \to tH^{\pm}b$. As discussed in Section 3.4.3, the choice of the

simulated process at leading order depends on the goals of the analysis. In the case of triple b tagging, the $gb \to tH^{\pm}$ interaction is considered, with the initial b quark taken from the parton density functions. When four b jets are identified, the kinematics of the spectator b needs to be properly simulated with the $gg \to tH^{\pm}b$ process.

In both cases, complex final states arise from the decay $H^{\pm} \to tb$ and the additional associated top-quark decay. From an experimental point of view, the most interesting final state is the semi-leptonic one, since an isolated charged lepton (electron or muon) allows efficient triggering of the events and the branching fraction of this decay mode of the two top quarks is large ($\sim 30\%$). In the three b-jet mode, the semi-leptonic final state becomes

$$gb \to tH^{\pm} \to t\bar{t}b \to W^{+}W^{-}b\bar{b}b \to q\bar{q}'\ell\nu b\bar{b}b. \tag{7.1}$$

Backgrounds to this channel arise from top-quark pair production with additional jets. At leading order, this background consists of the irreducible $t\bar{t}b$ and the reducible $t\bar{t}j$ production, for which the generation aspects are discussed in Section 3.4.2. In the case of fourfold b-jet identification with charged Higgs-boson production in the gluon-initiated channel, the semi-leptonic final state takes the form

$$gg \to tH^{\pm}b \to t\bar{t}b\bar{b} \to W^{+}W^{-}b\bar{b}b\bar{b} \to q\bar{q}'\ell\nu b\bar{b}b\bar{b}. \tag{7.2}$$

For this channel the dominant backgrounds at leading order result from the irreducible $t\bar{t}b\bar{b}$ and the reducible $t\bar{t}jj$ production. Simulation aspects of these backgrounds at the parton level are as well discussed in Section 3.4.2.

In the analysis presented in this chapter, charged Higgs-boson detection is studied for both final states (7.1) and (7.2). In the case of three identified b jets, the parametrized CMS detector simulation is used, because of the need to simulate a large amount of background events. In the case of fourfold b-jet identification the detailed CMS detector simulation is employed. Both cases are only considered for final states with an isolated muon, since this is the lepton with the cleanest and most efficient reconstruction. The similar decay channels with an electron instead of a muon can be expected to give the same results, but with somewhat smaller statistics due to the lower trigger acceptance and the electron reconstruction efficiency.

Throughout the chapter the results are given in the MSSM, but without considering supersymmetric decay modes. Also the production of $H^{\pm}$ bosons through cascade decays of supersymmetric particles is not taken into account. Hence, using the relation (1.30) between the A, $H^{\pm}$ and W-boson masses, this study can be directly interpreted in a general 2HDM Type-II.

The potential of the decay channel $H^{\pm} \to tb$ for large Higgs-boson masses at the LHC has been considered at parton level in several phenomenological studies [175–179]. These studies emulated detector resolutions by smearing and parametrizing the partonic final state. They have shown the possibility of detecting the charged Higgs boson in certain regions of the MSSM ($\tan\beta, m_A$) parameter space during the low-luminosity run of the LHC, with either three or four b-tagged jets. The best performance is obtained at high $\tan\beta$, where the production cross section is enhanced, and at moderate values of $m_A \sim 300\,\mathrm{GeV}/c^2$, for which the cross section is high, and the b quark from the $H^{\pm} \to tb$ decay is sufficiently energetic to be efficiently reconstructed and identi-

fied. Crucial to these studies is a good b-identification capability to suppress the very large, and kinematically very similar $t\bar{t}$ + jets background.

7.2 Charged Higgs-Boson Identification with Three b-Tagged Jets

In this section the possibility to identify charged Higgs bosons with triple b tagging is analysed in the semi-leptonic final state with an isolated muon. The reconstruction of events is built on the techniques from Chapters 4 and 5. For events passing a basic event selection, the best jet association is determined using a likelihood-ratio approach. The reconstruction of the charged Higgs-boson mass is attempted, but found to be hampered by a large combinatorial background. Further likelihood-ratio based background suppression is then presented, which allows to optimize the statistical significance of the signal with respect to the signal-to-noise ratio. Possible systematic uncertainties are finally discussed, and the effect on the visibility of the signal is demonstrated with contour plots in the MSSM $(m_{\mathrm{A}}, \tan\beta)$ plane. The non-visibility in the MSSM parameter space is translated to a lowest cross section needed to observe MSSM-like charged Higgs bosons with the presented three b-jet analysis.

7.2.1 Event reconstruction and selection

The offline reconstruction of isolated muons is described in Section 5.3.2. The muon considered to originate from the $W^{\pm}$ decays is chosen as the candidate with the highest $\mathcal{L}_{\mathrm{muon}}$, defined in (5.13), satisfying $\mathcal{L}_{\mathrm{muon}} > 0.01$. Jets are clustered using the iterative-cone algorithm with cone size $\Delta R = 0.5$ and a jet seed cut $E_{\mathrm{T}} > 2\,\mathrm{GeV}$, as described in Section 5.1.1, using the underlying-event input thresholds for the calorimeter towers, as defined in Section 4.1.4. The jet clustering is performed excluding the calorimeter deposits around the selected muon. Monte-Carlo jet energy-scale corrections, introduced in Section 5.1.3, are applied to the reconstructed jets. Jets emerging from pile-up collisions are vetoed using the vertex-association criterion from Section 5.1.5. The secondary-vertex based combined b-tagging algorithm, described in Section 5.2.2, is used on all jets for b-jet identification. The missing transverse energy, finally, is calculated as the vectorial sum of the transverse calorimeters deposits, correcting for the muons' transverse energy, as explained in Section 5.4.

The event selection starts with the application of the single-muon Level-1 trigger threshold and the corresponding High-Level Trigger requirement, as discussed in Sections 4.3.1 and 4.3.2. Events passing these trigger criteria are required to have

- at least one muon with $p_{\mathrm{T}} > 20\,\mathrm{GeV}/c$, $|\eta| < 2.5$ and $\mathcal{L}_{\mathrm{muon}} > 0.01$;
- at least five jets with $E_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|\eta| < 2.5$;
- at least three b-tagged jets among the five jets with largest E_{T}, applying the secondary-vertex based algorithm with discriminator $\zeta_{\mathrm{b}} > 1.0$.

The relative efficiencies of these cuts and the remaining number of events after the complete event selection are summarized in Table 7.1, for the considered signal and background samples at $\tan\beta = 30$, and for an integrated luminosity of $30\,\mathrm{fb}^{-1}$.

$\tan\beta = 30$ $L = 30\,\mathrm{fb}^{-1}$ Details on the simulation in Section 4.2.4	$t\bar{t}b/t\bar{t}j$	$gb \to tH^{\pm}$ $m_{H^{\pm}} = 263\,\mathrm{GeV}/c^2$	$gb \to tH^{\pm}$ $m_{H^{\pm}} = 311\,\mathrm{GeV}/c^2$	$gb \to tH^{\pm}$ $m_{H^{\pm}} = 359\,\mathrm{GeV}/c^2$	$gb \to tH^{\pm}$ $m_{H^{\pm}} = 408\,\mathrm{GeV}/c^2$	$gb \to tH^{\pm}$ $m_{H^{\pm}} = 457\,\mathrm{GeV}/c^2$	$gb \to tH^{\pm}$ $m_{H^{\pm}} = 506\,\mathrm{GeV}/c^2$
cross section (pb) $\times$ BR($H^{\pm} \to tb$)	678	0.850	0.570	0.377	0.251	0.169	0.116
# ev. before cuts	20.3M	25 489	17 088	11 319	7 529	5 063	3 472
single-μ L1&HLT	17%	16%	16%	16%	16%	16%	16%
1 muon	95%	95%	95%	95%	96%	96%	96%
5 jets	18%	35%	42%	44%	46%	49%	51%
3 b-tagged jets	6%	27%	29%	30%	32%	31%	29%
# ev. remaining	32 880	364	314	230	171	116	80

Table 7.1: Selection yield for $\tan\beta = 30$ and an integrated luminosity of $30\,\mathrm{fb}^{-1}$.

The applied event-selection criteria are based on the expected final state of the semi-leptonic channel with an isolated muon, three b jets and two non-b jets. For the muon trigger and selection criteria, the yield corresponds to the expectations from the W-boson branching fractions and the acceptance cuts. The requirement to find five reconstructed jets within the tracker acceptance is motivated by the need to apply the b-identification algorithm on all the jets. The lower efficiency of the jet finding for the background is due to the softer spectrum and the more forward production of the extra jet, as shown in Section 3.4.2. The increase of the efficiency of the five-jet selection as a function of the Higgs-boson mass is a consequence of the resulting increase of the overall event kinematics.

The most important selection cut is related to the b-tagging. Both signal and background produce at least two b quarks from top-quark decays, therefore at least three b jets need to be identified to significantly suppress the background with respect to the signal. The obtained improvement of the signal-to-noise ratio is a factor of about five, after the b-tag requirement. Further improvements with more stringent b-identification cuts are not applied in this stage of the analysis, but are part of a more global event identification further on. The signal-to-noise ratio, although increased by about an order of magnitude, is only about 1% or less at $\tan\beta = 30$ after the event selection.

On the selected events, a kinematic fit which imposes both W-boson and both top-quark mass constraints is applied to all possible jet combinations, using the technique described in Section 5.5. The fitting procedure returns a χ^2 probability if the fit converged, along with the fitted kinematical variables. This output is used in Section 7.2.2

to choose the best jet association. Events for which the fit does not converge for any jet association are discarded.

7.2.2 Selection of the optimal jet association

Five jets are present in the final state (7.1), of which three are b jets. With these jets, 60 possible associations can be made to the final-state quarks, taking into account the interchangeability of the jets from the hadronic W-boson decay. The b-tagging information can be used to reduce this number of combinations to six. To tackle the combinatorial ambiguities, several observables are identified that differentiate between the correct and wrong jet associations:

- observables associated with the output of the kinematic fit: the χ^2 probability of the fit (Figure 7.1) and the difference in mass $m(\mathrm{t_{hadr,fit}}) - m(\mathrm{t_{hadr,rec}})$ between the hadronically-decaying top quark after and before the fit (Figure 7.2). The behaviour of the distributions can be understood similarly as for the kinematic-fit related variables in Section 6.1.3;

- the transverse momentum and pseudorapidity of the b jet from the $H^{\pm}$ decay: $p_{\mathrm{T}}(\mathrm{b_{H^{\pm}}})$ (Figure 7.3) and $|\eta(\mathrm{b_{H^{\pm}}})|$ (Figure 7.4). This b jet becomes more energetic with increasing $H^{\pm}$ mass, as shown in Figure 3.18, causing at the same time a rise in the separation power of the p_{T} observable;

- the combined b-identification probability

$$P_{3\mathrm{b}} = \frac{P_{\mathrm{b}}(\mathrm{b}_1) + P_{\mathrm{b}}(\mathrm{b}_2) + P_{\mathrm{b}}(\mathrm{b}_3)}{P_{\mathrm{b}}(\mathrm{b}_1) + P_{\mathrm{b}}(\mathrm{b}_2) + P_{\mathrm{b}}(\mathrm{b}_3) + P_{\mathrm{b}}(\mathrm{j}_1) + P_{\mathrm{b}}(\mathrm{j}_2)},$$

  obtained from the b-jet probabilities of the three b jets and of the two jets from the W decay (Figure 7.5). The b-jet probability $P_{\mathrm{b}}(\zeta_{\mathrm{b}})$ associated with the b-jet discriminator ζ_{b} is introduced in Section 5.2.2. The shape of the distribution is a consequence of the large b-probability separation between b and non-b quarks, resulting in peaks for one, two or three correctly-associated b jets.

The jet associations used to obtain the distributions for these observables include also the ones disfavoured by the b-tagging, except for the associations for which the kinematic fit fails. For each of the observables, a corresponding likelihood-ratio function $\mathcal{L}_i(x_i)$ is obtained by calculating bin-by-bin the ratio $S/(S+B)$, where S and B denote respectively the number of correct and wrong jet associations, and by fitting the resulting distribution. The fit functions, being sigmoid, Gaussian and polynomial functions, are chosen to depend on only a few parameters, to keep the fit quality acceptable. The individual $\mathcal{L}_i(x_i)$ distributions are shown in Figures 7.1 through 7.5 for $m_{\mathrm{H^{\pm}}} = 311\,\mathrm{GeV}/c^2$.

For each jet association a combined likelihood-ratio value is calculated as

$$\mathcal{L}_{\mathrm{sol}}(\vec{x}) = \prod_i \mathcal{L}_i(x_i), \tag{7.3}$$

ignoring possible correlations between the observables. The solution in an event with the highest value for $\mathcal{L}_{\mathrm{sol}}$ is chosen as the best possible jet association. This choice results in all five jets correctly associated in 13% of the events for $m_{\mathrm{H^{\pm}}} = 311\,\mathrm{GeV}/c^2$.

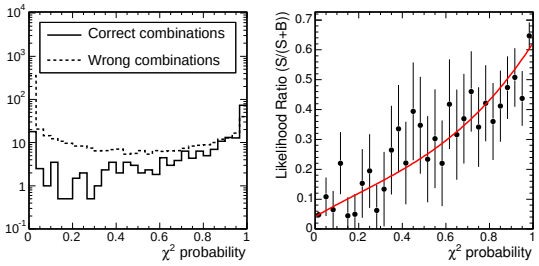


Figure 7.1: χ^2 probability for $m_{H^{\pm}} = 311\,\mathrm{GeV}/c^2$.

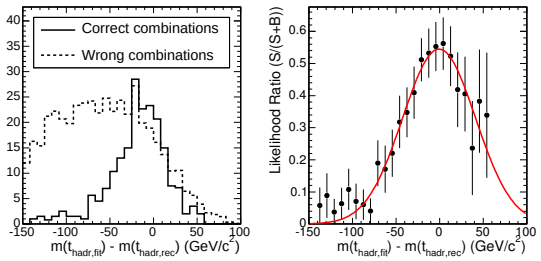


Figure 7.2: $m(\mathrm{t_{hadr,fit}}) - m(\mathrm{t_{hadr,rec}})$ for $m_{H^{\pm}} = 311\,\mathrm{GeV}/c^2$.

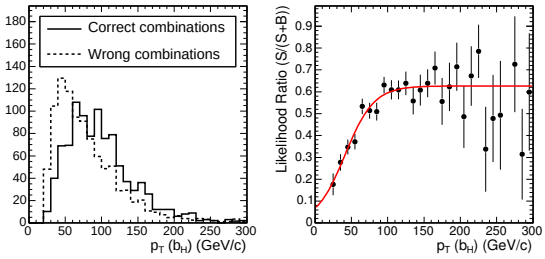


Figure 7.3: $p_{\mathrm{T}}(\mathrm{b_{H^{\pm}}})$ for $m_{H^{\pm}} = 311\,\mathrm{GeV}/c^2$.

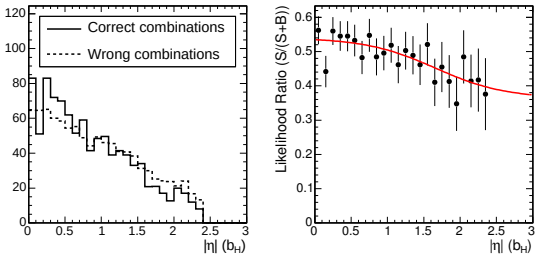


Figure 7.4: $|\eta(\mathrm{b_{H^{\pm}}})|$ for $m_{H^{\pm}} = 311\,\mathrm{GeV}/c^2$.

7.2.3 Mass reconstruction

Since the longitudinal momentum of the neutrino is determined from the kinematic fit, it is possible to reconstruct the charged Higgs-boson mass. The use of the fit additionally improves the resolution of the mass estimator compared to conventional techniques, as is shown in the framework of a top-quark mass measurement in semi-leptonic $\mathrm{t\bar{t}}$ events in Section 5.5. An ambiguity in the reconstruction remains, however, as it is a priori not possible to know which top quark the additional b jet should be combined with. For large charged Higgs-boson masses, the top quark for the combination can be chosen on the basis of its momentum, which is in this case on average larger when originating from the charged Higgs-boson decay. In the region of interest, $m_{H^{\pm}} \sim 300\,\mathrm{GeV}/c^2$, the kinematical properties of both top-quark candidates are however very similar. Therefore both combinations are taken into account in the following.

Figures 7.6 and 7.7 present the reconstructed Higgs-boson mass in signal events for the correct jet associations and for the chosen jet associations, respectively, with $m_{H^{\pm}} = 311\,\mathrm{GeV}/c^2$. The events considered are required to have the Higgs boson decaying into the hadronically-decaying top quark. Figures 7.8 and 7.9 present the same distributions for the charged Higgs bosons in signal events with leptonic decays of the top quark. The observed resolution of a few tens of GeV on the correctly-paired charged Higgs-boson reconstructions corresponds to what can be expected from the combination of the jet and missing-energy resolutions in these complex objects. The resolution for the leptonic decay is somewhat smaller since the lepton's momentum is accurately measured and the large resolution on the missing energy is reduced in the

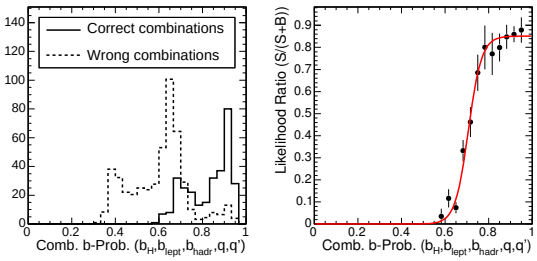


Figure 7.5: Combined b probability P_{3b} for $m_{H^{\pm}} = 311\,\mathrm{GeV}/c^2$.

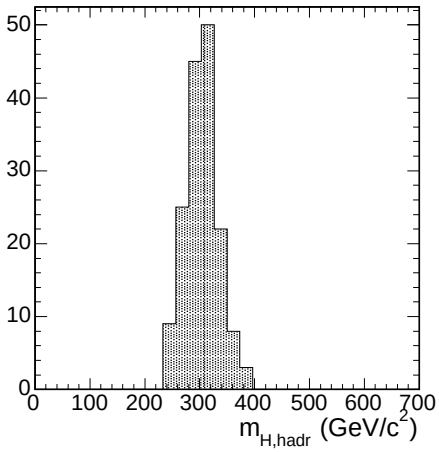


Figure 7.6: Reconstructed $H^{\pm}$ mass with hadronically-decaying top quark for correct jet associations ($m_{H^{\pm}} = 311\,\mathrm{GeV}/c^2$).

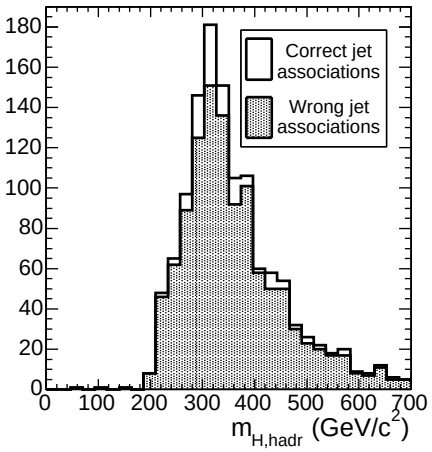


Figure 7.7: Reconstructed $H^{\pm}$ mass with hadronically-decaying top quark for the chosen jet association ($m_{H^{\pm}} = 311\,\mathrm{GeV}/c^2$).

kinematic fit.

Due to the many possible jet associations and the ambiguity in the choice of the top quark candidate originating from the charged Higgs-boson decay, the combinatorial background is very large. In Figures 7.7 and 7.9 the events for which jets are correctly associated to the decay products of the Higgs boson are shown on top of the combinatorial background. Because of the likelihood-ratio based selection, which optimizes the way the reconstructed information in the events is used to choose the best jet association, it can be safely concluded that the reconstructed events do not contain enough discriminative information to efficiently select the correct association. Because of the large combinatorial background no mass reconstruction is attempted, and the analysis is further on considered as a counting experiment of the events remaining after the various selection criteria. This way all events are retained and wrongly-chosen jet associations in signal events also contribute to the visibility of the signal.

7.2.4 Background suppression

To suppress further the large $t\bar{t}b/t\bar{t}j$ background that remains after the initial event selection, observables are identified that have different distributions for signal and back-

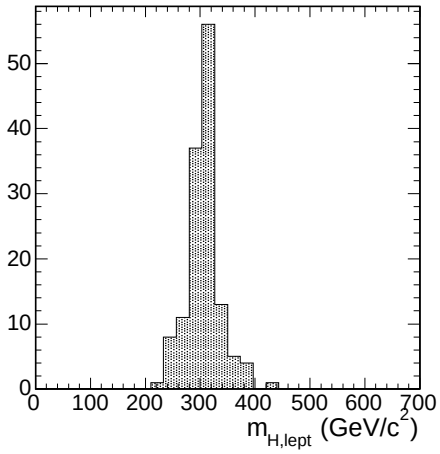


Figure 7.8: Reconstructed $H^{\pm}$ mass with leptonically-decaying top quark for correct jet associations ($m_{H^{\pm}} = 311\,\mathrm{GeV}/c^2$).

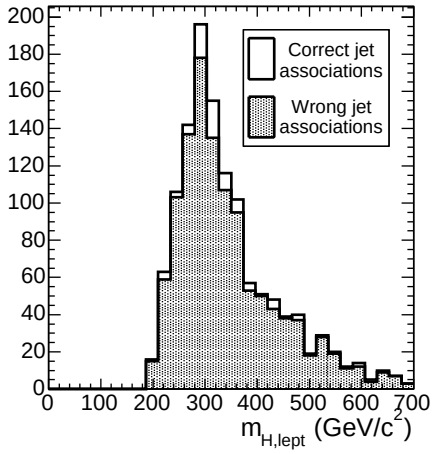


Figure 7.9: Reconstructed $H^{\pm}$ mass with leptonically-decaying top quark for the chosen jet association ($m_{H^{\pm}} = 311\,\mathrm{GeV}/c^2$).

ground events. The following quantities are considered for the chosen jet association:

- the χ^2 probability of the kinematic fit (Figure 7.10). The χ^2 probability of the fit is being used already to select the best jet association. Some discrimination power between signal and background remains, however, especially for higher Higgs-boson masses, where the differences in kinematics in the $t\bar{t}$ part of the events increase;

- the transverse momentum $p_T(q')$ of the softest jet q' from the W-boson decay (Figure 7.11). The q' jet, being the softest jet from the hadronic W-boson decay, is the jet with the highest probability of being mis-associated with extra jets in the event. Because of the rather soft and forward spectrum of extra jets in the $t\bar{t}b/t\bar{t}j$ background events, shown in Figures 3.11 and 3.12, the spectrum of $p_T(q')$ is also softer for background than for signal events;

- the b-jet discriminator ζ_b for the b-jet originating directly from the $H^{\pm}$ decay (Figure 7.12);

- the ratio of the E_T of the sixth over the fifth jet, provided a sixth jet is found, otherwise zero (Figure 7.13). This observable is similar to the corresponding variables in Chapter 6, but has another interpretation. The distribution reveals more extra hard jets in the signal, which is explained by the generally harder and more central kinematics compared to the background.

Using these observables x_i, likelihood-ratio functions $\mathcal{L}_i(x_i)$ are constructed in a similar way as for the jet association, and combined in a single likelihood ratio

$$\mathcal{L}_{SB}(\vec{x}) = \prod_i \mathcal{L}_i(x_i), \tag{7.4}$$

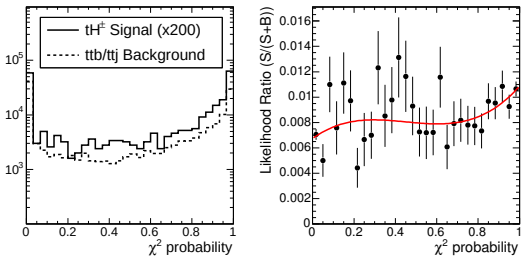


Figure 7.10: χ^2 probability for $m_{H^{\pm}} = 311\,\mathrm{GeV}/c^2$.

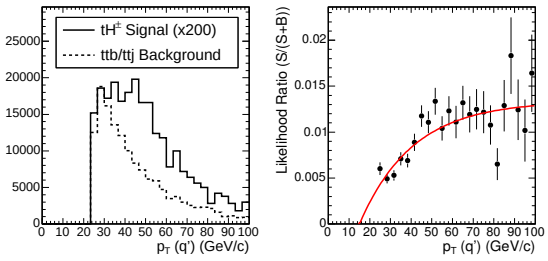


Figure 7.11: $p_{\mathrm{T}}(\mathrm{q}')$ for $m_{H^{\pm}} = 311\,\mathrm{GeV}/c^2$.

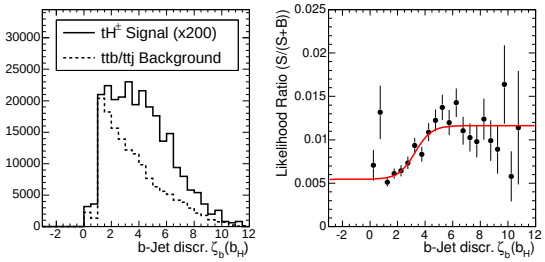


Figure 7.12: b-jet discriminator ζ_{b} of $\mathrm{b}_{H^{\pm}}$ for $m_{H^{\pm}} = 311\,\mathrm{GeV}/c^2$.

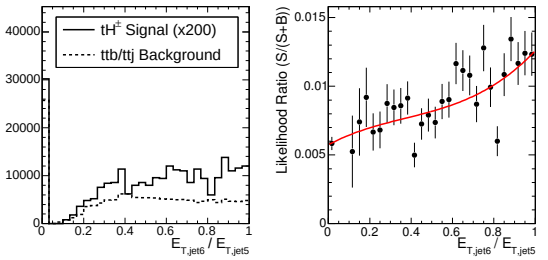


Figure 7.13: $E_{\mathrm{T},6}/E_{\mathrm{T},5}$ for $m_{H^{\pm}} = 311\,\mathrm{GeV}/c^2$.

neglecting correlations between the observables x_i. The distribution of $\mathcal{L}_{\mathrm{SB}}$ is shown in Figure 7.14 for both the signal with $m_{H^{\pm}} = 311\,\mathrm{GeV}/c^2$ and the background. To make the small signal visible, it is multiplied with a factor 200 in the figure. Although the description of the individual functions $\mathcal{L}_i(x_i)$ and the combined likelihood-ratio distributions is limited by the available signal simulation statistics, it is supposed that the main differences between the signal and the background are accounted for.

7.2.5 Statistical significance and discovery potential

The statistical significance of the observation of the signal is calculated using the S_{cP} tool [180]. This procedure translates the probability that a number of background events with Poissonian statistics mimics an additional number of signal events, into the number of Gaussian standard deviations that corresponds to this probability.

By construction, a cut on the combined likelihood ratio $\mathcal{L}_{\mathrm{SB}}$ increases the signal-to-background ratio. The significance of the signal observation is calculated as a function of this cut for an integrated luminosity of $30\,\mathrm{fb}^{-1}$, and its maximum is determined, corresponding to an optimized cut on $\mathcal{L}_{\mathrm{SB}}$ at which the analysis is to be performed. In Figure 7.15 the statistical significance as a function of the cut on the combined likelihood ratio is shown for $m_{H^{\pm}} = 311\,\mathrm{GeV}/c^2$.

With the knowledge of the maximal statistical significance for a certain integrated luminosity, the signal cross section for a 5σ discovery is derived for a given charged Higgs-boson mass, or equivalent m_{A}. Using the cross-section dependence on $\tan\beta$, this 5σ discovery limit for the cross section can be translated into a lower limit on $\tan\beta$.

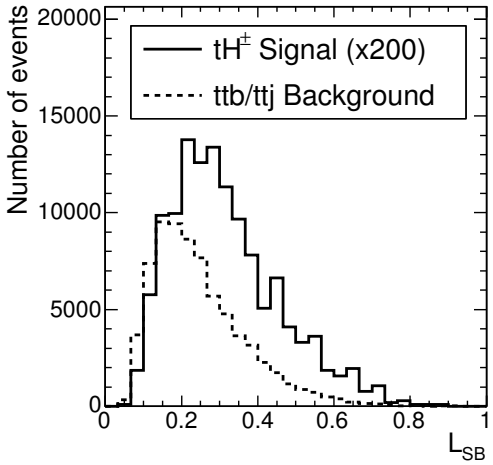


Figure 7.14: Distribution of the combined likelihood ratio $\mathcal{L}_{\mathrm{SB}}$ used to distinguish between signal and background, for $m_{\mathrm{H}^{\pm}} = 311\,\mathrm{GeV}/c^2$ and $30\,\mathrm{fb}^{-1}$.

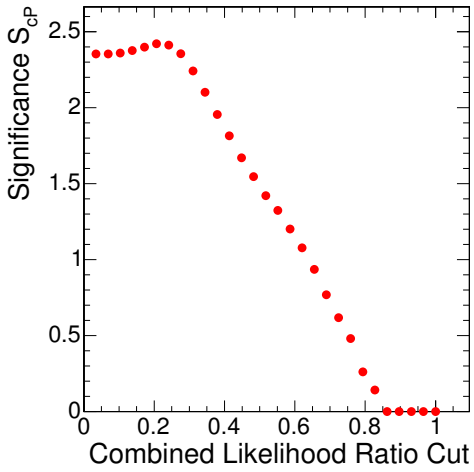


Figure 7.15: Statistical significance as a function of the cut on $\mathcal{L}_{\mathrm{SB}}$ for $m_{\mathrm{H}^{\pm}} = 311\,\mathrm{GeV}/c^2$ and $30\,\mathrm{fb}^{-1}$.

Performing the complete analysis chain, including the described maximization of the significance, at different values of m_{A}, an optimized discovery contour is obtained in the MSSM $(m_{\mathrm{A}}, \tan\beta)$ plane, as shown in Figure 7.16 for an integrated luminosity of $30\,\mathrm{fb}^{-1}$ and $60\,\mathrm{fb}^{-1}$. On the same figure the exclusion limit at 95% confidence level for $30\,\mathrm{fb}^{-1}$ is drawn. Systematic uncertainties are not yet included at this point. For comparison, the previous result [181], obtained in `CMSJET` [182] is also shown for $30\,\mathrm{fb}^{-1}$.

7.2.6 Systematic uncertainties

The significance calculated above corresponds to the ideal case of a perfect knowledge of the background cross section. The background is large, however, and hence the effect of systematic uncertainties on the knowledge of the background needs to be estimated. In order to measure the background cross section and uncertainty from the data, a signal-free control sample should be obtained. For this analysis, however, the signal and background are kinematically very similar, and the combinatorial background is omnipresent in the studied distributions.

A possible way to estimate the background level from data is to impose the same selection cuts on the data as in Section 7.2.1, but to require minimally only two b-tagged jets among the five selected jets. Even with two b-tagged jets, the main background remains only $\mathrm{t\bar{t}}$ production. After such a selection with only two b-tagged jets, the signal-to-background ratio is further reduced. It is then possible to calculate the expected number of background events plus its uncertainty, when tagging a third b jet.

For this method, the b-tagging efficiency and purity need to be measured from data. As shown in Chapter 6, the use of enriched b-jet samples from $\mathrm{t\bar{t}}$ events allows the b-tagging efficiency in CMS to be known to about 5% relative uncertainty with $10\,\mathrm{fb}^{-1}$ of integrated luminosity. For mistag probabilities, such estimations are only currently being explored. In the CDF [183] and DØ [184] experiments an uncertainty

on the mistag probability of 10% has been found for a secondary-vertex technique. Supposing an optimistic case where CMS obtains a 5% relative uncertainty on the mistag probability, then the uncertainty on the expected $t\bar{t}b/t\bar{t}j$ background level will at best be of the order of 5%. Possibly large theoretical uncertainties could also come into play using this method, like the ratio of $t\bar{t} + b$ events with real extra b jets and $t\bar{t} + j$ events with only jets from light quarks and gluons accompanying the top quarks.

The systematic effects on the discovery potential are evaluated by adding the systematic uncertainty in quadrature to the statistical contribution. The value of S/B has to be sufficiently large for the systematic contribution of the background uncertainty to be kept under control. Figure 7.17 shows the result of the significance calculation as a function of the $\mathcal{L}_{SB}$ threshold for a systematic uncertainty on the background level of only 1%. Comparing this with the case of perfect background knowledge in Figure 7.15, a large drop in significance for a low cut on $\mathcal{L}_{SB}$ is observed. This is expected from the large number of background events in this region, which makes the systematic uncertainty grow quadratically above the statistical uncertainty.

Depending on the expected systematic uncertainty on the background level, the maximal significance and its corresponding optimized cut on $\mathcal{L}_{SB}$ is determined. In Figure 7.18 the discovery contours are plotted, for perfect knowledge of the $t\bar{t}b/t\bar{t}j$ cross section, for a 1% uncertainty and for a 3% uncertainty. From the above estimate of the systematic uncertainty on the number of background events, it is concluded that with this analysis no sensitivity in this channel with triple b-tagging is obtained in the relevant MSSM parameter space during the low-luminosity phase of LHC.

Assuming the same signal kinematics as for the MSSM charged Higgs-boson production, the required $\sigma(pp \to tH^{\pm}X) \times BR(H^{\pm} \to tb)$ for a 5σ discovery by tagging three b jets is extracted. It is shown in Figure 7.19 as a function of m_A.

7.3 Charged Higgs-Boson Identification with Four b-Tagged Jets

This section presents the identification of charged Higgs-boson production with four identified b jets in the final state. This analysis is performed in the semi-leptonic final state (7.2) with an isolated muon. For events passing the event reconstruction and selection criteria, the best jet association is determined using a likelihood ratio approach. Further background suppression is obtained with a similar technique, such that the statistical significance of the signal is optimized with respect to the signal-to-background ratio. Systematic uncertainties are discussed, and their effect on the visibility of the signal is demonstrated with contour plots in the MSSM $(m_A, \tan\beta)$ parameter plane. The non-visibility in the MSSM parameter space is finally translated to the minimal cross section needed to observe MSSM-like charged Higgs bosons with the presented four b-jet analysis.

7.3.1 Event reconstruction and selection

The reconstruction of physics objects in the event is performed in the same way as for the three b-jet channel in Section 7.2.1. As a result isolated muon candidates, a list of

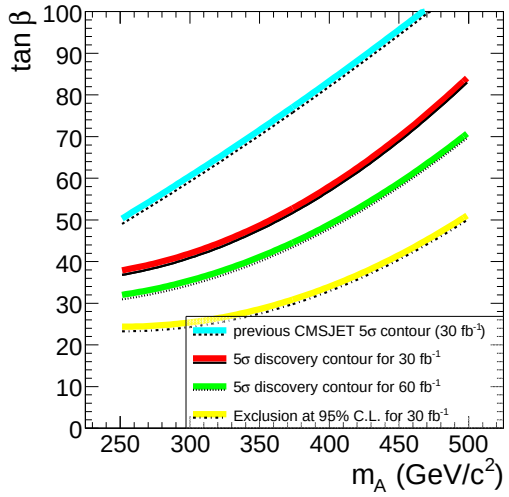


Figure 7.16: Discovery contours for the charged Higgs boson in the $H^{\pm} \to tb$ decay, applying 3 b-tags, for $30\,\mathrm{fb}^{-1}$ and $60\,\mathrm{fb}^{-1}$. Also shown are the exclusion limit at 95% C.L. and the previous CMSJET result [181] for $30\,\mathrm{fb}^{-1}$. The discovery region lies above the curves. Systematic uncertainties are not included.

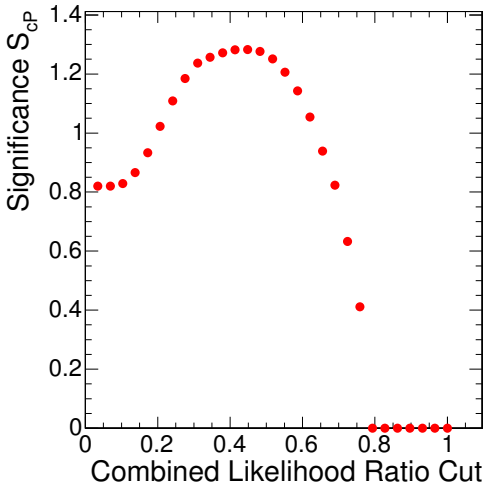


Figure 7.17: Significance as a function of the cut on $\mathcal{L}_{\mathrm{SB}}$ for $m_{\mathrm{H}^{\pm}} = 311\,\mathrm{GeV}/c^2$ and $30\,\mathrm{fb}^{-1}$, taking a systematic uncertainty of 1% on the background cross section into account.

calibrated jets, and the estimated missing transverse energy are obtained. In addition, the b-tagging discriminator ζ_{b} is calculated for each of the jets.

Only events are considered which pass the single-muon Level-1 and HLT trigger thresholds. These events are required to have

- at least one muon with $p_{\mathrm{T}} > 20\,\mathrm{GeV}/c$, $|\eta| < 2.5$ and $\mathcal{L}_{\mathrm{muon}} > 0.01$;
- at least six jets with $E_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|\eta| < 2.5$;
- at least four b-tagged jets out of the six jets with the largest E_{T}. The ζ_{b} discriminator threshold to tag a jet as b with the secondary-vertex based algorithm is fixed loosely at 0.5.

The relative efficiencies of these cuts and the remaining number of events after the complete event selection are summarized in Table 7.2 for the considered signal and background samples at $\tan\beta = 30$, and for an integrated luminosity of $30\,\mathrm{fb}^{-1}$.

The event selection criteria follow from the expected final state of the semi-leptonic decay channel with an isolated muon, four b jets and two non-b jets. For the muon trigger and selection criteria, the yield corresponds to the expectations from the W-boson branching fractions and the acceptance and reconstruction efficiencies. The difference with the selection in the three b-jet case is to be attributed to the HLT implementation in the parametrized CMS detector-simulation program FAMOS. The somewhat higher efficiencies on the background samples stem from the extra jets, which enhance the overall muon content. The six reconstructed jets are again demanded

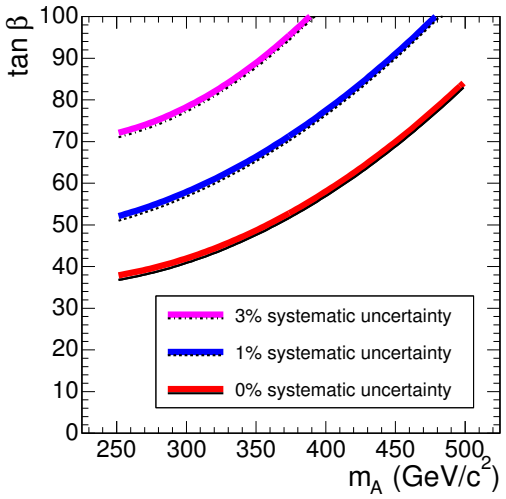


Figure 7.18: Discovery contour for the charged Higgs boson in the $H^{\pm} \to tb$ decay, applying 3 b-tags, for $30\,\text{fb}^{-1}$; systematic uncertainties on the background of 0%, 1% and 3% are taken into account. The discovery region lies above the curves.

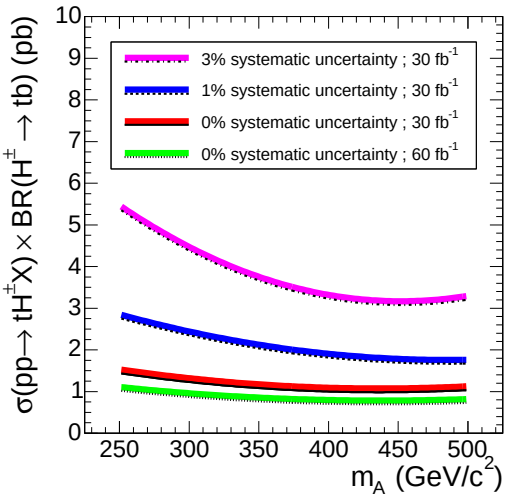


Figure 7.19: Minimally required $\sigma(pp \to tH^{\pm}X) \times BR(H^{\pm} \to tb)$ for discovery of a signal with the same kinematics as in the MSSM, tagging 3 b jets, as a function of m_A.

$\tan\beta = 30$ $L = 30\,\text{fb}^{-1}$ Details on the simulation in Section 4.2.4	$t\bar{t}b\bar{b}$	$t\bar{t}jj$	$gg \to tbH^{\pm}$ $m_{H^{\pm}} = 263\,\text{GeV}/c^2$	$gg \to tbH^{\pm}$ $m_{H^{\pm}} = 311\,\text{GeV}/c^2$	$gg \to tbH^{\pm}$ $m_{H^{\pm}} = 359\,\text{GeV}/c^2$	$gg \to tbH^{\pm}$ $m_{H^{\pm}} = 408\,\text{GeV}/c^2$	$gg \to tbH^{\pm}$ $m_{H^{\pm}} = 457\,\text{GeV}/c^2$	$gg \to tbH^{\pm}$ $m_{H^{\pm}} = 506\,\text{GeV}/c^2$
cross section (pb) $\times$ BR($H^{\pm} \to tb$)	2.386	235.8	0.850	0.570	0.377	0.251	0.169	0.116
# ev. before cuts	71 580	7.07M	25 489	17 088	11 319	7 529	5 063	3 472
single-μ L1&HLT	19%	19%	13%	13%	13%	13%	13%	13%
1 muon	96%	97%	96%	95%	97%	97%	97%	97%
6 jets	19%	23%	19%	23%	25%	26%	28%	31%
4 b-tagged jets	7%	0.55%	6%[1]	5%[1]	7%[1]	7%[1]	5%[1]	6%[1]
# ev. remaining	179	1 623	37	24	25	18	9	8

Table 7.2: Selection yield for $\tan\beta = 30$ and an integrated luminosity of $30\,\text{fb}^{-1}$.

within the tracker acceptance to be able to apply the b-identification algorithm on these jets. The increase of the efficiency of the six-jet selection as a function of the charged Higgs-boson mass is a consequence of the resulting increase of the overall hardness of the event kinematics.

[1]Large uncertainty due to low number of simulated and selected events.

As in the case of the three-b final state, the most important selection cut is related to the b identification. In this case both the third and fourth b tag suppress the dominant $t\bar{t}$ + jets background, and the signal loss due to these cuts is expected to be compensated by the simultaneous increase in the signal-to-noise ratio. The resulting efficiency of the fourfold b tagging at the working point defined by $\zeta_b > 0.5$ is found to be about 6% for the signal. The threshold on the b-tagging discriminator is relaxed compared to the case of triple tagging in order to select a sufficient amount of events for the next steps in the analysis. When compared to the efficiencies of triple b-tagging in Table 7.1, the difficulty of identification of the fourth b-jet in the signal stands out. This behaviour, however, can be expected from the forward and soft production of this spectator jet, illustrated in Figures 3.16 and 3.17 in Section 3.4.3.

Two opposite effects influence the b-tagging performance in the $t\bar{t}b\bar{b}$ background compared to the signal. The two extra b quarks in $t\bar{t}b\bar{b}$ events are produced more centrally, shown in the distributions in Sections 3.4.2 and 3.4.3, making them easier to tag as b compared to the spectator jet in the signal. At the same time these quarks are softer and less central than the b quark from the charged Higgs-boson decay, inducing the opposite effect. All in all the efficiency for quadruple b-tagging for the $t\bar{t}b\bar{b}$ background is about the same as for the signal. The prime reason for the requirement of all four b jets is the suppression of the large $t\bar{t}jj$ background. The signal-to-noise ratio is indeed increased by roughly an order of magnitude due to the b-tagging, reaching over 2% at low charged Higgs-boson masses. From naive expectations, a much higher reduction can be hoped for, however. The difference is explained by high mis-identification rates in the detailed analysis for the jets accompanying the $t\bar{t}$ pair in $t\bar{t}jj$ events. These high rates are due to the forward and soft nature of these extra jets, shown in Figures 3.13 and 3.14, which strongly increases the rate of tagged non-b jets, as discussed in Section 5.2.4. Another reason for the high mistag rate is the gluon content of these jets, over 90%, which gives rise to a large overall fraction of jets with genuine heavy flavour. With an average energy $E \approx 80\,\text{GeV}$ of the extra jets in $t\bar{t}jj$ events, the fractions of these jets containing a $c\bar{c}$ or $b\bar{b}$ pair from gluon splitting is expected to be about 9% and 4% respectively, as shown in Section 3.3.2.

Further improvements with more stringent b-identification cuts are possible but not applied. On the contrary, the b-tag working point in the event selection is chosen loose enough to retain at least some signal events. The b-tag information is used again when the analysis is optimized further on.

A kinematic fit which imposes both W-boson and both top-quark mass constraints is also applied to all possible jet combinations of the four b-jet final state, using the technique described in Section 5.5. The returned χ^2 probability and fitted kinematical variables are used in Section 7.3.2 to choose the best jet association. Events for which the fit does not converge for any jet association are discarded.

7.3.2 Selection of the optimal jet association

In the final state (7.2), having six jets in total of which four are identified as b jets, the jets can be associated to the final-state partons in 360 possible ways, taking into account the interchangeability of the jets from the hadronic W-boson decay. The b-tagging information can be used to reduce this number to 24. Several observables are

identified that discriminate between correct and wrong associations:

- observables associated with the output of the kinematic fit: the χ^2 probability of the kinematic fit (Figure 7.20), and the difference in mass $m(t_{hadr,fit}) - m(t_{hadr,rec})$ of the hadronically-decaying top quark after and before the fit (Figure 7.21). The shape of the distributions can again be explained as in Section 6.1.3;

- the transverse momentum and pseudorapidity of the b jet from the $H^{\pm}$ decay: $p_T(b_{H^{\pm}})$ (Figure 7.22) and $|\eta(b_{H^{\pm}})|$ (Figure 7.23). Also in this case the discrimination power increases with rising charged Higgs-boson mass;

- the b-jet discriminator ζ_b for the b jet originating directly from the $H^{\pm}$ decay (Figure 7.24), and for the jet associated with the spectator b quark (Figure 7.25). These variables are treated separately from the remainder of the b-tag information, because of the limitations in simulation statistics.

- the combined b-identification probability

$$P_{2b} = \frac{P_b(b_1) + P_b(b_2)}{P_b(b_1) + P_b(b_2) + P_b(j_1) + P_b(j_2)},$$

  obtained combining the b-jet probabilities P_b for all the jets coming from top-quark decays (Figure 7.26). The peak structure is again caused by the presence of zero, one and two correctly-identified b jets.

The jet associations used to obtain the distributions for these observables include also the ones disfavoured by the b-tagging, except for the associations for which the kinematic fit fails. Because of the small number of simulated events the distributions are made using events from the same samples with only two b-tags applied. A likelihood-ratio function $\mathcal{L}_i(x_i)$ is obtained for each of the observables x_i by calculating bin-by-bin the ratio $S/(S+B)$, where S and B denote respectively the number of correct and wrong jet associations, and fitting the resulting distribution. Simple sigmoid, Gaussian and polynomial functions are chosen as fit functions. The individual $\mathcal{L}_i(x_i)$ distributions are shown in Figures 7.20 through 7.26 for $m_{H^{\pm}} = 311\,\mathrm{GeV}/c^2$.

For each jet association the combined likelihood ratio $\mathcal{L}_{sol}$ is calculated using the relation (7.3). The jet pairing with the highest value for $\mathcal{L}_{sol}$ is chosen as the best possible jet association. The problem of the combinatorial background, additionally complicated by the ambiguity in the choice of the top quark in the charged Higgs-boson decay, is even more pronounced than in the case with three b-tagged jets. For the same reasons as given in Section 7.2.3 the analysis is continued as a counting experiment of the events remaining after the various selection criteria, and no mass reconstruction of the charged Higgs-boson candidates is attempted.

7.3.3 Background suppression

In the case of the final state with four b quarks, the background consists of the irreducible, but relatively small, $t\bar{t}b\bar{b}$ process, and the reducible, but very large, $t\bar{t}jj$ process. After the event selection, which already improves the signal-to-noise ratio by

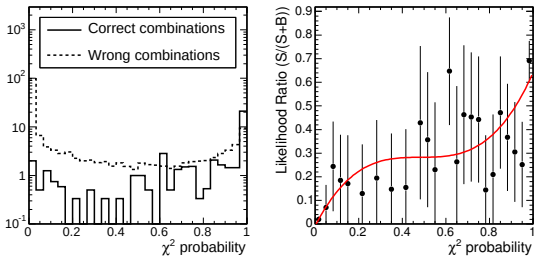


Figure 7.20: χ^2 probability for $m_{\mathrm{H}^{\pm}} = 311\,\mathrm{GeV}/c^2$.

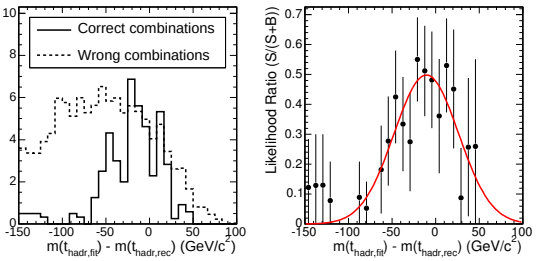


Figure 7.21: $m(\mathrm{t}_{\mathrm{hadr,fit}}) - m(\mathrm{t}_{\mathrm{hadr,rec}})$ for $m_{\mathrm{H}^{\pm}} = 311\,\mathrm{GeV}/c^2$.

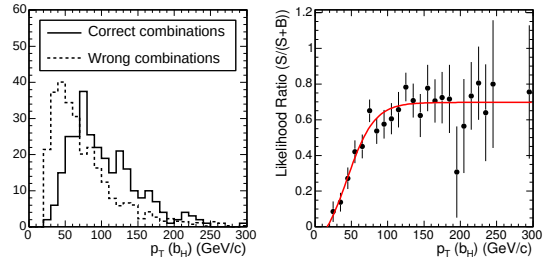


Figure 7.22: $p_{\mathrm{T}}(\mathrm{b}_{\mathrm{H}^{\pm}})$ for $m_{\mathrm{H}^{\pm}} = 311\,\mathrm{GeV}/c^2$.

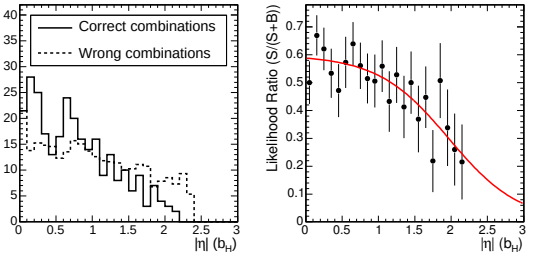


Figure 7.23: $|\eta(\mathrm{b}_{\mathrm{H}^{\pm}})|$ for $m_{\mathrm{H}^{\pm}} = 311\,\mathrm{GeV}/c^2$.

an order of magnitude, the dominant background is still formed by the $\mathrm{t\bar{t}jj}$ component. The optimization of the background suppression therefore mostly still relies on the b-tagging of the jets accompanying the $\mathrm{t\bar{t}}$ pair. Other observables, like the momentum and pseudorapidity of the top-quark candidates and of the extra jets, present some additional differences between signal and background, especially for large charged Higgs-boson masses. Due to the low simulation statistics, however, they could not reliably be shown to have a significant impact on background rejection, and are hence not used in the following.

In light of the limited statistics, an effective variable to discriminate between the signal and background events is found to be the sum of the b-tag discriminators $\zeta_{\mathrm{b}}(\mathrm{b}_{\mathrm{H}^{\pm}})$ and $\zeta_{\mathrm{b}}(\mathrm{b}_{\mathrm{spect}})$ of the b jets associated to the Higgs-boson decay and to the spectator jet respectively. The likelihood variable $\mathcal{L}_{\mathrm{SB}}$, defined as $(2 + \zeta_{\mathrm{b}}(\mathrm{b}_{\mathrm{H}^{\pm}}) + \zeta_{\mathrm{b}}(\mathrm{b}_{\mathrm{spect}}))/26$, such that $0 < \mathcal{L}_{\mathrm{SB}} < 1$, is shown in Figure 7.27 for $m_{\mathrm{H}^{\pm}} = 311\,\mathrm{GeV}/c^2$. To make the small signal stand out, it is multiplied by a factor 100.

7.3.4 Statistical significance and discovery potential

By cutting on the observable $\mathcal{L}_{\mathrm{SB}}$ it is possible to increase the signal-to-background ratio. Calculating the significance of the signal as a function of this cut, the point of maximal significance is determined for a given integrated luminosity, taken to be $30\,\mathrm{fb}^{-1}$. To accommodate the small signal statistics and large scale factor (~ 8) needed for the background simulation, the number of signal events and the corresponding number of background events after a cut on $\mathcal{L}_{\mathrm{SB}}$ is parametrized, as shown in Figure 7.28.

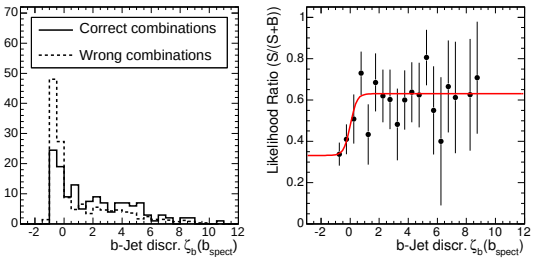


Figure 7.24: b-jet discriminator ζ_b of b_{spect} for $m_{H^{\pm}} = 311\,\mathrm{GeV}/c^2$.

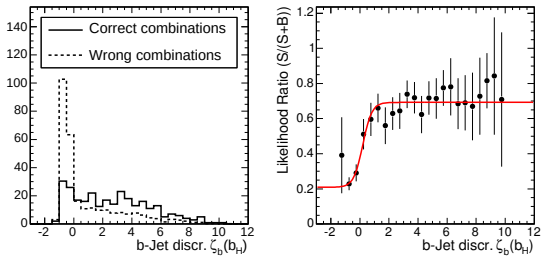


Figure 7.25: b-jet discriminator ζ_b of $b_{H^{\pm}}$ for $m_{H^{\pm}} = 311\,\mathrm{GeV}/c^2$.

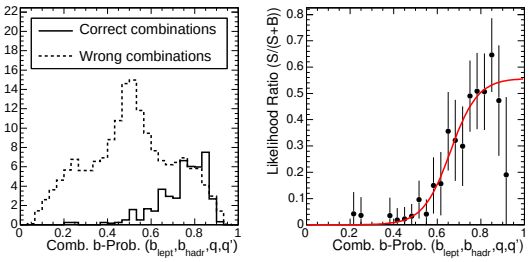


Figure 7.26: Combined b probability P_{2b} for $m_{H^{\pm}} = 311\,\mathrm{GeV}/c^2$.

This parametrized dependency is used when calculating the significance as a function of $\mathcal{L}_{SB}$, rather than the direct dependence on $\mathcal{L}_{SB}$ of the number of signal and background events. This causes the significance calculation to be smoothed, avoiding a large additional simulation of signal and background events. In Figure 7.29 the resulting statistical significance as a function of the cut on $\mathcal{L}_{SB}$ is shown for $m_{H^{\pm}} = 311\,\mathrm{GeV}/c^2$.

With the knowledge of the maximal statistical significance for a certain integrated luminosity, the signal cross section for a 5σ discovery is derived for a given charged Higgs-boson mass or equivalent m_A. As for the triple b-tag case, this corresponds to a minimal value of $\tan\beta$. Performing the complete analysis chain, including the described maximization of the significance and the parametrization of the number of signal versus the number of background events, at different values of m_A, an optimized discovery contour is obtained in the MSSM $(m_A, \tan\beta)$ plane, as shown in Figure 7.30 for an integrated luminosity of $30\,\mathrm{fb}^{-1}$ and $60\,\mathrm{fb}^{-1}$. On the same figure the exclusion limit at 95% confidence level for $30\,\mathrm{fb}^{-1}$ is drawn. Systematic uncertainties are not included at this point.

7.3.5 Systematic uncertainties

The significance calculated above corresponds to the ideal case of a perfect knowledge of the background cross section. In order to measure the background cross-section level and uncertainty, a similar technique can be used as in the case of triple b-tagging. Also in this case an optimistic uncertainty of 5% on both b-jet identification and mistag probability measurements is assumed, neglecting possible theoretical uncertainties.

As in the three b-jet case, the significance calculation is repeated for a hypothetical

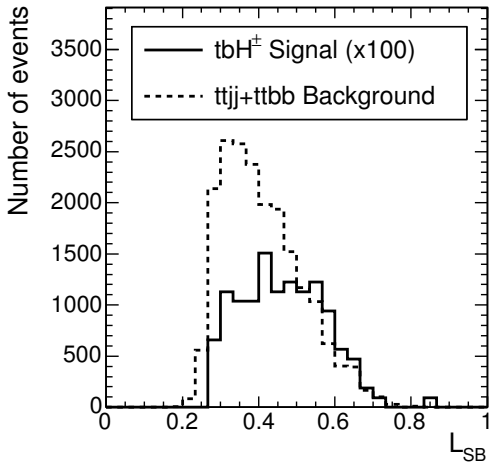


Figure 7.27: Distribution of the variable $\mathcal{L}_{\mathrm{SB}}$ used to distinguish between signal and background, for $m_{\mathrm{H}^{\pm}} = 311\,\mathrm{GeV}/c^2$ and $30\,\mathrm{fb}^{-1}$.

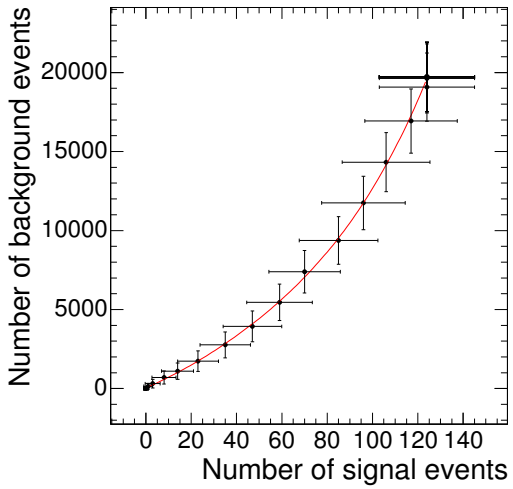


Figure 7.28: Number of background events as a function of the number of signal events, for various cuts on $\mathcal{L}_{\mathrm{SB}}$, for $m_{\mathrm{H}^{\pm}} = 311\,\mathrm{GeV}/c^2$ and $30\,\mathrm{fb}^{-1}$.

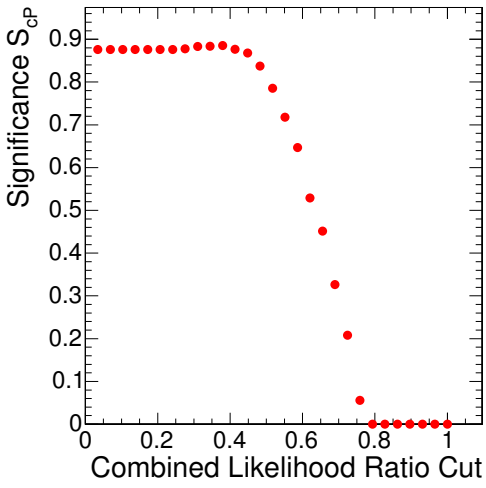


Figure 7.29: Statistical significance as a function of the cut on $\mathcal{L}_{\mathrm{SB}}$ for $m_{\mathrm{H}^{\pm}} = 311\,\mathrm{GeV}/c^2$ and $30\,\mathrm{fb}^{-1}$.

systematic uncertainty on the level of the background of 1% and 3%. The resulting significance as a function of $\mathcal{L}_{\mathrm{SB}}$ for 1% uncertainty is shown in Figure 7.31. The corresponding discovery contours are plotted in Figure 7.32. A value of 3% is no longer visible in the figure. From the above estimate of the systematic uncertainty on the number of background events, it is concluded that with the presented analysis, no sensitivity in this channel with four b-tagged jets is obtained in the relevant MSSM parameter space during the low-luminosity phase of LHC.

Assuming the same signal kinematics as for the MSSM charged Higgs-boson production, the required $\sigma(\mathrm{pp} \to \mathrm{tH}^{\pm}\mathrm{X}) \times \mathrm{BR}(\mathrm{H}^{\pm} \to \mathrm{tb})$ for a 5σ discovery by tagging four b jets is extracted. It is shown in Figure 7.33 as a function of m_{A}.

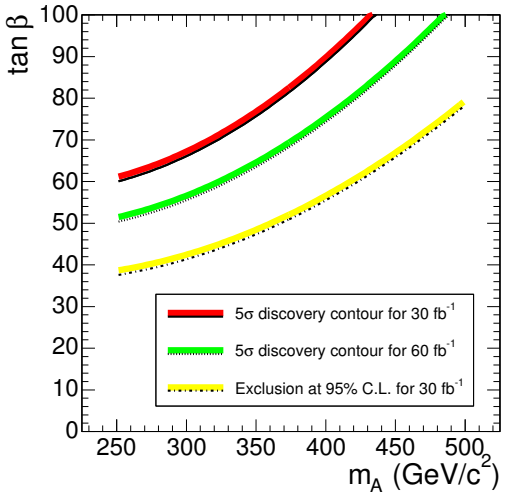


Figure 7.30: Discovery contour for the charged Higgs boson in the $H^{\pm} \to tb$ decay, applying 4 b-tags, for $30\,fb^{-1}$ and $60\,fb^{-1}$, and the exclusion limit at 95% C.L. for $30\,fb^{-1}$. The discovery region lies above the curves. Systematics are not included.

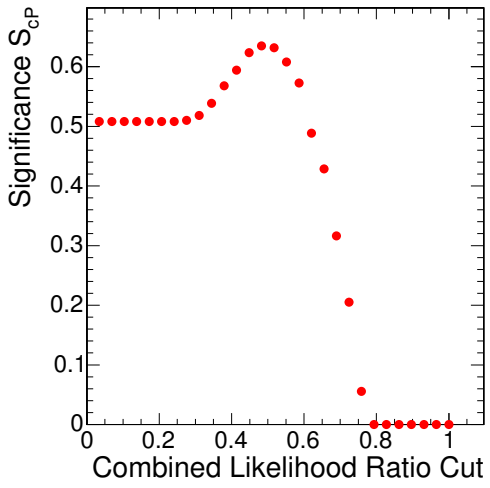


Figure 7.31: Significance as a function of the cut on $\mathcal{L}_{SB}$ for $m_{H^{\pm}} = 311\,GeV/c^2$ and $30\,fb^{-1}$, taking a systematic uncertainty of 1% on the background cross section into account.

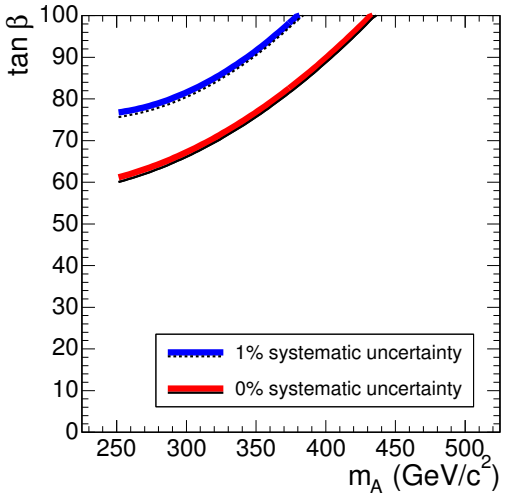


Figure 7.32: Discovery contours for the charged Higgs boson in the $H^{\pm} \to tb$ decay, applying 4 b-tags, for $30\,fb^{-1}$; systematic uncertainties on the background of 0%, 1% and 3% are taken into account. The discovery region lies above the curves.

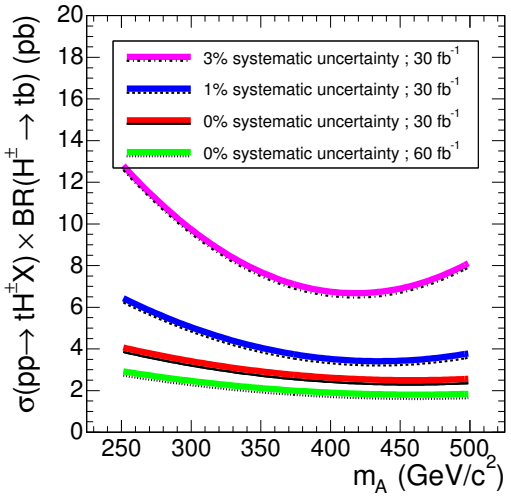


Figure 7.33: Minimally required $\sigma(pp \to tH^{\pm}X) \times BR(H^{\pm} \to tb)$ for discovery of a signal with the same underlying physics, tagging 4 b jets, as a function of m_A.

Chapter 8

Conclusions and Perspectives

Many theoretical and experimental indications favour the interpretation of the Standard Model of elementary particles as a low-energy effective theory. One of the most appealing extensions of the Standard Model, that deals with many of its shortcomings, relates bosons and fermions by a so-called supersymmetry. Such theories require an extension of the Standard Model Higgs sector with an extra complex doublet, such that three neutral Higgs bosons h, H and A and two charged Higgs bosons $H^{\pm}$ are expected after electroweak symmetry breaking, preferably in the energy range that can be probed with the LHC accelerator. If only a neutral Higgs boson is discovered at the LHC, it will be difficult to derive information on the nature of the Higgs sector. The observation of a charged Higgs boson, on the other hand, yields immediate proof of new physics in the form of an extended Higgs sector.

In this thesis, an optimized method for charged Higgs-boson identification in the $H^{\pm} \to tb$ decay channel has been proposed and studied for the CMS experiment at the LHC. It has been demonstrated that b-tagging techniques are crucial experimental tools to extract the charged Higgs-boson signal from the large top-quark background. In this context of b-identification a novel technique has been designed to measure the b-tagging performance on b jets from data. This method provides currently the best estimate of the achievable precision for the calibration of b-identification algorithms.

In this chapter, the results on the $H^{\pm} \to tb$ decay channel are summarized and compared to the discovery prospects for other decay channels in the CMS and ATLAS experiments. The $H^{\pm} \to \tau\nu$ decay is shown to be the most promising channel for light and heavy charged Higgs-boson discovery at the LHC, while supersymmetric decay modes provide perspectives for charged Higgs-boson observation in regions of MSSM parameter space not accessible by Standard Model decay channels. The discovery potential for charged Higgs bosons is finally put into the broader perspective of the overall observability of the charged and the neutral MSSM Higgs bosons at the LHC.

8.1 Charged Higgs Bosons in the $H^{\pm} \to tb$ Decay

In Chapter 7 a detailed analysis is performed of the observability of heavy charged Higgs bosons in the $H^{\pm} \to tb$ decay channel in a general Type-II two-Higgs-doublet model and an interpretation in the MSSM parameter space is given. The channel is analysed

for the final states $tH^{\pm} \rightarrow qq'\ell\nu bbb$ and $tH^{\pm}b \rightarrow qq'\ell\nu bbbb$, where respectively three or four b jets are tagged. Additional event-selection requirements are the presence of an isolated muon, which allows the events to be triggered, and two additional non-b jets.

The combinatorial problem of the association of the jets in an event to the originating partons is tackled with a least-square kinematic fit using Lagrange multipliers to constrain the top-quark and W-boson candidates. The output of this fit is combined with information on the additional jets by a multivariate likelihood technique. The large number of jet combinations, however, and the contaminations arising from final-state gluon radiation and pile-up collisions, limit the fraction of events with correctly associated jets. In the case of three b jets in the final state, 13% of the events is found to have all jets correctly associated. Therefore no charged Higgs-boson mass reconstruction is attempted and the analysis is considered as a counting experiment.

In the next step an optimization is performed of the signal significance. Several variables that discriminate between background and signal are combined into a global observable that allows to balance the signal-to-noise ratio and the signal selection efficiency. By scanning this observable, the point of maximal signal significance is obtained. This optimization is repeated for several input masses $m_{H^{\pm}}$ and for several assumptions on the systematic uncertainty. The resulting discovery contours are shown in the MSSM $(m_A, \tan\beta)$ plane in Figures 7.18 and 7.32, respectively for the final states with three and four b-tagged jets, and are summarized in Figure 8.1 for a general 2HDM Type-II.

The limiting systematic uncertainty in the analysis proves to be the knowledge of the background cross section, which reduces the potential of the $H^{\pm} \rightarrow tb$ channel drastically. A survey of the expected systematic uncertainty on the background level showed that the uncertainties associated with the efficiency of the b-identification algorithms, discussed in Chapter 6, play a dominant role. In both cases with three or four b-tagged jets, it is concluded that the $H^{\pm} \rightarrow tb$ decay channel is not observable during the low-luminosity phase of the LHC, unless nature chooses a very large value for $\tan\beta$, which has been argued in Section 1.2.4 to lead to perturbativity problems in the theory. The minimal cross sections for discovery of the $H^{\pm} \rightarrow tb$ channel with the presented analyses are derived, shown in Figures 7.19 and 7.33 as a function of m_A, independent of $\tan\beta$. These minimal cross-section values can be used as a benchmark for other models in which charged Higgs-bosons are produced with enhanced cross sections and similar kinematics.

Throughout the thesis, emphasis is put on many experimental and phenomenological aspects that hamper the use of the $H^{\pm} \rightarrow tb$ channel for charged Higgs-boson identification. The production of $t\bar{t}$ events with extra jets is discussed in Section 3.4.2, and it is noted that at the LHC two out of three $t\bar{t}$ events are accompanied by at least one extra jet from gluon radiation with $p_T > 20\,\text{GeV}$ and $|\eta| < 5$, giving rise to a similar topology as the charged Higgs-boson signal. In Section 3.3.2 it is shown that a significant amount of such gluon jets splits into a heavy-flavour quark pair, limiting the rejection power of b-tagging tools on these extra jets. A final problem for the charged Higgs-boson identification, discussed in Section 5.2.4, is the reduction of the b-tag efficiency for soft and forward jets and the simultaneous increase up to an order

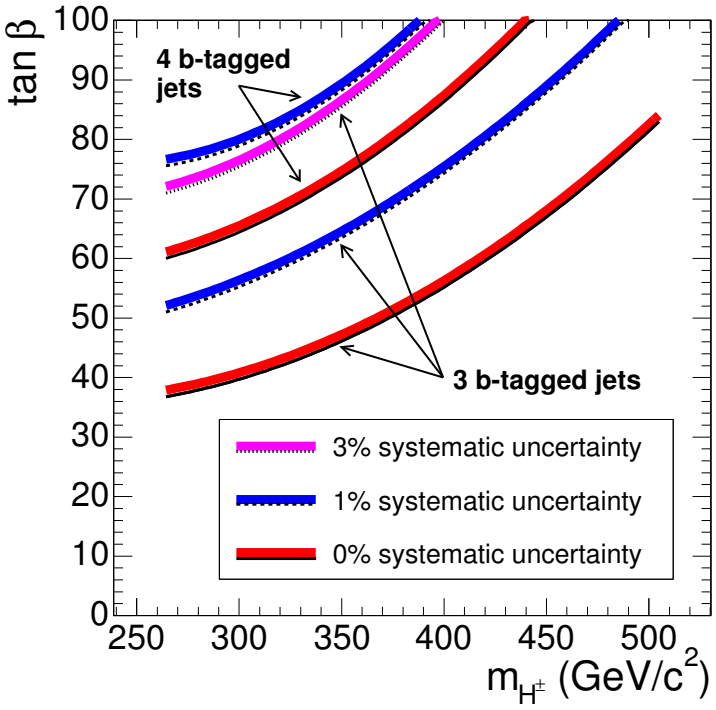


Figure 8.1: General 2HDM Type-II discovery contours for the charged Higgs boson in the $H^\pm \to tb$ decay, asking for 3 and 4 b-tagged jets, for 30 fb^{-1}; systematic uncertainties on the background of 0%, 1% and 3% are taken into account. The discovery region lies above the curves.

of magnitude of the mis-identification probabilities for non-b jets. This large deviation from the ideal b-tagging environment is especially important in the analysis with four b-tagged jets, where the fourth b jet is expected to be soft and very forward and hence very difficult to correctly identify.

8.2 Charged Higgs Bosons in the $H^\pm \to \tau\nu$ Decay

In Section 1.2.2 the phenomenology of the Higgs sector in two-Higgs-doublet models is described. In a 2HDM Type-II, of which the MSSM is an extended example, the charged Higgs boson decays as shown in Figure 1.7, provided no additional decay modes to for example supersymmetric particles are allowed. Taking into account the difficulties in observing the charged Higgs boson in the $H^\pm \to tb$ channel, only the $H^\pm \to \tau\nu$ decay seems a viable Standard Model decay channel for charged Higgs-boson discovery.

Background hadronic τ leptons originate either from genuine τ's in for example the $t\bar{t}$ background process, or from fake τ's in the W + jets or QCD backgrounds. The latter can be efficiently suppressed by exploiting the characteristic signatures of hadronic τ decays complemented with signal-dependent cuts. The former and main source of background τ's from W decays, seems at first sight only kinematically different from τ leptons from charged Higgs-boson decays. Hence, such τ's seem rather difficult to

suppress, especially for a low charged Higgs-boson mass. An additional difference exists, however, between the τ leptons from W and $H^{\pm}$ decays. The W boson, being a vector boson of spin 1, and the charged Higgs boson, being a scalar spin-0 particle, give rise to differently polarized τ's. This property is reflected experimentally in a leading charged pion taking a large part of the τ momentum in a significant fraction of the $H^{\pm} \to \tau^{\pm}\nu \to$ hadrons decays, compared to almost none in the $W^{\pm} \to \tau^{\pm}\nu \to$ hadrons decays [185]. Hence, a cut on the ratio of the momentum of the leading track in the τ jet and the jet's energy allows to suppress τ jets arising from W-boson decays.

For masses $m_{H^{\pm}}$ below the top-quark mass, the charged Higgs-boson production happens in $t\bar{t}$ events through top-quark decays, as discussed in Section 3.4.3, and the subsequent decay $H^{\pm} \to \tau\nu$ dominates. In CMS a study has been performed of the observability of a light charged Higgs boson in this channel with the τ lepton decaying hadronically [186]. The analysis is considered in the MSSM m_{h}^{max} benchmark scenario [45, 187], also called the maximal-mixing scenario, which gives rise to maximal values of the mass of the lightest Higgs boson and leads to conservative bounds on $\tan\beta$ and m_A [188]. The signal of interest is

$$pp \to t\bar{t} \to H^{\pm}W^{\mp}b\bar{b} \qquad \text{with} \qquad H^{\pm} \to \tau^{\pm}\nu, \;\; W^{\mp} \to \ell^{\mp}\nu, \tag{8.1}$$

with the τ lepton decaying to highly-collimated and low-multiplicity hadron jets. This final state contains an isolated lepton from the W decay, being an electron or a muon, which ensures a relatively high trigger efficiency, as has been shown in Section 4.3.2. Backgrounds arise mainly from $t\bar{t} \to W^{\pm}W^{\mp}b\bar{b}$ decays with the same final state or misidentified τ's, and to a lesser extent from W + jets production.

The event selection requires an isolated electron or muon, a highly collimated τ jet, and at least three jets of which exactly one is tagged as b jet. Since the charged Higgs-boson mass is constrained to be larger than the W-boson mass, the accompanying b is softer in top-quark decays to a charged Higgs boson than in decays to a W boson. Hence, asking for a second b-tagged jet does not help to suppress the dominant $t\bar{t}$ background and can be avoided, provided the W + jets background can be efficiently suppressed. This is ensured by a hard cut on the transverse missing energy ($E_{T}^{miss} > 70\,\text{GeV}$) and by the demand that the isolated lepton and the τ jet have opposite charge. Finally, the τ-polarization difference with τ's from W decays is exploited by asking the momentum of the leading track in the jet to account for at least 80% of the total jet energy.

Two neutrinos are present in the final state for this channel, which hinders the reconstruction of the charged Higgs-boson mass. Considering the analysis as a counting experiment, good control of the background cross sections is needed to interpret a possible excess of events as a charged Higgs-boson signal. In Figure 8.2 the resulting discovery reach is shown in the $(m_A, \tan\beta)$ plane in the m_{h}^{max} scenario, with and without all expected background systematic uncertainties taken into account.

When the charged Higgs-boson mass is larger than the top-quark mass, the $H^{\pm} \to \tau\nu$ decay becomes subleading and is additionally suppressed for decreasing $\tan\beta$. In this range of large $m_{H^{\pm}}$, a study has been performed to determine the visibility of

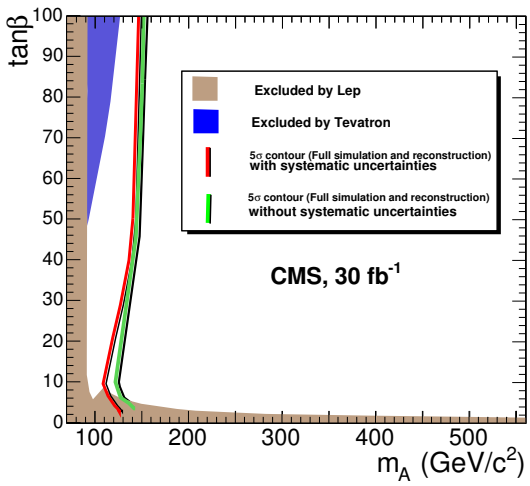


Figure 8.2: $H^{\pm}$ discovery reach in the $(m_A, \tan\beta)$ plane in the MSSM m_h^{max} scenario, for the $H^{\pm} \to \tau\nu$ decay with $m_{H^{\pm}} < m_t$ and using a lepton trigger. The discovery region lies on the left side of the curves.

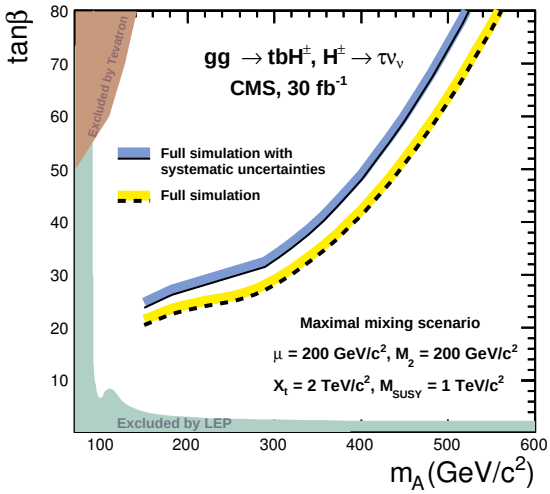


Figure 8.3: $H^{\pm}$ discovery reach in the $(m_A, \tan\beta)$ plane in the MSSM m_h^{max} scenario, for the $H^{\pm} \to \tau\nu$ decay with large $m_{H^{\pm}}$ and using the hadronic-τ trigger. The discovery region lies above the curves.

$gg \to tH^{\pm}b$ production in the previously-defined maximal-mixing MSSM scenario for the fully hadronic final state [189]. In this final state only one neutrino is present, and hence the charged Higgs-boson mass can be determined. The trigger is provided in this case by the large missing energy and the hard τ jet.

The event selection in this analysis requires both the p_T of the reconstructed τ jet and the transverse missing energy to exceed 100 GeV. In addition, exactly three jets are demanded, of which exactly one needs to be identified as a b jet. The veto on extra jets or b jets suppresses the $t\bar{t}$ background which has on average a higher jet multiplicity. These jets need to be compatible with the hypothesis that they are produced in a hadronic top-quark decay. Also the absence of isolated electrons and muons is required. Finally, the polarization correlations in τ decays are exploited by asking the leading track in the jet to carry at least 80% of the τ-jet's energy.

With the proposed event selection, the distribution of the transverse invariant mass of the τ jet and the missing energy is almost background free in the signal region above $\sim 100\,\mathrm{GeV}/c^2$. Because of the statistical limitation of the available signal simulation, no mass reconstruction has been attempted yet. A survey of possible systematic influences on the background revealed that the systematic uncertainty is currently dominated by the limited size of the simulated background samples. Therefore, the results of this study are conservative when taking into account the systematic uncertainties. In Figure 8.3 the discovery reach is shown in the $(m_A, \tan\beta)$ plane, with and without the influence of the uncertainty on the background cross section.

8.3 Charged Higgs-Boson Observability at the LHC

Using the results from the studies described in the previous sections and in [85, 186, 189, 190], the charged Higgs-boson discovery potential for CMS in the MSSM $m_{\mathrm{h}}^{\mathrm{max}}$ scenario is summarized in Figure 8.4 for an integrated luminosity of $30\,\mathrm{fb}^{-1}$ [86]. The quoted studies supersede all previous CMS results, in particular [40, 181]. Major improvements are the use of a realistic detailed detector simulation, background simulations from dedicated matrix-element generators, the application of realistic b-tagging algorithms and trigger scenarios and the inclusion of an estimate of all relevant systematic uncertainties.

Several studies of the charged Higgs-boson observability have also been carried out for the ATLAS experiment. All these studies have been performed with parametrized detector simulation, recently also including estimates of systematic uncertainties. For the $\mathrm{H}^{\pm} \rightarrow \tau\nu$ channels [191, 192] similar strategies have been adopted as in the previously-discussed CMS studies. In addition, the use of a hadronic-τ trigger has been investigated in the low-mass region, which allows to concentrate on a hadronically-decaying associated top quark and hence it becomes possible to reconstruct the charged Higgs-boson mass. A relative precision on this mass determination of less than 4% is envisaged with $10\,\mathrm{fb}^{-1}$ of integrated luminosity.

In the case of the $\mathrm{H}^{\pm} \rightarrow \mathrm{tb}$ decay, both the final states with three and four b jets have been considered for ATLAS [193, 194]. These studies, however, have been performed with a b-tagging efficiency of 50%, assuming the b-jet mis-identification probabilities to be 10% for jets originating from c quarks and 1% for udsg jets. In addition, all these efficiencies have been taken fixed with respect to the jets' transverse energy and pseudorapidity. As a result, the signal-to-background ratio is relatively high, since only the $\mathrm{t\bar{t}b}$ or $\mathrm{t\bar{t}b\bar{b}}$ is found to be an important background. In this thesis, however, it has been shown that such a b-tagging performance is not realistic, especially in the presence of soft, forward jets and gluon jets. Consequently, the signal-to-background ratio has been largely overestimated so far in the ATLAS studies, and the $\mathrm{t\bar{t}}$ background with additional non-b jets needs to be reconsidered.

In Figure 8.5 the summary is shown of some of the results of the ATLAS studies for an integrated luminosity of $300\,\mathrm{fb}^{-1}$ [88]. This overview could further be updated with the recent $\mathrm{H}^{\pm} \rightarrow \tau\nu$ study using a hadronic-τ trigger in the $m_{\mathrm{H}^{\pm}} < m_{\mathrm{t}}$ range, which improves the visibility in the intermediate $\tan\beta$ region, and with the results on the $\mathrm{H}^{\pm} \rightarrow \mathrm{tb}$ decay with four b-tagged jets in the final state. A re-evaluation of most channels is needed, however, taking into account a detailed detector simulation, all relevant backgrounds, realistic trigger scenarios, systematic uncertainties, etc. Such an effort is currently ongoing within the ATLAS collaboration, and new results can be expected before the LHC start-up foreseen in 2007.

From Figure 8.4 it is concluded that for a large part of the MSSM parameter space the detection of charged Higgs bosons will not be possible using the dominant Standard Model decay channels. The $\mathrm{H}^{\pm} \rightarrow \mathrm{W}^{\pm}\mathrm{h}$ and $\mathrm{H}^{\pm} \rightarrow \mathrm{W}^{\pm}\mathrm{H}$ decays, which can reach significant branching fractions in small regions of the MSSM parameter space, have been shown not to be viable channels because of the large irreducible backgrounds [195–197]. The subleading $\mathrm{H}^{\pm} \rightarrow \mu^{\pm}\nu$ decay, which gives rise to a distinct

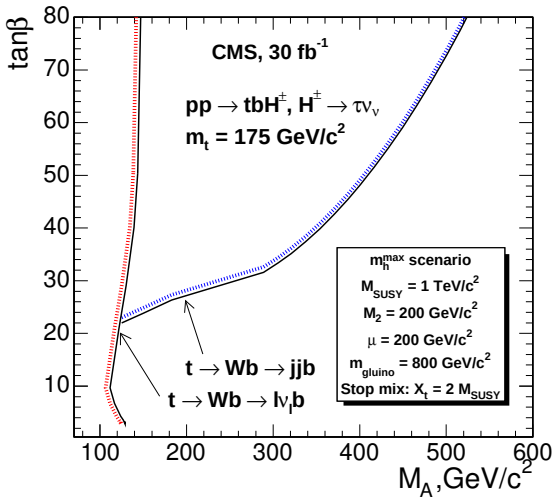


Figure 8.4: Summary of the $H^{\pm}$ discovery potential for CMS in the MSSM m_h^{max} scenario for an integrated luminosity of $30\,fb^{-1}$. The discovery region lies at the dashed side of the curves.

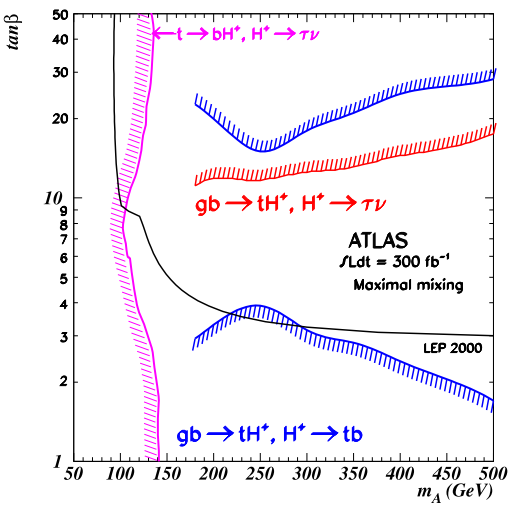


Figure 8.5: Summary of the $H^{\pm}$ discovery potential claimed by ATLAS in the MSSM m_h^{max} scenario for an integrated luminosity of $300\,fb^{-1}$. The discovery region lies at the dashed side of the curves.

experimental signature with a hard muon, has a tiny branching fraction of $\mathcal{O}(10^{-3})$ and is therefore also expected to be overwhelmed by Standard Model backgrounds.

Apart from Standard Model and Higgs-boson decay modes, also decays into supersymmetric particles can be exploited in the MSSM, if kinematically allowed. A channel of particular interest is the decay $H^{\pm} \to \tilde{\chi}_1^{\pm}\tilde{\chi}_2^0$ into a chargino–neutralino pair, giving rise to a final state with three isolated leptons and missing energy from two escaping stable neutralinos. Backgrounds originate mainly from other supersymmetric processes and $t\bar{t}$ production. The presence of such signatures, however, depends strongly on all MSSM parameters, not only m_A and $\tan\beta$. Therefore, current analyses have selected a particular MSSM scenario and have determined the accessible region in the $(m_A, \tan\beta)$ plane assuming the other MSSM parameters fixed. It has been shown that, for a particular choice of MSSM parameters, the discovery of the charged Higgs boson using sparticle decay modes is possible in a large part of the yet uncovered low-$\tan\beta$ region in the $(m_A, \tan\beta)$ plane [198, 199]. These studies prove the usefulness of supersymmetric decay modes of Higgs bosons, and have initiated the determination of benchmark points in the MSSM parameter space that characterize particular signatures from supersymmetric charged Higgs-boson decays [200].

8.4 The MSSM Higgs Sector at the LHC

It has been indicated before that the observation of charged Higgs bosons at the LHC would yield immediate proof of physics beyond the Standard Model. The interpretation of such a signal in terms of a Higgs boson in a certain model with two complex Higgs doublets, on the other hand, is far from trivial. In the previous sections it has been

emphasized that the detection of the charged Higgs boson in the MSSM is difficult, except for very low m_{A} or very high $\tan\beta$, and seems only possible in one decay channel for large parts of the parameter space. Hence the translation of the observed signal to estimates of model parameters is even more challenging. It is clear that many additional experimental inputs will be needed to unravel the structure of the model underlying the new physics.

Assuming the MSSM is valid and the charged Higgs boson is discovered, then further information on the parameters m_{A} and $\tan\beta$ can be obtained directly from measurements on the neutral Higgs bosons, since the complete Higgs sector only depends on these two parameters at the tree level. Inversely, observation of the neutral Higgs bosons constrains the available parameter space for the charged Higgs bosons. In the following the discovery prospects for the neutral Higgs bosons in the MSSM are summarized [86].

As has already been briefly discussed in Section 2.3.3, the lightest Higgs boson decays predominantly into b-quark and τ-lepton pairs, except close to its upper mass limit of about 130 GeV, where it behaves like the Standard Model Higgs boson. At this limit, called the decoupling limit, m_{A} is large and also the other neutral scalar H is heavy. The lightest Higgs boson h is Standard Model-like, and hence can be discovered in the $\mathrm{h} \to \gamma\gamma$ decay channel, while all other Higgs bosons are heavy and therefore difficult to observe. For m_{h} smaller than its upper limit, the $\mathrm{h} \to \mathrm{b\bar{b}}$ decay is extremely difficult to detect at the LHC due to the large QCD multi-jet background, but the $\mathrm{h} \to \tau^+\tau^-$ channel can be observed in either the vector-boson fusion production mode $\mathrm{pp} \to \mathrm{q\bar{q}W^+W^-/q\bar{q}ZZ} \to \mathrm{q\bar{q}h}$ tagging the extra forward jets, or at large $\tan\beta$ in the associated production $\mathrm{pp} \to \mathrm{b\bar{b}h}$.

This production with associated b quarks $\mathrm{pp} \to \mathrm{b\bar{b}}\phi$ is insensitive to the nature of the produced neutral Higgs boson $\phi = \mathrm{h, H, A}$ and therefore the previously mentioned $\tau^+\tau^-$ decay channel can be used at large $\tan\beta$ in searches for all three neutral Higgs bosons. The vector-boson fusion channel, on the other hand, vanishes for the CP-odd scalar A, for which no tree level couplings are allowed to the gauge bosons in two-Higgs-doublet models. This production channel remains useful, however, for the heavy CP-even Higgs boson H when it is close to its lower mass limit.

The discovery potential of the neutral Higgs bosons in the production and decay channels discussed above is summarized for CMS in Figure 8.6 for the inclusive $\mathrm{h} \to \gamma\gamma$ search and the h and H-boson search in the vector-boson fusion mode, and in Figure 8.7 for all neutral Higgs bosons in the production with associated b quarks [86]. For completeness, a few more channels can be exploited, involving for instance decays of H or A bosons in top-quark pairs at sufficiently large m_{A} and small $\tan\beta$.

8.5 Concluding remarks

Throughout this chapter an overview has been given of the observability at the LHC of the charged and neutral Higgs bosons in the MSSM. Several interesting observations can be made from this survey. First of all, the LHC is capable of observing at least one neutral Higgs boson over the complete MSSM parameter space with $30\,\mathrm{fb}^{-1}$ of integrated luminosity. At low m_{A} or large $\tan\beta$ many channels are complementary in

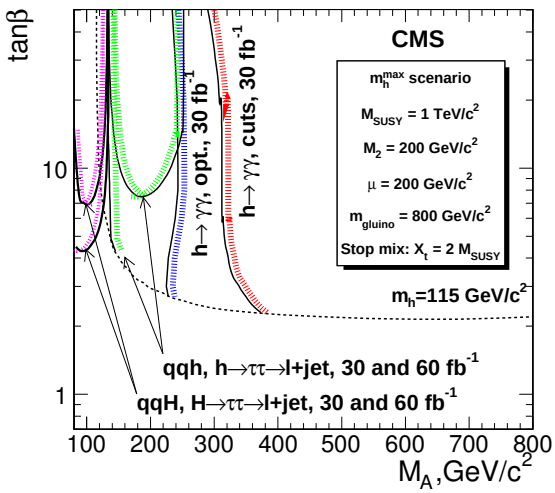


Figure 8.6: The CMS discovery potential for neutral Higgs bosons in the inclusive $h \to \gamma\gamma$ search and in the vector-boson fusion production mode. The discovery region lies at the dashed side of the curves.

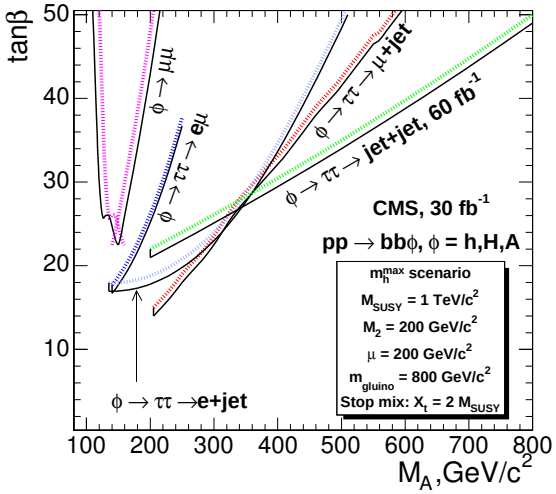


Figure 8.7: The CMS discovery potential for neutral Higgs bosons in the production mode with associated b quarks. The discovery region lies at the dashed side of the curves.

observing the complete MSSM Higgs sector. Efficient τ- and b-tagging are essential experimental techniques for such measurements and rely on the performance of the central tracking detector. The interpretation of the observed signals, however, might prove to be a difficult task, as some or all Higgs bosons can be degenerate in mass. Especially in such cases of degeneracy the charged Higgs boson can play an important role. For intermediate values of tanβ and/or high m_{A}, only one Higgs boson might be observed, of which the behaviour cannot be distinguished from the Standard Model Higgs boson. In this scenario extra input from supersymmetric decay channels, or from supersymmetric cascade decays involving new Higgs-boson production channels, will be needed to provide insight in the model parameters underlying the MSSM Higgs sector.

In the previous sections, all interpretations have been given in the framework of the MSSM, restricted to the particular $m_{\mathrm{h}}^{\mathrm{max}}$ scenario. The presented studies, however, are devised as analyses that look for deviations from Standard Model expectations in given topologies, possibly inspired by the MSSM. As such, these analyses are far more general than a specific scenario. Many of the analysed topologies are also expected in other models with an extended Higgs sector, although the absence of the MSSM constraints complexifies the exploration of the parameter space of such models. The interpretation in the MSSM can therefore be seen as a general demonstration of the potential of Higgs-boson physics in the CMS and ATLAS experiments at the LHC accelerator. Since many indications point towards new physics at the TeV energy scale, these studies serve as ideal preparations for the analyses that will explore this energy frontier in the coming years. Whatever the phenomena that await discovery in the TeV energy domain, the LHC and its experiments represent a new milestone on the journey towards a deeper comprehension of the foundations of nature.

Bibliography

[1] S. Weinberg, *The Quantum Theory of Fields. Vol. 1: Foundations*, Cambridge, UK: Univ. Pr. (1995) 609 p.

[2] M. E. Peskin and D. V. Schroeder, *An Introduction to Quantum Field Theory*, Reading, USA: Addison-Wesley (1995) 842 p.

[3] F. Halzen and A. D. Martin, *Quarks and Leptons: an Introductory Course in Modern Particle Physics*, New York, Usa: Wiley (1984) 396p.

[4] G. L. Kane, *Modern Elementary Particle Physics*, Redwood City, USA: Addison-Wesley (1987) 344 P. (The Advanced Book Program).

[5] S. L. Glashow, Nucl. Phys. **22** (1961) 579–588.

[6] A. Salam and J. C. Ward, Phys. Lett. **13** (1964) 168–171.

[7] S. Weinberg, Phys. Rev. Lett. **19** (1967) 1264–1266.

[8] ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaboration, SLD Collaboration, LEP Electroweak Working Group, SLD Electroweak Group and SLD Heavy Flavour Group, Phys. Rept. **427** (2006) 257.

[9] DØ Collaboration, S. Abachi *et al.*, Phys. Rev. Lett. **78** (1997) 3818–3823.

[10] CDF Collaboration, A. A. Affolder *et al.*, Phys. Rev. Lett. **84** (2000) 835–840.

[11] `http://www-cdf.fnal.gov/physics/new/top/top.html`

[12] G. 't Hooft and M. J. G. Veltman, Nucl. Phys. **B44** (1972) 189–213.

[13] F. Englert and R. Brout, Phys. Rev. Lett. **13** (1964) 321–322.

[14] P. W. Higgs, Phys. Rev. Lett. **13** (1964) 508–509.

[15] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, Phys. Rev. Lett. **13** (1964) 585–587.

[16] UA2 Collaboration, M. Banner *et al.*, Phys. Lett. **B122** (1983) 476–485.

[17] UA1 Collaboration, G. Arnison *et al.*, Phys. Lett. **B122** (1983) 103–116.

[18] UA1 Collaboration, G. Arnison *et al.*, Phys. Lett. **B126** (1983) 398–410.

[19] UA2 Collaboration, P. Bagnaia *et al.*, Phys. Lett. **B129** (1983) 130–140.

[20] CDF Collaboration, F. Abe *et al.*, Phys. Rev. Lett. **74** (1995) 2626–2631.

[21] DØ Collaboration, S. Abachi *et al.*, Phys. Rev. Lett. **74** (1995) 2632–2637.

[22] The LEP Working Group for Higgs Boson Searches, R. Barate *et al.*, Phys. Lett. **B565** (2003) 61–75.

[23] M. J. G. Veltman, Acta Phys. Polon. **B8** (1977) 475.

[24] `http://lepewwg.web.cern.ch/LEPEWWG/`

[25] ALEPH Collaboration, *A Combination of Preliminary Electroweak Measurements and Constraints on the Standard Model*, `hep-ex/0511027`.

[26] Tevatron Electroweak Working Group, *Combination of CDF and DØ Results on the Mass of the Top Quark*, `hep-ex/0608032`.

[27] Higgs Working Group, M. Carena *et al.*, *Report of the Tevatron Higgs Working Group*, `hep-ph/0010338`.

[28] `http://tevnphwg.fnal.gov/`

[29] G. Altarelli (Ed.) and M. L. Mangano (Ed.), *Standard Model Physics (and more) at the LHC. Proceedings, Workshop, Geneva, Switzerland, May 25-26, October 14-15, 1999*, CERN-2000-004.

[30] Particle Data Group Collaboration, W. M. Yao *et al.*, J. Phys. **G33** (2006) 1–1232.

[31] T. Hambye and K. Riesselmann, *SM Higgs Mass Bounds from Theory*, `hep-ph/9708416`.

[32] D. N. Spergel *et al.*, *Wilkinson Microwave Anisotropy Probe (WMAP) Three Year Results: Implications for Cosmology*, `astro-ph/0603449`.

[33] G. L. Fogli *et al.*, *Observables Sensitive to Absolute Neutrino Masses: A Reappraisal after WMAP-3y and First MINOS Results*, `hep-ph/0608060`.

[34] J. F. Gunion, H. E. Haber, G. L. Kane, and S. Dawson, *The Higgs Hunter's Guide*, SCIPP-89/13.

[35] A. Djouadi, J. Kalinowski, and M. Spira, Comput. Phys. Commun. **108** (1998) 56–74.

[36] S. P. Martin, *A Supersymmetry Primer*, `hep-ph/9709356`.

[37] D. Z. Freedman, P. van Nieuwenhuizen, and S. Ferrara, Phys. Rev. **D13** (1976) 3214–3218.

[38] S. Deser and B. Zumino, Phys. Lett. **B62** (1976) 335.

[39] M. Dine and A. E. Nelson, Phys. Rev. **D48** (1993) 1277–1287.

[40] P. Salmi, R. Kinnunen, and N. Stepanov, *Prospects of Detecting Massive Charged Higgs from Hadronic Decay $H^{\pm} \to tb$ in CMS*, CMS-NOTE-2002-024.

[41] V. D. Barger, J. L. Hewett, and R. J. N. Phillips, Phys. Rev. **D41** (1990) 3421.

[42] B. Schrempp and M. Wimmer, Prog. Part. Nucl. Phys. **37** (1996) 1–90.

[43] A. Sopczak, *Review of Current Limits from Charged Higgs Searches at LEP*, presented at the *Charged Higgs Workshop 2006*, Uppsala, Sweden, 2006.

[44] DELPHI Collaboration, J. Abdallah *et al.*, Eur. Phys. J. **C34** (2004) 399–418.

[45] ALEPH Collaboration, *Search for Neutral MSSM Higgs Bosons at LEP*, `hep-ex/0602042`.

[46] DØ Collaboration, B. Abbott *et al.*, Phys. Rev. Lett. **82** (1999) 4975–4980.

[47] DØ Collaboration, V. M. Abazov *et al.*, Phys. Rev. Lett. **88** (2002) 151803.

[48] CDF Collaboration, A. Abulencia *et al.*, Phys. Rev. Lett. **96** (2006) 042003.

[49] UTfit Collaboration, M. Bona *et al.*, *The Unitarity Triangle Fit in the Standard Model and Hadronic Parameters From Lattice QCD: A Reappraisal After the Measurements of Δm_s and BR($B \to \tau\nu_\tau$)*, `hep-ph/0606167`.

[50] `http://www.slac.stanford.edu/xorg/hfag/`

[51] Heavy Flavor Averaging Group (HFAG), *Averages of b-Hadron Properties at the End of 2005*, `hep-ex/0603003`.

[52] T. Browder, for the Belle collaboration, *Rare B-Decays with Missing Energy*, presented at the *33rd International Conference on High Energy Physics (ICHEP 06)*, Moscow, Russia, 2006.

[53] BABAR Collaboration, B. Aubert, *A Search for $B^+ \to \tau^+\nu$ Recoiling Against $B^- \to D^0 l^- \bar{\nu} X$*, `hep-ex/0608019`.

[54] T. Hurth, E. Lunghi, and W. Porod, Nucl. Phys. **B704** (2005) 56–74.

[55] CLEO Collaboration, S. Chen *et al.*, Phys. Rev. Lett. **87** (2001) 251807.

[56] Belle Collaboration, P. Koppenburg *et al.*, Phys. Rev. Lett. **93** (2004) 061803.

[57] BaBar Collaboration, *Measurement of the Branching Fraction and Photon Energy Moments of $B \to X_s\gamma$ and $A_{CP}(B \to X_{s+d}\gamma)$*, `hep-ex/0607071`.

[58] H. Itoh, S. Komine, and Y. Okada, Prog. Theor. Phys. **114** (2005) 179–204.

[59] S. Sekula, for the BaBar collaboration, *Constraints on the Charged Higgs Boson from the BaBar Experiment*, presented at the *Charged Higgs Workshop 2006*, Uppsala, Sweden, 2006.

[60] T. Appelquist, H.-C. Cheng, and B. A. Dobrescu, Phys. Rev. **D64** (2001) 035002.

[61] N. Arkani-Hamed, S. Dimopoulos, and G. R. Dvali, Phys. Lett. **B429** (1998) 263–272.

[62] L. Randall and R. Sundrum, Phys. Rev. Lett. **83** (1999) 3370–3373.

[63] L. Randall and R. Sundrum, Phys. Rev. Lett. **83** (1999) 4690–4693.

[64] A. Leike, Phys. Rept. **317** (1999) 143–250.

[65] C. T. Hill and E. H. Simmons, Phys. Rept. **381** (2003) 235–402.

[66] K. D. Lane, *Technicolor 2000*, `hep-ph/0007304`.

[67] N. Arkani-Hamed, A. G. Cohen, and H. Georgi, Phys. Lett. **B513** (2001) 232–240.

[68] N. Arkani-Hamed, A. G. Cohen, E. Katz, and A. E. Nelson, JHEP **07** (2002) 034.

[69] `http://www.fnal.gov/pub/now/tevlum.html`

[70] O. Bruning (Ed.) *et al.*, *LHC Design Report. Vol. I: The LHC Main Ring*, CERN-2004-003-V-1.

[71] J. List, *Physics at the International Linear Collider*, `hep-ex/0605087`.

[72] `http://lhc.web.cern.ch/lhc/`

[73] `http://lhc-commissioning.web.cern.ch/lhc-commissioning/`

[74] `http://atlasinfo.cern.ch/ATLAS/`

[75] `http://aliceinfo.cern.ch/`

[76] `http://lhcb.web.cern.ch/lhcb/`

[77] CMS Collaboration, *CMS, The Compact Muon Solenoid: Technical Proposal*, CERN-LHCC-94-38.

[78] ATLAS Collaboration, W. W. Armstrong *et al.*, *ATLAS: Technical Proposal for a General-Purpose pp Experiment at the Large Hadron Collider at CERN*, CERN-LHCC-94-43.

[79] LHCb Collaboration, S. Amato *et al.*, *LHCb Technical Proposal*, CERN-LHCC-98-4.

[80] ALICE Collaboration, *ALICE: Technical Proposal for A Large Ion Collider Experiment at the CERN LHC*, CERN-LHCC-95-71.

[81] TOTEM Collaboration, *TOTEM: Total Cross Section, Elastic Scattering and Diffraction Dissociation at the LHC: Technical Proposal*, CERN-LHCC-1999-007.

[82] LHCf Collaboration, O. Adriani *et al.*, *Technnical Proposal for the CERN LHCf Experiment: Measurement of Photons and Neutral Pions in the Very Forward Region of LHC*, CERN-LHCC-2005-032.

[83] TOTEM Collaboration, V. Berardi *et al.*, *TOTEM: Technical Design Report. Total Cross Section, Elastic Scattering and Diffraction Dissociation at the Large Hadron Collider at CERN*, CERN-LHCC-2004-002.

[84] COMPETE Collaboration, J. R. Cudell *et al.*, Phys. Rev. Lett. **89** (2002) 201801.

[85] CMS Collaboration, D. Acosta (Ed.), *CMS, Physics Technical Design Report, Vol. 1: Detector Performance and Software*, CERN-LHCC-2006-001.

[86] CMS Collaboration, A. De Roeck (Ed.), *CMS, Physics Technical Design Report, Vol. 2: Physics Performace*, CERN-LHCC-2006-021.

[87] *ATLAS: Detector and Physics Performance Technical Design Report. Vol. 1*, CERN-LHCC-99-14.

[88] *ATLAS Detector and Physics Performance. Technical Design Report. Vol. 2*, CERN-LHCC-99-15.

[89] R. K. Ellis, W. J. Stirling, and B. R. Webber, Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol. **8** (1996) 1–435.

[90] G. Dissertori, I. G. Knowles, and M. Schmelling, *Quantum Chromodynamics: High Energy Experiments and Theory*, Oxford, UK: Clarendon (2003) 538 p.

[91] CTEQ Collaboration, R. Brock *et al.*, Rev. Mod. Phys. **67** (1995) 157–248.

[92] T. Sjostrand *et al.*, Comput. Phys. Commun. **135** (2001) 238–259.

[93] F. Maltoni and T. Stelzer, JHEP **02** (2003) 027.

[94] CompHEP Collaboration, E. Boos *et al.*, Nucl. Instrum. Meth. **A534** (2004) 250–259.

[95] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa, JHEP **07** (2003) 001.

[96] H. L. Lai *et al.*, Phys. Rev. **D51** (1995) 4763–4782.

[97] A. D. Martin, R. G. Roberts, W. J. Stirling, and R. S. Thorne, Eur. Phys. J. **C23** (2002) 73–87.

[98] `http://www.desy.de/`

[99] M. A. Dobbs *et al.*, *Les Houches Guidebook to Monte Carlo Generators for Hadron Collider Physics*, `hep-ph/0403045`.

[100] CDF Collaboration, D. Acosta *et al.*, Phys. Rev. **D70** (2004) 072002.

[101] P. Bartalini, R. Chierici, and A. De Roeck, *Guidelines for the Estimation of Theoretical Uncertainties at the LHC*, CERN-CMS-NOTE-2005-013.

[102] D. Acosta *et al.*, *The Underlying Event at the LHC*, CERN-CMS-NOTE-2006-067.

[103] V. N. Gribov and L. N. Lipatov, Sov. J. Nucl. Phys. **15** (1972) 438–450.

[104] L. N. Lipatov, Sov. J. Nucl. Phys. **20** (1975) 94–102.

[105] G. Altarelli and G. Parisi, Nucl. Phys. **B126** (1977) 298.

[106] Y. L. Dokshitzer, Sov. Phys. JETP **46** (1977) 641–653.

[107] G. Corcella *et al.*, JHEP **01** (2001) 010.

[108] A. H. Mueller and P. Nason, Nucl. Phys. **B266** (1986) 265.

[109] M. L. Mangano and P. Nason, Phys. Lett. **B285** (1992) 160–166.

[110] M. H. Seymour, Z. Phys. **C63** (1994) 99–110.

[111] B. Andersson, Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol. **7** (1997) 1–471.

[112] C. Peterson, D. Schlatter, I. Schmitt, and P. M. Zerwas, Phys. Rev. **D27** (1983) 105.

[113] DELPHI Collaboration, P. Abreu *et al.*, Z. Phys. **C73** (1996) 11–60.

[114] P. Demin, S. De Visscher, A. Bocci, and R. Ranieri, *Tagging b Jets with Electrons and Muons at CMS*, CERN-CMS-NOTE-2006-043.

[115] CDF Collaboration, E. Shabalina, *Top Quark Production Cross Section at the Tevatron*, `hep-ex/0605045`.

[116] M. Beneke *et al.*, *Top Quark Physics*, `hep-ph/0003033`.

[117] F. Maltoni, Private communication.

[118] T. Plehn and M. Spira, Private communication.

[119] J. Alwall, *MATCHIG: A Program for Matching Charged Higgs Boson Production at Hadron Colliders*, `hep-ph/0503124`.

[120] T. Plehn, Phys. Rev. **D67** (2003) 014–018.

[121] CMS Collaboration, *CMS, Tracker Technical Design Report*, CERN-LHCC-98-6.

[122] CMS Collaboration, *Addendum to the CMS Tracker TDR*, CERN-LHCC-2000-016.

[123] CMS Collaboration, P. Sphicas (Ed.), *CMS: The TriDAS project. Technical Design Report, Vol. 2: Data Acquisition and High-Level Trigger*, CERN-LHCC-2002-026.

[124] W. Adam, B. Mangano, T. Speer, and T. Todorov, *Track Reconstruction in the CMS Tracker*, CERN-CMS-NOTE-2006-041.

[125] T. Speer, K. Prokofiev, R. Fruhwirth, W. Waltenberger, and P. Vanlaer, *Vertex Fitting in the CMS Tracker*, CERN-CMS-NOTE-2006-032.

[126] T. Muller, C. Piasecki, G. Quast, and C. Weiser, *Inclusive Secondary Vertex Reconstruction in Jets*, CERN-CMS-NOTE-2006-027.

[127] P. Vanlaer *et al.*, *Impact of CMS Silicon-Tracker Misalignment on Track and Vertex Reconstruction*, CERN-CMS-NOTE-2006-029.

[128] CMS Collaboration, *CMS: The Electromagnetic Calorimeter Technical Design Report*, CERN-LHCC-97-33.

[129] CMS Collaboration, *CMS: The Hadron Calorimeter Technical Design Report*, CERN-LHCC-97-31.

[130] L. Agostino, L. Malgeri, G. Daskalakis, P. Govoni, and M. Paganoni, *Inter-calibration of the CMS Electromagnetic Calorimeter with Isolated Electrons*, CERN-CMS-NOTE-2006-021.

[131] V. Elvira, *Validation of Geant4 Physics Using the CMS HCAL Test Beam 2002 Experiment*, CERN-CMS-NOTE-2004-021.

[132] E. Meschi, T. Monteiro, C. Seez, and P. Vikas, *Electron Reconstruction in the CMS Electromagnetic Calorimeter*, CERN-CMS-NOTE-2001-034.

[133] A. Heister *et al.*, *Measurement of Jets with the CMS Detector at the LHC*, CERN-CMS-NOTE-2006-036.

[134] *CMSIM*, `http://cmsdoc.cern.ch/cmsim/cmsim.html`

[135] *OSCAR: Object-Oriented Simulation for CMS Analysis and Reconstruction*, `http://cmsdoc.cern.ch/oscar/`

[136] GEANT4 Collaboration, S. Agostinelli *et al.*, Nucl. Instrum. Meth. **A506** (2003) 250–303.

[137] *CMS Fast Simulation*, `http://cmsdoc.cern.ch/famos/`

[138] CMS Collaboration, *CMS: The Computing Project. Technical Design Report*, CERN-LHCC-2005-023.

[139] `http://www.eu-egee.org/`

[140] I. Bird (Ed.) *et al.*, *LHC Computing Grid. Technical Design Report*, CERN-LHCC-2005-024.

[141] `http://cmsdoc.cern.ch/cms/ccs/wm/www/Crab/`

[142] CMS Collaboration, S. Dasu *et al.*, *CMS. The TriDAS project. Technical Design Report, Vol. 1: The Level-1 Trigger*, CERN-LHCC-2000-038.

[143] C. Buttar *et al.*, *Les Houches Physics at TeV Colliders 2005, Standard Model, QCD, EW, and Higgs Working group: Summary Report*, `hep-ph/0604120`.

[144] R. Harris, *Jet Calibration from Di-jet Balancing*, CMS-AN-2005-034.

[145] V. Konoplyanikov, O. Kodolova, and A. Ulyanov, *Jet Calibration Using γ+jet Events in the CMS Detector*, CERN-CMS-NOTE-2006-042.

[146] J. D'Hondt, S. Lowette, J. Heyninck, and S. Kasselmann, *Light Quark Jet Energy-Scale Calibration Using the W-Mass Constraint in Single-Leptonic $t\bar{t}$ Events*, CERN-CMS-NOTE-2006-025.

[147] J. Rohlf and C. Tully, *Recommendations for Jet and Missing Transverse Energy Reconstruction Settings and Systematics Treatment*, CMS-IN-2006-025.

[148] N. Ilina, *Association of Jets with the Signal Vertex*, CMS-AN-2005-006.

[149] A. Rizzi, F. Palla, and G. Segneri, *Track Impact-Parameter Based b-Tagging with CMS*, CERN-CMS-NOTE-2006-019.

[150] M. Vos and F. Palla, *b-Tagging in the High-Level Trigger*, CERN-CMS-NOTE-2006-030.

[151] C. Weiser, *A Combined Secondary-Vertex Based b-Tagging Algorithm in CMS*, CERN-CMS-NOTE-2006-014.

[152] C. Weiser, Private communication.

[153] J. D'Hondt, S. Lowette, and J. Heyninck, *Electron and Muon Reconstruction in Single-Leptonic $t\bar{t}$ Events*, CERN-CMS-NOTE-2006-024.

[154] S. Baffioni *et al.*, *Electron Reconstruction in CMS*, CERN-CMS-NOTE-2006-040.

[155] G. Davatz, M. Dittmar, and A. S. Giolo-Nicollerat, *Standard Model Higgs Discovery Potential of CMS in the $H \to WW \to \ell\nu\ell\nu$ Channel*, CERN-CMS-NOTE-2006-047.

[156] S. Baffioni *et al.*, *Discovery Potential for the SM Higgs Boson in the $H \to ZZ^{(*)} \to e^+e^-e^+e^-$ Decay Channel*, CERN-CMS-NOTE-2006-115.

[157] J. D'Hondt, S. Lowette, G. Hammad, J. Heyninck, and P. Van Mulders, *Observability of Same-Charge Lepton Topology in Di-leptonic $t\bar{t}$ Events*, CERN-CMS-NOTE-2006-065.

[158] H.-F. Pi, P. Avery, D. Green, J. Rohlf, and C. Tully, *Measurement of Missing Transverse Energy with the CMS Detector at the LHC*, CERN-CMS-NOTE-2006-035.

[159] LEP Collaboration, *A Combination of Preliminary Electroweak Measurements and Constraints on the Standard Model*, `hep-ex/0412015`.

[160] BABAR Collaboration, B. Aubert *et al.*, Phys. Rev. **D69** (2004) 111103.

[161] BABAR Collaboration, B. Aubert *et al.*, Phys. Rev. Lett. **93** (2004) 011803.

[162] BABAR Collaboration, B. Aubert *et al.*, Phys. Rev. Lett. **92** (2004) 071802.

[163] J. D'Hondt, S. Lowette, *et al.*, *Fitting of Event Topologies with External Kinematic Constraints in CMS*, CERN-CMS-NOTE-2006-023.

[164] R. K. Bock (Ed.), K. Bos (Ed.), S. Brandt (Ed.), J. Myrheim (Ed.) and M. Regler (Ed.), *Formulae and Methods in Experimental Data Evaluation with Emphasis on High-Energy Physics. Vol. 3: Articles on Statistical and Numerical Methods*, Petit Lancy, Switzerland: Eps (1984) Nonconsec. Pag.

[165] J. Heyninck, J. D'Hondt, and S. Lowette, *Top-Quark Mass Measurement in Single-Leptonic $t\bar{t}$ Events*, CERN-CMS-NOTE-2006-066.

[166] `http://www-cdf.fnal.gov/physics/new/top/public/btag/`

[167] S. Lowette, J. D'Hondt, J. Heyninck, and P. Vanlaer, *Offline Calibration of b-Jet Identification Efficiencies*, CERN-CMS-NOTE-2006-013.

[168] J. D'Hondt, J. Heyninck, and S. Lowette, *Measurement of the Cross Section of Single-Leptonic $t\bar{t}$ Events*, CERN-CMS-NOTE-2006-064.

[169] M. Davids *et al.*, *Measurement of Top-Pair Cross Section and Top-Quark Mass in the Di-Lepton and Full-Hadronic Channels with CMS*, CERN-CMS-NOTE-2006-077.

[170] N. Kidonakis and R. Vogt, Phys. Rev. **D68** (2003) 114014.

[171] J. Campbell, R. K. Ellis, F. Maltoni, and S. Willenbrock, Phys. Rev. **D69** (2004) 074021.

[172] J. M. Campbell, *$W/Z + b\bar{b}$/jets at NLO Using the Monte Carlo MCFM*, `hep-ph/0105226`.

[173] J. M. Campbell and R. K. Ellis, Phys. Rev. **D60** (1999) 113006.

[174] T. Sjöstrand, Private communication.

[175] V. D. Barger, R. J. N. Phillips, and D. P. Roy, Phys. Lett. **B324** (1994) 236–240.

[176] J. F. Gunion, Phys. Lett. **B322** (1994) 125–130.

[177] S. Moretti and D. P. Roy, Phys. Lett. **B470** (1999) 209–214.

[178] D. J. Miller, S. Moretti, D. P. Roy, and W. J. Stirling, Phys. Rev. **D61** (2000) 055011.

[179] D. Cavalli *et al.*, *The Higgs Working Group: Summary Report*, `hep-ph/0203056`.

[180] *Programs for Significance Calculations in CMS*, `http://cmsdoc.cern.ch/~bityukov/`

[181] S. Lowette, P. Vanlaer, and J. Heyninck, *Heavy Charged MSSM Higgs Bosons in the $H^{\pm} \to tb$ Decay in CMS*, CERN-CMS-NOTE-2004-017.

[182] S. Abdullin, *CMSJET*, CMS TN/94-180 (version of 10/10/2001).

[183] CDF Collaboration, T. Affolder *et al.*, Phys. Rev. **D64** (2001) 032002.

[184] DØ Collaboration, V. M. Abazov *et al.*, Phys. Lett. **B626** (2005) 35–44.

[185] D. P. Roy, Phys. Lett. **B459** (1999) 607–614.

[186] M. Baarmand, M. Hashemi, and A. Nikitenko, J. Phys. **G32** (2006) N21.

[187] M. Carena, S. Heinemeyer, C. E. M. Wagner, and G. Weiglein, Eur. Phys. J. **C26** (2003) 601–607.

[188] S. Heinemeyer, W. Hollik, and G. Weiglein, JHEP **06** (2000) 009.

[189] R. Kinnunen, *Search for the Heavy Charged MSSM Higgs Bosons with the $H^{\pm} \to \tau^{\pm}\nu_{\tau}$ Decay Mode in the Fully Hadronic Final State*, CERN-CMS-NOTE-2006-100.

[190] S. Lowette, J. D'Hondt, and P. Vanlaer, *Charged MSSM Higgs Boson Observability in the $H^{\pm} \to tb$ Decay*, CERN-CMS-NOTE-2006-109.

[191] C. Biscarat and M. Dosil, *Charged-Higgs Search in Top-Quark Decays with the ATLAS Detector*, ATLAS-PHYS-2003-038.

[192] K. A. Assamagan and Y. Coadou, *The Hadronic τ Decay of a Heavy Charged Higgs in ATLAS*, ATLAS-PHYS-2000-031.

[193] K. A. Assamagan, *The Charged Higgs in Hadronic Decays with the ATLAS Detector*, ATLAS-PHYS-99-013.

[194] K. A. Assamagan and N. Gollub, Eur. Phys. J. **C39S2** (2005) 25–40.

[195] K. A. Assamagan, Acta Phys. Polon. **B31** (2000) 881–893.

[196] R. Kinnunen, M. Hashemi, and M. Baarmand, *Study of the* $W^{\pm}h^0$ *Decay Channel of the Charged Higgs Boson in CMS*, CMS-IN-2005-032.

[197] B. Mohn, N. Gollub, and K. A. Assamagan, *Study of the* $H^{\pm} \to W^{\pm}H^0$ *Decay in a Large Mass-Splitting MSSM Scenario with ATLAS*, ATLAS-PHYS-PUB-2005-017.

[198] M. Bisset, F. Moortgat, and S. Moretti, Eur. Phys. J. **C30** (2003) 419–434.

[199] C. Hansen, N. Gollub, K. Assamagan, and T. Ekelof, Eur. Phys. J. **C44S2** (2005) 1–9.

[200] Action item in the conclusions of the *Charged Higgs Workshop 2006*, Uppsala, Sweden, 2006.

List of Abbreviations

2HDM	Two-Higgs-Doublet Model
ADD	Arkani-Hamed, Dimopoulos and Dvali, used to refer to a particular type of extra-dimension scenarios
ALICE	A Large Ion Collider Experiment at the LHC
ALPGEN	Monte-Carlo event generator for multi-parton processes in hadronic collisions
ATLAS	A Toroidal LHC ApparatuS
BABAR	Experiment at SLAC's PEP-II B-factory, USA
BELLE	Experiment at the KEK B-factory, Japan
BR	Branching Ratio
BSM	Beyond the Standard Model
C++	Object-oriented programming language
CDF	Collider Detector Facility at the FNAL Tevatron
CERN	Conseil Européen pour la Recherche Nucléaire, European laboratory for particle physics
CKM	Cabibbo, Kobayashi and Maskawa, used to refer to the quark-mixing matrix
CMKIN	CMS KINematics package (legacy `FORTRAN`)
CMS	Compact Muon Solenoid
CMSIM	CMS SIMulation package (legacy `FORTRAN`)
CMSJET	CMS fast simulation package (legacy `FORTRAN`)
COBRA	Coherent Object-oriented Base for Reconstruction, Analysis and simulation
CRAB	Cms Remote Analysis Builder
CompHEP	Package for evaluation of Feynman diagrams and integration over multi-particle phase space
CP	Charge and Parity transformations
CSC	Cathode Strip Chamber
CTEQ	Coordinated Theoretical-Experimental Project on QCD
DØ	Experiment at the FNAL Tevatron
DAQ	Data AcQuisition
DELPHI	An experiment at LEP
DESY	Deutsches Elektronen SYnchrotron laboratory, Hamburg

DGLAP	Dokshitzer, Gribov, Lipatov, Altarelli and Parisi, used to refer to the parton-evolution equations
DT	Drift Tube
ECAL	Electromagnetic CALorimeter
ED	Extra Dimensions
EM	ElectroMagnetic
EWSB	ElectroWeak Symmetry Breaking
FAMOS	CMS fast simulation package
FCNC	Flavour-Changing Neutral Current
FNAL	Fermi National Accelerator Laboratory, USA
FORTRAN	Low-level procedural programming language for numeric and scientific computation
FSR	Final-State Radiation
GEANT	Detector simulation framework and toolkit
GRID	Infrastructure for distributed computing
GUT	Grand Unified Theory
HCAL	Hadron CALorimeter
HDECAY	Program for the calculation of branching fractions in Higgs-boson decays
HEP	High-Energy Physics
HERA	Electron-proton collider at DESY
HERWIG	Hadron Emission Reactions With Interfering Gluons, a Monte-Carlo event generator for high-energy physics collisions
HLT	High-Level Trigger
IIHE	Inter-university Institute for High Energies, Brussels
ISR	Initial-State Radiation
JES	Jet Energy Scale
L1	Level-1 Trigger
LEP	Large Electron-Positron collider at CERN
LHC	Large Hadron Collider
LHCb	Large Hadron Collider beauty experiment
LHCf	Large Hadron Collider forward experiment
LO	Leading Order
LSP	Lightest Supersymmetric Particle
MadGraph	Multi-particle matrix-element generator
MadEvent	Event generator powered by MadGraph
MB	Minimum Bias collisions
MC	Monte-Carlo simulation program or technique
ME	Matrix Element
MRST	Martin, Roberts, Stirling and Thorne, used to refer to a particular set of parton density functions
MS	Minimal Subtraction
MSSM	Minimal SuperSymmetric Model

mSUGRA	Minimal SUperGRAvity model of supersymmetry
NLO	Next-to-Leading Order
ORCA	Object-oriented Reconstruction for CMS Analysis
OSCAR	Object-oriented Simulation for CMS Analysis and Reconstruction
PDF	Parton Density Function
PS	Parton Shower or Proton Synchrotron
PV	Primary Vertex
PYTHIA	Monte-Carlo event generator for high-energy physics collisions
QCD	Quantum ChromoDynamics
QED	Quantum ElectroDynamics
RF	Radio Frequency
RGE	Renormalization-Group Equations
RICH	Ring-Imaging Čerenkov Hodoscope
ROOT	An object-oriented data-analysis framework
RPC	Resistive Plate Chamber
RS	Randall and Sundrum, used to refer to a particular type of extra-dimension scenarios
SCRAM	Software Configuration, Release And Management tool
SM	Standard Model
SPS	Super Proton Synchrotron collider at CERN
TEVATRON	TeV-range proton–anti-proton accelerator at FNAL
TOTEM	LHC experiment for the measurement of the total cross section, elastic scattering and diffraction dissociation
TPC	Time Projection Chamber
UA1/2/5	Experiments at the CERN SPS collider
UE	Underlying Event
UED	Universal Extra Dimensions
VBF	Vector-Boson Fusion
VEV	Vacuum Expectation Value

Summary

The Standard Model describes extraordinary well the known elementary particles and their interactions. Several strong arguments, however, like the hierarchy problem, favour the interpretation of the Standard Model as a low-energy effective limit of a more fundamental theory at higher energy scales. Several of the proposed extensions of the Standard Model predict new physics to be observed at the TeV energy scale. This energy domain will be explored by the Large Hadron Collider (LHC) at CERN, of which the start-up is foreseen at the end of 2007.

An appealing example of such a Standard Model extension introduces supersymmetry, which relates bosons and fermions and solves many of the shortcomings of the Standard Model. Among the predicted new particles are several Higgs bosons that are needed to provide mass terms in the theory. In a minimal scenario two Higgs doublets are needed, giving rise to five Higgs bosons of which two are charged.

In this thesis the observability for a heavy charged Higgs boson is studied in the $H^{\pm} \to tb$ decay channel with the CMS experiment at the LHC. Two strategies based on the tagging of either three or four b jets are considered. In this context a novel method is presented to calibrate b-tagging algorithms on data. It consists of using $t\bar{t}$ pairs, abundantly produced at the LHC, to isolate jet samples with a highly-enriched b-jet content. On these selected samples the b-tagging performance can be measured.

A multivariate technique is used to balance the systematic and statistical uncertainty, resulting in an optimal expected precision on the measurement of the b-tagging performance. For $1\,\mathrm{fb}^{-1}$ ($10\,\mathrm{fb}^{-1}$) of integrated luminosity the relative accuracy on the b–jet identification efficiency is expected to be about 6% (4%) in the barrel region and about 10% (5%) in the endcaps. The obtained result reflects the best current estimate of the expected b-tagging uncertainties in the CMS experiment, and it is used as a benchmark for systematic uncertainties in all analyses employing b-jet identification.

The analysis of the charged Higgs-boson identification via the $H^{\pm} \to tb$ decay illustrates the importance of the calibration of b-tagging algorithms. In this study the combinatorial complexity in the signal is tackled with a multivariate approach, but despite the optimization the charged Higgs-boson mass reconstruction proves to be very challenging. Additional difficulties arise from the $t\bar{t}$ + jets background, which is very large and dominated by events with additional non-b jets that are mis-identified as b jet. The inclusion of systematic uncertainties on the background level leaves no relevant sensitivity for this channel in two-Higgs-doublet models during the low-luminosity phase of the LHC. This result is interpreted in the context of the minimal supersymmetric extension of the Standard Model including alternative decay channels and neutral Higgs-boson searches at the LHC.

Samenvatting

Het Standaard Model beschrijft de gekende elementaire deeltjes en hun interacties uitzonderlijk goed. Verscheidene tekortkomingen, zoals het hiërarchieprobleem, duiden er echter op dat het Standaard Model dient gezien te worden als een effectieve limiet bij lage energie van een meer fundamentele theorie bij een hogere energieschaal. Verschillende van de voorgestelde uitbreidingen van het Standaard Model voorspellen de waarneming van nieuwe fysica aan de TeV energieschaal. Dit energiedomein zal worden verkend door de Large Hadron Collider (LHC) aan het CERN, voorzien om opgestart te worden aan het einde van 2007.

Een aantrekkelijk voorbeeld van een dergelijke uitbreiding van het Standaard Model introduceert de zogenaamde supersymmetrie, die bosonen met fermionen relateert en veel van de tekortkomingen van het Standaard Model oplost. Enkele van de vele nieuwe deeltjes die voorspeld worden zijn Higgs bosonen, dewelke nodig zijn om massatermen te voorzien in de theorie. In een minimaal scenario zijn twee Higgs-doubletten vereist, wat leidt tot vijf Higgs bosonen, waarvan er twee geladen zijn.

In deze thesis wordt de mogelijkheid bestudeerd tot het waarnemen van geladen Higgs bosonen met grote massa in het $H^{\pm} \to tb$ vervalkanaal met het CMS experiment aan de LHC. Twee strategieën worden onderzocht, gebaseerd op de identificatie van ofwel drie, ofwel vier b-jets. In deze context wordt een nieuwe methode voorgesteld voor de calibratie van algoritmes voor b-identificatie. Top-quark paren, die in grote hoeveelheden zullen worden geproduceerd aan de LHC, worden gebruikt om stalen van jets af te zonderen met een hoge zuiverheid aan b-jets. Op deze geselecteerde stalen kan de performantie van de b-identificatie worden gemeten.

Een multivariate techniek wordt gebruikt om de systematische en de statistische fout in evenwicht te brengen, wat resulteert in een optimale verwachte precisie op de meting van de performantie van de b-identificatie. Voor $1\,\mathrm{fb}^{-1}$ ($10\,\mathrm{fb}^{-1}$) geïntegreerde luminositeit wordt de verwachte relatieve precisie op de efficienctie van b-jet identificatie geschat op ongeveer 6% (4%) in de centrale "barrel" regio, en ongeveer 10% (5%) in het voorwaardse "endcap" gedeelte van de detector. Het behaalde resultaat is op dit ogenblik de beste schatting van de verwachte b-identificatie onzekerheden in het CMS experiment, en het wordt gebruikt als systematische onzekerheid in alle analyses die b-jet identificatie gebruiken.

De analyse van de geladen Higgs-boson identificatie in het $H^{\pm} \to tb$ vervalkanaal illustreert het belang van de calibratie van algoritmes voor b-identificatie. In deze studie wordt de combinatorische complexiteit in het signaal behandeld met een multivariate aanpak, maar ondanks deze optimalisatie blijkt de reconstructie van de geladen Higgs boson massa bijzonder moeilijk. Extra complicaties komen van de $t\bar{t}$+jets achtergrond,

die zeer groot is en wordt gedomineerd door gebeurtenissen met extra niet-b jets die verkeerdelijk worden geïdentificeerd als b-jet. Het in rekening brengen van systematische onzekerheden op de achtergrond laat geen sensitiviteit over voor dit kanaal in twee Higgs-doublet modellen tijdens de fase van lage luminositeit van de LHC. Dit resultaat is geïnterpreteerd in de context van de minimale extensie van het Standaard Model, en wordt in het kader geplaatst van alternatieve vervalkanalen en de zoektocht naar neutrale Higgs-bosonen aan de LHC.

Acknowledgements

The completion of a thesis does not only involve research, studying and writing, but is as much a personal exploration and learning process. A process which is impossible to bring to an end without the expertise, help and friendship of many. In these few paragraphs I would like to show my gratitude towards all those people who have been important to me during the realization of this thesis.

First I would like to thank my promotor Prof. Stefaan Tavernier, who gave me the opportunity to enter this exciting world of particle physics, and who has always provided me with the freedom and means to explore and learn. Many thanks also to the members of the jury, the Professors Catherine De Clercq, Gigi Rolandi, Fabio Maltoni, Ben Craps, and Robert Roosen, for their interest and efforts and the resulting constructive comments and lively discussions. A very special thank you goes to my co-promotor Prof. Jorgen D'Hondt, who has been during these years an amazing source of guidance, motivation, creativity, patience and friendship. And delicious meals.

A doctorate in physics involves a certain passion for science, which often implies a great deal of stimulation from several people in earlier years. In this context I remember with great pleasure the years at the Volkssterrenwacht Mira that stirred up my interest in astronomy and physics. In secondary school I had the luck to be taught physics by Henk Foriers, a very inspiring and gifted teacher, who put me on the track towards physics studies. At the university the challenges and corresponding pleasure only increased further, and I'd like to thank especially Stijn Nevens for those fine years, Prof. Franklin Lambert for his involvement in my diploma thesis and Prof. Henri Eisendrath for bearing all those delays for the teacher's education.

Working in high-energy physics means working in a surprisingly warm and social community. Many colleagues have also become good friends, with whom I often share very good memories of both work and leisure. At the IIHE, I would in particular like to thank Pascal Vanlaer, Barbara Clerbaux, Olivier Devroede, Benoît Roland, Daan Hubert, Cedric Lemaître, Stijn De Weirdt and especially Jan Heyninck, a friend that is very valuable to me, and from whom I have still a lot to learn. Also a big thanks to Rosine, Marleen, and all others who provided me with invisible but invaluable logistic support. In the Belgian community many colleagues from Louvain-la-Neuve, Antwerp and Mons are to be thanked, in particular Christophe Delaere and Pavel Demin. And let's not forget about the large but scattered international group of colleagues and friends: the CMS b-τ and b-tag groups, with Christian Weiser and Fabrizio Palla; the Higgs group, with Sasha Nikitenko; the many individuals within CMS that helped me out, and whom I sometimes only know from emails; the theoreticians from Les Houches and elsewhere, in particular Tilman Plehn; the fellow summer-school students,

especially the “JJC” and the “Cernia” crowd; and finally so many individuals, in particular Filip Moortgat, Alexander Schmidt, Daniele Benedetti, Caroline Collard and of course Stephanie Beauceron.

Life is not all about work, and many friends have contributed as well to these years in an indirect but significant way. There is the VUB’s orchestra, this large group of people sharing the passion for music and great fun. Also many WK-people are to be thanked, preferably in a very nostalgic way. And so many others... Thanks a lot Martijn, Anke, Marianne, Gabriël, Eveline, Rembert, Viki, Pierre, Dimitri, Els, Feije, Dirk, Jo, Julie, Veronique, Bram, Evi, Eileen, Annelies,...

To conclude I would like to thank my parents and my sisters, who have been the most valuable support of all. So many times you have been ready to listen, help and support in difficult moments. You have given much more than I ever asked for. Thank you so much.

Het denken mag zich nooit onderwerpen,
noch aan een dogma,
noch aan een partij,
noch aan een hartstocht,
noch aan een belang,
noch aan een vooroordeel,
noch aan om het even wat,
maar uitsluitend aan de feiten zelf,
want zich onderwerpen betekent het einde van alle denken.

Henri Poincaré
21 november 1909

Uit een redevoering ter gelegenheid van de 75ste verjaardag van de Université Libre de Bruxelles.